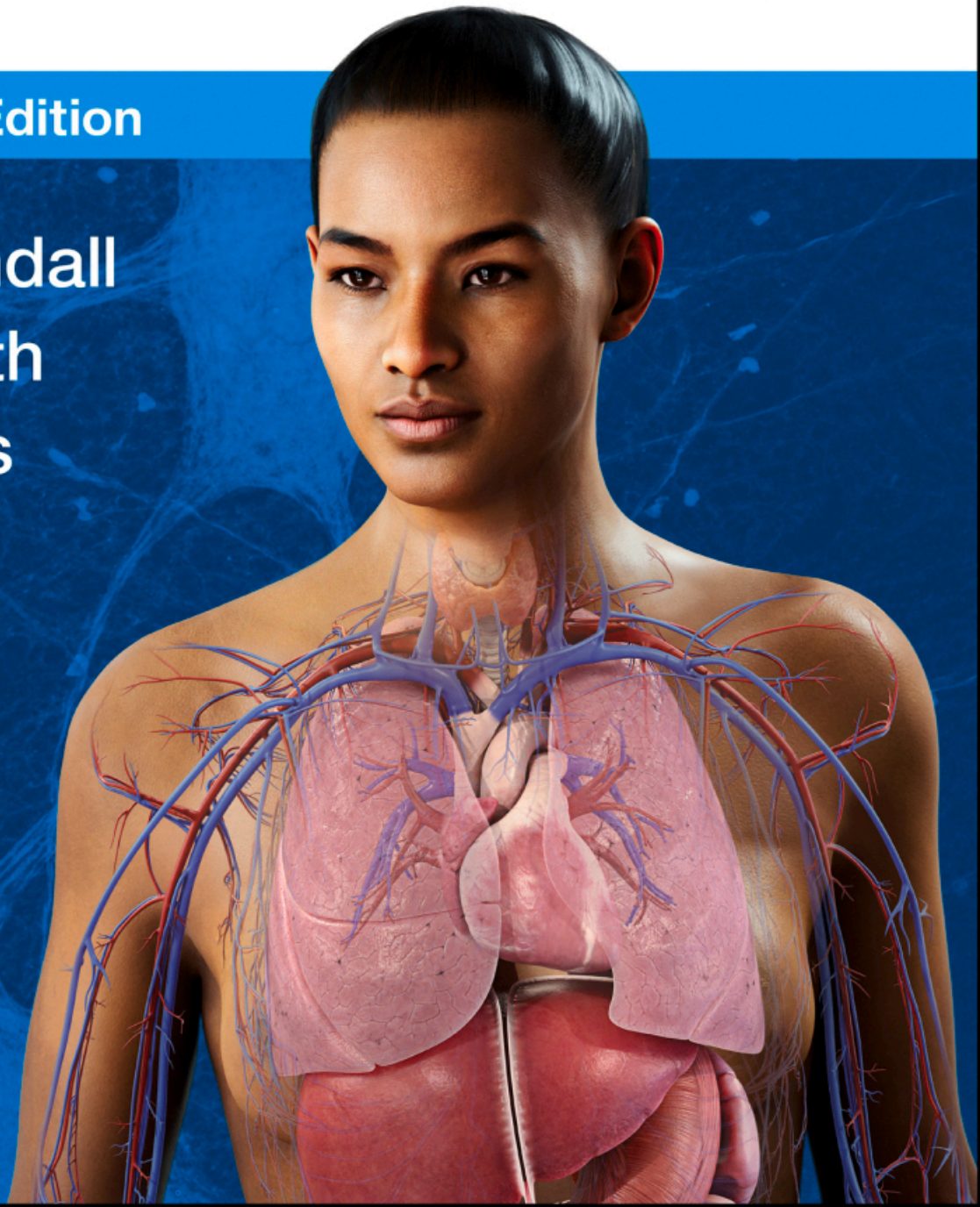
KUMAR & CLARK'S

Clinical Medicine

Enhanced DIGITAL VERSION Included

Eleventh Edition

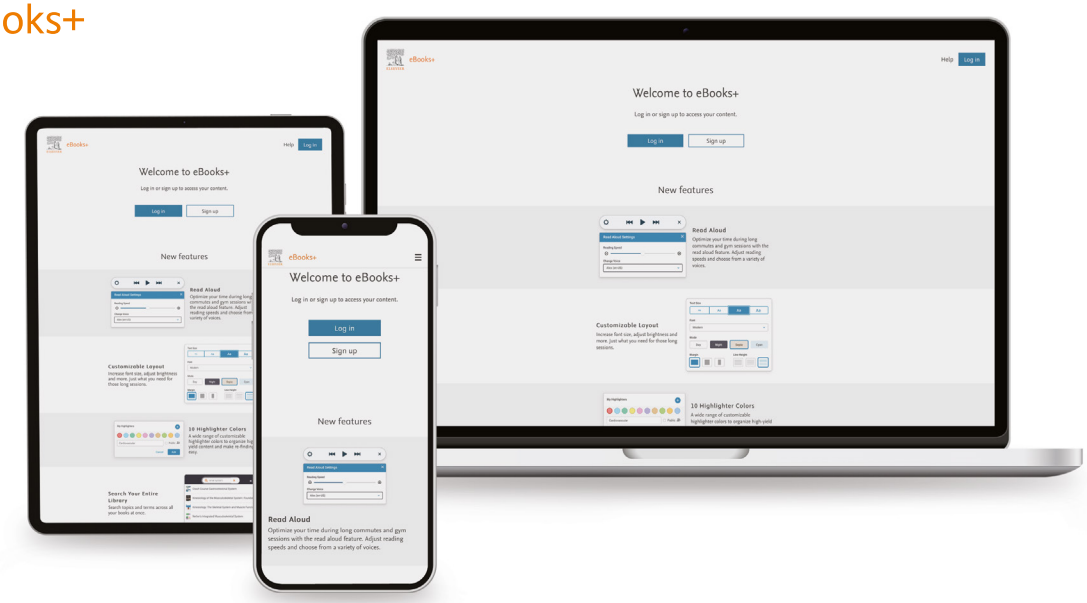
David Randall
John Booth
Kate Wiles



ELSEVIER

eBooks+

ELSEVIER



Activate your eBook today!

Elsevier eBooks+ allows you to search, browse, make notes, highlight, and have content read aloud.

1. To sign up, visit http://ebooks.health.elsevier.com/
2. Scratch to reveal your code below, and enter in **Redeem Access Code** box
3. Click **Redeem**

Contact Customer Support via
https://service.elsevier.com/app/home/supporthub/elsevierebooksplus/

PLACE PINCODE
STICKER HERE

By buying this genuine product you are guaranteed that Elsevier's high-quality online resources are included in this purchase.

Use of the current edition of the electronic version of this book (eBook) is subject to the terms of the non-transferable, limited license granted at https://www.elsevier.com/legal/elsevier-website-terms-and-conditions. Access to the eBook is limited to the first individual who redeems the access code located on the inside cover of this book at http://ebooks.health.elsevier.com/. The access code is valid for 8 years post publication date. The code may not be transferred to another party.

KUMAR & CLARK'S

Clinical Medicine

This page intentionally left blank

KUMAR & CLARK'S

Clinical Medicine

Eleventh Edition

Edited by

David Randall, PhD, MRCP

Honorary Clinical Senior Lecturer, Barts and the London School of Medicine and Dentistry, QMUL;
Consultant Physician and Nephrologist, Barts Health NHS Trust, London, United Kingdom

John Booth, PhD, MRCP

Honorary Clinical Senior Lecturer, Barts and the London School of Medicine and Dentistry, QMUL;
Consultant Physician and Nephrologist, Barts Health NHS Trust, London, United Kingdom

Kate Wiles, PhD, MRCP

Honorary Senior Lecturer, Wolfson Institute of Population Health, Queen Mary University of London;
Consultant Obstetric Physician, Department of Women's Health, Barts Health NHS Trust, London, United Kingdom

Foreword by

Professor Dame Parveen J. Kumar, DBE, BSc, MD, DM(hc), DEd, DSc, FRCP, FRCPG, FRCPI, FRCPath

Emerita Professor of Medicine and Education, Barts and The London School of Medicine and Dentistry, Queen Mary University of London
Honorary Consultant Physician and Gastroenterologist, Barts Health NHS Trust and Homerton University Hospital NHS Foundation Trust, London, United Kingdom

Michael L. Clark, MD, FRCP

Honorary Senior Lecturer, Barts and The London School of Medicine and Dentistry, QMUL, and Princess Grace Hospital, London, United Kingdom

Use the PIN code on the inside front cover to access additional online content.

London New York Oxford Philadelphia St Louis Sydney 2026

Copyright © 2026 by Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

For accessibility purposes, images in electronic versions of this book are accompanied by alt-text descriptions provided by Elsevier. For more information, see https://www.elsevier.com/about/accessibility.

Publisher's note: Elsevier takes a neutral position with respect to territorial disputes or jurisdictional claims in its published content, including in maps and institutional affiliations.

Previous editions copyrighted 2021, 2017, 2012, 2009, 2005, 2002, 1998, 1994, 1990, and 1987.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Details on how to seek permission, further information about the Publisher's permissions policies and our arrangements with organizations such as the Copyright Clearance Center and the Copyright Licensing Agency, can be found at our website: www.elsevier.com/permissions.

This book and the individual contributions contained in it are protected under copyright by the Publisher (other than as may be noted herein).

Notice

Practitioners and researchers must always rely on their own experience and knowledge in evaluating and using any information, methods, compounds or experiments described herein. Because of rapid advances in the medical sciences, in particular, independent verification of diagnoses and drug dosages should be made. To the fullest extent of the law, no responsibility is assumed by Elsevier, authors, editors or contributors for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

ISBN: 978-0-443-12594-2
E-ISBN: 978-0-443-12675-8
INTERNATIONAL EDITION ISBN: 978-0-443-12828-8

Senior Content Strategist: Alex Mortimer
Senior Content Development Specialist: Susan Callahan
Publishing Services Manager: Catherine Jackson
Senior Project Manager: Daniel Fitzgerald
Designer: Amy Buxton
Illustration Manager: Akshaya Mohan
Marketing Manager: Deborah Watkins

Printed in Poland

Last digit is the print number: 9 8 7 6 5 4 3 2 1

Contents

Foreword to the Eleventh Edition ix
Preface to the Eleventh Edition xi
How to use this book xii
List of Contributors xiii
International Advisory Board xix
Acknowledgments for the Eleventh Edition xxi

PART 1 THE PRACTICE OF MEDICINE

1. **The art of being a doctor** 1
   Kate Wiles, John Booth, and David Randall
2. **Clinical communication** 14
   Lorraine M. Noble
3. **Medical ethics** 25
   June Jones
4. **Evidence-based practice** 35
   Rosalind Ann Raine and Peter Martin
5. **Human genetics** 52
   Neil Rajan and David Peter Kelsell
6. **Immunity** 79
   Matthew Buckland and James Emmanuel Thaventhiran
7. **Therapeutics and prescribing** 107
   Vikas Kapil
8. **Radiology** 117
   Cindy Chew, Andrew A. Plumb, and Patrick O'Dwyer

PART 2 MEDICINE AND THE WORLD

9. **Public health** 153
   Sarah Elizabeth Winchester, Simona Baracaia, and Ian Basnett
10. **Global and planetary health** 167
    Babulal Sethia, Dominik Zenner, and Parveen J. Kumar

PART 3 ACUTE AND EMERGENCY MEDICINE

11. **Acute medicine** 175
    Nicholas Robert Murch, Barbara Lachana Onen, John Booth, John Patrick Donaghy, and Laura Kay Harrington
12. **Sepsis and the treatment of bacterial infection** 278
    Mark Melzer, Andrew W. Leitch, and David W. Wareham
13. **Critical care medicine** 297
    Stephen J. Shepherd, Colette Coyle, Andrew W. Leitch, and Parjam Zolfaghari
14. **Anaesthesia and perioperative medicine** 334
    Sohail Bampoe, Peter M. Odor, Stuart Connal, and Elliott Ridgeon
15. **Toxicology** 353
    Mark Pucci and Sally Marie Bradberry
16. **Environmental medicine** 371
    Michael Greene and Peter David Hodkinson

PART 4 MEDICAL SPECIALTIES

17. **Cardiology** 384
    Robin Ray and Hitesh C. Patel
18. **Hypertension** 496
    Vikas Kapil
19. **Lipid and metabolic disorders** 510
    Anthony S. Wierzbicki and Wiaam Al-Hasani
20. **Respiratory medicine** 521
    Veronica White and Prina Ruparelia
21. **Ear, nose, and throat** 582
    Elinor Warner and Kay Seymour
22. **Gastroenterology** 601
    Noor Jawad and Charlotte I. Skinner
23. **Hepatology** 679
    Graham R. Foster and Alastair O'Brien
24. **Biliary tract and pancreatic disease** 735
    Christopher Antony Wadsworth
25. **Nutrition** 762
    Shameer J. Mehta and Sarah Faloon
26. **Kidney and urinary tract disease** 789
    M. Magdi Yaqoob and Neil Ashman
27. **Water balance, fluids, and electrolytes** 853
    M. Magdi Yaqoob and Kieran McCafferty
28. **Neurology** 885
    Apeksha M. Shah

29. Stroke medicine **977**
Ganesh Subramanian, Margaret O'Connor, Soma Banerjee, and Abhijit Das

30. General hospital psychiatry **1001**
Peter Byrne, Sotirios Posporelis, and Sridevi Siramahalingappa

31. Ophthalmology **1044**
Robin D. Hamilton

32. Rheumatology **1061**
Mohammed Abbas Anisur Rahman and Ian Philip Giles

33. Bone disease **1121**
Richard Conway

34. Endocrinology **1137**
Helena Gleeson, Narendra L. Reddy, and Miles J. Levy

35. Diabetes mellitus **1200**
Richard Ian Gregory Holt

36. Metabolic disease **1240**
Radha Ramachandran and Raphael Buttigieg

37. Obesity **1248**
Anjali Zalin, Ritwika Mallik, and Beenish Masood

38. Haematology **1270**
Michael F. Murphy, K. John Pasi, and Noemi Roy

39. Haematological oncology **1328**
Joost de Vos and Vanessa Foggo

40. Venous thromboembolic disease **1363**
Peter K. MacCallum, K. John Pasi, and Suthesh Sivapalaratnam

See your e-book for MCQs, videos, and audio.

41. Malignant disease **1375**
Hamoun Rozati, Emily Scott, Piers Nicholas Plowman, and Jonathan Shamash

42. Infectious disease **1417**
Michael Brown, Philip Gothard, and William L. Irving

43. Sexually transmitted infections **1519**
Janet D. Wilson

44. Human immunodeficiency virus **1537**
Jane Anderson and Chloe M. Orkin

45. Dermatology **1564**
Anthony Paul Bewley, Malvina Cunningham, Anshoo Sahota, and Catherine Anne Harwood

PART 5 MEDICINE THROUGHOUT LIFE

46. Young adult medicine **1631**
Arvind Nagra, Gary James Connett, Matthew Harmer, and Yincent Tse

47. Sex and gender in medicine **1639**
Eoin Dinneen, Michael Dinneen, J. Michael Dixon, Elizabeth Egbase, Leighton John Seal, Sophie Michelle Strong, and Rebecca Swan

48. Obstetric medicine **1680**
Catherine Nelson-Piercy and Melanie Nana

49. Medicine of the older adult **1694**
Adam Harper, Iain Stewart Ayrton Wilkinson, and Joanna Preston

50. Palliative care and symptom control **1720**
Chloe Chin, Catherine Moffat, and Richella Ryan

Index **1735**

Video Contents

17. Cardiology

Video 17.1: The cardiovascular system

18. Hypertension

Video 18.1: Measurement of blood pressure

20. Respiratory medicine

Video 20.1: Examination of the respiratory system

21. Ear, nose, and throat

Video 21.1: Examination of the neck

22. Gastroenterology

Video 22.1: Examination of the gastrointestinal system

27. Water balance, fluids, and electrolytes

Video 27.1: Assessment of fluid balance

28. Neurology

Video 28.1: The timed up and go test

Video 28.2: Examination of the sensory system of the lower limbs

Video 28.3: Femoral nerve stretch test

Video 28.4: Patellar tap test

Video 28.5: Examination of the optic (II), oculomotor (III), trochlear (IV) and abducens (VI) nerves

Video 28.6: Examination of the trigeminal (V) nerve

Video 28.7: Examination of the facial (VII) nerve

Video 28.8: Examination of the vestibulocochlear (VIII) nerve

Video 28.9: Examination of the cranial (IX, X, XI and XII) nerves

Video 28.10: Examination of the motor system of the upper limbs

Video 28.11: Examination of the motor system of the lower limbs

Video 28.12: Examination of the sensory system of the upper limbs

32. Rheumatology

Video 32.1: Examination of the hand and wrist

Video 32.2: Examination of the foot and ankle

Video 32.3: Schober's test

35. Diabetes mellitus

Video 35.1: Examination of the peripheral vascular system and diabetic foot

47. Sex and gender in medicine

Video 47.1: Gynaecological examination

48. Obstetric medicine

Video 48.1: Obstetric examination

This page intentionally left blank

Foreword to the Eleventh Edition

Much has happened in the years since the last edition, not least the rapid spread of disease leading to the pandemic which stopped the world. We were all reminded that the world is intricately linked and that diseases can spread with great rapidity. Medicine is now viewed on a global landscape with the emergence of new diseases, and re-emergence of others. We live in a world with a changing and unpredictable climate, influencing the mass movement of populations due to, and leading to, difficult socioeconomic factors. This has led to a change in epidemiology and presentation of some diseases.

This book has always been written for students and doctors from around the world and each edition has kept up with the changes in diseases and technology. When the first edition of this book appeared on the market, it was said to have 'changed the face of medical publishing' by its clear format, accessibility, and the succinct way in which evidence-based facts were presented. The old turgid presentations with long unstructured prose were substituted by diagrams, algorithms, and tables to explain the scientific concepts and clinical aspects of diseases. Colour was introduced to help the reader find medicine fun and easy to learn. Subsequent editions of *Kumar & Clark's Clinical Medicine* have followed this basic structure but with added innovative ways of presentation, particularly in the online books. As new curricula evolve, new chapters have been added to keep up with the way that medicine is being taught.

This 11th edition is no exception. It has had to keep up with the extremely rapid pace of recent changes, some triggered by the pandemic, and others due to new discoveries in medicine. These include advances in technology, artificial intelligence, imaging for diagnostics, genomics, and personalised medicine. Modern physicians now have the difficult task of adapting their diagnostic skills and clinical acumen with the use of continually changing modern technology. They need to be compassionate, empathic, and adaptable, to face a new world of machines. With the advance of mechanical diagnostic aids, robots, and exciting new therapies, the role of the doctor often becomes blurred, making it ever important to keep the patient central.

In this edition, all the chapters have been carefully updated by the authors. David Randall has continued as editor, and John Booth and Kate Wiles have come on board as new editors. New authors have also been appointed with chapters added to follow the trends of modern medicine.

The multidisciplinary care of the patient is now the norm. This book continues to be used by pharmacists, nurses, and other allied professionals. We hope that they will find this useful for their respective roles as this book provides comprehensive knowledge for the practise of medicine.

We wish all our readers the very best in their chosen careers and hope they find medicine as exciting and enjoyable as we have. It has been an immense privilege to be a doctor.

Parveen J. Kumar and Michael L. Clark

Dr. Michael L. Clark and Professor Dame Parveen J. Kumar on the occasion of the publication of the first edition of *Clinical Medicine* in 1987.

This page intentionally left blank

Preface to the Eleventh Edition

First published in 1987, *Kumar & Clark's Clinical Medicine* has established itself as a core textbook for medical students, doctors in training, and allied health professionals worldwide. Born out of Professor Dame Parveen Kumar and Dr. Michael Clark's desire to transmit their passion for high-quality and evidence-based practice of medicine in an engaging and accessible format to the next generation of doctors, the editorial team for the eleventh edition of *Clinical Medicine* have worked hard to remain true to these fundamental aims. As ever, we are grateful for the extensive feedback and suggestions received on a regular basis, enabling us to revise and reshape the book with sensitivity to its most important stakeholder: yourself, the reader.

The practice of medicine continues to evolve at a staggering pace. Many hospitals are now 'paper free', with electronic records and real-time data transfer ushering in a new era of remote and 'virtual' consultation. Advances in clinical genomics are bringing next-generation genetic testing and gene therapies to the clinic, together with new ethical and moral responsibilities. Artificial intelligence and machine learning are no longer just the subject of science fiction but are already providing opportunities to collate, analyze, and interpret clinical data at a speed and scale previously considered unimaginable. Despite these breath-taking advances, the essence of being a doctor remains unchanged: to evaluate, diagnose, and treat each individual patient within their own unique clinical, social, and emotional context, to do so with empathy and compassion, and to navigate them through an increasingly complex and many-layered 21st-century healthcare system.

With this in mind, we have reshaped the book to resemble, as near as possible, a 'blueprint' for medical practice in the modern world. We have restructured the book into five sections, corresponding to core themes of medical knowledge and patient care, and introduced a number of new chapters to better reflect the structure of the contemporary hospital, and the knowledge and skills required by those working within it:

The practice of medicine: Drawing together the rudiments of medical practice—the 'Clinical Method' of information gathering and diagnosis, legal and ethical aspects of medicine, principles of evidence-based medicine, and relevant basic science applied to hands-on patient care. 'Evidence-based practice' (Chapter 4) includes an extensive new section on 'Medical statistics' to equip readers with the skills to analyze and appraise the medical literature. 'Radiology' (Chapter 8) is a new addition, covering the full spectrum of diagnostic imaging and interventional modalities.

Medicine and the world: 'Zooming out' to examine the determinants of health at the population level including chapters on 'Public health' (Chapter 9) and 'Global and planetary health' (Chapter 10).

Acute and emergency medicine: Pulling together key knowledge required for 'first contact' with patients in emergency and acute medical settings and management of the unwell or critically ill patient. 'Acute medicine' (Chapter 11) includes a 'ground up' approach to assessing patients presenting with 25 of the most common medical symptoms, concise pictorial guides to medical procedures and overviews of management of a broad range of medical emergencies. 'Anaesthesia and perioperative medicine' (Chapter 14) covers the contemporary approach to pre-, peri-, and postoperative management of the surgical patient. 'Environmental medicine' (Chapter 16) explores the interface between human health and extremes of the natural world.

Medical specialties: Covering the full spectrum of organ and system-based hospital specialty medicine. New standalone chapters include 'Stroke medicine' (Chapter 29), 'General hospital psychiatry' (Chapter 30), 'Obesity' (Chapter 37), and 'Human immunodeficiency virus' (Chapter 44), reflecting the growth and importance of these areas and better mapping to modern hospital teams and clinics.

Medicine throughout life: Focusing on the principles and practice of medicine contextualized to age, sex and gender. New chapters on 'Young adult medicine' (Chapter 46) and 'Sex and gender in medicine' including transgender health (Chapter 47) have been added to revised and updated chapters on 'Obstetric medicine' (Chapter 48), 'Medicine of the older adult' (Chapter 49), and 'Palliative care and symptom control' (Chapter 50).

Each chapter is formatted to include 'clinical skills' content at the beginning, providing a practical 'toolkit' for the reader, summaries of specialty-specific patient assessments, and a pictorial overview of key clinical examination findings. Online, readers will find additional resources, including self-assessment questions, clinical skills videos, and a table listing the location of conditions relevant to the Medical Licensing Assessment (MLA).

We hope you will find this edition helpful in your efforts to learn and practise medicine. Any suggestions you may wish to make are warmly welcomed and will help us to ensure that this book continues to meet the needs of its readers.

David Randall, John Booth, and Kate Wiles
2025

HOW TO USE THIS BOOK

The book is organized into five major sections (see 'Preface to the Eleventh Edition', previous page). Chapters focused on individual medical subspecialties begin by offering a tailored approach to basic clinical skills such as history-taking and examination. We have tried to encourage generic skills to be applied thoughtfully to the specialty in question: to encourage learners to adopt the approach taken by experts in the field, so that relevant questions are asked and key physical signs are carefully sought. Most chapters then cover basic sciences, including the anatomy and physiology of the relevant organs in question, before dealing with individual conditions in order. Within the text:

- **Pink subheadings** introduce individual conditions,
- **Green subheadings** introduce the scientific background to the condition (e.g., 'Aetiology' or 'Clinical features'), and
- **Orange subheadings** introduce clinical material, such as 'Investigations' or 'Management'.

Boxes have been classified by the type of information they provide and have the following symbols:

Emergency management

Clinical information

Differential diagnosis

Practical procedures

'Emergency' boxes are in red.

In the Index Medical Licensing Assessment (MLA) conditions have been highlighted in colour to make them easier to find. An MLA conditions contents table can also be found in the e-book.

E-Book extras

A selection of examination videos and assessment content are available online, accessible via the PIN page in the front of this book.

Prescribing

We have used the Recommended International Non-proprietary Names (rINNs) for all drugs. In some diseases where a particular formulation of a drug is required, the proprietary name is used. Drugs spellings follow international usage, e.g., bendroflumethiazide and not bendrofluazide, and amfetamine and not amphetamine. For adrenaline and noradrenaline, we have added epinephrine and norepinephrine in brackets, as these names are often used in emergency guidelines across the world.

Dosages have been given where appropriate, but we recommend that all readers check with their national formularies for the exact dosages.

Units of measurement

We have used the International System of Units (SI units) throughout the book. On occasion, if there is a possibility for confusion, we have also used non-SI units and given a conversion factor.

Also in the Kumar & Clark family of books:

- *Essentials of Kumar & Clark's Clinical Medicine,* 7e
- *Kumar & Clark's Cases in Clinical Medicine,* 5e
- *Clinical Surgery,* 4e

List of Contributors

The editors would like to acknowledge and offer grateful thanks for the input of all previous editions' contributors, without whom this new edition would not have been possible.

Wiaam Al-Hasani MBChB, MRCS, MSc
Specialist Registrar, Chemical Pathology, Guy's and St Thomas' Hospital, London, United Kingdom
Lipid and metabolic disorders

Jane Anderson BSc, MBBS, PhD, FRCP
Consultant Physician, Jonathan Mann Centre, Homerton Healthcare NHS Foundation Trust; Honorary Professor, Wolfson Institute of Population Health, Barts and The London School of Medicine and Dentistry, Queen Mary University of London; Consultant Physician, HIV and Sexual Health, Barts Health NHS Trust, London, United Kingdom
Human immunodeficiency virus

Neil Ashman FRCP, PhD
Consultant Nephrologist, Renal Medicine & Transplantation, Royal London Hospital, London, United Kingdom
Kidney and urinary tract disease

Sohail Bampoe MB BS, BSc, MSc, FRCA
Consultant, Anaesthesia and Perioperative Medicine, University College London Hospitals; Clinical Professor, Cleveland Clinic London, Cleveland Clinic Lerner College of Medicine, Case Western Reserve University; Honorary Associate Professor, Perioperative Medicine, University College London, London, United Kingdom
Anaesthesia and perioperative medicine

Soma Banerjee MBBS, BSc(Hons), MD(Res), FRCP(UK)
Consultant in Stroke Medicine, Department of Stroke Medicine, Imperial College Healthcare NHS Trust; Honorary Senior Lecturer, Department of Brain Sciences, Imperial College London, London, United Kingdom
Stroke medicine

Simona Baracaia BSc (Hons), BM BCh, MPH
Consultant in Public Health Medicine, Public Health, NHS England, London, United Kingdom
Public health

Ian Basnett MBBS, MSc, FFPH, PGcertEdu
Public Health Director, Public Health, Barts Health; Honorary Professor of Public Health, Public Health, QMUL, London, United Kingdom
Public health

Anthony Paul Bewley BA(Hons); MBChB, FRCP
Consultant Dermatologist, Dermatology, Barts Health NHS Trust; Honorary Professor, Dermatology, Queen Mary University, London, United Kingdom
Dermatology

John Booth PhD, MRCP
Honorary Clinical Senior Lecturer, Barts and the London School of Medicine and Dentistry, QMUL; Consultant Physician and Nephrologist, Barts Health NHS Trust, London, United Kingdom
The art of being a doctor
Acute medicine

Sally Marie Bradberry BSc, MD, FRCP, FAACT, FEAPCCT
Director, West Midlands Poisons Unit, City Hospital, Birmingham, United Kingdom
Toxicology

Michael Brown BM BCh, FRCP, PhD, DTM&H
Consultant Physician, Hospital for Tropical Diseases, UCLH, London, United Kingdom
Infectious disease

Matthew Buckland MBBS, MSc, PhD, FRCP, FRCPath
Consultant, Department of Clinical Immunology, Barts Health NHS Trust; Honorary Associate Professor, Department of Infection, Immunity and Inflammation, University College London, London, United Kingdom
Immunity

Raphael Buttigieg MD, MRCP, FRCPath
Consultant in Inherited Metabolic Disease, Chemical Pathology and Metabolic Medicine, Guy's and St Thomas' NHS Foundation Trust, London, United Kingdom
Metabolic disease

Peter Byrne MA, MD, MRCPsych
Consultant, Psychological Medicine, Royal London Hospital, London, United Kingdom; Visiting Professor, Health Sciences, University of Strathclyde, Glasgow, United Kingdom
General hospital psychiatry

Cindy Chew MBChB, MRCS (Glasg, 2003), FRCR (2008), MSc(Research), PhD
Professor, School of Medicine, Dentistry & Nursing, University of Glasgow, Glasgow, United Kingdom
Radiology

Chloe Chin MBBS, MRCP, MPhil (Cantab)
Consultant in Palliative Medicine, Department of Palliative Medicine, Cambridge University Hospitals NHS Foundation Trust, Cambridge, United Kingdom
Palliative care and symptom control

Stuart Connal MBChB, BMedSci(Hons), MSc, AFHEA, MRCP(UK), FRCA
Fellow in Perioperative Medicine, Department of Anaesthesia & Perioperative Medicine, University College London Hospitals NHS Foundation Trust; Specialty Registrar in Anaesthesia, London School of Anaesthesia; Honorary Lecturer, Department of Targeted Intervention, University College London, London, United Kingdom
Anaesthesia and perioperative medicine

Gary James Connett MBChB, MD, FRCPCH
Professor, Southampton Children's Hospital, University Hospital Southampton NHS Foundation Trust, Southampton, United Kingdom
Young adult medicine

Richard Conway MB BCh, BAO, PhD, CCD, FRCPI
Consultant Rheumatologist, Rheumatology, St James's Hospital; Associate Professor, Clinical Medicine, Trinity College Dublin, Dublin, Ireland
Bone disease

Colette Coyle MBBS, BSc (Hons) physiotherapy, MRCP, FFICM
Consultant in Critical Care, Adult Critical Care Unit, The Royal London Hospital, London, United Kingdom
Critical care medicine

Malvina Cunningham MBPhD, MRCP (Derm)
Consultant Dermatologist, Royal London Hospital, Barts Health NHS Trust; Honorary Senior Lecturer, Queen Mary University, London, United Kingdom
Dermatology

Abhijit Das MBBS, MD, DM, MRCP (Neurology), FRCP (Edin)
Consultant, Neurology, Lancashire Teaching Hospitals NHS Foundation Trust; Honorary Associate Professor, University of Central Lancashire, Preston, United Kingdom
Stroke medicine

Joost de Vos MD, FRCPath
Consultant Haemato-Oncologist, Department of Haemato-Oncology, St Bartholomew's Hospital, London, United Kingdom
Haematological oncology

Eoin Dinneen BA (Hons), PhD, FRCS (Urol)
Specialist Registrar, Department of Urology, University College London Hospitals; Associate Principal Investigator NeuroSAFE PROOF, Division of Surgery & Interventional Sciences, University College London, London, United Kingdom
Sex and gender in medicine

Michael Dinneen MD, FRCSI, FRCS(Urol)
Urological Surgeon, Urology, Chelsea & Westminster Hospital, London, United Kingdom
Sex and gender in medicine

J. Michael Dixon OBE, BSc (Hons), MBChB, MD, FRCS, FRCSEd, FRCPEd
Professor of Surgery at Edinburgh University, Consultant Breast Surgeon, Edinburgh Breast Unit, Edinburgh, Scotland, United Kingdom
Sex and gender in medicine

John Patrick Donaghy MB BCh, BAO
Consultant, Department of Acute Medicine, Whittington Health NHS Trust, London, United Kingdom
Acute medicine

Elizabeth Egbase MBBS, MRCOG, MA
Consultant in Obstetrics and Gynaecology, Women's Division, Royal London Hospital; Honorary Clinical Senior Lecturer, Wolfson Institute of Population Health, Queen Mary University of London, London, United Kingdom
Sex and gender in medicine

Sarah Faloon MBChB, MRCP
Gastroenterology Registrar, Department of Gastroenterology, Barts Health NHS Trust, London, United Kingdom
Nutrition

Vanessa Foggo MA, MBBS, MRCP, FRCPath
Consultant, Haemato-Oncology, St Bartholomew's Hospital, London, United Kingdom
Haematological oncology

Graham R. Foster FRCP
Professor of Hepatology, Barts Liver Centre, Blizard Institute, QMUL, London, United Kingdom
Hepatology

Ian Philip Giles BSc, MBBS, PhD, FRCP, SFHEA
Professor, Centre for Rheumatology, Division of Medicine, University College London; Professor, Department of Rheumatology, University College London Hospital, London, United Kingdom
Rheumatology

Helena Gleeson MBBS, MD, FRCP
Consultant, Endocrinology, Queen Elizabeth Hospital, Birmingham, United Kingdom
Endocrinology

Philip Gothard FRCP
Consultant Physician, Hospital for Tropical Diseases, University College London Hospitals, London, United Kingdom
Infectious disease

Michael Greene MB BS, FRCSEd, FRCEM, Dip Mount Med, Dip Med Ed
Emergency Medicine Consultant (Retired), Kendal Mountain Rescue Team, Mountain Rescue England and Wales, Kendal, United Kingdom; Delegate, International Commission for Mountain Emergency Medicine, ICAR MedCom, Zurich, Switzerland; Lecturer, International Diploma of Expedition and Wilderness Medicine, Royal College of Physicians and Surgeons, Glasgow, United Kingdom
Environmental medicine

Robin D. Hamilton MB BS, DM (Dist), FRCOphth
Consultant Ophthalmic Surgeon, Medical Retina Service, Moorfields Eye Hospital NHS Trust; Honorary Senior Lecturer, NIHR Biomedical Research Centre, Institute of Ophthalmology, London, United Kingdom
Ophthalmology

Matthew Harmer BSc(Hons), BM, PhD, MRPCH
Consultant in Paediatric Nephrology, Paediatric Nephrology, Southampton Children's Hospital, Southampton, United Kingdom
Young adult medicine

Adam Harper MBBS, FRCP, MSc
Consultant Geriatrician, Department of Geriatric Medicine, University Hospitals Sussex NHS Trust, Brighton, United Kingdom
Medicine of the older adult

Laura Kay Harrington MBChB, MRCP (AIM), DipObsMed (RCP), PG Cert (Education)
Acute Internal Medicine Registrar, Acute Internal Medicine, Barts Health NHS Trust, London, United Kingdom
Acute medicine

Catherine Anne Harwood BA, MA, MBBS, PhD, FRCP
Professor of Dermatology, Dermatology, Barts Health NHS Trust and Queen Mary University of London, London, United Kingdom
Dermatology

Peter David Hodkinson MBChB, PhD, MRCP, DAvMed, FRAeS, FAsMA
Consultant in Aviation and Space Medicine, Head of Aerospace Medicine (Clinical and Education), Centre for Human and Applied Physiological Sciences, King's College London, London, United Kingdom
Environmental medicine

Richard Ian Gregory Holt MA, MB BChir, PhD, FRCP
Professor in Diabetes and Endocrinology, Human Development and Health, Faculty of Medicine, University of Southampton; Honorary Consultant Physician, Diabetes and Endocrinology, University Hospital Southampton NHS Foundation Trust, Southampton, United Kingdom
Diabetes mellitus

William L. Irving MB BChir, MRCP, PhD, FRCPath
Professor, School of Life Sciences, University of Nottingham, Nottingham, United Kingdom
Infectious disease

Noor Jawad MBBS, BSc, PhD, FRCP, MRCP (Gastro)
Consultant Gastroenterologist, Barts Health NHS Trust, London, United Kingdom
Gastroenterology

June Jones PhD, MSc
Senior Lecturer in Professional Ethics and Law, Medical School, Edge Hill University, Ormskirk, United Kingdom
Medical ethics

Vikas Kapil MA (Cantab), MBBS, PhD, FRCP
Clinical Senior Lecturer in Clinical Pharmacology and Therapeutics, William Harvey Research Institute, Faculty of Medicine and Dentistry, Queen Mary University London; Consultant Physician in Cardiovascular Medicine, Barts BP Centre of Excellence, Barts Heart Centre, Barts Health NHS Trust, London, United Kingdom
Therapeutics and prescribing
Hypertension

David Peter Kelsell BSc, PhD
Professor of Human Molecular Genetics, Blizard Institute, Faculty of Medicine and Dentistry, Queen Mary University of London, London, United Kingdom
Human genetics

Parveen J. Kumar DBE, BSc, MD, DM(hc), DEd, DSc, FRCP, FRCPG, FRCPI, FRCPath
Emerita Professor of Medicine and Education, Barts and The London School of Medicine and Dentistry, Queen Mary University of London; Honorary Consultant Physician and Gastroenterologist, Barts Health NHS Trust and Homerton University Hospital NHS Foundation Trust, London, United Kingdom
Global and planetary health

Andrew W. Leitch MB BS, FRCA, FFICM
Consultant in Intensive Care Medicine and Anaesthesia, Adult Critical Care Unit, Royal London Hospital—Barts Health NHS Trust, London, United Kingdom
Sepsis and the treatment of bacterial infection
Critical care medicine

Miles J. Levy MD, FRCP, MSc
Professor, Endocrinology, University Hospitals of Leicester, Leicester, United Kingdom
Endocrinology

Peter K. MacCallum MD, FRCP, FRCPath
Professor, Wolfson Institute of Population Health, Queen Mary University of London; Consultant, Haematology, Barts Health NHS Trust, London, United Kingdom
Venous thromboembolic disease

Ritwika Mallik MBBS, DNB, MRCP (London), SCOPE Certified
Clinical Research Fellow, Centre for Obesity Research, University College London, London, United Kingdom
Obesity

Peter Martin PhD, GradStat
Associate Professor, Department of Primary Care and Population Health, University College London, London, United Kingdom
Evidence-based practice

Beenish Masood MBBS, FCPS, MRCP, MSc
Clinical Teaching Fellow, Diabetes & Endocrinology, Royal Free NHS Foundation Trust; Clinical Research Fellow Diabetes and Metabolism, Diabetes, Barts Health NHS Trust, London, United Kingdom; Locum Specialty Registrar, Endocrinology, Mid and South Essex NHS Foundation Trust, Basildon, United Kingdom
Obesity

Kieran McCafferty MB BChir, MRCP, MD(Res)
Consultant Nephrologist, Barts Health NHS Trust; Honorary Senior Lecturer, Translational Medicine and Therapeutics, Queen Mary University London, London, United Kingdom
Water balance, fluids, and electrolytes

Shameer J. Mehta MBBS, BSc, MRCP, AFHEA, MD(Res)
Consultant Gastroenterologist, Department of Gastroenterology, Barts Health NHS Trust; Honorary Senior Lecturer, Centre for Immunobiology, Queen Mary University of London, London, United Kingdom
Nutrition

Mark Melzer FRCP, FRCPath
Consultant, Infection, Barts Health, London, United Kingdom
Sepsis and the treatment of bacterial infection

Catherine Moffat BSc(Hons) Physiotherapy
Specialist Physiotherapist, Palliative Care, Cambridge University Hospitals NHS Foundation Trust, Cambridge, United Kingdom
Palliative care and symptom control

Nicholas Robert Murch MB BCh, FRCP
Consultant Physician, Acute Medicine, Royal Free Hospital; Consultant Physician, Acute Admissions Unit, Cleveland Clinic London; Honorary Clinical Associate President, University College London Medical School, London, United Kingdom
Acute medicine

Michael F. Murphy MD, FRCP, FRCPath, FFPath
Consultant Haematologist, NHS Blood & Transplant; Consultant Haematologist, Department of Haematology, Oxford University Hospitals NHS Foundation Trust; Professor of Transfusion Medicine, Nuffield Department of Clinical and Laboratory Sciences, University of Oxford, Oxford, United Kingdom
Haematology

Arvind Nagra MBChB, MRCPCH (UK)
Consultant Paediatric Nephrologist, Paediatric Nephro-Urology Department, Southampton Children's Hospital, Southampton, United Kingdom
Young adult medicine

Melanie Nana MD, MRCP
Clinical Research Fellow, Department of Obstetric Medicine, Guy's and St Thomas' NHS Foundation Trust, London, United Kingdom
Obstetric medicine

Catherine Nelson-Piercy MD, FRCP, FRCOG
Professor, Department of Obstetric Medicine, Guy's and St Thomas' Foundation Trust; Consultant Obstetric Physician, Women's Health, Guy's and St Thomas' Foundation Trust, London, United Kingdom
Obstetric medicine

Lorraine M. Noble PhD, MPhil, Dip Clin Psychol, C Psychol, FBPsS, PFHEA
Professor, UCL Medical School, University College London, London, United Kingdom
Clinical communication

Alastair O'Brien MBBS, BSc, PhD
Professor, Institute of Liver Disease and Digestive Health, University College London, London, United Kingdom
Hepatology

Margaret O'Connor FRCPI, FRCP (UK)
Adjunct Associate Clinical Professor, Department of Ageing & Therapeutics, University of Limerick Hospital Group; Intern Network Programme Director, MidWest Intern Network, University of Limerick, Limerick, Ireland; BIASP Executive Lead for Ireland, British & Irish Association of Stroke Physicians, Ireland
Stroke medicine

Peter M. Odor BM BCh, PhD, MA, MSc, FRCA
Consultant, Department of Anaesthesia and Perioperative Medicine, University College London Hospital; Honorary Associate Professor, University College London, London, United Kingdom
Anaesthesia and perioperative medicine

Patrick O'Dwyer MCH, FRCSI, FRCS(Glas)
Emeritus Professor, School of Medicine, Dentistry and Medicine, University of Glasgow, Glasgow, United Kingdom
Radiology

Barbara Lachana Onen MBChB, MSc, MRCP, DTM&H
Acute & General Internal Medicine Consultant, Department of Acute Medicine, Royal London Hospital, Barts Health NHS Trust, London, United Kingdom
Acute medicine

Chloe M. Orkin MBBCh, MSc Infections Disease, LSHTM, FRCP
Professor of Infection and Inequities, SHARE Research Collaborative for Health Equity, Department of Immunobiology, Blizard Institute, Faculty of Medicine & Dentistry, Queen Mary University of London, London, United Kingdom
Human immunodeficiency virus

K. John Pasi MBE, MBChB, PhD, FRCP, FRCPath, FRCPCH, FMedSci
Honorary Professor of Haemostasis and Thrombosis, Blizard Institute, Barts and the London School of Medicine and Dentistry, London, United Kingdom
Haematology
Venous thromboembolic disease

Hitesh C. Patel MBBS, PhD, MRCP, FRACP, CCDS
Deputy Director of Cardiology, Head Heart Failure and Cardiac Transplant, Cardiology, Alfred Hospital; Associate Professor, Monash University, Melbourne, Victoria, Australia
Cardiology

Piers Nicholas Plowman MA, MD, FRCR, FRCP
Consultant Oncologist, Department of Clinical Oncology, St Bartholomew's Hospital, London, United Kingdom
Malignant disease

Andrew A. Plumb BM, BCh, MRCP, PhD, FRCR (2010)
Consultant Radiologist, Department of Imaging, University College London Hospitals NHS Foundation Trust, London, United Kingdom
Radiology

Sotirios Posporelis MD, MRCPsych
Consultant Liaison Neuropsychiatrist, Psychological Medicine at King's College Hospital, South London and Maudsley NHS Foundation Trust; Honorary Senior Clinical Lecturer, Neuropsychiatry, Institute of Psychiatry, Psychology and Neuroscience, London, United Kingdom
General hospital psychiatry

Joanna Preston MBBS, MRCP(Geris), MSc
Consultant Geriatrician, Senior Health, St George's NHS University Hospitals Foundation Trust, London, United Kingdom
Medicine of the older adult

Mark Pucci MBChB, MRCP
Consultant, Department of Clinical Toxicology, Sandwell and West Birmingham NHS Trust, Birmingham, United Kingdom
Toxicology

Mohammed Abbas Anisur Rahman PhD, FRCP
Professor of Rheumatology, Division of Medicine, University College London, Watford, United Kingdom
Rheumatology

Rosalind Ann Raine BSc, MBBS, MSc, PhD, FFPH, FMedSci
Professor, Department of Primary Care and Population Health, UCL, London, United Kingdom
Evidence-based practice

Neil Rajan MBBS, PhD
Professor, Translational and Clinical Research Institute, Newcastle University; Honorary Consultant Dermatologist, Department of Dermatology, Royal Victoria Infirmary, Newcastle upon Tyne, United Kingdom
Human genetics

Radha Ramachandran MBBS, FRCP, MSc, FRCPath, PhD
Consultant, Adult Inherited Metabolic Diseases, Guy's and St Thomas' Hospitals NHS Foundation Trust; Consultant, Chemical Pathology and Metabolic Medicine, Guy's and St Thomas' Hospitals NHS Foundation Trust; Honorary Clinical Senior Lecturer, School of Life Course and Population Sciences, King's College London, London, United Kingdom
Metabolic disease

David Randall PhD, MRCP
Honorary Clinical Senior Lecturer, Barts and the London School of Medicine and Dentistry, QMUL; Consultant Physician and Nephrologist, Barts Health NHS Trust, London, United Kingdom
The art of being a doctor

Robin Ray MB BS, MA(Oxon), FRCP, FHFA, PhD
Cardiology Consultant and Honorary Senior Lecturer, Cardiology Department, St George's University Hospitals NHS Foundation; Honorary Senior Lecturer, Molecular and Clinical Sciences Research Institute, St George's University, London, United Kingdom
Cardiology

Narendra L. Reddy MD, FRCP, CCT (Oxf)
Consultant Endocrinologist, Diabetes & Endocrinology, University Hospitals of Leicester; Honorary Senior Lecturer, University of Leicester, Leicester, United Kingdom
Endocrinology

Elliott Ridgeon BA, BMBCh, MSc, FRCA
Consultant, Department of Anaesthesia and Perioperative Medicine, Wexham Park Hospital, Slough, United Kingdom; Honorary Tutor in Physiology and Pharmacology, Magdalen College, University of Oxford, Oxford, United Kingdom
Anaesthesia and perioperative medicine

Noemi Roy MBChB, DPhil
Honorary Senior Clinical Lecturer, Weatherall Institute of Molecular Medicine, University of Oxford; Consultant Haematologist, Department of Haematology, Oxford University Hospitals NHS Foundation Trust, Oxford, United Kingdom
Haematology

Hamoun Rozati MBBS, PGCert(HE), MSc, MD(Res), MRCP(UK), FRCR
Consultant Clinical Oncologist, Oncology, Barts Cancer Centre, London, United Kingdom
Malignant disease

Prina Ruparelia MBChB, MD, FRCP
Respiratory Consultant, Department of Respiratory Medicine, St Bartholomew's Hospital, London, United Kingdom
Respiratory medicine

Richella Ryan MB BCh, BAO, PhD, MRCPI
Consultant, Palliative Medicine, Arthur Rank Hospice Charity, Cambridge, United Kingdom
Palliative care and symptom control

Anshoo Sahota BSc, MBBS, FRCP
Consultant Dermatologist, Department of Dermatology, Barts Health NHS Trust, London, United Kingdom
Dermatology

Emily Scott BMBS, MRCP, FRCR
Specialty Doctor, Department of Oncology, Mount Vernon Cancer Centre, London, United Kingdom
Malignant disease

Leighton John Seal BSc, MBBS, PhD
Professor and Consultant in Endocrinology, Adult Gender Identity Clinic, Tavistock and Portman NHS Foundation Trust; Professor and Consultant in Diabetes and Endocrinology, Thomas Addison Unit, St George's University Hospitals NHS Foundation Trust; Professor, Institute of Biomedical Education, St George's Hospital Medical School University of London, London, United Kingdom
Sex and gender in medicine

Babulal Sethia FRCS, FRCP (Hon)
Past President, Royal Society of Medicine; Royal Society of Medicine, London, United Kingdom
Global and planetary health

Kay Seymour MB BChir, MA (Cantab), DOHNS, FRCS (ORL-HNS)
Consultant ENT Surgeon, Department of ENT, Royal London Hospital, Barts Health NHS Trust, London, United Kingdom
Ear, nose, and throat

Apeksha M. Shah MA (Cantab), FRCP
Consultant Neurologist, Neurology Department, Barts Health NHS Trust; Honorary Research Fellow, Centre for Preventive Neurology, Wolfson Institute of Population Health, Queen Mary University of London, London, United Kingdom
Neurology

Jonathan Shamash MBChB, MD, FRCP
Consultant, Department of Medical Oncology, St Bartholomew's Hospital, London, United Kingdom
Malignant disease

Stephen J. Shepherd MBBS, FRCP, FRCA, FFICM, EDIC, FEACVI
Consultant in Anaesthesia & Intensive Care Medicine, Department of Perioperative Medicine, St Bartholomew's Hospital; Consultant in Anaesthesia, Department of Perioperative Medicine, Royal London Hospital; Consultant in Transfer Medicine, Adult Critical Care Transfer Service, London, United Kingdom
Critical care medicine

Sridevi Siramahalingappa MBBS, FRCPsych, FRANZCP
Consultant Psychiatrist, Psychiatry, South London and Maudsley NHS Foundation Trust, London, United Kingdom
General hospital psychiatry

Suthesh Sivapalaratnam MD, PhD, FRCP, FRCPath
Clinical Senior Lecturer, PHURI, QMUL; Honorary Consultant, Clinical Haematology, Royal London Hospital, London, United Kingdom
Venous thromboembolic disease

Charlotte I. Skinner BM BCh, MD (Res)
Specialty Doctor, Department of Gastroenterology, Imperial Healthcare NHS Trust, London, United Kingdom
Gastroenterology

Sophie Michelle Strong MRCOG, MBBS, BSc(Hons)
Consultant Gynaecologist, Department of Women's Health, Barts Health NHS Trust; Obstetrics & Gynaecology, Women's Health, London, United Kingdom
Sex and gender in medicine

Ganesh Subramanian MBBS, FRCP, MEd (Med Ed), MRes (Med)
Consultant in Stroke Medicine, Department of Stroke Medicine, Nottingham University Hospitals NHS Trust, Newark, United Kingdom; Clinical Associate Professor, University of Nottingham, Nottingham, United Kingdom
Stroke medicine

Rebecca Swan MBChB, BSc (MedSci)
Specialty Doctor, Western General Hospital, Edinburgh, United Kingdom
Sex and gender in medicine

James Emmanuel Thaventhiran BSc, MBBS, PhD
MRC Investigator, MRC Toxicology Unit, University of Cambridge; Honorary Consultant Clinical Immunologist, Clinical Immunology, Cambridge University Hospitals, Cambridge, United Kingdom
Immunity

Yincent Tse MBChB, FRCPCH
Consultant Paediatric Nephrologist, Department of Paediatric Nephrology, Great North Children's Hospital; Honorary Senior Clinical Lecturer, Newcastle University, Newcastle upon Tyne, United Kingdom
Young adult medicine

Christopher Antony Wadsworth MBBS, PhD, FRCP
Clinical Director, Specialist Medicine, Department of Gastroenterology, Imperial College Healthcare NHS Trust, London, United Kingdom
Biliary tract and pancreatic disease

David W. Wareham MB BS, MSc, PhD, MRCP, FRCPath
Consultant Microbiologist, School of Medicine and Dentistry, Queen Mary University London, London, United Kingdom
Sepsis and the treatment of bacterial infection

Elinor Warner MBBS, MA(Oxon), FRCS (ENT)
Specialty Doctor, Department of ENT, University College London Hospitals, London, United Kingdom
Ear, nose, and throat

Veronica White BSc, MBBS, MSc, MD, FRCP
Respiratory Consultant, Department of Respiratory Medicine, Barts Health NHS Trust, London, United Kingdom
Respiratory medicine

Anthony S. Wierzbicki DM, DPhil, FRCPath
Consultant, Department of Metabolic Medicine, Guy's and St Thomas' Hospitals; Professor, Cardiometabolic Disease, King's College London, London, United Kingdom
Lipid and metabolic disorders

Kate Wiles PhD, MRCP
Honorary Senior Lecturer, Wolfson Institute of Population Health, Queen Mary University of London; Consultant Obstetric Physician, Department of Women's Health, Barts Health NHS Trust, London, United Kingdom
The art of being a doctor

Iain Stewart Ayrton Wilkinson MBBS, FRCP
Consultant Geriatrician, Medicine for the Elderly, Surrey and Sussex Healthcare NHS Trust, Redhill, United Kingdom; Honorary Senior Lecturer, Brighton and Sussex Medical School, Brighton, United Kingdom
Medicine of the older adult

Janet D. Wilson MBChB, FRCP
Consultant in Genitourinary Medicine and HIV, Genitourinary Medicine and HIV, Leeds Teaching Hospitals Trust, Leeds, United Kingdom
Sexually transmitted infections

Sarah Elizabeth Winchester MB BChir, MSc, MA (Cantab), FFPH
Consultant in Public Health Medicine, Public Health Team, Buckinghamshire Council, Aylesbury, United Kingdom
Public health

M. Magdi Yaqoob MD, FRCP
Professor, Renal Medicine and Transplantation, Barts Health NHS Trust and Queen Mary University of London, London, United Kingdom
Kidney and urinary tract disease
Water balance, fluids, and electrolytes

Anjali Zalin MBBS, BMedSci, FRCP, SCE (Diabetes and Endocrinology), PhD
Consultant Physician, Obesity Medicine/ Tier 3, Barts Health NHS Trust/ Bedfordshire Hospitals NHS Foundation Trust, London/Bedfordshire, United Kingdom
Obesity

Dominik Zenner MA, MSc, FHEA, MD, MRCGP, FFPH
Clinical Reader in Infectious Disease Epidemiology, Wolfson Institute of Population Health, Queen Mary University, London, United Kingdom
Global and planetary health

Parjam Zolfaghari MBBS, PhD, FRCA, FFICM, SFHEA
Consultant, Adult Critical Care Unit, The Royal London Hospital, Barts Health NHS Trust; Senior Lecturer, William Harvey Research Institute, Queen Mary University London, London, United Kingdom
Critical care medicine

International Advisory Board

The editors would like to acknowledge and thank all previous edition members of our International Advisory Board for their contributions.

AUSTRALIA

Clinical Associate Professor Peter Katelaris MB BS, MD (Lond), FRACP, FRCP (Lond), FGESA, AGAF
Director, Gastroenterology Department, Concord Hospital, University of Sydney, Sydney

BANGLADESH

Dr. A.B.M. Abdullah
Emeritus Professor, Bangabandhu Sheikh Mujib Medical University, Dhaka

CANADA

Dr. Bradley Little BSc, MD, CCFP
Family Physician, Spectrum Health; Clinical Instructor, University of British Columbia—Department of Family Practice; Medical Staff, Providence Health Care—Department of Family Practice, Vancouver, British Columbia

INDIA

Professor Richa Dewan
Ex Director, Professor, and Head, Department of Medicine, Maulana Azad Medical College, New Delhi

MALAYSIA

Dr. Anita Devi Krishnan Thantry
Manipal University College Malaysia (MUCM), Jalan Batu Hampar, Bukit Baru, Melaka

PAKISTAN

Prof. Dr. M. Akbar Chaudhry
Dean, Faculty of Medical Sciences, Azra Naheed Medical College, Superior University Lahore-Pakistan, Department of Medicine, Azra Naheed Medical College, Lahore

Dr. Nazish Imran
Professor of Child & Adolescent Psychiatry, King Edward Medical University, Lahore

Dr. Somia Iqtadar
Associate Professor Medicine, King Edward Medical University, Lahore; Secretary General, Pakistan Society of Internal Medicine and Dengue Expert Advisory Group, Punjab

SRI LANKA

Professor Christeine Ariaranee Gnanathasan
Senior Professor in Medicine and Consultant Physician, University of Colombo, Colombo

Professor Senaka Rajapakse
Senior Professor & Chair of Medicine Director, Postgraduate Institute of Medicine University of Colombo, Colombo

Professor Udaya K. Ranawaka
Department of Medicine, University of Kelaniya, Ragama

UNITED STATES OF AMERICA

Dr. Ahmed Abd Almalik Daak
Clinical Development-Immunohematology, Sanofi, Cambridge, Massachusetts

This page intentionally left blank

Acknowledgments for the Eleventh Edition

The editors would like to thank Parveen Kumar and Mike Clark for their innovation, inspiration, and longstanding commitment to medical education over nine previous editions of *Clinical Medicine*. We remain grateful to them for their ongoing guidance and wisdom. As training doctors, we were educated by them, and it is now our privilege to be able to contribute to the education of others. It is upon their shoulders that we stand.

This book would not exist without the chapter authors who have worked hard to ensure that the content is up to date and reflective of modern practice. We are very grateful to them for their efforts and expertise.

As well as welcoming new contributors to this edition, we would like to thank those who have stepped down, especially those who have had a long association with the book. The text of this edition continues to reflect the contributions made by them:

- Ajit Abraham (Surgery)
- Annie Cushing (Ethical practice and clinical communication)
- Damiete Harry (Male health)
- Deborah Bowman (Ethical practice and clinical communication)
- Edward Mullins (Female health)
- Francis Chinegwundoh (Male health)
- Francis Vaz (Ear, nose, and throat)
- Gavin Barlow (Infectious diseases)
- Helen Barratt (Public health)
- Jonathan Sive (Haematological oncology)
- Julius Bourke (Liaison psychiatry)
- Klaartje Bel Kok (Nutrition)
- Lesley Regan (Female health)
- Marinos Elia (Nutrition)
- Mark Peakman (Immunity)
- Michael Clark (Environmental medicine)
- Munir Pirmohamed (Prescribing, therapeutics, and toxicology)
- Nicholas Bunce (Cardiology)
- Nishchay Mehta (Ear, nose, and throat)
- Paul Jarman (Neurology)
- Peter Moss (Infectious diseases)
- Rachel Lewis (Oncology)
- Sara Booth (Palliative medicine)
- Sarah Wakelin (Dermatology)
- Silvie Cooper (Public health)
- Susan Lanham-New (Nutrition)
- Umesh Vivekananada (Neurology)

We would like to acknowledge the editorial contribution of Adam Feather and Mona Waterhouse to the 10th edition of *Kumar & Clark's Clinical Medicine*. Thank you for moving the book forward whilst keeping it relevant for modern medical practice. We wish you well in your new endeavours.

The team at Elsevier have worked extremely hard to bring this project to fruition. We are particularly grateful to Alex Mortimer, Susan Callahan, and Daniel Fitzgerald.

Finally, to our families including Abigail, Robbie, Thomas, Emily, Hayden, Nick, George, Fred, Meg, Ophelia, Amy, Oliver, and Timothy. This edition would not have been possible without your support. Thank you for allowing Kumar and Clark to live alongside you for a while.

Rapid MLA Conditions Contents Table

Conditions	Chapter	Page Number
Acid-base abnormality	13: Critical care medicine	307–308, 308b
Acne vulgaris	45: Dermatology	1570f
Acoustic neuroma	28: Neurology	946
Acute bronchitis	20: Respiratory medicine	545
Acute cholangitis	24: Biliary tract and pancreatic disease	741
Acute coronary syndromes	17: Cardiology	444–452
Acute glaucoma	31: Ophthalmology	1047b
Acute kidney injury	13: Critical care medicine	327–328, 228b, 228–231, 835–837
Acute pancreatitis	24: Biliary tract and pancreatic disease	749
Acute stress reaction	30: General hospital psychiatry	1023
Addison's disease	34: Endocrinology	1156, 1605
Adverse drug effects	7: Therapeutics and prescribing	114–115, 115b
Alcoholic hepatitis	23: Hepatology	724
Allergic disorder	6: Immunity	101b
Anaemia	38: Haematology	1277–1279
Anal fissure	22: Gastroenterology	667–668
Anaphylaxis	11: Acute medicine	198
Ankylosing spondylitis	32: Rheumatology	1074
Anxiety disorder: generalised	30: General hospital psychiatry	1021–1022
Anxiety disorder: post-traumatic stress disorder	30: General hospital psychiatry	1021–1022
Anxiety, phobias, OCD	30: General hospital psychiatry	1021–1022
Aortic aneurysm	17: Cardiology	492
Aortic dissection	18: Hypertension	509b
Aortic valve disease	17: Cardiology	456–462
Appendicitis	22: Gastroenterology	675
Arrhythmias	17: Cardiology	411–429
Arterial thrombosis	40: Venous thromboembolic disease	1364–1365
Arterial ulcers	45: Dermatology	1609
Asbestos-related lung disease	20: Respiratory medicine	578
Ascites	23: Hepatology	692, 226
Asthma	20: Respiratory medicine	540–545
Asthma COPD overlap syndrome	20: Respiratory medicine	540–551
Atopic dermatitis and eczema	45: Dermatology	1577
Atrophic vaginitis	47: Sex and gender in medicine	1649
Attention deficit hyperactivity disorder	30: General hospital psychiatry	1040
Autism spectrum disorder	30: General hospital psychiatry	1040b
Bacterial vaginosis	43: Sexually transmitted infections	1529, 1529f
Basal cell carcinoma	45: Dermatology	1618–1619, 1619f
Bell's palsy	28: Neurology	899
Benign eyelid disorders	31: Ophthalmology	1050–1060
Benign paroxysmal positional vertigo	21: Ear, nose, and throat	590, 900
Benign prostatic hyperplasia	47: Sex and gender in medicine	1663f, 1664b, b0135
Bipolar affective disorder	30: General hospital psychiatry	1019–1021, 1021b
Bladder cancer	41: Malignant disease	1396b
Blepharitis	31: Ophthalmology	1051–1052, 1053f
Brain abscess	41: Malignant disease	944, 1412
Brain metastases	41: Malignant disease	1412
Breast abscess/mastitis	47: Sex and gender in medicine	1644–1645
Breast cancer	41: Malignant disease	1399–1402, 1399f
Breast cysts	47: Sex and gender in medicine	1642b, 1641–1642

Rapid MLA Conditions Contents Table—cont'd

Conditions	Chapter	Page Number
Bronchiectasis	20: Respiratory medicine	565–566
Bronchiolitis	20: Respiratory medicine	575
Bursitis	32: Rheumatology	1077
Candidiasis	43: Sexually transmitted infections	1529,
Cardiac arrest	11: Acute medicine	192–198
Cardiac failure	11: Acute medicine	192–198, 193f
Cataracts	31: Ophthalmology	1055, 1056b, 1056f
Cellulitis	47: Sex and gender in medicine	1570f, 1589, 256b–257b, 275–276, 275f
Central retinal arterial occlusion	31: Ophthalmology	1059f, 1059
Cerebral palsy and hypoxic-ischaemic encephalopathy	28: Neurology	924, 960
Cervical cancer	47: Sex and gender in medicine	1396b, 1410f, 1410
Cervical screening (HPV)	47: Sex and gender in medicine	1396b, 1410f, 1410
Chlamydia	43: Sexually transmitted infections	1050–1060
Cholecystitis	24: Biliary tract and pancreatic disease	744
Chronic fatigue syndrome	30: General hospital psychiatry	1029, 1458
Chronic glaucoma	31: Ophthalmology	1057f, 1589f, 1055–1057
Chronic kidney disease	26: Kidney and urinary tract disease	837–851
Chronic obstructive pulmonary disease	20: Respiratory medicine	545–549, 327–328
Cirrhosis	23: Hepatology	708–722
Coeliac disease	22: Gastroenterology	645, 732
Colorectal tumours	22: Gastroenterology	669–674
Compartment syndrome	32: Rheumatology	1077
Conjunctivitis	31: Ophthalmology	1052–1054, 1054b
Constipation	22: Gastroenterology	615–617, 992
Contact dermatitis	45: Dermatology	1582
COVID-19	13: Critical care medicine	327
Croup	20: Respiratory medicine	539, 1451
Crystal arthropathy	32: Rheumatology	1103–1106, 1103f, 256b–257b
Cushing's syndrome	34: Endocrinology	1054b, 1147, 1153–1154, 1156
Cutaneous fungal infection	45: Dermatology	1595–1598
Cutaneous warts	45: Dermatology	1593
Cystic fibrosis	20: Respiratory medicine	566–567, 730
Deep vein thrombosis	11: Acute medicine	264–272
Dehydration	11: Acute medicine	258f
Delirium	11: Acute medicine	250–253
Dementias	28: Neurology	953b, 952–959
Depression	30: General hospital psychiatry	1017–1019
Diabetes in pregnancy (gestational and pre-existing)	48: Obstetric medicine	1688–1689
Diabetes insipidus	34: Endocrinology	1191, 1691
Diabetes mellitus type 1 and 2	35: Diabetes mellitus	1200–1239, 1201b, b0015, 1202f, 1604f, 1604–1605, 966–967
Diabetic eye disease	35: Diabetes mellitus	1225f
Diabetic ketoacidosis	11: Acute medicine	258b, 257–258, 260b
Diabetic nephropathy	26: Kidney and urinary tract disease	807–808, 808b, 1227–1229, s0815
Diabetic neuropathy	35: Diabetes mellitus	1229–1231
Disease prevention/screening	42: Infectious disease	1438–1440
Disseminated intravascular coagulation	38: Haematology	1326b–1327b, 1326f, 1326
Diverticular disease	22: Gastroenterology	666–667, f0250
Down's syndrome	5: Human genetics	54–56

Rapid MLA Conditions Contents Table—cont'd

Conditions	Chapter	Page Number
Eating disorders	30: General hospital psychiatry	1633, 1036–1039
Ectopic pregnancy	22: Gastroenterology	676
Encephalitis	28: Neurology	941
Endometrial cancer	47: Sex and gender in medicine	1410–1411
Endometriosis	43: Sexually transmitted infections	1528
Epididymitis and orchitis	47: Sex and gender in medicine	1667–1668
Epiglottitis	42: Infectious disease	1472
Epilepsy	28: Neurology	1687, 1027, 914, 921–927
Epistaxis	21: Ear, nose, and throat	b0015, 593–595,
Essential or secondary hypertension	26: Kidney and urinary tract disease	819
Essential tremor	28: Neurology	932
Extradural haemorrhage	11: Acute medicine	245, 245f
Febrile convulsion	28: Neurology	924
Fibroadenoma	47: Sex and gender in medicine	1641b, 1641
Fibroids	47: Sex and gender in medicine	1657–1658, 1658f
Fibromyalgia	32: Rheumatology	808f, 1079
Fibrotic lung disease	20: Respiratory medicine	536
Folliculitis	45: Dermatology	1589–1590
Gallstones and biliary colic	24: Biliary tract and pancreatic disease	736–743
Gangrene	11: Acute medicine	276
Gastric cancer	41: Malignant disease	1396b, 1402–1403
Gastrointestinal perforation	22: Gastroenterology	601b–602b, 601–678, b0015
Gastro-oesophageal reflux disease	22: Gastroenterology	621–625, 622b, 622f–624f
Gonorrhoea	43: Sexually transmitted infections	1527f, 121f, 1526
Haemochromatosis	23: Hepatology	720, 720f
Haemoglobinopathies	38: Haematology	1291–1292
Haemophilia	5: Human genetics	72–73, 1319
Haemorrhoids	22: Gastroenterology	667
Head lice	45: Dermatology	1599
Henoch-Schonlein purpura	26: Kidney and urinary tract disease	816, 1321
Hepatitis	23: Hepatology	692–704
Hernias	22: Gastroenterology	601b–602b, 601–678, b0015
Herpes simplex virus	42: Infectious disease	1445–1446, 1592f, 1592
Hiatus hernia	22: Gastroenterology	622b, 622
Hospital acquired infections	12: Sepsis and the treatment of bacterial infection	283
Human immunodeficiency virus	44: Human immunodeficiency virus	1537–1563, 1525
Human papilloma virus infection	42: Infectious disease	1462
Hypercalcaemia of malignancy	8: Radiology	145f, 456–462
Hyperlipidemia	19: Lipid and metabolic disorders	1050–1060, 1183
Hyperosmolar hyperglycaemic state	11: Acute medicine	258b, 261–262, 261b
Hyperparathyroidism	32: Rheumatology	1119
Hyperthermia and hypothermia	28: Neurology	973, 1050–1060
Hypoglycaemia	36: Metabolic disease	1244, 1196–1197, 263, 263b
Hypoparathyroidism	34: Endocrinology	1196
Hyposplenism/splenectomy	38: Haematology	1291f, 1307
Hypothyroidism	34: Endocrinology	1604–1605
Idiopathic arthritis	32: Rheumatology	1118
Impetigo	45: Dermatology	1589f, 1589
Infectious colitis	22: Gastroenterology	662
Infectious diarrhoea	22: Gastroenterology	642

Continued

Rapid MLA Conditions Contents Table—cont'd

Conditions	Chapter	Page Number
Infectious mononucleosis	23: Hepatology	1402
Infective endocarditis	29: Stroke medicine	998, 1106, 465–469
Infective keratitis	31: Ophthalmology	1688–1689, 1292f–1293f
Inflammatory bowel disease	22: Gastroenterology	724, 222–223
Influenza	20: Respiratory medicine	228–231, 1453b, 545
Intestinal ischaemia	22: Gastroenterology	838–839
Intestinal obstruction and ileus	22: Gastroenterology	677b, 677
Intussusception	22: Gastroenterology	668–669
Iritis	31: Ophthalmology	1047b, 1592f, 1058
Irritable bowel syndrome	22: Gastroenterology	1664b, b0135, 642, 1664b–1665b, b0145
Ischaemic heart disease	14: Anaesthesia and perioperative medicine	341–342
Kawasaki disease	32: Rheumatology	465–469
Leukaemia	39: Haematological oncology	590, 817
Liver failure	23: Hepatology	710
Lower respiratory tract infection	42: Infectious disease	1472, 1451–1454
Lung cancer	41: Malignant disease	p0210, 1396–1399
Lyme disease	45: Dermatology	1590–1591, 1482, 1057f
Lymphoma	39: Haematological oncology	198
Macular degeneration	31: Ophthalmology	1060f, 212f, 899
Malabsorption	22: Gastroenterology	1326b–1327b, 644–650
Malaria	42: Infectious disease	1497–1502
Malignant melanoma	41: Malignant disease	644–650, 1059f
Malnutrition	25: Nutrition	763–764, 1201b, b0015
Measles	42: Infectious disease	193f, 1449
Ménière's disease	21: Ear, nose, and throat	590, 900
Meningitis	28: Neurology	941, 1191–1193
Menopause	47: Sex and gender in medicine	1649
Metastatic disease	41: Malignant disease	1401
Migraine	28: Neurology	915–921, s1060, 242b, 242
Mitral valve disease	17: Cardiology	452f, 452–456
Motor neurone disease	28: Neurology	960b, 959–960
Multiorgan dysfunction syndrome	13: Critical care medicine	1056f, 899
Multiple myeloma	39: Haematological oncology	1056f, 1358–1362
Multiple sclerosis	28: Neurology	934–941
Mumps	42: Infectious disease	1457, 1458f
Muscular dystrophies	28: Neurology	974–975
Myasthenia gravis	48: Obstetric medicine	1688, 973
Myeloproliferative disorders	38: Haematology	1323
Myocardial infarction	14: Anaesthesia and perioperative medicine	348, 444
Myocarditis	42: Infectious disease	710, 480–481
Necrotising fasciitis	45: Dermatology	1574, 276
Nephrotic syndrome	26: Kidney and urinary tract disease	1029–1030
Non-accidental injury	30: General hospital psychiatry	1013–1014
Notifiable diseases	42: Infectious disease	1230b
Obesity	37: Obesity	1248–1269
Obesity and pregnancy	37: Obesity	1259
Obstructive sleep apnoea	20: Respiratory medicine	710, 622b, 1589f
Occupational lung disease	20: Respiratory medicine	677b, 577–578
Oesophageal cancer	41: Malignant disease	1402, 630
Optic neuritis	28: Neurology	894, 115b
Osteoarthritis	32: Rheumatology	1053f, 1605, 228b

Rapid MLA Conditions Contents Table—cont'd

Conditions	Chapter	Page Number
Osteomalacia	33: Bone disease	1134b, 1134–1135, 840
Osteomyelitis	33: Bone disease	1135
Osteoporosis	33: Bone disease	590, 840
Otitis media	21: Ear, nose, and throat	587f, 588–589
Ovarian cancer	47: Sex and gender in medicine	1396b, 915–921
Pancreatic cancer	24: Biliary tract and pancreatic disease	756–761, 1396–1399, 1405–1406
Pancytopenia	38: Haematology	1056b
Parkinson's disease	28: Neurology	904, 1027, 929
Pathological fracture	39: Haematological oncology	1361
Patient on anticoagulant therapy	17: Cardiology	446–447, 341–342, 1366f
Patient on antiplatelet therapy	29: Stroke medicine	995
Pelvic inflammatory disease	43: Sexually transmitted infections	1528
Peptic ulcer disease and gastritis	22: Gastroenterology	633
Perianal abscesses and fistulae	22: Gastroenterology	601b–602b, 601–678, b0015
Pericardial disease	17: Cardiology	487–490
Periorbital and orbital cellulitis	21: Ear, nose, and throat	593
Peripheral nerve injuries/palsies	28: Neurology	964, 964b
Peripheral vascular disease	17: Cardiology	490–494
Peritonitis	22: Gastroenterology	676
Personality disorder	30: General hospital psychiatry	1033–1036, 1035b–1036b, b0200, 950b
Pituitary tumours	34: Endocrinology	1688–1689, 1054b
Pneumonia	20: Respiratory medicine	222–223, 552–553, 1229–1231, 212f, 1472
Pneumothorax	8: Radiology	763–764, f0395, f0400, 121f, 216–218
Polycythaemia	38: Haematology	1609
Polymyalgia rheumatica	32: Rheumatology	1075, 1326
Postpartum haemorrhage	48: Obstetric medicine	1692
Pre-eclampsia, gestational hypertension	48: Obstetric medicine	1047b, 1683–1684, b0025, 509b
Prostate cancer	47: Sex and gender in medicine	1664b–1665b, 749, 193f, 592f, b0145, 1407
Psoriasis	45: Dermatology	1584–1586, 1657–1658
Pulmonary embolism	11: Acute medicine	216, 264–272
Pulmonary hypertension	40: Venous thromboembolic disease	1366, 477–480, 960b
Pyloric stenosis	22: Gastroenterology	601b–602b, 601–678, b0015
Radiculopathies	28: Neurology	963
Raised intracranial pressure	13: Critical care medicine	101b, 273–274, 329b
Reactive arthritis	32: Rheumatology	1102f, 1102, 256b–257b
Respiratory arrest	49: Medicine of the older adult	1704
Respiratory failure	13: Critical care medicine	318–325, 807–808
Retinal detachment	31: Ophthalmology	1059f, 1059
Rheumatoid arthritis	32: Rheumatology	1609
Rhinosinusitis	21: Ear, nose, and throat	1056b, 593
Right heart valve disease	17: Cardiology	462–463
Rubella	42: Infectious disease	1450
Sarcoidosis	45: Dermatology	1603–1604, 1472, 1118
Scabies	43: Sexually transmitted infections	1592, 1560, 1598f, 1598–1599, 622b, 1517
Schizophrenia	30: General hospital psychiatry	1001–1043
Scleritis	32: Rheumatology	1093
Self-harm	30: General hospital psychiatry	1013–1014
Sepsis	11: Acute medicine	230
Septic arthritis	32: Rheumatology	1091, 255–257, 256b–257b
Sickle cell disease	5: Human genetics	65

Continued

Rapid MLA Conditions Contents Table—cont'd

Conditions	Chapter	Page Number
Somatisation	30: General hospital psychiatry	1001–1043
Spinal cord compression	39: Haematological oncology	1328, 911f, 909, 950b, 950f, 273
Spinal cord injury	11: Acute medicine	188b
Spinal fracture	11: Acute medicine	188b
Squamous cell carcinoma	22: Gastroenterology	1075
Stroke	29: Stroke medicine	977–1000, 1525–1563, 1027
Subarachnoid haemorrhage	11: Acute medicine	243b, 243, 244f
Subdural haemorrhage	11: Acute medicine	245, 245f
Substance use disorder	30: General hospital psychiatry	1031b
Surgical site infection	14: Anaesthesia and perioperative medicine	350
Syphilis	43: Sexually transmitted infections	1533
Systemic lupus erythematosus	26: Kidney and urinary tract disease	815, 1607, 1107f, 842, 575, 941
Tension headache	28: Neurology	918, 243b, 242–243
Termination of pregnancy	47: Sex and gender in medicine	615–617
Testicular cancer	47: Sex and gender in medicine	1451
Testicular torsion	47: Sex and gender in medicine	692–704
Thyroid eye disease	34: Endocrinology	1168
Thyroid nodules	45: Dermatology	1604
Thyrotoxicosis	34: Endocrinology	1167, 1117b
Tonsillitis	21: Ear, nose, and throat	596–597, 597b
Toxic shock syndrome	42: Infectious disease	1470, 276
Transfusion reactions	38: Haematology	1310–1311
Transient ischaemic attacks	29: Stroke medicine	1201b, 1050–1060, 308b
Trichomonas vaginalis	43: Sexually transmitted infections	1530
Trigeminal neuralgia	28: Neurology	749, 920
Tuberculosis	20: Respiratory medicine	1664b, 553–557, 650, 1489
Unstable angina	17: Cardiology	440
Upper respiratory tract infection	42: Infectious disease	245
Urinary incontinence	47: Sex and gender in medicine	1657b, 998, 228b
Urinary tract calculi	26: Kidney and urinary tract disease	822–823
Urinary tract infection	26: Kidney and urinary tract disease	1021b, 1618–1619
Urticaria	45: Dermatology	1667–1668, 1057f
Uveitis	31: Ophthalmology	1057–1058
Varicella zoster	45: Dermatology	1592
Varicose veins	17: Cardiology	494
Vasovagal syncope	11: Acute medicine	238
Venous ulcers	45: Dermatology	244f, 1608–1609
Viral exanthema	45: Dermatology	1592
Viral gastroenteritis	42: Infectious disease	1472
Viral hepatitides	23: Hepatology	1033–1036
Visual field defects	28: Neurology	893, 1259
Vitamin B_{12} and/or folate deficiency	38: Haematology	1284, 1284b, 1134b
Volvulus	22: Gastroenterology	677
VTE in pregnancy and puerperium	48: Obstetric medicine	1054b, 1686
Wernicke's encephalopathy	28: Neurology	897
Whooping cough	9: Public health	115b, 161f

This page intentionally left blank

1 The art of being a doctor

Kate Wiles, John Booth, and David Randall

The clinical consultation 1
The initial interaction 2
Forming a rapport 2
Information-gathering 3
The structure of history-taking 3
The presenting symptom 3
History of the presenting symptom 4
Medical history 4
Drug history 4
Family history 4
Social history 4
Physical examination 4
Investigation 5
Interpretation of test results 7
When not to investigate 7
Diagnostic reasoning 7
Models of diagnosis 7
Clinical intuition 8
Bias and error 8
Levels of diagnostic depth 9
The role of watchful waiting 9
Diagnostic criteria 10
Overdiagnosis: too much medicine 10
Uncertainty 10
Communicating a diagnosis 11
Inequalities in medicine 11
Race and ethnicity 11
Sex and gender 12
People living with disabilities 12
Health inequality and intersectionality 12
Artificial intelligence and medicine 12

CORE SKILLS AND KNOWLEDGE

Being a doctor is a privilege. Patients open their lives to us, sharing troubles and triumphs; for this they should be met with kindness, judgement, and skill. It is the interactions with others that make medicine the most humane of sciences and the most scientific of humanities.

In the 21st century, professional roles and responsibilities are changing and defining what it means to be a doctor can be challenging. Healthcare is often provided by teams of people from different professional backgrounds, with the potential to leave individual doctors unsure of their role and patients without a constant name and face to guide them: a state of dis-'ease' for both.

Artificial intelligence, which utilizes vast datasets to produce algorithms for diagnosis and management, challenges the traditional role of a doctor. However, human beings have not yet been replaced and remain key within a diagnostic and treatment partnership. The definition of a clinical problem and selection of the best treatment require more than assimilation of information, technical skills, data interpretation, and a standardized framework for decision-making; they require compassion, empathy, trust, respect, and humour: attributes that are exclusively human, at least for now.

THE CLINICAL CONSULTATION

A clinical consultation is a complex process including at least two unique individuals, with potentially very different knowledge and expectations. In simplistic terms, the clinical consultation is an interaction that achieves a diagnosis. Diagnosis forms the linchpin of the therapeutic journey, shifting the direction of travel from experience in the past (medical history) to the future (treatment plans). However, in practice, a clinical consultation is multifaceted, and in addition to diagnosis, it is enriched by a transfer of knowledge between individuals and the development of a complex shared understanding. No textbook can give you a complete understanding of the experience of disease for the individual in front of you.

The complexity of the medical consultation depends upon whether priorities are aligned between clinician and patient and the difficulty of the medical problem (Fig. 1.1).

- When diagnosis and/or treatment options are clear and priorities are shared, the clinician and patient work in ***coordination***.
- Where diagnosis and treatment options are clear, but patient and clinician priorities do not align despite best efforts (e.g., a request for antibiotics to treat a viral illness), ***competition*** occurs until one set of priorities prevails.
- When diagnoses are evolving or unclear, or when treatment options have unknowable or unintended consequences, or when definitions of success vary, the consultation is one of ***negotiation*** and ***evolution***.

Such models offer a guide as to why some consultations are more difficult than others. Negotiation is harder than coordination, and uncertainty and change are difficult for patient and clinician alike. A consultation includes both a person and a disease, the first being uniquely complex, and the latter potentially complex. Recognition of differences, adaptation, and skill are needed.

2 The art of being a doctor

This chapter describes the information-gathering steps of the medical consultation according to separate phases: history-taking, investigation, and diagnosis. Fig. 1.2 shows how diagnosis is an iterative process: information is gathered, interpreted, and integrated to form a diagnosis, which is communicated and acted upon. A diagnosis may be refined when there is new information including the patient's response to treatment. This structure is traditional and convenient and means that key aspects of doctoring are not forgotten. However, clinicians should be wary of the limitations of such a structured interaction designed by doctors for doctors. Although, for the clinician, this structure is everyday discourse, the consultation 'map' is not visible to patients. Patients may only become aware of this structure when it has progressed without them: the moment or opportunity to provide important information has gone with the doctor's focus shifting to the next step in their process.

Signpost and share your consultation agenda with your patients and change it when it does not work for your patient. At the end of the consultation, both patient and doctor should feel that they gave a full and honest account of themselves. If you are not sure whether your patient feels this, ask them. A successful consultation is more than the traditional ritual surmised in textbooks such as this.

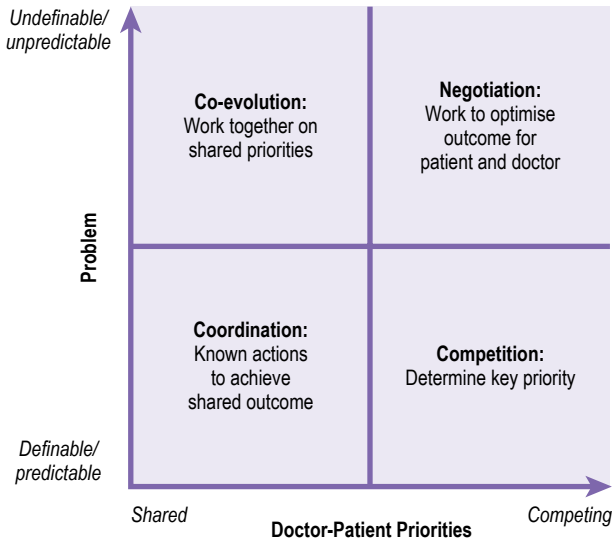


Fig. 1.1 The clinical consultation: variation according to doctor-patient priorities and the clinical problem.

THE INITIAL INTERACTION

Forming a rapport

First impressions matter. A clinical interaction that begins badly often leads to later problems. Patients need to be able to share intimate and worrying problems, which requires confidence and trust in clinicians. Though you may have consulted with many patients already that day, you may still have the privilege of being the first doctor that patient has met. Ask yourself, 'How would I wish my own doctor to behave?' Treat patients and their carers with respect. Ensure the physical environment supports interaction whilst also providing privacy, dignity, and comfort.

Introduce yourself (and any other health professionals present) and explain your role and the purpose of the encounter. A healthcare system can be busy and confusing, and it is important that patients understand where they are in their journey through it. The 'Hello, my name is…' campaign highlights the importance of introductions (Box 1.1). Appropriate expressions of empathy, support, and care can help establish rapport, allowing more successful subsequent discussion about the problem that led the patient to seek medical help. Non-medical discourse is the common language of both doctor and patient, often forgotten by those fluent in 'medicalese'. It may be safest to start first consultations with the assumption of no

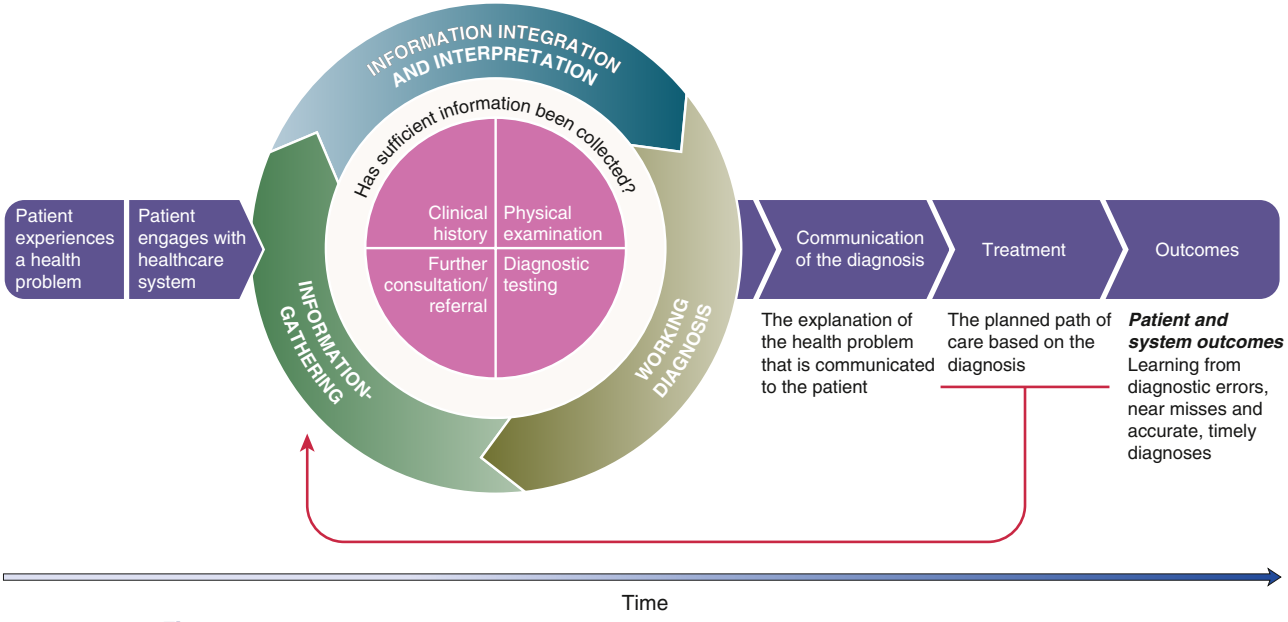


Fig. 1.2 A conceptualization of the diagnostic process. *(From National Academies of Sciences, Engineering, and Medicine.* Improving Diagnosis in Health Care. *Washington, DC: The National Academies Press; 2015. https://doi.org.10.17226/21794.)*

prior medical knowledge of disease (whilst remembering that the patient remains the expert on the impact of their disease on their life). The risk of impaired understanding contributing to dissatisfaction and poor engagement with care is far greater than the risk that lay descriptions of symptoms or disease processes will be considered offensive. Indeed, the expert patient may be very happy to help educate clinicians.

The worlds inhabited by doctors and patients have different cultures and different levels of power. The bridge between these worlds may be very wide, and you should be prepared to meet your patient beyond the middle to help them across.

Information-gathering

A full and detailed history from the patient is essential for accurate diagnosis. It is estimated that up to 80% of diagnoses can be made based on a careful history alone. In addition to information-gathering, eliciting the history forms a key part of the therapeutic relationship (**Box 1.2**). If it goes well, the patient feels listened to and has had an opportunity to convey their concerns. This builds mutual trust and respect.

The golden minute

At the start of the consultation the clinician should avoid interrupting the patient. An early interruption disrupts the patient's description of their symptoms and concerns, leaving the patient justifiably dissatisfied with the interaction and the doctor with less information about the presenting problem. The term '***the golden minute***' has been coined to encourage clinicians to allow patients to tell their story uninterrupted for at least 1 minute. This gives time for them to describe their symptoms as they have experienced them, including information that might be missed if the clinician jumps in too soon with closed, focused questions.

Allowing the patient freedom to relate their account of events also forms and reinforces the therapeutic relationship and increases both patient and clinician satisfaction. **Box 1.3** summarizes techniques that can be used to help patients share information.

A meeting of two experts

A medical consultation is a meeting of two experts: patient and doctor. A good consultation is based on mutual respect, rejecting a traditional paternalistic view of medicine that disempowers the patient ('doctor knows best'). The aim is shared understanding, where clinician and patient jointly grasp what is wrong, what impact it is having on the patient's life, what the patient expects from medical intervention, and which options might be best for investigating and treating the problem.

Box 1.1 The 'Hello, my name is…' campaign

In 2011, geriatrician Dr Kate Granger was diagnosed with metastatic sarcoma, from which she eventually died in 2016. Noticing that many of the healthcare staff treating her during her illness failed to introduce themselves, she created the 'Hello, my name is…' campaign. Describing why an initial introduction is so important, she said, 'I firmly believe it is not just about common courtesy, but it runs much deeper. Introductions are about making a human connection between one human being who is suffering and vulnerable, and another human being who wishes to help. They begin therapeutic relationships and can instantly build trust in difficult circumstances.'

The campaign encourages healthcare staff to greet and introduce themselves to patients and encourages the wearing of easy-to-read name badges like this one.

THE STRUCTURE OF HISTORY-TAKING

Over the last 150 years a formal structure for the recording of the patient's history has evolved. This has several subsections (**Box 1.4**). While facts may be recorded in this very stylized manner the patient will rarely, if ever, present them according to this structure. It is up to the interviewing clinician to assimilate and interpret information into a diagnostic narrative so that diagnostic reasoning and conclusions make sense.

The presenting symptom

What has prompted the patient to seek medical attention today? Although tempting to paraphrase a patient's presenting symptom using medical terms, it is preferable to record this in the patient's own words (e.g., 'vomiting blood' rather than 'haematemesis'). Care should be taken to understand exactly what the patient experienced. If a patient applies a medical term to their symptoms, it is important to ensure that term has the same meaning for both patient and clinician. For example, a patient reporting 'palpitations' may in fact be experiencing breathlessness or chest pain, and a patient with 'acute' symptoms may actually have severe symptoms that have been present for a long time.

Box 1.2 Good practice for using the medical history to build a therapeutic relationship

- Allow the patient to tell their story. Avoid jumping in prematurely with questions.
- Ask the patient specifically about their ideas, concerns, and expectations (ICE):
  - ***Ideas:*** What do they think might be going on? Have they done any reading about their symptoms? Do they know anyone who had similar symptoms?
  - ***Concerns:*** Are they feeling anxious or worried about their symptoms? What are they worried about? Are their symptoms causing any specific difficulties?
  - ***Expectations:*** What are they hoping for from this consultation?
- Try to develop an understanding of who the patient is as a person. Where, and with whom, do they live? What is their occupation? What things do they enjoy? Moments of human connection, such as a shared interest in a place or activity, can be powerful in building a relationship between clinician and patient.
- Try to convey empathy and concern. Reinforce to the patient that you are their advocate and will do your best to help them.

4 The art of being a doctor

Box 1.3 Strategies in history-taking

- ***Begin with open questions:*** 'Could you tell me more about the pain?' Do this before moving to closed questions that may help rule diagnoses in or out: 'Did the pain get worse after eating?'
- ***Active listening:*** Maintain eye contact, nod, acknowledge key points, respond to comments the patient might make that are humorous or sad.
- ***Respond to body language:*** Body language may reveal distress, anxiety, or embarrassment. Acknowledge this and address it.
- ***Empathize:*** Try to show the patient that you care about what they are going through. Put yourself in their shoes: how might they be feeling? Communicate this: 'Thank you for sharing this. It must be difficult not knowing what is going on.'
- ***Summarize:*** Run through what the patient has told you, to make sure nothing has been misunderstood or missed out.
- ***Signpost:*** Explain what you have covered, what happens next, and why.
- ***Use a common language:*** Avoid medical jargon or complicated vocabulary, unless it is clear that the patient is able to understand this.

Box 1.4 A structured approach to information-gathering in the medical history

- Presenting symptom: why has the patient sought medical advice?
- History of the presenting symptom: the information the patient wants to share, and the information needed to help make a diagnosis
- Medical and surgical history
- Drug history and allergies
- Family history
- Social history: information on the patient's present living arrangements and relevant risk factors

The patient's presenting symptom is the initial springboard towards diagnosis. For the experienced clinician a symptom will conjure a list of potential diagnoses, which will then be narrowed and refined (see 'Diagnostic reasoning' section later).

History of the presenting symptom

The aim is to provide a thorough account of symptoms. It is vital to listen closely and not to miss any clues that require follow-up or clarification. For each symptom, it should be possible to understand the following:

- ***Time course:*** When did the problem begin? Does it come and go? Is there anything that triggers it? Is there any variation e.g., time of day, eating, exertion? Has the patient ever had anything like this in the past? Symptom pattern is helpful for diagnosis (**Box 1.5**).
- ***Associated symptoms:*** What else has the patient noticed? Begin with an open question, then clarify the presence or absence of other symptoms that help with diagnosis.
- ***Pain:*** Severity, site, radiation, and character. How does the patient rate it on a scale from 1 to 10? How does the pain feel to them? Tight, dull, electric, and burning can be helpful when voiced by the patient. Is the pain getting worse, staying the same, or starting to improve?
- ***Responses:*** Have they sought medical advice or used medication that they have at home? Did it make things better or worse?

The 'Clinical skills' sections at the beginning of many of the chapters in this book present symptoms that are relevant to particular conditions within different medical specialties. However, most symptoms (at least at the start of the patient's story) do not tell you which chapter to go to. A review of systems may be valuable in eliciting additional symptoms that the patient may not have mentioned or thought relevant (**Box 1.6**).

Box 1.5 Using the time course of symptoms to suggest an underlying cause

Onset of symptoms	Possible causes
Immediate (seconds to minutes)	Vascular—thrombotic or embolic Anatomical—e.g., perforation of a viscus Electrical—e.g., dysrhythmias, seizures
Hours to days	Bacterial or viral infections Inflammatory and autoimmune diseases
Weeks to months	Malignant disease Inflammatory and autoimmune diseases Chronic infections, e.g., mycobacterial
Months to years	Degenerative conditions Fibrotic diseases

Medical history

This should include all significant conditions: hospital admissions, long-term conditions, important investigations, procedures, and therapeutic interventions (operations, endoscopies, chemo- or radiotherapy). If the patient has had surgery, an anaesthetic history should be sought and recorded. Five useful questions to ask about chronic conditions are listed in **Box 1.7**.

Drug history

Record all medications taken, including prescribed, over-the-counter, herbal, and traditional treatments. Record when each was started, the dose used, and adherence. Ask about side-effects. Medication error is a common cause of morbidity and mortality. Ask about and record ***drug allergies***, including the timing and nature of any reactions.

Family history

This is particularly relevant when the differential diagnosis includes genetic conditions or genetic risk. The medical history of first-degree relatives (parents, siblings, offspring) and consanguinity should be recorded. Diseases in the wider family are recorded in a pedigree.

Social history

The social history has two key purposes:

- Establishing whether ***environmental factors*** may be causative or exacerbating the patient's symptoms. Consider housing, occupation, smoking, alcohol, and illicit drugs. Where relevant, ask about travel history, animal exposure and insect bites, sexual history, and pregnancy history.
- Understanding more about ***the experience of the patient.*** Where and how do they live? How active are they? Are they limited in any daily activities, including mental health? Do they have informal or formal carers? If so, how often do carers attend and what do they do for the patient?

PHYSICAL EXAMINATION

Physical examination, including clinical observations, is used to confirm or refute the initial diagnosis/diagnoses suggested by the

Box 1.6 Review of systems

Conducting a brief but comprehensive 'review of systems' may be particularly useful when symptoms are non-specific and do not point to the diagnosis, for example, weight loss, tiredness, weakness, 'I just do not feel right'.

Respiratory
Cough: chronicity, nocturnal
Sputum: volume, frequency, consistency, colour, offensive taste or smell
Haemoptysis: volume, frequency, fresh or altered, clots
Shortness of breath: exercise tolerance, orthopnoea, wheeze
Chest pain: 'pleuritic' (with breathing)

Cardiovascular
Chest pain: character, precipitants, frequency, radiation
Shortness of breath: exercise tolerance, orthopnoea, paroxysmal nocturnal dyspnoea
Swelling: peripheral oedema, abdominal swelling (ascites)
Palpitations: precipitants, character (ask the patient to tap out what they feel)

Gastrointestinal (GI)
General: weight, appetite, oral intake
Upper GI:
Dyspepsia, early satiety
Nausea/vomiting: frequency, consistency (food/regurgitation)
Haematemesis: volume, frequency, consistency, fresh/altered, clots
Dysphagia: level (base of neck, mid or lower chest), consistency of food tolerated/not tolerated
Pain: odynophagia, upper abdominal pain
Lower GI:
Pain: lower abdominal pain, anal pain
Altered bowel habit: constipation, diarrhoea (volume, frequency, consistency, colour), blood per rectum, mucus per rectum
Hepatobiliary and pancreatic:
Jaundice, pain, pruritus, symptoms of encephalopathy, abdominal swelling (ascites)

Renal
Urine: frequency, volume, colour, offensive odour, dysuria
Haematuria: volume, frequency, colour, fresh/altered, clots
Bladder outflow tract obstruction: hesitancy, frequency, volume, dribbling

Musculoskeletal
Stiffness: timing during the day
Pain: acute/chronic, swelling, erythema, patterns of joints involved
Muscular: pain, weakness

Dermatological
Rashes, blisters, ulcers

Endocrine
Diabetes: thirst, vision, peripheral sensation
Thyroid: weight, tremor, cold/heat intolerance, palpitations, skin or eye changes
Sexual function, menstruation, pregnancy

Neurological
Movement: seizures, involuntary movements, loss of movement, altered gait, swallowing difficulties, fatigue
Sensation: increased or loss of sensation, balance
Higher neurological function: speech, vision, hearing

Ophthalmic
Eye pain, redness, dryness or grittiness, change in vision, flashing lights

Ear, nose, and throat
Pain, changes in smell, taste or hearing, pain in ears, discharge/crusting

Haematological
Easy bleeding or bruising, tiredness, lymph node swelling, abdominal fullness

Mental health
Mood: anxiety, depression, suicidality
Altered perceptions (hallucinations)
Abnormal beliefs (delusions)

Genitourinary
Discharge: urethral or vaginal, itch
Pain: abdominal, pelvic, testicular, anal, pain during sexual intercourse
Sexual function: menstrual cycle, contraception, pregnancy

Box 1.7 Five questions for the patient with a long-term condition

1. ***Diagnosis***—HOW, WHY, WHERE, and WHEN?
2. ***Progression***—How has it progressed since?
3. ***Control***—How do you monitor your condition? When did you last have a check-up? What medications do you take? What are your biggest challenges in managing your condition?
4. ***The good and the bad***—What is the BEST and WORST your condition has ever been? How does it impact your life?
5. ***Today/recently***—How are things today (compared to last week/year/last time you were unwell)? What happened last time you were unwell?

history. Extent and focus will be guided by the presenting problem. Even in the era of complex technological investigations, physical examination performs several key functions, including:

- Providing objective evidence (physical signs) to complement subjective evidence (symptoms). Rarely, diagnosis can be made with examination alone, e.g., skin disorders.
- Assessment of severity.
- Identification of unexpected findings that patients may have not noticed.
- Building rapport with patients. Physical examination can reinforce that assessment of their problem is thorough and compassionate.

An overall assessment is made as the history is elicited: does the patient look well or unwell? Are there any obvious clinical signs from the 'end of the bed?' Formal observations of physiological parameters are recorded routinely in the hospital setting: temperature, heart rate, blood pressure, oxygen saturation, respiratory rate, and level of consciousness.

A detailed assessment is then carried out, including:

- a general examination of the hands, face, and neck
- examination of the likely affected system or systems
- a wider examination of other relevant organ systems if indicated.

Fig. 1.3 outlines a typical general and systematic examination routine suitable for use in patients presenting with a wide range of medical conditions.

INVESTIGATION

There are a huge number of investigations available to the clinician, ranging from bedside tests including spirometry and urine dipstick analysis, to complex radiological imaging and invasive procedures. Some tests combine diagnostic and therapeutic potential, for example, endoscopy and angiography. A good test is appropriate, cost-effective, has high diagnostic value, and little burden to the patient or cost to providers. The following should be considered when investigations are requested:

- ***What question will this test help to answer?***
  It is helpful to have a clear diagnostic question in mind. Specialists can help select the most appropriate investigation to answer a relevant clinical question. Beware

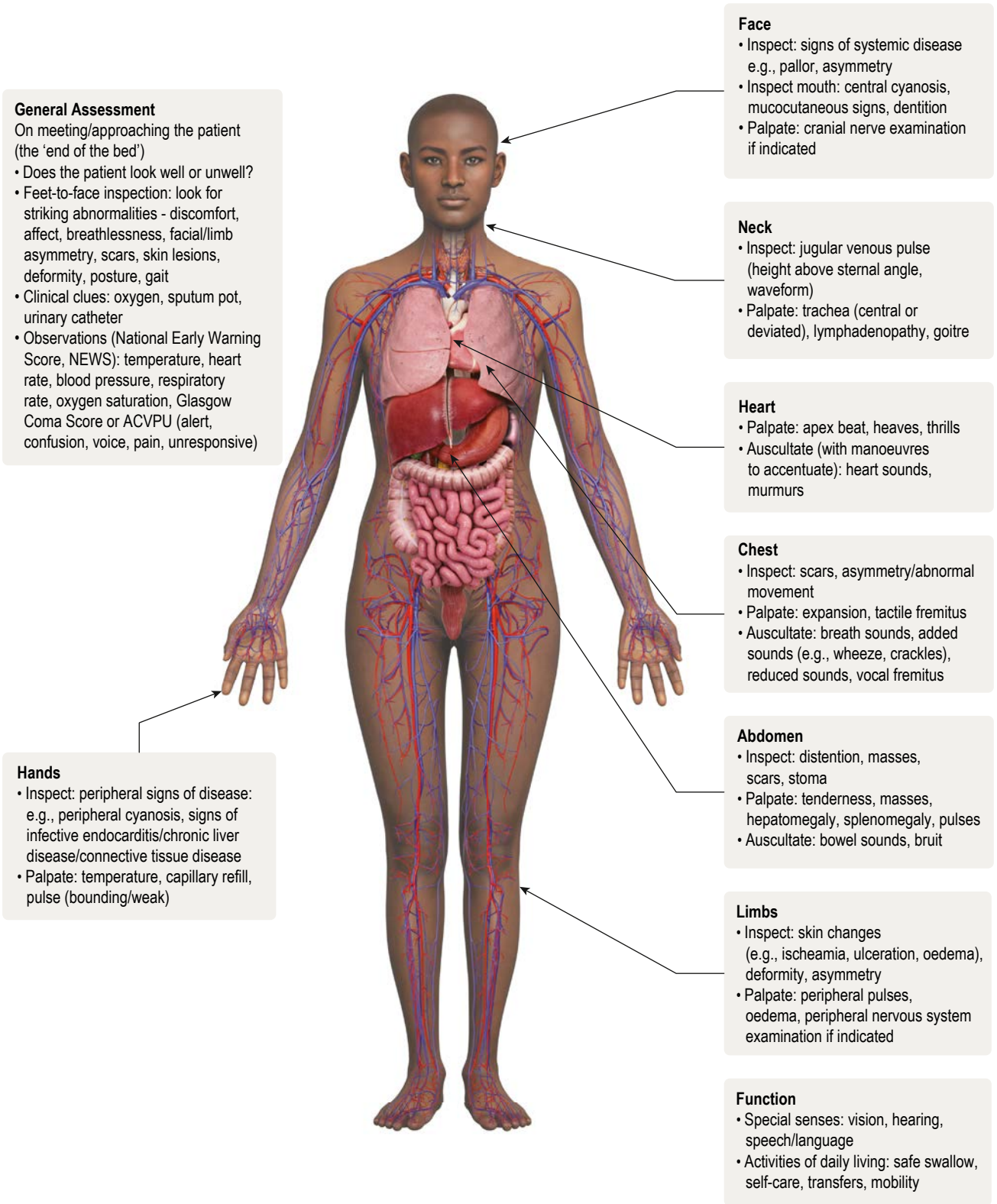


Fig. 1.3 **A basic approach to clinical examination.**

of requesting tests without a clinical reason: results may lead to more questions (and unnecessary further tests) than answers.

- ***What are the sensitivity and specificity of the test?***

A sensitive test correctly identifies patients with a disease and a negative result is useful in reassuring patients that they do not have the condition ('rule out'). For example, D-dimer is a highly sensitive test for thrombus: in patients with a low pre-test probability, a negative D-dimer test accurately excludes thrombus.

A specific test is accurate when the result is positive. For example, D-dimer is not a specific test for thrombus as

the result can also be positive in infection, autoimmune conditions, pregnancy, and following surgery. A specific test such as CT-pulmonary angiography is therefore needed to confirm ('rule-in') pulmonary embolus.

- ***What are the risks of the test to the patient?***
  Invasive procedures may cause bleeding, infection, or organ injury. Ionizing radiation carries a small risk of future malignant disease (<1 in a million for a chest X-ray). The risks of investigations and benefit of knowing the result need to be discussed and weighed with patients to allow them to make the best decision about proceeding with investigation.
- ***How much certainty is needed?***
  The need for certainty is often proportional to the burden of treatment: when treatment is complex or has significant risk of harm the need for certainty in the diagnosis is greater. For example, surgery or chemotherapy for cancer usually requires biopsy diagnosis.
- ***Will the test result make a difference?***
  If all treatment options have a greater risk of harm than benefit, undergoing investigation may not be appropriate.

Interpretation of test results

Diagnoses typically result from history and examination. However, diagnostic testing is often required to confirm a provisional clinical diagnosis and/or to rule out alternatives; at least 10% of all diagnoses are not considered final until testing is complete. Laboratory test results are also used to monitor disease progression and response to treatment.

Interpretation of test results is often undertaken by comparison to a reference interval. This reference interval may be erroneously termed the 'normal range', erroneous in that it does not define normal, and it is not a range. For example, a serum sodium concentration of 134 mmol/L (reference interval 135–145 mmol/L) may not be pathological, and the range of serum sodium is 10 mmol (the difference between highest and lowest reference limits).

A reference interval describes blood analyte concentrations in defined groups of individuals, estimating what can be expected in the population. It provides a reference point against which to interpret an individual patient's result. However, a reference interval may not encompass differences that can significantly impact an analyte, e.g., age, gender, pregnancy.

The decision to include 95% of results from a healthy population in a reference interval is purely pragmatic. It produces less false positives than the inclusion of 90% (when more people sit outside the reference interval) and less false negatives than the inclusion of 99% (when fewer people sit outside the reference interval). But if 95% are included, by default, 5% of people should be expected to be outside of all reference intervals in the absence of disease.

It is rare to request one test in isolation. In clinical practice it is common to request a series of tests on the same sample, for example, a 'metabolic' panel, a 'tired-all-the-time' screen. As, by pragmatic definition, 5% of test results sit outside the reference interval, the more tests that are requested, the greater the likelihood of an 'abnormal' result. If 20 (independent) test results are requested, there is a 38% chance that one result will be outside a 95% reference interval. It is therefore important that clinical suspicion guides test requests and not the appeal of tick box on a request form.

Reference intervals also do not consider intraindividual variation. For example, in a bleeding patient, a fall in haemoglobin from 170 g/L to 130 g/L may be significant even though both results sit within the reference interval. Repeated results may also conform to the phenomenon of 'regression to the mean'. This is a statistical tendency where extreme results tend to be followed by measurements that are closer to the mean when a test is repeated, meaning that the assumption that change is caused by disease or treatment may be flawed.

Reference intervals do not, and cannot, define normality. A result within the reference interval may occur with pathology and a result outside the interval can be seen in health.

When not to investigate

Decisions about whether and how to pursue a diagnosis are often complex and should be made in conjunction with the patient. Individuals differ in their willingness to tolerate uncertainty, with some wanting to seek a definitive diagnosis at all costs and others happy to accept a presumed diagnosis that may or may not be proven wrong. Clinicians can guide and help patients to understand likely and possible outcomes of different tests, offering 'best- and worst-case scenarios'. In general, investigations may not be undertaken when:

- the patient is too frail to derive any benefit from confirming a diagnosis, for example, if the burden is too great, or treatment is likely too ineffective.
- the patient agrees to have an initial investigation, but tests for definitive diagnosis are unacceptable. For example, the patient agrees to a myocardial perfusion scan but does not want to undergo subsequent coronary angiography.

Patients can make an informed decision to decline an investigation or intervention, but there is no entitlement to investigations or tests that are of no overall benefit.

DIAGNOSTIC REASONING

Making the correct diagnosis is key to safe and effective medical care. However, diagnosis is complex. Large amounts of information gathered from history, examination, and investigation results need to be assimilated and synthesized. Each piece of evidence should be weighed according to the confidence you have in its accuracy based on the information obtained. If significant findings remain unexplained, diagnosis should be re-evaluated. Each patient has a unique story and examination findings may vary, for example, with body mass index, posture, or the provision of analgesia. Diagnostic tests that are 100% sensitive and specific, with 100% positive predictive value do not exist (except for life ultimately predicting death). Be humble and insightful enough to seek help from others.

MODELS OF DIAGNOSIS

In a traditional model of medical diagnosis the clinician begins by considering all possible causes of a presenting symptom. Information gathered is then used to include and exclude causes, using probabilistic reasoning, until only one remains. This requires 'type 2 thinking', which is analytical and deductive. Type 2 thinking is the kind of thinking we would like to carry out all the time if only it were not so effortful and slow (**Fig. 1.4A**).

An alternative model is diagnosis by pattern recognition. This is 'type 1 thinking': automatic, intuitive, quick, and requiring little effort. It relies on ***heuristics:*** mental shortcuts or rules of thumb drawn from previous experiences and learned associations. These heuristics allow quick judgements of new situations by comparing

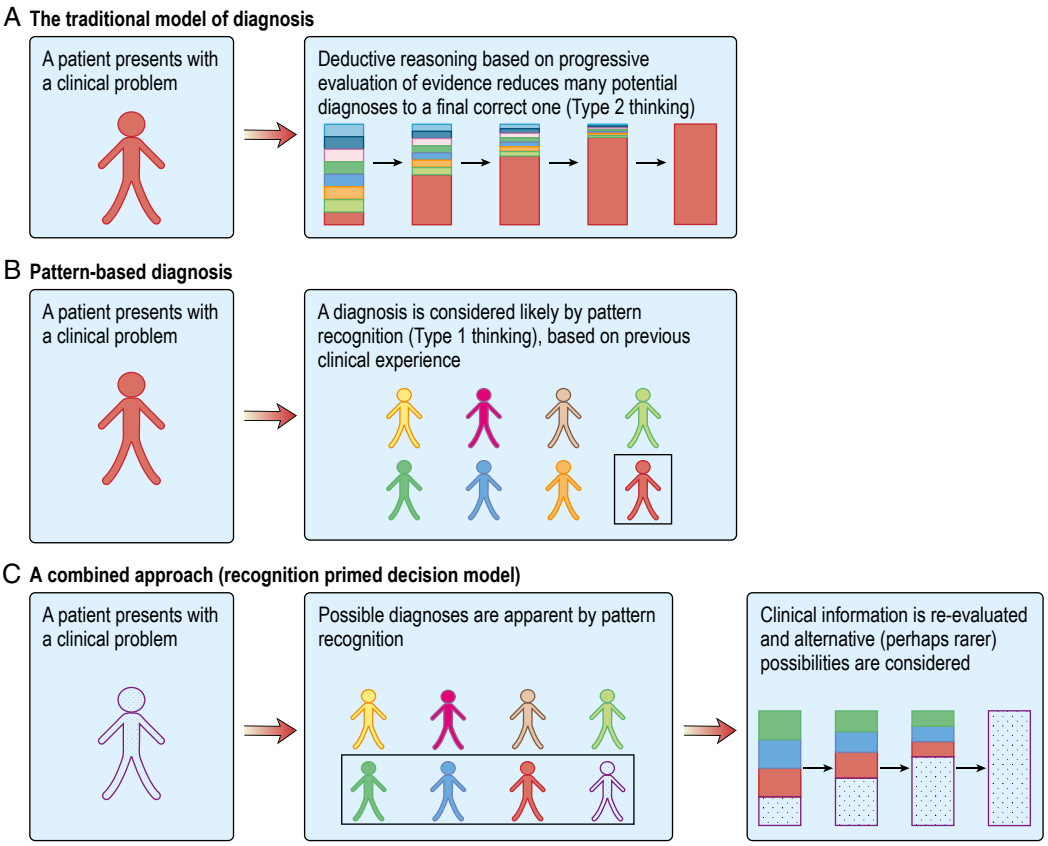


Fig. 1.4 Approaches to diagnosis. **(A)** The traditional model of diagnosis. **(B)** Pattern-based diagnosis. **(C)** A combined approach.

them with similar situations encountered in the past. In this form of diagnostic reasoning, clinicians rapidly compare the presenting patient with the many other patients they have seen, subconsciously drawing on similarities and differences that lead efficiently to a likely diagnosis (Fig. 1.4B).

The best model for efficient and safe diagnosis is likely to be a combination of these two approaches. This is a recognition-primed decision model: the rapid impression generated by type 1 thinking is revised by slower, reason-based type 2 thinking. This model draws on the strengths of both types of thinking (Fig. 1.4C).

CLINICAL INTUITION

Clinical intuition is often described, and aspired to, in medicine. It exists in the apparent 'magic' of the senior clinician to focus on a stable patient during a ward round due to apparent psychic knowledge that, by the end of the day, it will be that patient who will be most unwell. It occurs in their capacity to ask the single question that clinches the diagnosis, despite the extensive answers provided to the 50 questions asked by the trainee.

Yet there is no magic here, and to view clinical intuition as magic is to do a disservice to the complex scaffold of knowledge and expertise on which it is founded. The development of clinical intuition is like learning to read; a complex skill in which the letters and words of clinical information are suddenly read as whole sentences. This looks like a magic leap to those who are still learning letters or words. But with practice, and a degree of clinical predictability, complex analytical type 2 thinking can start to look as if it is type 1 thinking, with apparent near-instantaneous answers to clinical problems.

Clinical medicine is full of opportunities for diagnostic practice. Most clinicians will eventually learn to read in sentences, with a rare few going on to develop the expertise to devour a whole paragraph or even chapter at a time. Clinical medicine is, however, limited in predictability. Patients are unique. They present in different settings, with different experiences of disease, and at different time courses within their disease. This inherent unpredictability means that slower, effortful thinking will always be required. Although apparent 'intuition' looks easy, some cognitive effort is mandatory to keep it in check. Take comfort that on the days that you feel the burden of cognitive strain for it is likely to be protecting you from error.

BIAS AND ERROR

It has been estimated that 10% to 15% of medical diagnoses are wrong. Mis- or delayed diagnosis causes harm, including failure to receive timely intervention, and the adverse effects of unnecessary treatment. While misdiagnoses may occur because of inadequate knowledge or lack of diagnostic resources (particularly in resource-limited settings), they also arise as a result of cognitive bias. Box 1.8 lists cognitive biases which can lead to error.

Various strategies can be employed to reduce the chance of incorrect diagnosis:

- ***Working conditions:*** An environment with sufficient time, information, resources, and sleep is required for the cognitive load of analytical thought. Working conditions which promote reflex rather than analytical thinking should be addressed.
- ***Awareness of bias:*** Increasing clinicians' familiarity with cognitive biases may decrease bias-related errors. Guided reflection

Box 1.8 Cognitive biases leading to misdiagnosis

- ***The affect heuristic:*** A context or patient-specific response due to clinicians experiencing positive or negative feelings towards a patient, based on prior experiences. For example, the symptoms of someone viewed as a 'complainer' fail to trigger appropriate investigation.
- ***Anchoring:*** The tendency to rely on features in the patient's early presentation, with failure to adjust in the light of later information. It is a form of confirmation bias (see below). For example, a patient is sent to hospital by their GP with a suspected urinary tract infection; the doctor in the emergency department fails to consider other possible causes for the patient's confusion.
- ***Aggregate bias:*** The belief that conclusions drawn from aggregate data do not apply to the patient being evaluated due to the possibility of rare or atypical disease. For example, a clinician orders a CT scan despite decision-making tools showing it is not indicated.
- ***Ascertainment bias:*** A clinician's thinking is shaped by prior expectation. This includes stereotyping and race/ethnicity and gender biases. The routine use of race and ethnicity descriptors before the provision of clinical information in case presentations and medical assessments should be questioned.
- ***Availability:*** Explanations which spring easily to mind are presumed more likely to be correct. For example, a doctor assumes that acute kidney injury is due to dehydration (despite an absence of clinical signs) and administers intravenous fluids due to unfamiliarity with renal parenchymal disease.
- ***Confirmation bias:*** The tendency to look for confirming evidence to support a diagnosis rather than evidence that refutes a diagnosis. This occurs despite refuting evidence often being more definitive. For example, a mild neutrophilia is viewed as confirmation of a chest infection, and the clinical history consistent with pulmonary embolus is not sought.
- ***Framing:*** Making decisions in different ways, depending on whether a choice is presented in positive or negative terms. For example, a doctor may respond differently to an incidentally discovered adrenal mass that is reported to be 'probably benign', compared to a report reading 'unable to exclude malignancy'.
- ***Omission and commission biases:*** A tendency towards inaction (omission) or action (commission) depending on whether the clinician's focus is on harm arising from diagnosis/intervention, or conversely, a belief that active intervention is required to prevent harm. Omission bias is more common due to acceptance of the natural progression of disease over iatrogenic harm, for example, the unnecessary avoidance of chest X-rays in pregnancy.
- ***Optimism:*** The assumption that negative outcomes occur less often than they really do. For example, a junior doctor may fail to call for senior help with a clearly deteriorating patient because of an unrealistic expectation that the interventions they have instituted will improve the situation.
- ***Outcome bias:*** A belief that good or bad results are always attributable to prior decisions. Although a relationship between decisions and outcomes might seem logical, the outcome of a decision is not the only determinant of its quality; sometimes, a bad outcome can happen despite a good clinical decision. For example, a tissue biopsy is not requested because a previous biopsy resulted in a life-threatening bleed.
- ***Recency bias:*** Remembering most easily things that have happened most recently. For example, a doctor may be distracted away from a common diagnosis by a much rarer one, which was covered in a recent teaching session.
- ***Substitution:*** An easier (and related) question is solved in place of a more difficult one. For example, a patient with previous intravenous drug misuse presents with worsening headache and a CT scan of their head is normal. Rather than address a difficult question (why does this patient still have a headache?), a simpler question is solved (might they simply be seeking opiate analgesia?).
- ***The sunk-cost fallacy:*** Continuing to invest in a failing idea because of significant prior investment. For example, following an unsuccessful operation that resulted in significant complications, a clinician recommends a repeat procedure rather than conservative management to try to salvage something from the situation.

involves searching for and being open to alternative diagnoses. 'Cognitive forcing' involves conscious consideration of alternative diagnoses that do not come intuitively.

- ***Use language devoid of assumption:*** Patient's problems should be summarized succinctly and objectively. Do not assume previous diagnostic reasoning was correct.
- ***Awareness of the wider differential:*** Aides-mémoirs encourage consideration of the wide differential of any presenting problem. Diagnostic sieves remind clinicians of the spectrum of disease processes that can affect humans, for example, ***VITAMIN C DEF***iciency = **V**ascular, **I**nfective/**I**nflammatory, **T**rauma, **A**utoimmune, **M**etabolic, **I**atrogenic, **N**eoplastic, **C**ongenital, **D**rug related, **E**ndocrine/**E**nvironmental, **F**unctional.
- ***An iterative approach:*** All previous diagnoses are subjected to appropriate re-evaluation, looking specifically for evidence that brings them into question.
- ***Team discussion:*** All team members (including the patient) are empowered to challenge a diagnosis, including that of senior clinicians. Multidisciplinary team meetings bring together clinicians and allied healthcare professionals from different specialties so that relevant knowledge and expertise is shared in a collaborative environment. Healthcare leaders have a responsibility to flatten the authority gradient in team discussions.
- ***Diagnostic criteria:*** Evidence-based standards may be available to help guide diagnosis.

LEVELS OF DIAGNOSTIC DEPTH

Clinicians may assign an umbrella term that categorizes a collection of signs and symptoms into a clinical entity, for example, 'acute coronary syndrome', 'delirium', 'upper respiratory tract infection', or 'sepsis'. This umbrella term describes an initial impression; it is a step towards the final, specific diagnosis. Some patient presentations do not fall neatly under umbrella terminology and the initial impression remains a narrative description of the patient's problem.

The impression formed at the end of a clinical consultation may be confirmed, refined, or refuted by tests. For some conditions, complex, expensive, or potentially hazardous investigations are required for diagnosis. Tests might also be needed to demonstrate the extent, severity, or treatment responsiveness of the condition (Fig. 1.5).

It might be appropriate to stop at the level of a syndrome or clinical impression rather than continuing with investigations to define the histopathological or mechanistic cause if:

- the patient is satisfied with this level of explanation and declines further diagnostic work-up
- the problem is mild or self-limiting
- further investigation is unlikely to yield a specific cause, e.g., resolving diarrhoea due to likely viral gastroenteritis
- further investigation is unlikely to influence management as treatment will be the same (or not effective) regardless of further diagnostic clarification
- the risks of investigations and/or treatments are too great for the patient, for example, due to frailty or co-morbidity.

THE ROLE OF WATCHFUL WAITING

All diseases have a natural history: some are self-limiting, some progress, and some relapse and remit. Where there is diagnostic uncertainty, there can be a role for waiting to see how events

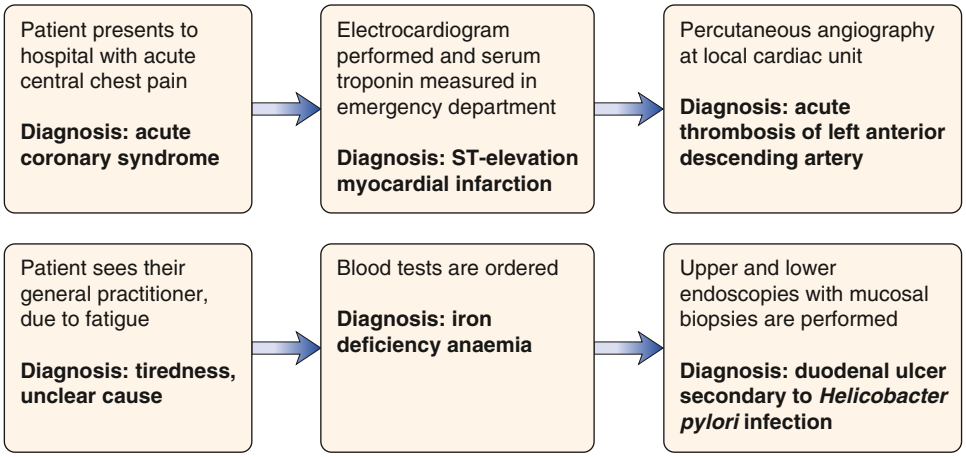


Fig. 1.5 **Examples of different levels of diagnosis.**

i Box 1.9 The role of watchful waiting in diagnosis

- **Incidental finding of a small pulmonary nodule.** A 77-year-old female undergoes CT scanning of the coronary arteries after an episode of chest pain. Incidentally, a small (<10 mm) pulmonary nodule is detected. She is at risk for lung cancer because of a long smoking history. In accordance with the British Thoracic Society's Guidelines for the Investigation and Management of Pulmonary Nodules, she is offered repeat computed tomography screening at 3 months and 1 year. These reveal stable appearances with no increase in size. She is discharged after 1 year.
- **Abnormal liver blood tests.** A 34-year-old male is found to have abnormal liver blood test results. Blood tests for viral hepatitis and autoantibodies associated with liver disease are negative. A liver biopsy is discussed with the patient, but because his liver blood tests are not rapidly deteriorating, a decision is made in favour of watchful waiting. Two months later, his liver blood tests are within the reference interval and remain normal. He is discharged without a diagnosis for his transient liver injury.
- **Relapse of nephrotic syndrome in adulthood.** An 18-year-old female with a history of presumed minimal change disease as a child presents with recurrent nephrotic syndrome. She is commenced on high-dose corticosteroids and her response to treatment is monitored. After 2 months, remission has not been achieved. No response to corticosteroids is atypical for minimal change disease so a kidney biopsy is done. This reveals a different but related kidney disorder: focal segmental glomerulosclerosis.

develop if it is safe to do so and when beneficial treatment is not inappropriately delayed (Box 1.9).

Watching and waiting are sometimes necessary for diagnosis when diagnostic criteria stipulate symptoms must have been present for a certain length of time. It may be wise to defer risky or burdensome investigations until the clinical course of the disease justifies investigation. Discussion with the patient is crucial.

DIAGNOSTIC CRITERIA

Some conditions are diagnosed with a single pathognomonic investigation. For example, the presence of urate crystals on microscopic examination of synovial fluid is diagnostic of gout. In other situations, confirming a diagnosis is more complex and based on a combination of symptoms, physical signs, and investigation results. For example, a diagnosis of systemic lupus erythematosus is based on evidence of autoimmunity plus additional symptoms, signs, or test results demonstrating constitutional, hematologic, neuropsychiatric, mucocutaneous, serosal, musculoskeletal, renal, or immunological involvement.

Diagnostic criteria are useful for:

- Patients: enabling correct diagnosis and treatment
- Epidemiology and public health: the compilation of accurate statistics for monitoring trends in the incidence and distribution of disease.
- Research: to allow study of defined diseases and their treatments
- Remuneration and service planning: adequate provision of care and payment to healthcare providers based on diagnostic codes.

OVERDIAGNOSIS: TOO MUCH MEDICINE

Overdiagnosis refers to the use of a diagnostic label that confers no benefit. It occurs due to:

- Overdetection: identification of diagnoses that will never cause harm
- Overdefinition: the medicalization of normality

Examples are shown in Box 1.10. Overtreatment is the use of treatments which do not provide benefit. Overtreatment can occur in the absence of overdiagnosis, but the two phenomena often go hand-in-hand.

UNCERTAINTY

Sometimes, it is not possible to make a definitive diagnosis. The reasons for this are varied:

- Available investigations are unable to confirm a definitive diagnosis. This can occur when diagnoses are more complex than the tests available or when medicine is delivered in a resource-limited setting without access to investigations.
- Definitive investigation is avoided due to risk of harm.
- Investigations are declined.
- There is a time-lag between investigations being sent and results becoming available.

Without a confirmed diagnosis, treatment and prognostication may be uncertain. Uncertainty is inherent to medical practice, as it is in all situations that encompass complex behaviours, dependent variables, and shifting coefficients. Osler (1849–1919) stated that medicine is the science of uncertainty and art of probability. R.C.

Box 1.10 Examples of overdiagnosis in medicine

Overdiagnosis	Mechanism
Chronic kidney disease (CKD)	Overdefinition: The absence of an age-adjusted CKD definition increases the diagnosis of CKD in elderly patients with an indolent decline in kidney function.
Subsegmental pulmonary embolus	Overdetection: Increasingly sensitive cross-sectional imaging may reveal emboli that do not cause harm.
Hypercholesterolaemia in the very elderly	Overdefinition: Reference intervals for the very elderly may differ from those derived from younger cohorts. Overdetection: Treatment may not confer long-term benefit.
Cancer screening	Overdetection: Screening for thyroid and prostate cancers reveal small indolent tumours that are unlikely to progress. Intervention does not change mortality, though morbidity from treatment may occur.
Incidental findings	Overdetection: The pulmonary nodule or small adrenal lesion found on imaging undertaken for other reasons leads to further tests or surveillance imaging.

Fox, a medical sociologist writing in 1980, stated that uncertainty derives from the limitations of medical knowledge, the limitations of personal knowledge, and the difficulty distinguishing between the two.

Uncertainty in medicine is a challenge, especially in a technologically advanced world in which the expectation of certainty is increasing. Clinicians who experience anxiety in response to uncertainty have higher levels of burnout and decreased job satisfaction. There may be more uncertainty when the traditional pathophysiological model of disease is insufficient; anxiety and discomfort may therefore be greater when patients present with psychological disease or when patients are dying.

Adherence to evidence-based or consensus clinical guidelines is a helpful response to uncertainty: knowing what other people do in similar or comparable circumstances can be useful. However, randomized controlled trial data deals with cohorts, not the individual patient in front of you. Research data also rarely measures the variance of humanity beyond the scientific, including levels of desire and interest in the outcome. This meta-uncertainty can be hard for both patients and clinicians.

Uncertainty requires acknowledgement and acceptance because it is inevitable in medicine. Talk to colleagues, for it is helpful to know that your peers are also uncertain and because uncertainty stimulates different ideas from different people. And talk to your patients; they are likely to have always been more uncertain than you. Empathetic common ground can be a comfortable place, and their unique perspective on their own life might bring some certainty about the right thing for them.

Some options are always available even when a diagnosis is uncertain:

- Symptomatic treatments, including pain relief and anti-sickness drugs, are always appropriate.
- Empirical treatment can be instituted based on suspected pathology, provided potential side-effects and the possibility of harm if the diagnosis is wrong are acceptable to the patient.
- Follow-up. Change or evolution of disease may lead to certainty, or diseases may remit, meaning certainty is no longer required.
- What is important to the patient? How can you help them live their lives according to their priorities, for example, time with family or friends, the achievement of an ambition or goal, or according to their faith.
- Compassion is 'to suffer with'. Walk and talk with your patient. Partner them on their health journey. Expertise and certainty may be less important than this.

COMMUNICATING A DIAGNOSIS

Towards the end of a clinical consultation the clinician will need to explain their diagnostic reasoning to the patient, with plans for further investigation and intervention. Effective clinical communication is an evidence-based, learnable, professional skill, which demonstrably improves the quality and safety of healthcare, and improves health outcomes. Clinical communication is addressed in Chapter 2.

INEQUALITIES IN MEDICINE

Race and ethnicity

The terms 'race' and 'ethnicity' attempt to describe highly complex concepts including phenotypic, genetic, geographic, and cultural factors, alongside an individual's perceived identity. However, aggregate categories are poor indicators of underlying biological and physiological variation; the assumption that a specific race or ethnicity equates to genetic or physiological distinction is unfounded.

Racism exists in every modern society, including in medicine. There is a shameful history of medicine being complicit in racist practices within society including forced sterilization, unethical research, and Nazi atrocities. Healthcare practitioners have an ethical duty to remove racial bias: look for it (especially in your subconscious), acknowledge it, and change things.

Treating patients differently based on a categorization of race or ethnicity can result in substandard care. For example, a race correction for estimated glomerular filtration rate (eGFR) *presumed* a higher muscle mass in those of Black ethnicity. This has now been abandoned in contemporary eGFR calculations; it is a flawed presumption that led to delayed care in those of Black ethnicity. The prescription of antihypertensives in the UK based on a binary definition of race may delay prescription of angiotensin blockade to those of Black ethnicities who may have much to gain from these drugs.

Although disease prevalence can vary with ethnicity, the inappropriate use of ethnicity as a proxy for disease or disease risk is pervasive throughout medical training. Race and ethnicity are commonly presented in case histories and exam questions before any clinical information is given, teaching clinicians that racial bias is helpful or necessary for diagnosis.

However, an understanding of differences in health and healthcare by ethnicity can identify and, therefore, help address inequalities. Ethnic differences in disease outcomes remain troublingly common. During the SARS-CoV-2 pandemic in 2020 to 2021, minority ethnic groups in the UK were more likely to develop severe infection with increased risk due predominantly to non-biological factors: increased exposure due to front-line work, overcrowded accommodation, barriers to health services, co-morbidity, and a life course that increases the risk of non-communicable disease. Distinction of health provision and outcomes by race and ethnicity to identify disparities requires accurate race and ethnicity data beyond umbrella definitions and binary or simplistic classifications, as well

as a wider reflection on the multifactorial nature of health outcomes (see 'Health inequality and intersectionality' section later).

Everyone has biases and, as clinicians, we are taught to stereotype. Awareness of this is the first step to correcting inequalities that arise from these on an individual level. On a wider level, the dismantling of systems that perpetuate intersecting and compounding inequities and action to address health inequalities between different ethnic groups are needed.

Sex and gender

Sex describes the biological differences between male and female: chromosomes, hormones, and reproductive anatomy. Gender refers to the socially and culturally constructed differences between males and females: normative roles and expectations in society. Gender identity is the person's sense of self: whether they see themselves as male, female, other, or no gender. Where distinction between sex is important for pathophysiology, terminology should describe sex and not presume gender; male or female is therefore preferentially used.

Sexism is the differential treatment of individuals (either more or less favourably) based on sex. Females and those who identify as females are disproportionately affected by sexism.

Clinical practice and research have traditionally defaulted to the male patient. For example, females are less likely to be diagnosed with acute coronary syndrome and their chest pain is commonly described as 'atypical'. The fact that a presentation of pain which affects 50% of the population can be described as 'atypical' is striking in itself: for 'atypical', read 'not male'. Most medications used in pregnancy have insufficient safety information. Only six drugs are licensed for use in pregnancy in the UK, and in the last 30 years only a single medication has been specifically developed for use in pregnancy. Females are underrepresented in clinical trials, and pregnant and breastfeeding females are largely excluded from most research. Caution around medications in pregnancy leads to fear (in female patients and their healthcare providers) and undertreatment.

Conversely, males on average die younger than females, are less likely to access medical services, and tend to present later in the course of disease. Suicide is the leading cause of death in males under 50 in the UK, with societal perceptions of masculinity thought to be a contributor.

Examine your own gender biases and look for those that exist in healthcare and healthcare research. Was this female/male/gender-diverse patient represented in the published data for this disease? If not, highlight this and work to ensure inclusion in future research. Would you do or consider anything else if this patient were a different sex or gender? If so, make sure there is a very valid reason for not doing so now.

People living with disabilities

People living with physical or learning difficulties have poorer health outcomes compared to those with the same diseases without disability. They may struggle to access, understand, or adhere to care.

Healthcare facilities should be accessible, and appointment letters should be easy to read and understand. Resources should be made available in braille. Sign language should be included amongst the languages for which translation is made available. Specialist learning support services should be available for any for whom healthcare attendance is difficult. Individualized communication strategies may help patients that are neurodiverse to understand and engage with care. There should be a focus on barriers to care, as these may impact outcomes more than the disease. Ask service users what they think and how they would design their care.

Health inequality and intersectionality

Many groups in society experience worse health outcomes because of social, cultural, and demographic factors. As well as the differences according to ethnicity, gender, and ability, inequalities are also experienced by asylum seekers and other immigrant populations, those whose first language is not the main language of healthcare services, those with mental health problems or substance addiction, those living in poverty, older people and those with dementia, those with stigmatised conditions, those with low levels of educational attainment, and those without good social support networks.

Health inequalities are avoidable differences in health status or in the distribution of health resources that occur between different groups within society according to the conditions in which people are born, grow, live, work, and age. Health inequalities include differences in:

- health status, e.g., life expectancy, morbidity, mortality
- access to care, e.g., due to available health services (including contraception and safe termination), language, digital access
- quality and experience of care, e.g., patient satisfaction, understanding
- behavioural risks, e.g., smoking, sexual practices, alcohol/substance misuse
- wider determinants of health, e.g., income, housing, nutrition

Individuals may have multiple factors contributing to variation in health, for example, an unemployed female of Black ethnicity living with disability in a socially deprived area. The way in which these factors combine and interact is termed ***intersectionality***.

This is not an easy problem to solve. It requires attention, an intolerance of injustice, and a holistic policy-led approach that extends beyond healthcare providers. The individual clinician is not devolved of responsibility just because something is difficult; these are areas where a doctor's voice and advocacy can be powerful.

Clinicians have a responsibility to:

- pay attention to, measure, and talk loudly about inequality
- engage with public health colleagues and consider where there are opportunities to prevent disease, not just to diagnose and treat
- work with local communities to ensure equal access to services for all members of society
- empower and help all patients to be health literate.

ARTIFICIAL INTELLIGENCE AND MEDICINE

Artificial intelligence (AI) describes the ability of machines to perform tasks that require higher cognitive skills including natural language processing, perceiving, reasoning, problem-solving, creativity, executive planning, pattern recognition, and interacting with the environment. Although predictions are made about the development of self-aware technology and 'super' AI that out-performs humans, the only type of AI that currently exists is 'artificial narrow intelligence' (ANI): able to perform a single or narrow task (faster and better than a human) but with no capacity to function outside of this defined task. AI systems analyse and utilize trends in large datasets, 'seeing' structure and regularities, and 'learning' to perform tasks more efficiently than humans.

AI technologies have the potential to transform patient care. Several ANIs have been trialed:

- ***'Computer vision'.*** AI systems review large numbers of medical images with diagnostic labels supplied by experts. This produces algorithms that correlate image features with diagnosis.

Clinical utility has been shown in the analysis of electrocardiograms, diagnosis of skin cancers, interpretation of retinal screening images, and reporting of specific radiological images.

- ***Prediction.*** AI systems use large patient datasets and correlate demographics, physical signs, investigation results, and treatment responses with patient outcomes. This generates powerful predictors of outcomes including morbidity, mortality, long-term survival, and suicide risk. AI can reveal patterns overlooked by human researchers.
- ***Personalized treatment.*** AI systems identify genetic variants that predict responses to different treatments. These allow treatment regimens to be tailored to individual patients, maximizing efficacy and avoiding toxicity from treatments predicted to be ineffective.

All currently and imminently available AI presents clinicians with suggestions: the probability of a diagnosis, an assessment of risk, the identification of treatments most likely to be of benefit. AI requires training by expert humans, providing non-experts with a helpful, second opinion from which heuristics and biases are eliminated.

At present, AI that exceeds the intelligence of humans remains theoretical. But consider a future in which a patient tells a computer their symptoms, undergoes the investigations that a computer system determines to be most useful, receives an automated, algorithm-driven diagnosis, and is given a management plan based on a review of all available medical literature. Is this the perfect aspiration or a simplistic, mechanical representation of doctoring?

Whether AI will develop the capacity to read body language and verbal cues, be empathetic and compassionate, deal with uncertainty, hold a hand, share unique decision-making with every unique human, offer spiritual care to the dying, and raise its (automated) voice to try and address social determinants of health remains to be seen. For now, humans can take comfort in the predicted trajectory of AI that means doctors are not imminently likely to be replaced by AI. However, doctors that do not use AI may be replaced by those that do.

Further reading

Academy of Medicine Royal Colleges. *Artificial Intelligence in Healthcare*, 2018; https://www.aomrc.org.uk/wp-content/uploads/2019/01/Artificial_intelligence_in_healthcare_0119.pdf.

British Medical Association Sexism in Medicine Report 2022. Available at: https://www.bma.org.uk/advice-and-support/equality-and-diversity-guidance/gender-equality-in-medicine/sexism-in-medicine-report.

British Medical Association. *Reducing Health Inequalities in Your Local Area: A Toolkit for Clinicians.* 2022; Available at: https://www.bma.org.uk/what-we-do/population-health/addressing-social-determinants-that-influence-health-/reducing-health-inequalities-in-your-local-area-a-toolkit-for-clinicians.

Brodersen J, Schwartz LM, Heneghan C et al. Overdiagnosis: what it is and what it isn't. *BMJ Evid Based Med*. 2018;23(1):1–3.

Cerdeña JP, Plaisime MV, Tsai J. From race-based to race-conscious medicine: how anti-racist uprisings call us to act. *Lancet*. 2020;396(10257):1125–1128.

Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. *Future Healthc J*. 2019;6(2):94–98.

Department of Health. *'Hello, my name is…' Campaign*. https://www.health-ni.gov.uk/articles/hello-my-name.

Fischer M and Ereaut G. *When Doctors and Patients Talk: Making Sense of The Consultation*. The Health Foundation 2012; https://www.health.org.uk/publications/when-doctors-and-patients-talk-making-sense-of-the-consultation.

General Medical Council. *Good Medical Practice* 2024; https://www.gmc-uk.org/-/media/documents/good-medical-practice-2024---english-102607294.pdf.

Gopal DP, Chetty U, O'Donnell P et al. Implicit bias in healthcare: clinical practice, research and decision making. *Future Healthc J* 2021; 8(1):40–48.

Glynn M and Drake WM (eds). *Hutchinson's Clinical Methods: An Integrated Approach to Clinical Practice*. Edinburgh: Elsevier; 2022.

Hatch S. Uncertainty in medicine. *BMJ* 2017; 357:j2180.

Kahneman D. *Thinking, Fast and Slow*. New York: Strauss and Giroux; 2012.

Lewis T, Buck D, Wenzel L. *Equity and Endurance: How Can We Tackle Health Inequalities this Time?* The Kings Fund, 2022; https://www.kingsfund.org.uk/insight-and-analysis/long-reads/how-can-we-tackle-health-inequalities.

Smart A, Weiner K. Racialised prescribing: enacting race/ethnicity in clinical practice guidelines and in accounts of clinical practice. *Sociol Health Illn* 2018; 40(5):843–858.

Tong M, Artiga S. *Use of Race in Clinical Diagnosis and Decision Making: Overview and Implications*. Kaiser Family Foundation, 2021; Available at: https://www.kff.org/racial-equity-and-health-policy/issue-brief/use-of-race-in-clinical-diagnosis-and-decision-making-overview-and-implications/.

Whyte MB, Kelly P. The normal range: it is not normal and it is not a range. *Postgrad Med J* 2018;94:613–616.

World Health Organisation. *International Statistical Classification of Diseases and Related Health Problems (ICD)* 11, adopted 2019, effective 2022. Available at: https://www.who.int/standards/classifications/classification-of-diseases.

2 Clinical communication

Lorraine M. Noble

Patient-centred care 14
What is clinical communication? 14
Clinical communication is an evidence-based skill 15
What is effective clinical communication? 16
Learning clinical communication 18
Initiating a consultation 19
Ending a consultation 19
Gathering information 19
Sharing information 19
Discussing uncertainty and risk 20
Shared decision-making 20
Breaking bad news 21
Other difficult conversations 22
Remote consulting 22
Communication with colleagues 23

CORE SKILLS AND KNOWLEDGE

The goal of healthcare is to heal, save lives, and alleviate suffering. In contemporary healthcare, this is framed within the philosophy of patient-centred care. Patient-centred care places the person accessing healthcare at the centre of the health service they receive by:

- Providing care that is respectful of, and responsive to, individual patient preferences, needs, and values
- Ensuring that patient values guide clinical decisions.

Much of this is accomplished through communication. Effective clinical communication plays a critical role in enacting patient-centred care. It enables healthcare professionals to:

- Accurately identify problems
- Share information about the problem and options for intervention
- Learn the patient's perspective: values, needs, and concerns
- Understand the effects of the problem and its treatment on the patient, in their unique psychosocial context
- Empower the patient to make choices according to their values, needs, and circumstances
- Support the patient throughout the care process.

This chapter describes the core clinical communication skills that underpin these processes. Effective clinical communication is an evidence-based, learnable, professional skill that demonstrably improves the quality and safety of healthcare and affects health outcomes.

You are uneasy. You feel there is something wrong with your body. You navigate an unfamiliar website and receive a time and date for a phone call.

You are worried. Someone you don't know and can't see tells you that you need tests.

You are baffled. You try to orientate yourself in a strange building. It is cluttered with incomprehensible signs. You find your way to a corridor that looks the same as the others and are told to wait. You wait and wait. You are unsure if you are in the right place.

You are a body. You are passed through a series of rooms, where gowned figures and strange machines prod you, sample you, and take pictures. You answer the same questions again and again.

You are rebuffed. You are told to wait for results.

You are relieved. Another phone call, another voice you can't put a face to. You think you pick out, 'It's not cancer'. You hear yourself repeating those words. The voice repeats the words back and you appreciate its kindness.

You are alone. The conveyor belt you were on seems to have stopped. You wonder what will happen now.

PATIENT-CENTRED CARE

In patient-centred care, anyone accessing healthcare is treated as a person who makes autonomous choices based on their values and not as a bundle of symptoms or a disease. Values are what matters most to patients and are dependent on individual circumstances. In the United Kingdom, patient autonomy is enshrined in law. Enabling people accessing healthcare to make decisions that are right for them according to their unique circumstances often takes place during a medical consultation. A consultation can be defined as:

- the act of discussing something before deciding
- meeting with an expert to get advice

It can be helpful to understand the doctor–patient consultation as a 'meeting between experts'. Each participant brings expertise, working together to develop a feasible plan (**Box 2.1**).

Further reading

Montgomery v Lanarkshire Health Board. UKSC 2015.

WHAT IS CLINICAL COMMUNICATION?

Communication can be defined as the sharing of information, ideas, and feelings by speaking, writing, or other medium.

Emphasis is often placed on the importance of information exchange in healthcare, particularly verbal communication. But healthcare involves elements of cure (biomedical response) and

Box 2.1 Meetings between experts: what the patient and doctor bring to the consultation

Patient	Doctor
• Expert in their own life, values, and needs	• Expert in diagnosis and treatment
• Has autonomy: the right to make decisions about healthcare	• Must act in accordance with professional duties and responsibilities
• May have access to others' lived experiences relevant to this situation	• May have access to other patients' lived experiences relevant to this situation
• Has to integrate the agreed plan into their life	• Has to organize a feasible plan in the context of the healthcare system
• Has to live with the consequences of the healthcare decision	

care (impact on the person). A consultation to support a patient in deciding includes a discussion about:

- the evidence-based treatment options
- what matters to the patient according to individual circumstances.

Clinical communication is professional communication for and about patient care. This includes:

- direct communication between doctor and patient and those close to the patient: relatives, carers, and advocates
- communication about the patient's care: with colleagues and indirectly through record keeping.

This communication can be by any means: face to face, telephone, online, letter, and/or electronic patient record.

Clinical communication involves two key elements: tasks and process.

Tasks are the steps taken in the consultation to address the needs of the patient:

- Establish the nature of the problem: What is it? Is it serious?
- Gain information about available options: What is the treatment? What does it involve? Is there more than one option? How likely is it to work? What are the side-effects? What are the effects on other aspects of my life?
- Receive support in deciding: What does the doctor think based on their experience of others with this problem? How do I make the right decision for me?
- Devise a feasible plan: When, where, and how will this take place? What other information do I need?

Checklists, for example, for establishing medical history or breaking bad news, are universally written from a doctor's perspective, focusing primarily on the doctor's aims. This can marginalize or remove patient perspective and agency. Consider a meeting where the other person has set an agenda you do not have access to and the effect this has on your ability to prepare, participate meaningfully, and exercise control over decisions. Shared agenda setting is the key communicative strategy to facilitate patient agency.

Process is how communication is enacted to achieve the purpose of the meeting. This communication includes the following:

- Verbal: questioning, clarifying, summarizing, and explaining
- Para-verbal: tone, pauses, inflection, and volume
- Non-verbal: eye contact, facial expression, head movement, body posture, and gestures

The process facilitates the sharing of information, ideas, and feelings between the doctor and the patient. For example, a doctor may ask the patient about their concerns, while facing them and making eye contact. The doctor's verbal and non-verbal communication are congruent, signalling that they are listening. Alternatively, the doctor may ask the same question while looking away and typing at a computer. In this case, verbal and non-verbal incongruence signals a lack of interest in the answer.

A key principle of communication is that people understand the social rules of normal conversation. This means that people react not just to what is said (literal meaning) but also to what is communicated through the process of interaction (social function). This social function determines the effect of healthcare communication. It can empower, inform, comfort, and deliver patient-centred care, or do the opposite.

Medical encounters are a form of 'institutional communication': a conversation in an institutional setting with a professional, characterized by a format and language that may be unfamiliar to a lay person. Converting institutional communication to 'normal' effective communication requires the following:

- Orienting the patient to the care process: people need to know what is going to happen to have agency.
- Translation of technical language: people need comprehensible information.
- Compassion: health problems and receiving healthcare are worrying and people need support.

CLINICAL COMMUNICATION IS AN EVIDENCE-BASED SKILL

There is robust research literature, dating back nearly six decades, examining how doctor–patient communication affects care outcomes.

Clinical communication affects care outcomes

This includes outcomes that are:

- Immediate: accuracy of information, patient satisfaction, comprehension of information, medical error.
- Intermediate: follow-through of a treatment plan.
- Long term: health outcomes, litigation.

Communication is defined as 'effective' when it affects outcomes positively.

Clinical communication education is effective

This includes training in all aspects of clinical communication, from establishing a medical history to responding to difficult issues such as breaking bad news. There is evidence that communication deteriorates in the absence of training due to acculturation in healthcare systems that can depersonalize patients, prioritize targets that exclude patient agency, and use language saturated with medical jargon.

Key features of effective communication

- Experiential learning with feedback and reflection, e.g., simulated patients
- Observation of role models who communicate effectively
- Integration of clinical communication into summative assessments
- Valuing clinical communication and its education
- Revisiting skills through experiential learning (lifelong learning)
- Immersion in healthcare systems that value effective clinical communication

Further reading

Silverman J, Kurtz S, and Draper J. *Skills for Communicating with Patients,* 3rd edn. Boca Raton, FL: CRC Press; 2013.

WHAT IS EFFECTIVE CLINICAL COMMUNICATION?

Across the diversity of people accessing healthcare, medical specialties, and settings, the fundamental elements of effective communication are constant (**Box 2.2**). People want and need to:

- be listened to
- be respected
- have a sense of control.

These three elements are the foundation on which the doctor–patient consultation is built. Research has shown that the elements of a consultation needed to achieve these are straightforward, predictable, and replicable across healthcare settings. Effective communication fosters self-efficacy: an ability to influence outcomes, which has a positive effect on physical and mental health outcomes.

Responsiveness, efficiency, patient-centredness, and appropriateness of language are key principles of effective clinical communication.

Responsive

The goals depend on the individual, with different ways of achieving them. For example, medical empathy can be defined as the ability to understand a patient's situation and emotional state and convey that understanding back to the patient. This is an important prerequisite to patient-centred care. Conveying understanding back to a patient shows that the doctor has listened and the patient has been heard; it expresses care and respect, which are bedrocks of effective doctor–patient communication. But empathy is needed in differing degrees in different consultations, and the ways in which empathy is expressed differ too. Empathy can be conveyed through the use of silence, a verbal statement, eye contact, facial expression, and touch. It is sometimes conveyed directly, for example, through the doctor explicitly stating awareness of the patient's perspective or emotion. At other times, it may be conveyed more subtly, such as demeanour when breaking bad news. The skill is to respond to the patient's need. While one patient might value the opportunity to discuss their emotions, another might want to focus on practical elements. There is no 'one-size-fits-all' approach to empathy, as it is contextually dependent and driven by the individual needs of the patient. Appropriate empathy is less about whether the doctor has displayed specific communicative behaviours and more about whether the patient feels understood and cared for.

The constituents of a 'good consultation' and 'good clinical communication' are dependent on context. A consultation with a patient in the emergency department with a suspected heart attack will be very different from an outpatient consultation with a patient with longstanding, stable ischaemic heart disease. As the role of clinical communication is to enact quality care that is appropriate to the setting, communication will inevitably differ in terms of aims, tasks, processes, and duration.

Box 2.2 What do patients want?

- Greet me in a way that makes me feel comfortable.
- Treat me with respect.
- Show interest in my ideas about my health.
- Understand my main health concerns.
- Pay attention to me (look at me, listen carefully).
- Let me talk without interruptions.
- Give me as much information as I want.
- Talk in terms I can understand.
- Check to be sure I understand everything.
- Encourage me to ask questions.
- Involve me in decisions as much as I want.
- Discuss next steps, including any follow-up plans.
- Show care and concern.
- Spend the right amount of time with me.

Makoul G, Krupat E, Chang CH. Measuring patient views of physician communication skills: development and testing of the Communication Assessment Tool. *Patient Educ Couns.* 2007;67:333–342.

Time efficient

Effective communication does not take longer than ineffective communication. Good communication is not an additional task, rather it is how the tasks required for quality care are enacted. A core task for the doctor is to establish relevant clinical details, and there are different ways to approach this. Evidence has shown that listening to the patient's reason for attendance (which generally takes less than a minute) and then checking that the patient has finished ('Is there anything else?') is the most effective way to gather full and accurate information, facilitating accurate diagnosis and appropriate management. In contrast, interrupting the patient after the first symptom to ask a series of questions leads to omissions and the possibility of time spent on low-priority symptoms. This results in poorer quality of information and care. Inaccuracy of diagnosis and the resultant inappropriate treatment plans are costly in terms of time and resources and can often be traced back to communication.

Efficiency relates to how information is gathered. A checklist of closed questions that limit answers to 'yes' or 'no' responses is less efficient than open questions, which provide space for the answer. Asking closed questions may seem time-efficient, as topics can be cycled through quickly and the patient spends less time talking. However, evidence shows otherwise. Using open questions increases the doctor's ability to obtain complete diagnostic information and target discussion to priority topics. A balance of types of questions is the most time-efficient. Communication is time-efficient when a clinician can:

- share comprehensible information
- work with the patient to develop a feasible plan that meets patient need
- coordinate with colleagues to share accurate information
- record information accurately.

Communication that is incomprehensible, jargon-laden, and inadequately recorded is costly in terms of time and resources and leads to poorer health outcomes.

Patient-centred

From the patient's perspective, a consultation may represent only a fragment of the care they experience. In contemporary healthcare, the patient's journey can involve multiple meetings with different, multidisciplinary healthcare professionals on different days and in different settings (see **Fig. 2.1**). This is an increasing trend as more people are living longer with multiple medical conditions. The patient is the only person present in all their consultations and must piece the jigsaw together in order to make informed decisions about their care.

Communication needs to:

- provide an overview of the process of care
- orient the patient as to where they are at each stage

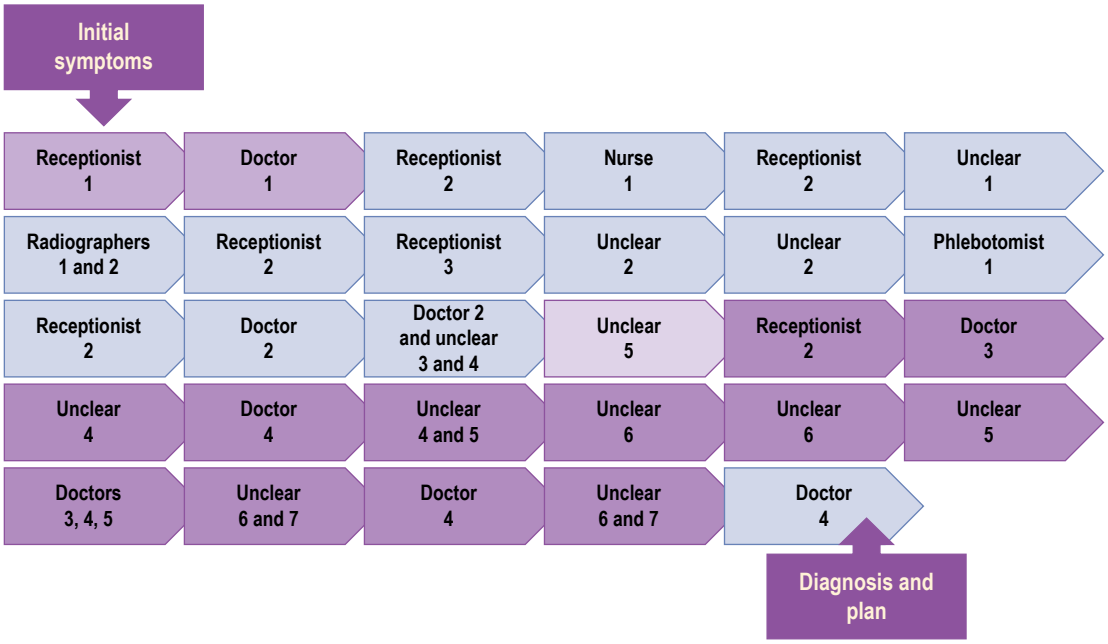


Fig. 2.1 An episode of care. This figure summarizes a patient's episode of care to establish *whether a symptom indicated that the patient had cancer.* It comprised 29 clinically relevant interactions between the patient and 18 different healthcare staff, of whom 4 were doctors. The episode comprised 5 events, on different days, with between 1 and 13 interactions per event. Three events were by phone (1st, 3rd, and 5th) and two (2nd and 4th) were in outpatient settings. The final event concluded the episode.

- foster self-efficacy and agency
- check that care meets need, which may change during the process
- empower the patient to engage in decision-making.

If the doctor's focus is on the immediate 'task at hand', with brief acknowledgement of previous and next steps, it can be unclear when, or whether, this necessary communication happens.

Language

From the first day at medical school, learners become immersed in a biomedical health service culture with its own language. This includes verbal communication (nouns, verbs, turns of phrase) and the way in which language is used to characterize the world and the people within it. This language is used both with and about patients and has a pervasive effect on doctor–patient communication. Traditional medical language, still used in contemporary healthcare, often relegates the patient to a passive role: the doctor **takes** a history, and the patient is expected to **comply** with treatment (Box 2.3). Such language reveals the underlying values of healthcare and reinforces these values in practice.

People accessing healthcare are confronted with strange medical language in all medical settings: on signs, on walls, and spoken by staff, rendering hospital visits akin to visiting a foreign country. This unfamiliar language is ubiquitous in medical settings: on signs, on walls, and spoken by staff, rendering hospital visits akin to visiting a foreign country. Language discordance is the difference between the patient's language and the language used in a consultation, and it applies equally to the language of health services as to discordance in national tongue. The language in medical consultations directly affects clarity and comprehensibility. Doctors' communication is imbued with medical jargon. There are words that the patient may have heard before but may not fully understand *('biopsy')*, phrases that have a different meaning in medicine *('the test results were positive')*, euphemisms that obscure meaning *('lesion')*, and language that is scary *('invasive procedure')* (Box 2.4).

Box 2.3 Examples of how traditional medical language can be rephrased to better represent patient-centred care

Traditional medical language	Patient-centred alternative
Doctor tasks	
Takes a history	Elicits a patient's medical history
Consents the patient	Records the patient's treatment choice
Deals with emotions	Responds to the patient's emotions
Manages the patient	Provides treatment
Discharges the patient	The patient leaves hospital
Patient tasks	
Gives consent	Makes a choice
Complies with treatment	Takes treatment
Tolerates treatment	Does not experience unacceptable side-effects
Description of patient	
This elderly primigravida	Mrs Olivia Leonard, who is pregnant for the first time
This multimorbid obese 42-year-old	Ms Rowena Rios, who is living with obesity and other long-term medical conditions
This poorly controlled diabetic	Mr Mason Harper, who is living with diabetes and has a high HbA1c

Box 2.4 Types of medical jargon

Type	Content	Examples
Technical terminology	Disease names, symptoms, anatomy, procedures, tests, treatments	*Adenocarcinoma, teratogenic, abdominal wall, endoscopy, stoma, radical prostatectomy, cholecystectomy*
Alphabet soup	Acronyms and abbreviations	*INR, CABG, BP, DNAR, DVT, PRN, U&E, TTA, PR, IM, AF, IV*
Medical vernacular	Words that seem familiar, but medical meaning may not be understood	*Biopsy, sepsis, sexually active, steroids, tumour, best interests, palliative, do your bloods, first-line treatment, conservative management*
Medicalized English	Words that have a different meaning in medicine	*Capacity, disease, chronic, drugs, theatre, occult, benign, F1* Includes 'accidental antonyms': *'positive' result, progress (tumour is progressing), advance (cancer is advanced)*
Unnecessary synonyms	Overcomplicating by replacing universally understood terms	*Upper/lower extremity, futile, take orally*
Euphemisms	Attempts to soften language that provide less clarity	*Shadow on X-ray, mass, lesion, growth, stool*
Judgmental jargon	Phrases appear derogatory or reflect bias	*Complains of, refuses treatment, denies, failure to progress, acopia*
Scary jargon	Words that imply violence or invoke troublesome imagery	*Strangulated, invasive treatment, organ harvesting, expose the abdomen*
Headings jargon	Tasks are presented without translation	*Does anything alleviate or exacerbate the pain? Is there anything associated with that? Does the pain radiate? We need to determine capacity.*
Depersonalizing jargon	Language that does not treat the patient as a person with agency	*You were diagnosed with, admit you, manage you, discharge you, send you for a procedure*
Official jargon	Presents the doctor as a detached professional, not in collaboration with the patient	*Most cases of, many patients with, upon further investigation, because of confidentiality issues, I must adhere to professional guidance*

Adapted from Pitt MB, Hendrickson MA. Eradicating jargon-oblivion: a proposed classification of medical jargon. *J Gen Intern Med.* 2020; 35(6):1861–1864.

The way doctors use medical jargon in the consultation can cause problems in communication. For example, a consultation that begins with, 'The biopsy showed non-Hodgkin lymphoma. Have you heard of that?' is foreign and does not provide the required information *('is it cancer?')*. The way the information is presented also implies that the patient should understand, and the preferred answer (following normal conversational rules) to such a closed question is: 'Yes'. The patient, already worried, now feels further disempowered. Explanations that inform and empower the patient start with what the patient already knows and take place in the language that they speak (lay language). This is termed 'given before new'. Providing the patient with the medical name of the condition/test/treatment is helpful (and empowering) in the right place in a consultation, such as following an explanation ('…this is called non-Hodgkin lymphoma'). When considering whether communication is respectful, clear, compassionate, and empowering, language must be considered in its entirety, not just according to the presence or absence of isolated jargon.

Further reading

Academy of Medical Royal Colleges. *Please, Write to Me: Writing Outpatient Clinic Letters to Patients, Guidance.* 2018; https://www.aomrc.org.uk/reports-guidance/please-write-to-me-writing-outpatient-clinic-letters-to-patients-guidance/.

Cox C and Fritz Z. Presenting complaint: use of language that disempowers patients. *BMJ* 2022; 377:e066720.

NHS England. *Language Matters: Language and Diabetes.* 2018; https://www.england.nhs.uk/wp-content/uploads/2018/06/language-matters.pdf.

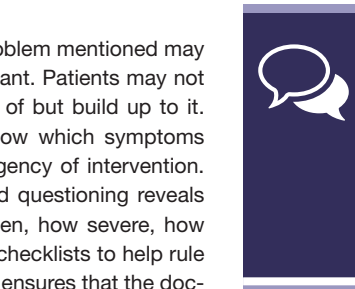
Pitt MB and Hendrickson MA. Eradicating jargon-oblivion—a proposed classification system of medical jargon. *J Gen Intern Med* 2020; 35:1861–4.

LEARNING CLINICAL COMMUNICATION

The development of effective clinical communication skills is a lifelong process, starting in medical school and ongoing throughout professional practice. It is iterative: building knowledge and skills and revisiting them over time. It draws on the learner's existing social and communication skills and emotional intelligence, as well as their lived experiences (e.g., cross-cultural communication, grief, parenthood, making difficult choices). The purpose of communication education is to enable the learner to use their natural skills to best effect in institutional communication. Clinical communication has a direct effect on healthcare outcomes with potentially long-lasting implications: consultations qualify as high-stakes interactions.

Key domains for learning include the following:

- *Core principles for an effective medical consultation.*
  Understanding the key features of effective communication and having the skills to achieve them. This includes building a working relationship, establishing the purpose of the meeting, developing a shared agenda, exchanging information effectively, supporting decision-making, and providing emotional care.
- *Responding to specific situations.*
  Having frameworks for conducting effective consultations in a variety of situations, including routine, sensitive, and difficult issues.
- *Being responsive.*
  Adapting communication style to meet the diverse needs of patients, relatives, and colleagues, in different settings and through different media (e.g., face to face, telephone, written, online).
- *Communicating beyond the patient.*



Communicating effectively with colleagues and healthcare systems.

Evidence-based core skills for some tasks are shown in Boxes 2.5–2.13.

Initiating a consultation

Investing in the beginning is important (**Box 2.5**). At the beginning of a consultation, the patient expresses their needs, helping the doctor understand what to address. A social greeting that includes introductions by name signals that the patient is a person (not just a body or medical condition). This is not the same as confirming the patient's name and date of birth. The successful *#hellomynameis* campaign, which encouraged all healthcare staff to introduce themselves, was initiated by a doctor who was also a patient who was terminally ill with cancer, who was shocked by the frequent absence of this basic component of communication.

Ending a consultation

The closing phase enables the patient and doctor to check that the purpose has been accomplished and confirm the decisions that have been agreed (**Box 2.6**). The time required will vary depending on the complexity of the content. Concise reiteration of important points (key information, decisions made, agreed next steps) capitalizes on the recency effect in memory, which aids recall. It also acts as a safety net to prevent misunderstanding. This is important for patients, often unfamiliar with medical settings, for whom information exchanged can be difficult to recall.

Gathering information

Gathering information to establish patient needs is best done in two stages: establishing the full problem list, followed by detailed questioning (**Box 2.7**). The first symptom or problem mentioned may not be the only problem, nor the most important. Patients may not start with the symptom they are most afraid of but build up to it. Furthermore, patients do not necessarily know which symptoms are diagnostically useful or determine the urgency of intervention. Once the problem list is established, detailed questioning reveals the comprehensive picture: what, where, when, how severe, how long, and how often. The doctor may deploy checklists to help rule in or out diagnoses. This two-stage approach ensures that the doctor captures the full problem, avoiding omissions through premature selective questioning. Sometimes called the 'open to closed question cone', this approach is reassuring to patients who feel listened to in the first stage and that the doctor is being thorough in the second stage. It is also effective in gathering information.

Sharing information

Patients generally expect information about (1) what is wrong, (2) what can be done, and (3) what the outcome will be. It helps to organize a consultation structure around these points. However, the patient may also have specific concerns; for example, whether the condition is serious or how soon they can get back to work, and more. Tailoring

Box 2.5 Core skills for initiating a consultation

- *Physical setting:* Ensure a suitable space with sufficient privacy to maintain confidentiality and dignity, with extraneous noise and distractions minimized.
- *Sit down:* Be at the same eye level, ideally both be sitting down. Do not stand over the patient talking down onto them.
- *Introductions:* Greet the person by their preferred name. Explain your name and role.
- *Agenda setting:* Ask the patient's aims for the consultation, explain your aim, and agree on joint aims. Clarify what will be covered and how (expected duration and format).
- *Establish rapport:* Begin to build a professional relationship, signal your curiosity, develop rapport, and use empathy.
- *Orient the patient:* Show consultation structure, e.g., when moving on to the next stage, signpost the transition.

Box 2.6 Core skills for ending a consultation

- *Orient the patient:* Signal transition to the closing stage; indicate that you have covered everything you need to and check whether there is anything more they need to add.
- *Revisit the purpose:* Check that the agenda has been addressed and note any aims not yet been accomplished.
- *Summarize:* Outline key information exchanged, decisions made, and plans agreed.
- *Next steps:* Ensure that what happens next is clear.
- *Close:* Thank the person and end the meeting.

Box 2.7 Core skills for gathering information

1. *Establish the full problem list*
   - Start with an open-ended question: ('Can you tell me why you've come to the clinic today?')
   - Listen in full to the person's opening statement.
   - Encourage a full description of the problem. Reflect back: ('…going on for a month', 'That's worrying you'). Use facilitative statements: ('Tell me more').
   - Listen actively and signal attention: eye contact, nodding head.
   - Encourage the person to hold the conversational floor. Use minimal encouragers: ('Mm-hmm.')
   - Focus on open-ended questions to achieve an overview.
   - Be careful about using the word 'symptom': ('Is there anything different that you have noticed about your health?', 'Are there any other problems you would like to mention?')
   - Note each problem to keep track without interrupting the flow.
   - Ask if there is anything else. Keep asking until the person says no.
   - Clarify the patient's goals.
   - Summarize key points.
2. *Detailed questioning*
   - Use signposting to orient the patient as to what you are going to ask about and when you are changing the topic.
   - Use a balance of open and closed questions.
   - Avoid jargon. Clarify jargon used by the patient or any part of their description that is not clear.
   - Adapt to the number of questions: different people require different numbers of questions to give you the same information.
   - Avoid leading questions *('You've never had this before?')* and multiple questions *('Have you ever had epilepsy, diabetes, rheumatic fever, stroke, heart attack, or any serious childhood illnesses?').*
   - Remember that people need time to digest each question and time to answer.
   - Listen to the answers.
   - Show receipt of information: acknowledging what the patient has said, reflecting it back, summarizing from time to time.
   - Before asking questions about sensitive issues or issues the patient might not expect to be asked about, give a brief explanation of why you need the information. Do not say that all patients are asked these questions unless it is true.
   - Identify the patient's perspective and goals and include worries and what they hope will be achieved by treatment.

information to the patient's starting point and responding to specific concerns are key. Sharing information is a process most effective as a dialogue, rather than a download (**Box 2.8**).

UK guidance states that doctors must:

- Give patients the information they want or need in a way they can understand.
- Work in partnership with patients, sharing the information they need to make decisions about their care.

It is helpful for the doctor to appreciate that:

- much of the information they are sharing (diagnosis, treatment, prognosis) will be new and unfamiliar to the patient
- medical information is often complex (new concepts and language used)
- the outlook may be uncertain and involve risk; there is rarely a straightforward route that provides a complete cure with no downsides
- being ill and navigating healthcare is stressful, generating fear, worry, and other emotions; this can make hearing and using information difficult

Box 2.8 Core skills for sharing information

Plan before the consultation:

- Know what information you need to share.
- Consider how to phrase it.
- Anticipate what information the patient may want and questions you might be asked.
- Identify a strategy to support the patient in recall, e.g., having a relative or friend present, recording the consultation, using a trusted resource.

During the consultation:

- Signpost the reason for the meeting and agree on shared aims for the consultation.
- Determine what the person knows at this point and what they want to know prior to any explanation.
- Give an overview of the agreed agenda, e.g., you are going to explain what the problem is, treatment options, and likely outcomes. Highlight if decisions are to be made in the consultation.
- Encourage the patient to guide the consultation by asking questions throughout (not just at the end).
- Given before new: start with what the person already knows and build up from there.
- Provide information about conditions, tests, procedures and treatments in lay terms first, then append with medical names. Write down or record medical names for them to take away.
- Ensure your language is understandable.
- Give information in small chunks and check that the information is clear and meets patient needs.
- Offer to repeat any or all information at key points, e.g., at the end of each segment and/or at the end.
- Check throughout that explanations are clear. Use non-verbal clues and ask directly: *('Would you like me to go over any of that again?')* Do not use loaded questions: *('Do you understand?')*
- Use signposting throughout to keep track of where you are in the consultation plan.
- Make sure the patient has all the information they need about the diagnosis before discussing treatment.
- Respect the patient's right to information; do not withhold information because you are worried it might upset the patient.
- Listen carefully to concerns and queries and respond fully to these.
- Towards the end, ask the patient to tell you what they have understood *('Would you like to go through how you will explain this to your partner/family when you get home?').*
- At the end, summarize the agreed plan including immediate next steps. If a plan has not been achieved, summarize to the point you have reached.

- the patient must make the right decision for themselves, often in an unfamiliar environment and sometimes under time pressure
- responding to patients' concerns and queries is integral to dialogue; a treatment plan requires understanding and agreement.

Discussing uncertainty and risk

Outcomes may be uncertain, and some treatments carry the risk of harm through side-effects or complications. Patients may therefore have to make decisions based on incomplete or probabilistic information including:

- a balance of likely benefits and risks for the individual
- the severity of any harmful effects.

There are different types of uncertainties that require different approaches:

1. Uncertainty that can be resolved.

   Explain what can be done to get the required information; for example, benign and malignant cancers distinguished by biopsy.
2. Uncertainty that cannot be resolved.

   It is not possible to know what will happen when the decision is made; for example, a choice of treatment options that may or may not provide a cure. It is important to acknowledge the stressful nature of uncertainty and to provide emotional support.

Uncertainty and potential harm are unsettling and can be distressing. People often hope for certain news of a positive outcome, and its absence is a form of bad news. Skills for discussing uncertainty and risk are required (**Box 2.9**).

Shared decision-making

Shared decision-making (**Box 2.10**) enables a person to develop an informed preference. People are often unused to making healthcare decisions. They may have expectations about what is to happen, and/or preferences about what they would like, but an informed decision cannot be made without the details of options and consequences. An informed preference is the patient's choice after being given the information they want and need, and the opportunity to consider implications in the context of their own circumstances and values. Depending on the complexity of the decision, this process may happen during a single consultation, across more than one

Box 2.9 Core skills for discussing uncertainty and risk

- State clearly that there is uncertainty and/or risk.
- Outline the decision clearly.
- Acknowledge that deciding in the face of uncertainty or risk is difficult and stressful.
- Offer partnership and support.
- Itemize options explaining likely benefits, harms, and uncertainties.
- Present information about risks and probabilities clearly:
  - Use natural frequencies not percentages, e.g., '3 out of 10 people' (not '30%').
  - Use small denominators where possible, e.g., '1 out of 10 people' (not '10 out of 100').
  - When comparing options, present 'like with like' comparisons, e.g., '2 out of 10 compared to 4 out of 10' (not '2 out of 10 compared to 2 out of 5').
  - Avoid relative risk (e.g., 'twice as big').
  - Be wary of describing risks only using words (such as usually, rarely, normally, or sometimes).
- Describe risk information about benefits and harms neutrally; avoid positive or negative bias.
- Supplement explanations with pictures such as icon arrays.
- Provide access to resources that the patient can take away.

consultation, or during blended care with access to decision aids either before, during, or after consultation. Decision aids/support tools are resources developed to support shared decision-making. They contain three components:

1. A clear statement of the decision to be made
2. Evidence-based information about options, benefits, harms, and uncertainties
3. Help to enable the person to clarify their values and preferences

Patient autonomy is facilitated by providing the required information about available options and supporting the patient in navigating options to make their choice. There can be a misalignment between a person's wish to decide and their confidence to do so, for example, due to an unfamiliar environment, new and complex information, worry about health, and worry about the 'right' decision. Doctors may query how to ensure autonomy when a patient is hesitant about deciding and wants to defer to a doctor's opinion. Most patients do want a professional opinion; the consultation is a meeting between experts (**Box 2.1**). If the patient concurs with the doctor's recommendation, it is not a failure of patient autonomy provided the patient has made an informed choice. Likewise, if the patient disagrees, it is not a failure of care provided when the patient has made an informed choice. More deliberation and support may be required if the choice is not straightforward. People may be more likely to defer to a clinician when seriously ill, but the patient must live with their decision in the long term, and it must be right for their needs. Uncertainty and fear reduce a person's confidence in decision-making; time and support increase confidence. Providing a support structure for decision-making (e.g., by using a decision aid and ensuring that the patient is not rushed) leads to better information and a choice that aligns with what matters most to the patient. Support is integral to this process; decision-making is not just about information. Patients appreciate reassurance that they do not have to make sense of options by themselves and decide on their own. When there is no treatment option that can deliver the goal the patient wishes, emotional support, acknowledgement, and empathy are essential.

Breaking bad news

Doctors have to share information that has a profound effect on patients, including many different forms of significant loss: news about death, serious illness, and life-changing injury. Patients and families need:

- information that is clear, personalized, and complete
- emotional support.

Box 2.10 Core skills for shared decision-making

Establish the decision-making team:
- Signpost: Explain the decision to be made.
- Support: State that you will work together to make a decision that is right for the patient.
- Explore: Find out what the patient knows and wants to achieve, and what questions or concerns they have.

Discuss the options:
- Share: Provide information about treatment options, potential benefits, potential harms, and uncertainties. Include the option of doing nothing.
- Supplement: Consider a written summary of information or decision support tool or decision aid.

Decision-making:
- Support: Ask the patient about their priorities and needs *('What matters most to you?')*, their thoughts and preferences based on the information provided, and their readiness to decide.

The principles of effective information sharing are amplified when the news is not good (**Box 2.11**). These consultations need more preparation, more attention to broaching the news, a slower delivery, continued vigilance about clarity, and a heightened sensitivity to emotions. The principle of patient autonomy applies; the patient has the right to be given information and to make decisions. Information should not be withheld because it might be upsetting (or because it is uncomfortable to deliver). In many consultations, including bad news consultations, patients and those close to them expect to be told key information without having to ask; it is important not to rely on the assumption that if a person has not asked about a particular topic (such as prognosis), they do not want to know about it.

Patients who receive bad news may remember little else about the rest of the consultation. The more serious the news, the less

Box 2.11 Core skills for breaking bad news

Prepare:
- What information do you have to give:
  - What is the bad news?
  - What are the implications for the person and their family?
  - What is the 'big picture', the overview that you need to convey?
  - What is the 'detail' that you will cover later?
- How will you phrase it?
- Anticipate questions.
- Prepare the setting: private, quiet, free from distractions and interruptions.

Broach:
- Establish the purpose of the consultation and signal its significance.
- Find out what the person already knows and what they want to know.
- Signal that there is bad news by using a warning shot *('The results of the tests are not good'; 'I'm afraid that it is serious')*.

Share:
- Given before new: start with what the patient already knows and provide new information one step at a time.
- Signpost each stage: bad news, implications, big picture, detail.
- 'Chunk, check, pause': give information in pieces, check that the pace and pitch of the explanation meet the person's needs, and pause after every segment leaving longer gaps than usual.
- Use clear words: *cancer, death, cannot be cured, not long to live.*
- Avoid jargon or euphemisms: *malignancy, tumour, lesion, shadow, not curative, palliative.*
- Check if a person is ready for a particular topic: *('There is more information I can give you about what will happen in the future. Would you like to talk about that now?')* Avoid vague questions: *('Are you the sort of person who would like to have all the information?').*
- Answer direct questions with direct answers.
- Be alert to cues that the person wants more information: indirect questions (which may be expressed as statements), expression of concern or other emotion, *('The doctor last time said that if it came back again it wouldn't be good news'); ('I'm worried about what this might be').*
- Give the person time to take the information in and time to be upset.
- Patients and relatives appreciate reassurance and hope, which must be honest.
- Sitting in silence with the person can be supportive.
- It is appropriate to say: *'I'm sorry'*. Do not say: *'I'm sorry to have to be the one to tell you'* (as this is about your needs, not theirs).

Plan:
- Offer time: ('We don't have to make a decision right now').
- Be responsive to cues about when to end the consultation; be flexible about what is covered.
- Keep any plans discussed clear and simple.
- Write information down.
- Provide contact information.

information a doctor will be able to provide in a single meeting. This impacts decision-making. Many patients are not in a position to make informed decisions immediately after receiving bad news. The patient may have heard the information but may not have fully processed it or be able to use it effectively. The doctor can provide support and reduce pressure by encouraging the person to take time and discuss the news with those close to them. However, this is not absolute; some patients may be immediately ready to make plans, either because they were expecting the news, or because the full emotional impact is delayed. Emotional reactions are not immediate and predictable but varied, including all kinds of emotions and none at all. The full emotional impact may come later, over time, rather than at a single point.

Other difficult conversations

Breaking bad news is typically viewed as the most difficult conversation a doctor may have, but other challenging topics include stigma, trauma, death, bereavement, and end-of-life care (**Box 2.12**).

These consultations can be difficult or sensitive due to the following:

- Topics that can elicit strong emotions, including shame, grief, fear, anger, and guilt. For example, sexual abuse, termination, terminal illness, addiction, sexual health, sexuality, domestic violence, suicide, obesity, and medical error (**Box 2.13**).
- Circumstances of the consultation. For example, absence of trust, worry about confidentiality, assumptions or judgment, or a doctor who is ill-equipped to respond to patient needs.

People can feel vulnerable when disclosing sensitive topics. The doctor may be the first person to whom the topic has been disclosed. In these conversations, it is particularly important to develop trust and rapport and to ensure that the patient feels safe, supported, and not judged. The basic elements of difficult conversations are the same as for any doctor–patient interaction: establish patient needs, listen, foster a respectful partnership, and work together to devise a plan. However, in these consultations, certain elements may be modified; for example, the patient may need more time for disclosure, or there may need to be a focus on listening and exploring to avoid premature solutions.

Remote consulting

Delivery of healthcare through remote means (telemedicine) has increased in recent years, in particular the number of doctor–patient consultations conducted at a distance, including video conferencing, phone, or written communication (email, messenger services, online platforms).

Remote consulting can be conducted:

- synchronously (in real time), e.g., by phone.
- asynchronously (not at the same time), e.g., by email.

Consultations can be conducted using one or more of the following methods:

- simple (single method), e.g., phone.
- augmented (more than one method), e.g., phone, with the patient sending a photo of a rash by email.

The question as to which is better, face-to-face or remote consulting, is a false dichotomy, as it depends on circumstances, including:

- the patient's clinical need
- whether physical examination is needed
- availability of medical records
- whether the care required can be provided through that method, e.g., sharing of information, tests, and treatments.

Box 2.12 Example framework: discussing end-of-life care

Before the consultation:	
Consider potential topics	Pain, dignity, the experience of dying, the effect of bereavement on loved ones.
During the consultation:	
Develop trust and rapport	Signal willingness to talk about issues that matter to the patient. Reassure about confidentiality. Spend time listening. e.g., *('You said you are anxious about what will happen nearer the time. Would you like to talk about that?')*
Pick up cues	Reflect back topics raised e.g., *('You said that you have never been there when someone has died. Are you worried about what dying will be like?')*
Explore	Ask about: • questions the patient has • worries • hopes • plans in the time left • who supports them • e.g., *('Have you talked with your family about who will be there with you?')*
Respond	• Listen and provide space for the patient to think things through (without providing advice) • Acknowledge difficulties or dilemmas • Provide/direct to information and/or practical support • e.g., *('You want to talk about dying, but you do not want to upset your family').*
Signal support	• Normalizing: signal that others have had the same experience. • Show that you are non-judgmental. • e.g., *('Many people worry that medication for the pain might affect their thinking'), ('You are angry that this has happened and that your life is being cut short.')*

Care may start with one method and then be augmented or switched. It may become apparent during a remote consultation that a physical examination is required, or care provided initially in a healthcare setting may be successfully followed up remotely. Neither face-to-face nor remote consultation is preferred by patients. Patients value choice, and what matters most will depend on the individual situation, including convenience, time off work, comfort in discussing sensitive issues, and privacy. Likewise, suitability varies, depending on the patient's digital access, sensory or cognitive impairments, and learning needs. Methods of remote consulting have advantages and disadvantages:

- Telephone is readily available, portable, and convenient but provides impoverished non-verbal communication.
- Video conferencing allows non-verbal communication, but it is more prone to technical and privacy issues and not everyone has access.

Box 2.13 Example of a framework for discussing medical error

Initiating the meeting
- Explain the purpose: to discuss a problem that has occurred with the patient's care.
- Listen to the patient's concerns about the care received.

Explanation
- Disclose clearly that a mistake or error has been made (not a 'complication', 'problem' or 'adverse event').
- Accept responsibility (do not be defensive, do not blame other staff or the patient's condition).
- Say sorry for the event; do not 'express regret' or say *'I am sorry to be the one to tell you.'*
- Provide a clear explanation of what happened.
- Explain the consequences of the mistake or error, both short term and long term.
- Explain what will be done to rectify the patient's situation.
- Provide reassurance about the steps being taken to ensure that the patient's future care will not be compromised.
- Explain what has been done to ensure the same mistake will not happen to anyone else.
- Provide clear information about complaint procedures and timescale.

Throughout the meeting
- Provide information that is honest, truthful, clear, and jargon-free.
- Answer questions honestly and in full.
- Empathize and acknowledge emotions that the patient expresses.
- Remain calm and professional.

Closing
- Check that the explanation has addressed the patient's needs.
- Summarize and agree on a plan.

- Written methods are convenient and give time to patients and doctors to prepare responses but may result in loss of nuance or focus or important omissions.

People communicate differently remotely, which is not inherently better or worse in terms of accomplishing the aims of doctor–patient communication. For example, phone consultations deprive the participants of eye contact and non-verbal cues. However, this may be compensated for by other means; for example, asking about the impact on the patient's life. Likewise, a patient may feel more comfortable discussing a sensitive issue in familiar surroundings. Remote conversations are often shorter than face-to-face equivalents due to natural differences in pacing and interaction style. This is not a problem in itself. However, shortcuts in care (e.g., a less thorough assessment of the patient's problem) do compromise quality. The key is to be vigilant to changes in communication and their impact (**Box 2.14**).

Electronic written communication tends to be used for:
- straightforward discussions: minor problems and reviewing existing treatment
- exchanging routine information: repeat prescriptions and routine results.

Written communication has the immediate advantage of being self-documented, so patients and doctors can refer back to the information. It can improve outcomes by enabling information exchange beyond the single 'event' of a consultation; for example, asking for additional information, confirming details and providing instructions. Helpful 'ground rules' include:
- an agreed turnaround time
- the types of topics it is to be used for, and when this mode of consulting is not appropriate
- how to seek help for urgent problems
- procedures to ensure confidentiality.

Box 2.14 Core skills for remote consulting

Before the consultation
- Check that the method selected is appropriate for the patient's needs, e.g., closed captioning for a patient with hearing loss.
- Ensure that the patient is given clear information about how to prepare including timing, duration, format, privacy and how to be 'called in'.

During the consultation
- Check technical quality: sound, picture, and time lag.
- Check confidentiality: does the patient have a private space? Is there anyone else in the room, e.g., a family member or children?
- Check consent: is the patient able to agree to a consultation in this format?
- Be aware of any differences in communication: pace, turn-taking, non-verbal communication, cues, and emotional tone that affect the achievement of the consultation aims.
- Consider how to support the patient to recall and use information in the absence of usual resources like information leaflets.

People also communicate differently in written form. Be mindful of the tone, clarity, level of language, and the way emotions may be expressed.

Further reading

Elwyn G, et al. A three-talk model for shared decision making: multistage consultation process. *BMJ* 2017; 359:j4891.
General Medical Council. *Remote Consultations.* 2023; https://www.gmc-uk.org/ethical-guidance/ethical-hub/remote-consultations.
#Hellomynameis: a Campaign for More Compassionate Care. https://www.hellomynameis.org.uk.
Lloyd M et al. *Clinical Communication Skills for Medicine,* 4th edn. Edinburgh: Elsevier; 2018.
Noble LM et al. Consensus statement on an updated core communication curriculum for UK undergraduate medical education. *Patient Educ Couns* 2018; 101:1712–19.
NHS England. *Principles for Supporting High Quality Consultations by Video in General Practice During COVID-19.* 2020; https://www.england.nhs.uk/coronavirus/wp-content/uploads/sites/52/2020/03/C0479-principles-of-safe-video-consulting-in-general-practice-updated-29-may.pdf.
NHS England. *Key Principles for Intimate Clinical Assessments Undertaken Remotely in Response to COVID-19.* 2020; https://www.gmc-uk.org/-/media/files/key_principles_for_intimate_clinical_assessments_undertaken_remotely_in_response_to_covid19_v1-(1).pdf.
Stacey D et al. Decision aids for people facing health treatment or screening decisions (review). *Cochrane Database Syst Rev* 2017.

Communication with colleagues

Communication about patient care with colleagues is as important as communication with patients. It facilitates the exchange of the right information, ideas and plans at the right time with the right people. Healthcare relies on communication between healthcare staff and within healthcare systems. Both are needed for safe and effective care.

Communication in healthcare systems

Systems must meet requirements for accurate and timely information exchange and recording and retrieval of information across healthcare providers. Contemporary healthcare is multidisciplinary and delivered over time. Delays, omissions, and loss of information lead to unsafe care, medical errors, and complaints. The system may fail to meet its goals due to:
- system design: lack of clarity in how healthcare teams share information
- system operation: inadequate technology, staff shortages, and delays, leading to additional work to compensate for system failures.

24 Clinical communication

Communication between healthcare staff

The principles of effective communication apply to interprofessional communication. Colleagues want and need to be listened to, to be respected, and to have self-efficacy. Communication is most effective when there is:

- a clear shared agenda
- a common language
- orientation through signposting
- exploration that establishes a full problem list
- sharing of information by 'chunking and checking' and summarizing
- an agreed plan.

Communication problems between colleagues can be traced back to a failure of these principles. Similar to doctor–patient communication, communication with colleagues is a form of institutional communication. Structured protocols adopted for these interactions improve accuracy and efficiency.

SBAR communication tool

This provides a structure for colleagues to exchange information about patient care, which can be used within and between healthcare teams and with external agencies. Necessary information is shared in four domains:

S—Situation: the problem or concern
B—Background: significant information from the medical history/admission
A—Assessment: clinical impression and test results
R—Recommendation: what is needed

These domains can also be used to receive information by using questions to elicit each domain.

WHO Surgical Safety Checklist

A structure to ensure that key tasks are completed during surgery/procedures. It includes three stages:

- Before anaesthesia: confirmation of identity, site marking, machine and medication check, allergies, and airway and blood loss risks
- Before an incision: team members, patient name and procedure, antibiotics, anticipated critical events, and imaging
- Before the patient leaves the operating room: instrument count, specimens, and key concerns for recovery

These protocols work to improve the quality and safety of care by:

- providing a structure that is clear to all participants and
- ensuring that all participants are heard

which are important features of all healthcare communication.

Further reading

NHS England. *SBAR Communication Tool: Situation, Background, Assessment, Recommendation.* 2018; https://www.england.nhs.uk/wp-content/uploads/2021/03/qsir-sbar-communication-tool.pdf.
World Health Organization WHO. *Surgical Safety Checklist.* https://www.who.int/teams/integrated-health-services/patient-safety/research/safe-surgery/tool-and-resources.

Questions for this chapter can be found on Elsevier eBooks+.

3 Medical ethics

June Jones

Ethical principles of medicine 25
Duties 25
Consequences 26
Values 26
Frameworks and tools 26
Moral injury and the ethics of self-care 27
Professional guidance and codes of practice 27

Ethics of inequality 28
Principles of equality, diversity, and inclusivity 28
Legal and ethical issues in medical practice 29
Respect for autonomy: capacity and consent 29
Confidentiality 32
Professional competence and mistakes 33

CORE SKILLS AND KNOWLEDGE

The practice of medicine is a human and moral endeavour concerned with forming therapeutic relationships with patients. It is crucial for everyone practising medicine to understand the moral and legal framework that underpins and regulates their work. In clinical practising, ethical and legal decisions are made daily by clinical teams working with patients and, where appropriate, involving their families. Increasingly, hospitals have clinical ethics committees to discuss complex ethical cases. Hospitals may also seek a legal opinion for more challenging cases that involve a point of law. If situations are contested and disagreement seems intractable, the courts will occasionally be asked to give a decision about whether a planned course of action is lawful or unlawful. All clinicians need to understand key ethical and legal principles to ensure that their clinical practice meets the standards of the profession.

Key learning in ethical practice includes:

- Recognition of the need for ethical judgement
- Identification of ethical and legal issues in modern clinical practice
- Identification of the relevant ethical values and principles from the perspective of each stakeholder
- Application of ethical values and principles in clinical decision-making.

ETHICAL PRINCIPLES OF MEDICINE

The subject of ethics incorporates knowledge, cognitive skills such as reasoning, critique, and logical analysis, and skills for integrating and applying abstract ethical learning with clinical learning. Students and clinicians need skills and knowledge in ethics to enhance daily practice and better serve their patients.

The practice of medicine is inherently moral:

- Biomedical expertise and clinical science must be applied by and to people
- Clinical decisions are underpinned by values and principles
- Potential courses of action will have implications and consequences that are often uncertain
- Technological advancements sometimes have unintended or unforeseen consequences
- Patients and clinicians come from diverse backgrounds, creating potential for different moral values and divergent opinions

Ethical reasoning involves thinking through the available options and weighing the relative merits logically and thoroughly. Becoming, and practising as, a clinician demands an awareness of, and reflection on, ethical attitudes. Each of us has personal ***values*** and moral intuitions. In the field of ethics, a necessary part of learning is becoming aware of the assumptions on which these personal values are based, reflecting on them critically, and listening and responding to challenging or opposing beliefs. There are some key ethical concepts that can help with this process.

Duties

The General Medical Council (GMC) creates the moral framework for medicine in *Good Medical Practice*. It provides a list of duties a doctor registered with the GMC is expected to uphold. Although the list is pragmatic, the **duties** of a doctor describe ethical practice including truth-telling, maintaining confidentiality, and acting in a trustworthy way. These are duties that have traditionally been accepted as part of the core values of the profession, duties that all doctors ascribe to by virtue of taking up the role of being a doctor. We should be able to justify these duties independently of any good consequences that come from having such duties, for if we rely on consequential justification, and the consequences can be framed negatively, then duty dissolves. For instance, in the paternalistic practice of medicine, the duty to tell the truth was often set aside due to the consequences of truth-telling and a belief that the patient would be harmed by the truth of a terminal diagnosis. Justification for the duties of a doctor includes the fundamental need to respect patient autonomy and dignity. Doctors have absolute duties, such as refraining from killing patients. They also have context-specific duties, like giving appropriate clinical care in accordance with evidence-based practice. A focus on the moral duties of a doctor comes from a philosophical tradition known as deontology.

26 Medical ethics

Consequences

One key aspect of being a clinician is the ability to **calculate benefits, risks, and harms** and communicate them in such a way that patients and colleagues can understand and make decisions based on the information provided. Each patient interaction has the potential to bring about a range of outcomes, some foreseeable, some not so foreseeable. Prescribed medication, invasive procedures, research, prolonged waiting lists, and miscommunication between colleagues each have consequences that clinicians need to be able to calculate. In an ethical sense, consequences need to be weighed in terms of maximizing benefits for the patient and minimizing harm. It is recognized that the practice of medicine exposes patients to potential harm, but clinicians have an ethical duty to minimize harm and to explain when potential harm exists so that patients can make informed choices about the level of risk acceptable to them. Research ethics procedures formalize the process of calculating and minimizing risks by requiring researchers to complete ethical evaluations of their intended study and submitting these to research ethics committee scrutiny in order for permission to be granted for the research to take place. The principles of informal weighing of risks and benefits apply to all aspects of clinical practice. A focus on the consequences of healthcare comes from a philosophical tradition known as consequentialism.

Values

We expect clinicians to have a strong set of ethical values, even if there is debate about which values these should be. Contemporary discussions about professional values include the use of social media, remote consultations, and working under pressure. In whatever context we practice, the professional values of a clinician stem from characteristics that are seen as key to being the sort of clinician we would want to treat us and our family. The focus on duties and consequences centres on 'What should I do?', but the focus on values centres on 'What kind of person should I be?'. Generally, we desire clinicians who are trustworthy, compassionate, critically engaged, and good communicators. We want to work with colleagues who exemplify these values. We seek these qualities not because they are a clinician's duty, nor because they are seen to bring about the best consequences, but because these qualities exemplify who a clinician is. A clinician may or may not fail in their duty and negative consequences may or may not result, but excellent procedural skills without honesty, integrity, and compassion leave us with fundamental questions about what kind of person this clinician is. Values are not prescriptive, like duties and consequences. They exist on a spectrum, making room for the wide variety of personality types who make up the medical profession. Each clinician can have their own set of personal values that still conform to an agreed-upon set of expected values. Values that are not acceptable sit outside this spectrum. A focus on the values of a clinician comes from a philosophical tradition known as virtue ethics.

Frameworks and tools

Frameworks and tools can be useful in applying ethical theory to clinical practice. Perhaps the best known is the ***'Four Principles'*** approach:

- ***autonomy:*** 'self-rule', letting competent patients make their own choices and decisions about what happens to them
- ***beneficence:*** doing good, acting in a patient's best interests to benefit them
- ***non-maleficence:*** avoiding harm
- ***justice:*** treating people equitably and fairly

While this captures the key concepts that are important in clinical care, it offers no way of deciding which principle is most important when two or more principles are in conflict. It may be better to think of the four principles as providing an important linguistic framework for describing ethical concepts, rather than helping to resolve an ethical dilemma. Engaging with an ethical issue in clinical practice depends on a range of factors. We must discern the relevant moral questions and select the relevant ethical tools and theories for each clinical setting.

In clinical ethics, there are often three different dimensions, each interacting together:

Procedural ethics

Procedural ethics identifies which policies and procedures govern clinical situations in your local area. Hospitals have policies on consent, maintaining confidentiality, and following data protection regulations. Research ethics committees provide forms prepopulated with ethical questions for the clinician to answer. Procedural ethics relies on identifying and following appropriate ethical policies and procedures, which have been established in advance, often in response to legal cases and evidence-based practice.

Regulatory ethics

Regulatory ethics are produced by regulators such as the GMC. They set the framework and boundaries of the expectations of doctors and create ethical guidance in a regulatory capacity. This guidance is developed in response to legal cases, public consultation, and select committees. Where procedural ethics sits at a local level, regulatory ethics sits at a national level, providing specific guidance that is to be applied in the local setting. Regulatory ethical guidance can be found on the GMC website. Failure to follow regulatory requirements can lead to legal and professional consequences.

Aspirational ethics

Aspirational ethics focus on the factors that motivate a clinician, alongside the procedures and regulatory guidance that apply to the case. Aspirational ethics are the personal values we bring to our professional practice and want to actualize in ethical dilemmas. It is likened to our internal moral compass, which we would follow even if policies, procedures, and ethical guidance did not exist. Aspirational ethics are the fundamental values we aspire to as we practice medicine. They are formed through a variety of experiences:

- We may have a personal experience of healthcare (positive or negative) that has given us insight into what values we want to exhibit.
- We may have family members who have received healthcare that has left a lasting impression on us, shaping our moral character.
- We may have worked with senior clinicians who have become role models, exhibiting qualities we want to emulate or avoid.
- We may have a faith position that motivates how we see our role as a clinician, providing a framework for personal values that are important to foster.
- We may have sociopolitical beliefs that inform the values we believe should be important in resolving ethical dilemmas.

One or more of these experiences help us to develop aspirational ethics, which are likely to include values such as compassion, fairness, respect, truthfulness, kindness, and patience.

Further reading

Beauchamp TL, Childress JF. *Principles of Biomedical Ethics,* 8th edn. Oxford University Press; 2019.
Doherty RF. *Ethical Dimensions in the Health Professions.* 7th edn. Elsevier Health Sciences; 2020.
General Medical Council. *Good Medical Practice.* London: GMC; 2013, updated 2024. https://www.gmc-uk.org/professional-standards/professional-standards-for-doctors/good-medical-practice.
Lidewij van Bruchem-Visser R, van Dijk G, de Beaufort I, Mallace-Raso F. Ethical frameworks for complex medical decision-making in older patients: A narrative review. *Arch Gerontol Geriatr* 2020; 90.
UK Clinical Ethics Network. https://www.ukcen.net/ethical_issues/ethical_frameworks.
World Medical Association. *Medical Ethics Manual*. https://www.wma.net/what-we-do/education/medical-ethics-manual/.

Moral injury and the ethics of self-care

Moral injury is defined as the profound psychological distress that results from actions, or the lack of actions, which violate one's personal moral code. Moral injury can be caused by a discrepancy between the care we want to provide and the care we are able to deliver, or by witnessing acts performed/not performed by others that go against our deeply held moral values. Resource allocation decisions must be made at local and national levels, but the impact these have on clinicians who must explain and justify them to patients and relatives is often overlooked. The daily stress experienced by many members of the healthcare team can lead to moral injury being more likely, as colleagues work under increasing pressure. Self-care in these working environments is an important moral duty for all clinicians. There is increasing awareness that moral injury has a serious impact on staff wellbeing, contributing to burnout. Regular debriefs, peer support and self-care are all important ethical requirements for working in increasingly pressured healthcare environments.

Professional guidance and codes of practice

Along with ethical theories and frameworks, there are professional guidelines and codes of practice. For example, in the UK, the standards set out by the GMC inform the regulation of doctors; if a doctor falls below these standards, disciplinary procedures may follow, irrespective of the harm caused or whether legal action ensues. In other countries, similar professional bodies license doctors and regulate healthcare. All clinicians should be aware of the regulatory framework and professional standards in the country within which they practice (**Box 3.1**).

Ethical practice and professionalism are significant requirements from the earliest days of medical study and training. In the UK, the GMC defines the standards expected of medical students, and all medical schools are required to have 'fitness to practice' procedures. Students should be aware of their professional obligations from the start of their training. Medical students commonly work with patients from the earliest days of their training and are privileged in the access they have to vulnerable people, confidential

Box 3.1 Vignette 1

A newly appointed consultant burns and plastics surgeon, specializing in breast reconstruction, is called to the emergency department to review Lisa, a 54-year-old patient. Lisa has returned to the UK from a European country where she had breast uplift and enlargement 5 days ago. She explains that after rapid weight loss, she was left with breast skin that had lost volume and sagged. She found this very distressing, losing her confidence, staying at home, and becoming withdrawn. A few of her family and friends contributed money so that she could have breast enlargements, but she couldn't afford it privately in the UK, so she went abroad to a clinic a friend had recommended. She reports that she had no immediate complications from the surgery, but on the flight home this morning she became 'uncomfortable'. When she arrived home, she took the dressings down and looked in the mirror. She was horrified to see 'bulging areas' under her breasts.

The clinician notes that Lisa has a temperature of 38.3°C and a heart rate of 124 beats per minute. Examination of the wound reveals that both suture lines are open, exposing the breast implants, which are protruding out of the wounds because they are too big for the pocket that was surgically created. The clinician has seen other patients who have arrived back from the same country with complications of breast enlargement procedures, where implants have been used that do not appear to meet UK safety standards.

The clinician explains to Lisa that the implants will need to be removed because they are causing significant health risks. Lisa is relieved that this will be done and asks that the replacement implants be as large as possible so that her body image will be close to what she was aiming for. The clinician now has the difficult task of explaining to Lisa that she can only remove the implants and she cannot replace them, as no NHS funding is available for enlargement surgery. Lisa becomes distraught. She cannot bear the thought of living without breast enlargement and cannot afford to have it done privately in the UK. The clinician knows that once the implants have been removed, future surgery might be more difficult if left for too long. While they are very sympathetic to Lisa's situation, they cannot comply with her request, as it is against hospital regulations. Lisa asks the clinician to recommend a private surgeon in the UK so that she can start the consultation process for enlargement surgery in the future and worry about finding the money later.

What ethical questions arise in this case?

- What **duties** does the clinician have, and to whom?
- What **consequences** should the clinician weigh, and for whom?
- What should the clinician communicate to Lisa to gain her **informed consent** for surgery?
- What **national and local policies and guidelines** apply to the management of this case?
- What guidelines about **conflicts of interest** apply to this case?
- What **professional values** should the clinician aim to display?
- What steps should the clinician take to deal with any **moral injury** they feel from dealing with this case?
- What principles of this case apply to your practice?

Commentary

This is a complex ethical dilemma. Leaving the implants in place will expose Lisa to risks of sepsis and death. Removing the implants without replacement will affect Lisa's mental health. The clinician consults NHS Trust policy and procedure to guide their decision-making. They access GMC guidance, including information on shared decision-making and conflicts of interest. They are particularly aware that when Lisa asks for a referral to a private surgeon, the clinician is in a difficult situation. Clinicians should not discuss private treatment during NHS consultations, but they can answer factual questions about the availability of private treatment and can inform the patient's GP about the request for information. Requests for second opinions in the private sector should be met unless refusal is clinically justified, which is not the case here.

They evaluate their duties and the consequences of each option they consider for Lisa. Each of those provides the clinician with the information they need to manage this consultation, but aspirational ethics play a deciding factor in how this consultation progresses. Their personal values of compassion, empathy, honesty, and respect for autonomy are expressed through their communication skills in this case. These personal values also help the clinician when they spend time reflecting on this case afterwards. They recognize the importance of using scarce resources appropriately and the need to prohibit breast enlargement on the NHS, but they would prefer to be able to make an exception in a case such as this. They experience moral injury as they reflect on the consultation with Lisa. They decide to refer the case to the hospital clinical ethics committee to ask whether an exception to hospital policy can be considered.

information, and sensitive situations. As such, medical schools have specific responsibilities to ensure that students behave professionally and are fit to study, as well as eventually practice, medicine.

Further reading

British Medical Association. *Moral Distress and Moral Injury: Recognising and Tackling it for UK Doctors.* London: BMA; 2021.
General Medical Council. *Achieving Good Medical Practice: Guidance for Medical Students.* London: GMC; 2016, updated 2022.
Jones-Bonofiglio K. *Health Care Ethics Through the Lens of Moral Distress.* Springer International Publishing, 2021.

ETHICS OF INEQUALITY

Principles of equality, diversity, and inclusion (EDI) are important in healthcare provision. There is a growing recognition that patients who come from minority communities have worse outcomes. Clinicians play an important role in reflecting on how healthcare can be delivered more fairly. While it is necessary to examine institutional responses to EDI principles, it is also necessary to reflect on choices made in individualized clinical care and research design.

Principles of equality, diversity, and inclusivity

The Equality Act 2010 lays out the legal framework that governs the duties healthcare providers have towards patients and colleagues. The Equality Act uses the concept of protected characteristics, identifying groups of people who are vulnerable to inequalities and need legal protection against discrimination. The protected characteristics are age, disability, gender reassignment, marriage and civil partnership, pregnancy and maternity, race, religion or belief, and sex and sexual orientation. While the law requires that healthcare provision addresses factors that are discriminatory, the ethical duty to consider EDI goes much further. It is not sufficient to avoid discriminatory practices; we need to embrace values that promote inclusivity. One way of doing this is to consider what the aims of healthcare ought to be.

We may assume that a goal of healthcare provision is to ensure that everyone has equal access to the same level of healthcare, no matter their background. In this assumption, we aim to treat everyone the same, without bias. **Equality of access** is an important ethical principle in the NHS. For instance, every person has access to a general practitioner and can attend an emergency department. Having equality of access does not ensure fairness. In some circumstances, patients may need to be treated differently to achieve **equality of outcome** rather than equality of access. Equality of outcome explores ways to assist patients and colleagues to reach the same goals as those who do not need additional support. This may include providing language services, offering 'out of hours' or longer appointments, ensuring correct pronouns are used, and ensuring diverse patient feedback is sought for service improvement. A focus on equality of outcome allows us to identify ways to respect and enhance patient autonomy and build better relationships that lead to better patient outcomes. It allows us to focus on inclusivity, where service provision improves for everyone, including patients with protected characteristics.

One key area that must be addressed to improve equality of outcome is research design, including a full diversity of patients to ensure better representation of participants. Data should be collected that truly represents patient/client groups to ensure that health improvements are experienced by all. If research is conducted with healthier participants who may have different responses to a therapy or service provision compared to groups most in need, the purpose of healthcare is undermined. NHS England has issued good practice guidance for researchers on how to engage with diverse participants and increase involvement with health research.

The concept of protected characteristics encourages us to think about categories of patients, but our true lived experiences rarely fit into a neat category. **Intersectionality** describes how a person's various identities combine to create experiences of discrimination and privilege, advantage and disadvantage (**Box 3.2**).

The NHS Race and Health Observatory has reported overwhelming evidence that racism affects healthcare. They identified five priority areas, namely mental health services, maternal and neonatal healthcare, digital inclusion and access to health services, genetic testing and genomic medicine studies, and ethnic inequalities in the NHS workforce, highlighting best quality evidence, and made recommendations for change. Among those making a substantial impact on inclusivity are medical students and junior doctors, who are raising awareness and providing resources to improve patient care and equality of outcome. *Mind the Gap: Online Clinical Handbook of Signs and Symptoms in Black and Brown Skin* is a collaboration between a medical

Box 3.2 Vignette 2

Monica consults her GP about back pain, which has been getting worse for around a week. Her GP takes a history, completes a physical examination, and reassures her it is nothing serious. They recommend a course of non-steroidal anti-inflammatory tablets taken regularly for 5 days and time off work to allow her back pain time to resolve.

Monica is a Somali refugee. She works as a cleaner. The other cleaners are male and earn more than her. She fears that if she takes time off work, she will be replaced. Her employer does not provide sick pay and she can't afford to take unpaid time off work to rest her back.

Monica lives in a flat with Jo. Monica's parents have recently learnt that Monica is in a relationship with Jo. They strongly disapprove of their daughter's same-sex relationship, as it is frowned upon in their community. They have become hostile to Monica and no longer speak to her.

What ethical questions arise in this case?

- What ethical **duties** does the GP have?
- What **consequences** should the GP weigh?
- What **national and local policies and guidelines** apply to the management of this case?
- What **professional values** should the GP aim to display?
- What factors should the GP consider with regards to **protected characteristics**?
- What issues of **intersectionality** does this case raise?
- What principles of this case apply to your practice?

Commentary

If the GP focuses on Monica's clinical presentation of back pain, without understanding her lived experience, Monica may have a worse outcome. Intersectionality enables us to understand Monica's situation beyond her initial presentation. In terms of protected characteristics, Monica's race, sex, and sexual orientation create multiple factors for discrimination and disadvantage. Being unaware of these can lead to Monica experiencing deteriorating mental health. While the GP cannot resolve each of the situations Monica faces, taking a broader social history can identify ways that Monica can talk about her lived experience and be referred for specific support at a local level. In this location, there was an active community voluntary sector support group for asylum seekers and refugees that ran social activities which Monica found interesting. She was keen to join a support group and talk with women who had similar experiences to her. This meets the ethical duty to respect patient autonomy and dignity more fully than solely concentrating on her presenting condition.

student and clinical skills teachers, which addresses the lack of diverse clinical images needed by clinicians to make accurate diagnoses in patients with darker skin tones. A new generation of medical students and junior doctors are being trained with a focus on inclusivity, with increasing access to inclusive resources. This will hopefully raise ethical standards of healthcare and improve patient outcomes.

Further reading

NHS England. *Increasing Diversity in Research Participation: A Good Practice Guide for Engaging with Underrepresented Groups.* NHS England; 2023.
NHS Race and Health Observatory. *Ethnic Inequalities in Healthcare: A Rapid Evidence Review.* NHS Race and Health Observatory; 2022.
Okenyi E, Walker L. Ethics and racial diversity in early phase trial design. *Prescriber* 2023; 34 (4) 27-30.
Black and Brown Skin—*Mind The Gap Handbook*. https://www.blackandbrownskin.co.uk/.
Equality and Human Rights Commission. https://www.equalityhumanrights.com/en/equality-act.
MBRRACE-UK: Mothers and Babies: Reducing Risk through Audits and Confidential Enquiries across the UK. https://www.npeu.ox.ac.uk/mbrrace-uk.
NHS Race and Health Observatory. *Review of Neonatal Assessment and Practice in Black, Asian, and Minority Ethnic Newborns: Exploring the Apgar Score, the Detection of Cyanosis, and Jaundice.* NHS Race and Health Observatory; 2023.

LEGAL AND ETHICAL ISSUES IN MEDICAL PRACTICE

The law establishes boundaries for what is deemed to be acceptable medical professional practice. In the UK, along with other jurisdictions, both ***statutes and common law*** apply to the practice of medicine. Statutes are made by parliament, while common law is built up out of precedent set by judges in their rulings on previous cases. There are some legal differences between England and Wales, Scotland and Northern Ireland. Most cases involving healthcare arise in the civil system, where patients seek financial compensation for harm they have experienced. Occasionally a case is subject to criminal law, for example, when a patient dies in circumstances that could constitute manslaughter.

Respect for autonomy: capacity and consent

Capacity

Capacity is central to ethical decision-making because it is the gateway to self-determination (**Box 3.3**). People can make choices only if they have the capacity to do so. The presumption is that patients over the age of 16 have the capacity to make their own decisions unless evidence exists to the contrary. Assessing capacity is a significant ethical undertaking: a patient's freedom to choose depends on it. In England and Wales, the Mental Capacity Act 2005 sets out the criteria for evaluating whether a person has the capacity to make a decision. In Scotland, the Adults with Incapacity (Scotland) Act 2000 applies. In Northern Ireland, the Mental Capacity Act (Northern Ireland) 2016 applies.

Evaluating capacity is not a one-off judgement. Capacity can fluctuate and assessments should be reviewed regularly. Capacity is decision-specific, time-specific, and task-oriented. People may have the capacity to make some choices but not others, and capacity is not precluded by specific diagnoses or impairments such as dementia. The communication style of a clinician can enhance or diminish a patient's capacity, as can pain, fatigue, and the environment.

Box 3.3 Principles of self-determination[a]

- Every adult has the right to make their own decisions and is assumed to have capacity unless proved otherwise.
- Everyone should be encouraged and enabled to make their own decisions, or to participate as fully as possible in decision-making.
- Individuals have the right to make eccentric or unwise decisions.
- Proxy decisions should consider best interests, prioritizing what the patient would have wanted, and should be the least restrictive of basic rights and freedoms.

[a]Principles of self-determination underlying the Mental Capacity Act 2005 (England and Wales), which applies to patients over the age of 16 years.

Box 3.4 GMC guidance *Decision Making and Consent* 2020 lists seven principles involved in gaining informed consent

1. All patients have the right to be involved in decisions about their treatment and care and be supported to make informed decisions if they are able.
2. Decision-making is an ongoing process focused on meaningful dialogue: the exchange of relevant information specific to the individual patient.
3. All patients have the right to be listened to, to be given the information they need to decide, and the time and support they need to understand the decision.
4. Clinicians must try to find out what matters to patients so that they can share relevant information about the benefits and harms of proposed options and reasonable alternatives, including the option to take no action.
5. Clinicians must start from the presumption that all adult patients have the capacity to make decisions about their treatment and care. A patient can only be judged to lack the capacity to make a specific decision at a specific time, and only after assessment in line with legal requirements.
6. The choice of treatment or care for patients who lack capacity must be of overall benefit to them, and decisions should be made in consultation with those who are close to them or advocating for them.
7. Patients whose right to consent is affected by law (for instance, patients who are subject to compulsory treatment orders) should be supported to be involved in the decision-making process and to exercise choice if possible.

Informed consent

Informed consent is integral to ethical and lawful practice (**Box 3.4**). To act without, or in opposition to, a patient's valid informed consent is, in many jurisdictions, to commit assault or battery. Obtaining consent fosters choice and gives meaning to autonomy. Valid consent is:

- given by a patient with capacity
- voluntary: free from undue pressure, coercion, or persuasion
- sufficiently informed
- reversible: patients can change their minds at any time

The clinician seeking informed consent for a particular intervention must be competent in applying the procedure or treatment and be familiar with its problems. While it is good practice for written information to be provided, it does not remove the responsibility to talk to the patient.

Patients should be encouraged to ask questions and express their concerns, values, and preferences. Since it is the health and lives of patients that are at stake, the moral focus of information disclosure should be on what is acceptable to the individual patient rather than to the clinician.

Consent in educational settings

Medical education and training often happen in clinical environments where future clinicians learn new skills and apply their knowledge to patients. However, patients must be given a choice

as to whether they wish to participate in educational activities. The principles of seeking consent for education are identical to those applied to clinical situations. The rise in remote consultations has generated ethical issues in the education setting. Consent should be gained from the patient at the start of the remote consultation for students to be present in the consultation. Students should be introduced to the patient so that they know who is listening to the consultation. Being a silent and anonymous listener to a remote consultation is a breach of patient consent and confidentiality.

Advance decisions

Many countries have laws that allow people to make decisions about their future care. In the UK, these are called advance decisions. An advance decision enables someone to express their wishes about future care and is made in anticipation of someone losing capacity. An advance decision may refuse treatment but cannot lawfully request specific treatment. Where the treatment is lifesaving, the advance decision must be made in a specific way (**Box 3.5**). Countries have differing approaches to advance decision-making so the relevant local law must be understood. For example, the Mental Capacity Act 2005 applies in England and Wales, where a legally valid advance decision must be:

- made by someone with capacity
- made voluntarily
- based on appropriate information
- specific and applicable to the situation in which it is being considered.

In practice, it is often the specificity requirement that is difficult to satisfy because inevitable uncertainty surrounds future illness and potential treatments or interventions. For some people, the difficulty in anticipating the future and how they are likely to feel about that future informs an ethical objection to advance decision-making. Advance decisions can be made orally or in writing. However, decisions about withdrawing or withholding life-sustaining treatment must be in writing and witnessed. The decision should state explicitly that it is intended to include lifesaving situations. The list of lifesaving situations the patient can anticipate will be based on their diagnosis and prognosis, which will require sensitive discussion with the clinicians leading their care, to ensure that an appropriate range of scenarios have been explored and covered in the advanced decision. The more informal and non-specific an advance decision is, the more likely it is to be challenged or disregarded. When working in a country where advance decisions are recognized, clinicians should make reasonable attempts to establish whether a valid advance decision exists. Where there is ambiguity about either the existence or the content of an advance decision, the default is to preserve life. Advance decisions should be periodically reviewed and recorded; amendments, revocations, or additions are possible, provided an individual still has capacity.

The ethical rationale for advance decisions is respect for ***patient autonomy*** and extension of the right to make choices about healthcare in the future. Respect for autonomy includes supporting people to make choices that others might consider misguided. Advance decisions can foster trust and effective clinical relationships, but it is important to recognize that making decisions about future care can be difficult because the person we become when we are ill can differ from who we are when we are healthy.

Advance statements

Advance statements are broader than advance decisions and allow patients to express preferences and priorities without specifically ruling out an intervention or treatment. For example, a patient may wish to be more lucid than not towards the end of life or may like to retain the ability to eat with their family for as long as possible. A birth plan is often comprised of advance statements expressing wishes for labour and delivery. These preferences are not about choosing or ruling out specific treatments but are advance statements of wishes regarding the details of imminent future care. Advanced statements are not legally binding.

Box 3.5 Vignette 3

Mark is brought to the emergency department profoundly unwell, semi-conscious, and disorientated. He has Parkinson's disease and has an advanced decision on file, which states that he refuses antibiotics. Clinical investigations show that he has a urinary tract infection, which is sensitive to antibiotics. Mark has no relatives to consult about the context of his advanced decision refusing antibiotics. The clinicians discuss the likelihood that Mark anticipated antibiotics for a chest infection when he created the advanced decision. Clinicians decide to treat the urinary tract infection with intravenous antibiotics, as the advanced decision was not specific about the site of infection it covered.

Ethical questions

- What factors should clinicians weigh when deciding whether to uphold the advanced directive?
- What principles of the Mental Capacity Act apply to this case?
- What should Mark be told when he recovers sufficiently to discuss his treatment?
- What principles of this case apply to your practice?

Commentary

Mark is assessed not to have capacity, meaning that clinicians must make a judgement about whether to respect his advanced decision. They recognize that to override his stated refusal, they must have substantive reason to believe Mark had not considered the situation he is currently in. They hold a multidisciplinary team meeting to discuss options and seek advice from the hospital ethics committee. Four days after starting intravenous antibiotics, Mark is well enough to ask why clinicians ignored his advanced decision. He is angry and upset, as he had carefully planned refusal of treatments at the end of his life. Senior clinicians meet with Mark to explain the decision, providing him with information about the options they had considered, including abiding by his advanced decision. They offer Mark the opportunity to make a complaint about the care he received. The following day, Mark asks to speak with the clinicians again and enquires about how to make his advanced directive more specific.

Lasting power of attorney

Some countries allow for the appointment of a ***proxy:*** someone else to make decisions for patients lacking capacity. In England and Wales, consent to or refusal of treatment can be expressed by someone who has been lawfully granted lasting power of attorney (LPA) by a patient at a point when the patient still had the capacity to decide on this. For an LPA to be legally valid, it must be registered with the Public Guardian. The attorney then gains the power to represent the patient's interests only when the patient loses capacity. In jurisdictions that have LPA, the clinician should establish whether there is a valid LPA for a patient lacking capacity and adhere to the attorney's wishes. The only exception is where the attorney appears not to be acting in the patient's best interests, in which case the matter should be referred to the Court of Protection. Like advance decisions, the ethical rationale for LPAs is that prospective autonomy is desirable and facilitates informed care, rather than second-guessing patient wishes.

Best interests of patients who lack capacity

Where an adult lacks the capacity to give consent and there is no valid advance decision or LPA, clinicians must act in the patient's

best interests. This encompasses more than someone's best *medical* interests. In practice, determining best interests is likely to involve several people: relatives and carers, members of the healthcare team, professionals with whom the patient had a long-term relationship, and clinical ethicists. In England and Wales, an Independent Mental Capacity Advocacy Service advocates for patients without capacity who have no family or friends to represent their interests. In such a situation, people, including Independent Mental Capacity Advocates (IMCAs), are not making decisions; rather, they are giving an informed sense of the patient and their likely preferences.

Further reading

General Medical Council. *Decision Making and Consent: Supporting Patient Choices about Health and Care.* London: GMC; 2020.
General Medical Council. *Making Decisions Together: The Implications of the Montgomery Judgment*. https://www.gmc-uk.org/professional-standards/learning-materials/blog---montgomery--judgement.
https://www.nhs.uk/conditions/end-of-life-care/advance-decision-to-refuse-treatment/.

Provision or cessation of life-sustaining treatment

For a patient who lacks capacity and has neither an advance decision nor an LPA, medical means to prolong life are not required when:

- the team believes, based on good evidence, that further treatment will not save the patient's life
- the patient is already imminently and irreversibly close to death
- the patient is so permanently or irreversibly brain-damaged that they will be incapable of any future self-directed activity or intentional social interaction.

Moral beliefs vary and, in general, decisions not to provide or continue life-sustaining treatment should be made with as much consensus as possible among the clinical team and those close to the patient. Where there is unresolvable conflict, hospital solicitors can apply for a court ruling. In the UK, judges are available in the relevant court, even in emergencies.

Where clinicians decide not to prolong the lives of imminently dying and/or extremely brain-damaged patients, the legal requirement is that they are acting in the patient's ***best interests*** by seeking to minimize suffering rather than intending to kill. Ethically, the significance of intention, along with the moral status of ***acts and omissions***, is integral to debates about assisted dying.

Assisted dying

In many countries assisted dying is illegal. In contrast, some jurisdictions, including New Zealand, Australia, Netherlands, Switzerland, Belgium, Portugal, and certain US states, allow assisted dying. England and Wales are in the process of implementing assisted dying following the House of Commons vote on the Terminally Ill Adults (End of Life) Bill. The Bill will go through committee and reading stages in the House of Commons and House of Lords before being implemented into practice in England and Wales. The timeframe for this is anticipated to be 2026. Prior to the legal change, assisted dying remains unlawful. Withholding and withdrawing treatment are usually acceptable in strictly defined circumstances, where the intention of the clinician is to minimize suffering, not to cause death. The ***doctrine of double effect*** remains in law, but its clinical application is limited because of improved practice and understanding in palliative and supportive care. Double effect is the legal doctrine that allows clinicians to prescribe medication that has, as its principal aim, the reduction of suffering but which may have side-effects that potentially shorten someone's life. Such prescribing is legally justifiable because the intention is benign and the side-effects, while foreseen, are not the primary aim. Once assisted dying is lawfully permitted in England and Wales, the doctrine of double effect will no longer apply to eligible patients, because it will become lawful for a doctor to intend to end a patient's life.

When assisted dying becomes lawful in England and Wales, no registered medical practitioner will be under a duty to participate in assistance. Conscientious objection will be protected. The Secretary of State will produce a Code of Practice along with guidance from the Chief Medical Officer. These will need to include:

- Whether doctors can initiate the discussion about assisted dying
- The scope of conscientious objection—who does it apply to and at what stage
- How to judge capacity if symptom control involves mediation which reduces consciousness
- What guidance to give Judges in the High Court about how to exercise their role
- Funding of assisted dying
- Likely numbers of patients
- Where assisted dying will be provided
- What medication will be used, in what quantity
- How to deal with complications of failed assisted dying
- Training and support for doctors/MDT

Whilst assisted dying is currently unlawful in the UK, ethical standards of clinical care requires that some questions are explored with the patient:

- What is most important to the patient at this point in their life?
- What symptoms are they experiencing and how effectively are these being managed?
- Has palliative care expertise been sought to support the patient journey?
- What involvement do the relatives/friends have in the patient journey?

Questions such as these allow clinicians to develop ethical relationships with patients which remain within legal boundaries and facilitate dignified care. Personal and professional moral values will inform how we relate to a patient asking for something currently unlawful but mindful that patients and relatives may be aware of the pending legal change. Once the change in law is implemented, doctors will need to be clear about whether they will take part in assisted dying or exercise their legal right to conscientiously object.

Further reading

General Medical Council. *Treatment and Care Towards the End of Life.* London: GMC; 2010, updated 2022. Terminally Ill Adults (End of Life) Bill. https://commonslibrary.parliament.uk/research-briefings/cbp-10123/.

Mental health and consent

Most people being treated for psychiatric illness have capacity. However, mental illness sometimes compromises an individual's capacity to make decisions. Many countries have specific legislation that enables people to be treated without consent on the basis that they pose a risk to themselves and/or to others.

People who have, or are suspected of having, a mental disorder may be detained for assessment and treatment in England and Wales under the ***Mental Health Act 2007***. The Act was reviewed in 2018, and a new Mental Health Bill 2022 is currently going through parliament. There are four principles that inform the Mental Health Bill 2022:

1. Choice and autonomy: ensuring service user's views and choices are respected.
2. Least restriction: ensuring the Mental Health Act powers are used in the least restrictive way.

3. Therapeutic benefit: ensuring patients are supported to get better so that they can be discharged from the Mental Health Act.
4. The person as an individual: ensuring patients are viewed and treated as individuals.

Implementation of the new Mental Health Act will be staggered over time, with the first duties to be introduced in 2024-2025. The first duties to be introduced include new detention criteria including for people with learning disabilities and autism, Nominated Person, and automatic referral to Independent Mental Health Advocacy (IMHA) services.

Further reading

Local Government Association. *Get in on the Act: Mental Health Act* 2022. https://www.rcpsych.ac.uk/improving-care/campaigning-for-better-mental-health-policy/reforming-the-mental-health-act?searchTerms=reform%20of%20mental%20health%20act.
Royal College of Psychiatry. *Reform of the Mental Health Act in England and Wales.* https://www.rcpsych.ac.uk/improving-care/campaigning-for-better-mental-health-policy/reforming-the-mental-health-act?searchTerms=reform%20of%20mental%20health%20act.

Consent and children

Where a child does not have the capacity to make decisions, treatment will usually depend on obtaining proxy consent. In the UK, consent is sought on behalf of the child from someone with parental responsibility. In the absence of someone with parental responsibility (e.g., in emergencies, where urgent treatment is indicated), clinicians proceed based on the child's best interests. Sometimes, parents and clinicians disagree about the care of a child who is too young to make their own decisions. Both national and international case law demonstrate that courts are prepared to override parental decisions if they are perceived to compromise the child's best interests. However, the courts have also emphasized that a child's best *medical* interests are not necessarily the same as a child's ***best overall interests or welfare***. Whenever the patient is a child, clinicians interact with a family unit. Sharing decisions and attending to the child as a member of a family are the most effective and ethical ways of practising.

As children mature, whether a young person has the capacity to make his or her own decisions is evaluated using principles derived from the case ***Gillick v. West Norfolk and Wisbech Area Health Authority***, which determined that a young person can make a choice about their own health where:

- the patient, although under 16, can understand medical information sufficiently
- the clinician cannot persuade the patient to inform, or give permission for the clinician to inform, the parents
- a young person is seeking contraception and is likely to have sexual intercourse with or without adequate contraception
- the patient's mental or physical health (or both) is likely to suffer if treatment is not provided
- it is in the patient's best interests for the clinician to treat without parental consent.

The *Gillick* case recognized that young people differ in their abilities to make decisions and established that decision-making capacity, not age, matters when considering whether a child can give consent. Situations should be approached on a case-by-case basis, considering the young person's level of understanding of a particular treatment. It is possible (and perhaps likely) that someone may be considered to have the capacity to consent to one treatment but not another. Even where a child does not have the capacity to make their own decision, clinicians should respect the child's dignity by discussing the proposed treatment with the child, even if parental consent must also be obtained. The Gillick principles now apply in a wide variety of paediatric settings. In the UK, once a child reaches the age of 16, the Mental Capacity Act 2005 states that they should be treated as an adult, except for the purposes of advance decision-making and appointing an LPA.

Further reading

General Medical Council. *0–18 Years: Guidance for All Doctors.* London: GMC; 2018.
General Medical Council. *Protecting Children and Young People: The Responsibilities of All Doctors.* London: GMC; 2012, updated 2018.

Confidentiality

Confidentiality is essential to therapeutic relationships. If clinicians violate the privacy of their patients, they disrespect autonomy, undermine trust, risk harm, and bring the medical profession into disrepute. Breaking the trust of a patient can lead to serious consequences for both the patient and the clinician. Within the UK, confidentiality is protected by common and statutory law. Some jurisdictions make legal provisions for privacy. Clinicians who breach confidentiality may face professional and legal sanctions.

Respecting confidentiality in practice

Patients should understand that information about them will be shared with other professionals involved in their treatment. Usually, by giving consent for investigations or treatment, patients are deemed to give their implied consent for information to be shared within the clinical team. Rarely, patients might object to information being shared, even within a team. In such situations, the advice is that the patient's wish should be respected, unless it compromises treatment. In almost all clinical circumstances, therefore, the confidentiality of patients must be respected. Confidentiality can be breached inadvertently by having clinical conversations in lifts, corridors, and cafés. Even on wards, confidentiality may be compromised by the proximity of beds and the visibility of whiteboards containing medical information. Everyone should be alert to incidental breaches of confidentiality and seek to minimize their role in unwittingly revealing sensitive information. In the UK, clinicians and trainee clinicians must ensure they practice in accordance with General Data Protection Regulations, which form part of hospital policy on handling and storing confidential data.

When confidentiality must or may be breached

Duty of confidence is not absolute. Sometimes, the law requires that clinicians reveal information about patients to others, even if the clinician is reluctant. Such instances include:

- Notifiable diseases, which, by virtue of public health legislation, must be notified to the relevant consultant in communicable disease control
- Court orders
- Road traffic accidents that lead to requests from the police
- Actual or suspected terrorist activity

There are also circumstances in which a clinician has the discretion to share confidential information within defined terms. For instance, not all communicable diseases are notifiable under law. A clinician could share confidential information about a communicable disease with colleagues who are directly involved in clinical care for that patient if they are at direct risk. It would be good practice to ask the patient's permission to share the information, but if the patient refuses permission, the clinician may judge that they have a duty to protect colleagues and other patients. In such a case, the clinician should inform the patient that they will share the information. Such

circumstances highlight the ethical tension between the rights of individuals and the public interest.

Aside from legal obligations, there are three situations in which information may be shared:

- The patient has given consent.
- It is in the patient's best interests to share the information, but it is impracticable or unreasonable to seek consent.
- It is in the public interest.

These three categories are a useful framework within which to think about the duty of confidentiality. They also require ethical judgement, particularly where sharing confidential information might be in the ***public interest***. In England and Wales, there is legal guidance on what constitutes sufficient 'public interest' to justify sharing confidential information, which is derived from the case of *W v. Egdell*. In that case, the Court of Appeal held that the 'most compelling circumstances' could justify a doctor acting contrary to the patient's perceived interest in the absence of consent. The court stated that it would be in the public interest to share confidential information where:

- there is a real and serious risk of harm
- the risk is of physical harm
- there is a risk to an identifiable individual or individuals.

Consent to disclose should be sought wherever possible, and disclosure based on 'public interest' should be a last resort. Each case must be weighed on its own merits, and a clinician who discloses confidential information on the grounds of 'public interest' must be able to justify their decision. Even where disclosure is justified, confidential information must be shared only with those who need to know.

Further reading

British Medical Association. *Confidentiality and Health Records Toolkit.* London: BMA; 2021.
General Medical Council. *Confidentiality: Good Practice in Handling Patient Information.* London: GMC; 2017.
General Medical Council. *Confidentiality: Disclosing for Education and Training Purposes.* London: GMC; 2017.

Professional competence and mistakes

Clinicians must practice to an acceptable professional standard. There are essentially three sources that inform what it means to be a 'competent' clinician:

- the law
- professional guidance from bodies such as the GMC and Royal Colleges
- local policy

In practice, there is often overlap and interaction between the categories; for example, a doctor may be both a defendant in a negligence action and the subject of fitness to practice procedures. Professional bodies are established by, and work within, a legal framework. In order for professional bodies to implement policy, legislation is required, and interpretative case law will often follow. Professional bodies have diverse but often overlapping roles in developing, defining, and revising standards for clinicians. The principal publications in which the GMC sets out standards and obligations relating to competence and performance are *Duties of a Doctor* and *Good Medical Practice*.

Standards and the law

In most countries, the law provides the statutory framework within which the medical profession is regulated. For example, in the UK, it is the function of the GMC to maintain the register of medical practitioners, provide ethical guidance, and guide and quality-assure medical education. The GMC defines standards of professional practice and has responsibility for investigation when a doctor's standard of practice is questioned. Fitness to practice is assessed by the Medical Protection Tribunal Service, who provide tribunal services that are separate from the investigatory processes of the GMC. Clinicians have a responsibility not only to reflect on their own practice but also to be aware of and, if necessary, raise concerns about the practice of colleagues, even in the absence of formal 'line management' responsibilities.

Conscientious objection

Clinicians will have their own views about what is and is not acceptable ethically. All clinicians should feel able to express concerns, raise questions, and challenge constructively. However, the GMC is clear that doctors must not allow their own beliefs or judgements to compromise clinical care or discriminate against patients. Clinicians who are exercising a conscientious objection should usually ensure that the patient is referred to another clinician who does not have a conscientious objection. Conscientious objection should not be used to prevent patient autonomy and access to lawful treatment.

In the UK (apart from Northern Ireland), there is one situation in which the law provides for a right of conscientious objection: the provision of termination of pregnancy and some reproductive medicine that involves embryos. The right of conscientious objection allows clinicians to choose not to be involved in providing terminations, including making referrals to specialist services. However, clinicians must not convey their beliefs in a judgmental way to a patient and should arrange for a colleague to see the patient without delay. In an emergency, which is likely to be extremely rare in contemporary medical practice, a doctor would have to act if no one else was available. The law is silent on whether the provision applies to medical students, although most medical schools will offer guidance and accommodate students who have a conscientious objection to termination of pregnancy.

Further reading

General Medical Council. *Personal Beliefs and Medical Practice.* London: GMC; 2013, updated 2020.

Clinical negligence

Negligence is a civil claim where damage or loss has arisen because of an alleged breach of professional duty, such that the standard of care was not, on balance, that which could be reasonably expected. To establish a case of negligence, the claimant must prove on the balance of probabilities that a doctor owed them a duty of care that breached the duty and that the breach caused them harm.

Of the components of negligence, duty is the simplest to establish: all doctors have a ***duty of care*** to their patients. Whether a doctor has discharged their professional duty adequately is determined by expert opinion based on standards that can reasonably be expected and an individual's conduct in relation to those standards. If doctors have acted in a way that is consistent with a reasonable body of their peers (the ***Bolam standard***) and their actions or omissions withstand logical analysis (the ***Bolitho standard***), they are likely to meet expected standards of care. Lack of experience is not considered a mitigating factor in legal determinations of negligence. The most common reason for a clinical negligence action to fail is ***causation***. For example, the alleged harm may have occurred against the background of a complex medical condition or course of treatment. In these cases, it can be very difficult to identify whether the alleged breach of duty caused the actual harm the

patient experienced, or whether disease progression played a role. Patients with chronic health conditions find it especially difficult to prove causation.

Clinical negligence remains relatively rare, but undue fear of litigation can lead to defensive medicine and poor practice. All clinicians make mistakes. Mistakes do not necessarily constitute negligence or indicate incompetence. Regulatory bodies and medical defence organizations require clinicians to be honest about their mistakes and to apologize, remembering that to do so is not an admission of negligence. Such honesty and humility, aside from its inherent moral value, have been shown to reduce the prospect of patient complaints or litigation. Patient complaints can be passed to *NHS Resolution*, who provide expertise to the NHS on resolving concerns and disputes fairly, share learning for improvements and preserve resources for patient care. They also administer the indemnity scheme for the NHS.

Professional accountability is important. The law, professional guidance, and policy documentation provide a starting point for clinicians. Complaints and possible litigation are often brought by patients who feel aggrieved for reasons that may be unconnected with the clinical care that they have received. When patients are asked about their decisions to complain or to sue clinicians, it is common for poor communication, insensitivity, administrative errors, and lack of responsiveness to be cited as motivation. Clinicians should offer an apology and explain fully and promptly what has happened, along with the likely short-term and long-term effects of any harm. Reluctance to say 'sorry' comes from a fear that it is an admission of fault, which later implies liability, but guidelines from medical defence organizations emphasize that this is not so. There is less to fear than clinicians sometimes believe. The courts and professional bodies are concerned neither with best practice nor with unfeasibly high standards of care. What is expected is that clinicians behave in a way that accords with the practice of a reasonable clinician, and the reasonable clinician is not perfect. Good habits begin in medical school. Effective communication is an important component in preventing and resolving complaints, reducing clinicians' stress, and developing the professional skills and reputation for handling difficult situations properly.

Further reading

General Medical Council. *Openness and Honesty When Things Go Wrong: The Professional Duty of Candour.* London: GMC; 2015, updated 2022.
Herring J. *Medical Law and Ethics,* 8th edn. Oxford Uni Press; 2020.

Significant websites

Medical Practitioners Tribunal Service. https://www.mpts-uk.org/about.
NHS Resolution. https://resolution.nhs.uk/.

Questions for this chapter can be found on Elsevier eBooks+.

4 Evidence-based practice

Rosalind Ann Raine and Peter Martin

Background 35
The need for EBP 35
Barriers to EBP 36
A framework for applying evidence to clinical problems 36
Asking answerable questions 36
Searching for the evidence 36
How to conduct a search 37
Appraising the evidence 37
The hierarchy of evidence 37
Critical appraisal 41
Understanding statistical results 43
Descriptive statistics 43
Inferential statistics 43
Worked examples 45
Summary and outlook 48
Applying the evidence 49
Challenges of implementing evidence 49
Improving the uptake of evidence 51

CORE SKILLS AND KNOWLEDGE

Estimates from various countries suggest that up to 40% of patients do not receive evidence-based healthcare. For example, despite the evidence for the effectiveness of prompt corticosteroids for Bell's palsy, almost half of patients (44%) do not receive this treatment, and provision is even lower for patients over the age of 60. Furthermore, 20% or more of healthcare is unnecessary or potentially harmful. For example, antibiotics continue to be widely prescribed for cough, despite being ineffective in the majority of cases.

Evidence-based practice offers a framework to assist clinicians in making efficient and appropriate use of the available literature and making informed decisions about its application.

All doctors need to develop key skills in applying this framework to clinical scenarios, including:

- Converting real-life clinical and population health problems into answerable questions
- Locating the best available evidence
- Appraising the evidence critically
- Applying the evidence in practice.

Opportunities for developing skills in evidence-based practice include attending (or setting up) a journal club where research papers are critically appraised, following the PICO approach—described later in this chapter—to apply evidence to particular dilemmas presenting in patients seen in clinical practice, visiting a local public health team to see how evidence is applied at a population level, or helping to write a local clinical guideline where evidence is applied at an institutional level.

BACKGROUND

The evidence-based medicine movement emerged over 30 years ago and is widely accepted as the gold standard for medical decision-making. It was defined in 1996 as 'the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients', with evidence in this case referring to peer-reviewed research literature. The focus has since expanded to include an emphasis on the importance of the patient and carer perspective in decision-making, alongside research evidence. We therefore refer in this chapter to evidence-based practice (EBP), which we define as '***an approach to decision-making in which the clinician uses the best scientific evidence available, in consultation with the patient, to decide upon the option that suits the patient best***'. This chapter will also embrace evidence-based public health practice, which applies evidence to wider populations, as well as individual patients.

The need for EBP

A number of factors have driven the rapid expansion of EBP. These include a growing awareness among health professionals of the gap between the large amount of good-quality research evidence available in the published literature and its often limited use in clinical practice. EBP also has a role in ensuring that treatments offered by healthcare providers are of high value and accrue the most benefits for patients and the public—especially where escalating demands for care result in constrained resources. For example, there is no national screening programme for pancreatic cancer in the UK, despite this disease often not being detected until a late stage and when already incurable. This is because pancreatic cancer is relatively uncommon and the population benefits are outweighed by costs (such as potential harms associated with follow-up invasive testing or surgeries in those with 'false positive' scan results, as well as financial costs).

Barriers to EBP

A multitude of barriers exist that lead to research evidence not being taken up into clinical practice, or to its being implemented in a delayed or patchy manner.

Knowledge barriers and the use of clinical guidelines

The most obvious barrier is lack of knowledge. It is estimated that general physicians would need to read 20 articles a day, all year round, to keep abreast of emerging knowledge in their specialist field. As research evidence continues to expand, clinical guidelines have become an important mechanism for informing day-to-day practice and helping clinicians keep up to date. Clinical guidelines identify, evaluate, and summarize the best available research evidence relating to the management of specific conditions. Consequently, they are a useful first port of call when seeking information on which to base decision-making. Guidelines are usually produced at the national or international level by medical associations or government bodies. In the UK, the National Institute for Health and Care Excellence (NICE) provides recommendations about care for patients with specific conditions. In the United States, the National Guideline Clearing House maintains a catalogue of guidelines published by various health and medical associations, including the Agency for Healthcare Research and Quality.

While clinical guidelines are not legally binding, doctors are expected to take them into account, and yet the extent to which this happens in practice varies. For example, UK clinical guidelines recommend that patients with high blood pressure and raised cholesterol receive antihypertensives, antiplatelet drugs, and lipid-lowering drugs for the secondary prevention of ischaemic stroke. However, one study demonstrated that only 25.6% of males and 20.8% of females receive appropriate treatment, and the likelihood of 80- to 89-year-olds receiving secondary prevention is nearly half that of 50- to 59-year-olds.

Other factors limiting the implementation of EBP

These include:

- doubts about the accuracy of the research evidence being promoted
- national and local political values
- priorities and pressures from industry, patients, the public, and clinicians
- cost.

These often interconnecting factors can lead to health policies being implemented *despite* contradictory or contested evidence. For example, a programme of 5-yearly preventative health checks for 40- to 74-year-olds was established in 2009 across the National Health Service in England. The programme aims to identify early signs of stroke, heart and kidney disease, type 2 diabetes, and certain types of dementia. However, findings of increased detection of heightened cardiovascular disease risk factors and diagnoses are inconsistent between studies, uptake varies between population subgroups, and the scientific evidence does not support the view that general health checks are effective in improving health outcomes. Some experts are concerned that universal assessment of risk could lead to unnecessary treatment and other harms, particularly in those known to be at low risk as they are less likely to benefit.

Consequently, this intervention could be considered inappropriate, as the potential harms outweigh the expected benefits.

A framework for applying evidence to clinical problems

EBP offers a framework to assist clinicians in making efficient and appropriate use of the available literature and in reaching considered decisions about its application. The goal of this chapter is to promote ways in which evidence can be used wisely, with a good understanding of its value and its limitations. In the following sections, we describe four essential steps:

- converting information needs into answerable questions
- locating the best available evidence
- appraising the evidence critically
- applying the evidence in practice.

Further reading

Krogsbøll LT, Jørgensen KJ, Grønhøj Larsen C et al. General health checks in adults for reducing morbidity and mortality from disease: Cochrane systematic review and meta-analysis. *BMJ* 2012; 345:e7191.
Muir Gray JA. *Evidence-Based Healthcare: How to Make Health Policy and Management Decisions*, 2nd edn. London: Churchill Livingstone; 2001.
Academy of Medical Royal Colleges. *Interventions A-Z*. URL: https://ebi.aomrc.org.uk/interventions/ [Accessed 2023-09-22].
Raine R, Wong W, Ambler G et al. Socio-demographic variations in the contribution of secondary drug prevention to stroke survival at middle and older ages: a cohort study. *BMJ* 2009; 338:b1279.
Sackett DL, Rosenberg WM, Gray JA et al. Evidence based medicine: what it is and what it isn't. *BMJ* 1996; 312:71–72.

ASKING ANSWERABLE QUESTIONS

Every day, doctors make multiple decisions that involve asking implicit and explicit questions. Evidence-based care protocols, pathways, and clinical guidelines are available that cover the management of a wide range of both common and rare conditions, and these should generally be a clinician's first port of call.

Sometimes, however, a literature search is required—perhaps where a patient does not fit well into the categories dealt with by clinical guidelines, or where a condition is rare and not covered by good evidence-based summaries. The first step is to turn the information need into a focused question. This must be directly relevant to the patient at hand and incorporate the key issues that need to be present in research articles to inform the decision-making process. Identification of these key issues will allow formulation of a focused literature search strategy.

The ***PICO approach*** facilitates this process (**Box 4.1**), where ***p***atient, ***i***ntervention, ***c***omparison, and ***o***utcome (PICO) are considered the four key parts of a well-built clinical question.

SEARCHING FOR THE EVIDENCE

The next step is to identify relevant research evidence that will provide an answer to the clinical question. There are two main sources of evidence available.

Primary sources of evidence

These are comprised of the thousands of original articles published every year in research journals. Research papers are catalogued in a variety of databases, which are searchable on the Internet. For many medical queries, Medline is the best-known database of health-related literature and makes a good starting place. The size and breadth of databases mean that a systematic searching approach is required to identify the best evidence available.

Box 4.1 Components of clinical questions structured using the PICO approach

Patient or problem (P)
- Who is the relevant patient?
- What is the condition or problem of interest?
- What are the most important characteristics of the patient?

Intervention (I)
- What is the management strategy of interest (e.g., diagnostic test, treatment)?

Comparison of interventions (C)
- What is the main alternative management strategy being considered, if any?

Outcome (O)
- What will the management strategy accomplish, measure, improve, or affect?
- What are the patient-relevant consequences?

Secondary sources of evidence

In order to deal with the growing volume of primary research, secondary sources of evidence are increasingly consulted. Secondary sources are summaries and analyses of evidence derived from and based on primary sources. The Cochrane Library (UK) was one of the first online collections of databases relevant to medicine and other healthcare specialties and includes the Cochrane Database of Systematic Reviews. Other organizations producing systematic reviews and other evidence syntheses to inform practice and policy include the Joanna Briggs Institute and the Campbell Collaboration.

How to conduct a search

Although biomedical databases vary in terms of their structure, the principles of using them apply to all. Using the PICO approach to generate a clinical question is helpful, as the components can then be employed to select search terms for conducting a search in Medline. Text-word searching involves seeking specific words or phrases, usually in the title or abstract of an article. The first step is to break the clinical question down into its individual components. **Box 4.2** contains a worked example, using the PICO approach.

Practical tips for searching databases include:

- Searching only for articles that include a full phrase, rather than individual words on their own, by putting double quotation marks around a group of words (e.g., "sore throat")
- Narrowing the search and reducing the number of citations identified by using AND in the database search function
- Avoiding narrowing too far. Using the example question in **Box 4.2**, entering "*sore throat*" AND *antibiotics* in the search function will identify articles that contain both terms, whereas including "*sore throat*" AND *antibiotics* AND "*symptom duration*", theoretically yielding the highest concentration of relevant articles, may also miss others. Consequently, for routine clinical queries, using two terms in a search strategy is usually sufficient to focus the search.
- Engaging the help of expert health sciences librarians. This may be valuable in developing search strategies for more complex queries.
- Remembering the dangers of publication bias (see also the section on 'Understanding statistical results' later in this chapter). Studies with negative results may not have made it as far as publication. On the other hand, studies may be in progress that could inform decision-making.

Box 4.2 Searching for evidence—worked example

Charlie is a 20-year-old student who has attended his GP practice and has a sore throat. You wonder whether antibiotics might help improve his symptoms.

Answerable clinical question
- Are antibiotics effective in reducing the duration of symptoms in patients with upper respiratory tract infections?

PICO
- ***Patient problem or population:*** Patients with a sore throat
- ***Intervention or exposure:*** Antibiotics
- ***Comparison or control:*** None
- ***Outcome measure:*** Duration of symptoms

Search terms in evidence/database (e.g., Medline)
- "sore throat" AND antibiotics—1032 articles
- "sore throat" AND antibiotics AND "symptom duration"—6 articles

Answer

Antibiotics shorten the duration of symptoms by about 16 hours overall

(Spinks A, Glasziou PP, Del Mar CB. Antibiotics for sore throat. *Cochrane Database of Systematic Reviews* 2013; 11. Art. No.: CD000023. https://doi.org/10.1002/14651858.CD000023.pub4.)

APPRAISING THE EVIDENCE

Once some potentially relevant research articles have been identified, the next step is to decide whether the studies described were undertaken in a way that makes their findings reliable and valid, and to consider the relevance of any articles to the clinical question identified. This process is known as critical appraisal.

The hierarchy of evidence

Ethically, patients should receive the best-evidenced treatment available, and the 'hierarchy of evidence' is a core principle of EBP that helps to ensure this. The hierarchy is formed by ***ranking study types based on the extent to which the reader can be confident that the results and estimates of a treatment effect are valid***. In most hierarchies, meta-analyses and systematic reviews are at the top of the pyramid, followed by experimental and observational studies, with case reports or case series at the bottom (**Fig. 4.1**). The hierarchy encourages a top-down approach to locate the best evidence, which involves first searching for evidence at the top (e.g., a recent meta-analysis) and, if that is not available, moving down to the next level of evidence.

Whatever the study design, when reading papers it is important to remain vigilant for poorly conducted research. Not all published research is of high quality, so it is necessary to critically appraise each study, as opposed to placing blind confidence in published results.

Systematic reviews and meta-analyses

Systematic reviews

These synthesize the results of all available studies in a particular area, identified via comprehensive, systematic literature searches. They therefore provide a summary of current evidence about a topic, including an analysis of study results, strengths, and weaknesses. Systematic reviews also highlight remaining gaps in knowledge or a lack of robust evidence. Data from all the included studies are combined to build up a broader picture of the evidence.

Systematic reviews aim to locate as much as possible of the research relevant to a particular research question. Clearly defined, methodical steps are then used to identify what can reliably be said

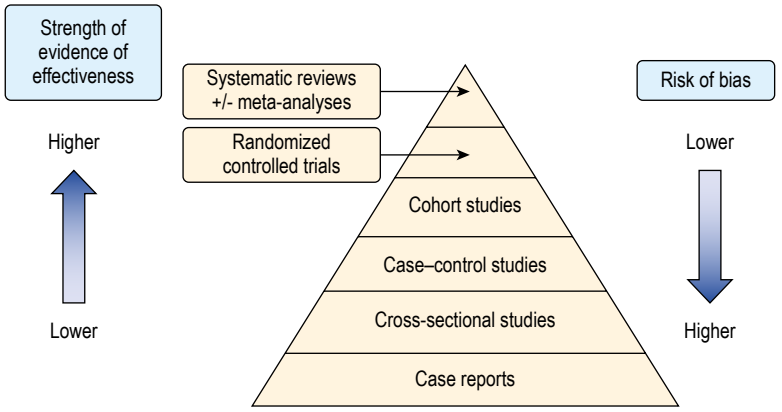


Fig. 4.1 **Hierarchy of evidence.**

about the topic, drawing on the findings of all the included studies. All study types can be included in a systematic review, including randomized controlled trials (RCTs) and observational studies, which are described later. A well-conducted review will assign a quality score to included studies so that the conclusions provide an overview of the status of the research field. For example, where there is a lot of robust evidence available to answer a question, clear recommendations can be made about patient care, or—in contrast—where there is a limited amount of low-quality evidence available, reliable recommendations cannot be made.

As systematic reviews draw together the findings of multiple studies, they lie at the top of the evidence hierarchy, though this position is not an absolute. For example, findings might quickly be superseded by evidence that is more recent. A large, well-conducted RCT may also provide more convincing evidence of effect than a systematic review of smaller RCTs.

Meta-analyses

A meta-analysis is an extension of a systematic review. Data from each of the included studies are combined statistically to give a single estimate of effect. To conduct a meta-analysis, the included studies must address the same research question and make use of the same outcome measure. Individual research papers can be included only when they report relevant numerical data in sufficient detail.

As meta-analyses pool findings from all the included studies, and therefore draw on data from a larger number of patients, they provide a stronger and more accurate estimate of a treatment effect than individual studies on their own.

Interventional studies

Randomized controlled trials (RCTs)

RCTs are generally considered to provide the most reliable evidence of whether an intervention is effective. A range of different types of intervention can be tested in an RCT, including therapeutic and prophylactic agents (e.g., drugs, vaccines), surgical and medical procedures, or health service and public health strategies.

Randomization

The strength of the RCT lies in the process of randomization, which is characteristic of this type of study design. Typically, study participants are randomly assigned to one of two study groups: the experimental group, who will receive the intervention being tested, and a comparison group (known as controls), who may receive:

- a conventional treatment (sometimes referred to as usual care)
- no intervention
- a placebo—a substance that resembles the intervention in all respects, except that it contains no active ingredients.

In the most common RCT design, every participant has an equal chance of being in the treatment group or the control group. The two groups are then followed up prospectively over time, to assess the effectiveness of the intervention compared with the conventional or placebo treatment.

The goal of randomization is to ensure that the two study groups are as similar as possible in all respects, apart from the treatment or intervention they receive. Randomization ensures that other factors that might also have an impact on the outcome, known as confounders, do not differ systematically between the two groups. Consequently, any observed differences between the study group outcomes are more likely to be caused by the treatment.

Those responsible for recruiting participants into a study must not influence the process of allocation to treatment groups. A common method of achieving this is randomization by Internet, where the investigator accesses a web-based randomization service to obtain information about which study arm the participant should be allocated to.

Confounding factors

A confounding factor is a variable that influences both the exposure and the outcome in a study. Failure to control for confounders can lead to biased results. For example, observational studies found that the proportion of people with coronary heart disease (CHD) was lower among those who take vitamin C supplements, compared to those who do not. This suggested that vitamin C may protect against CHD. However, randomized controlled trials that allocated participants randomly to taking or not taking vitamin C supplements found no evidence of a benefit of vitamin C. It is likely that socioeconomic position is a confounding factor: affluent people are more likely to take vitamin supplements and also have a lower risk of CHD than the less well-off (see *Lawlor DA et al. Those confounded vitamins: what can we learn from the differences between observational versus randomised trial evidence?* The Lancet *2004 363; 9422:1724–1727)*. This induces a spurious association between vitamin C and risk of CHD, and thus the observational studies which concluded a benefit of vitamin C were likely biased due to their failure to adequately control for confounding. In an RCT, randomization ensures that potential confounding factors are randomly distributed among the treatment and control groups. In sufficiently large trials, randomization ensures approximate balance between the groups with respect to any confounding factors, thus eliminating confounding bias.

Blinding

When a participant has been randomized, the investigator may or may not be aware of which group the participant has been allocated to. If they know which treatment the participant is receiving, this might influence (intentionally or unintentionally) the way in which they measure or interpret the outcome data, introducing bias into the study. For example, they may interpret information more favourably if it comes from participants who they know to be receiving the drug under investigation. Similarly, if a participant knows that they have been allocated to the placebo group, this may influence the way in which they participate in the trial. Therefore, blinding can be seen as a way to avoid potential bias.

- In a double-blind study, neither the patients nor the investigators know the study group allocation.
- In a single-blind study, patients do not know their study group allocation.

Blinding also involves ensuring that, as far as possible, the intervention and standard or placebo treatment both appear the same, although this may not be possible for non-drug interventions. For example, if the intervention involves physiotherapy exercises or a new way of delivering a service, it is often not feasible to conceal which group participants have been allocated to.

Intention-to-treat analysis

During an RCT, participants may be lost to follow-up for a range of reasons. Individuals may be unable to continue participating or may stop taking their allocated treatment—perhaps because of adverse side-effects or because they felt it was not working. If these individuals are subsequently excluded from the analysis of the results, the effectiveness of the drug may be overestimated. In the real world, patients do not always take their medication as prescribed, so RCT data are often examined using intention-to-treat analysis. This means that the final results are analysed based on the treatment arm to which participants were initially allocated, and not on the treatment they actually received.

Adaptive platform trials

Adaptive platform trials (APTs) are emerging RCT designs which use innovations in statistics and computing to enable trial protocols to be adapted and updated as evidence is generated. The design allows multiple treatments to be studied simultaneously in a continuous fashion, with therapies entering and leaving the platform on the basis of a predefined decision algorithm. The COVID-19 pandemic stimulated a rise in APTs because they are particularly appropriate for rapid evidence generation at a time of unprecedented pressures on healthcare systems worldwide (see Vanderbeek et al. in Further Reading). Their advantages are that they can be more time efficient and require fewer participants than standard RCTs. However, they can be logistically complicated and have varying adaptive designs and analytical approaches, in contrast to standard RCTs where design consensus is well established.

Limitations of RCTs

Although RCTs are considered to provide the most reliable evidence of effectiveness, they are not without their limitations.

Patient populations to whom the results might subsequently be applied in real clinical practice may often be excluded from RCTs in order to minimize differences between experimental groups. For example, strict inclusion and exclusion criteria may mean that patients with co-morbidities are excluded, as the presence of other conditions might affect their response to the intervention. This means, however, that the results of the RCT may not apply to patients with multiple morbidities, of which there are many in the real world.

Patients recruited to trials are also more likely to be receiving care in teaching hospitals, where care may differ from that provided in district general hospitals. This reduces the external validity or generalizability of the trial or the extent to which the findings can be applied in other settings. Adaptive platform trials may be able to overcome these limitations by widening inclusion criteria and by using routine healthcare data systems to identify eligible patients admitted to any hospital.

Additionally, although RCTs are the gold standard for evaluating the effectiveness of interventions, randomization may not always be required. For example, an RCT may be unnecessary when the effect of an intervention is dramatic (e.g., vaccination for smallpox or penicillin for bacterial infections). In these cases, the likelihood of unknown confounding factors affecting the outcome is so small that they can be ignored.

An RCT may be unfeasible when the outcome of greatest interest occurs rarely (e.g., death) and may be too costly if the outcome of greatest interest occurs far in the future (e.g., loosening of artificial hip joints). Observational methods, such as cohort studies, may be the only practical means of obtaining such information in these cases.

Finally, the use of RCTs raises several ethical issues. RCTs are based on the premise that there is sufficient doubt about the particular intervention being tested to justify not giving it to half of the participants. At the same time, there must be sufficient belief in its potential benefit to justify exposing the remaining half of the patients. This is known as ***clinical equipoise*** and requires careful consideration to be given to the intervention that is received by the control group. If an effective treatment already exists, participants in the control group should have this rather than a placebo in order to make a transparent assessment of the novel intervention in comparison to usual care. Informed consent is required in all study designs, but particularly so in RCTs. Participants need to understand that they are participating in an experiment and, consequently, may not receive the treatment under investigation.

Observational studies

Cohort studies

Cohort studies follow a group of individuals over a period of time (often years) to examine whether they benefit from a healthcare or public health intervention or develop a disease or other outcome of interest. A cohort is a group of individuals who have a common characteristic. Cohorts are selected on the basis that individuals share similar, relevant characteristics, such as receiving cancer screening or being exposed to cigarette smoking. For example, the British Doctors Study, which ran from 1951 to 2001, was a cohort made up of doctors, including both smokers and non-smokers. It provided convincing proof of the association between smoking and lung cancer and led to a dramatic decline in smoking from the 1970s onwards.

Cohort studies are a type of observational study, meaning that no interventions are made to influence participants' treatments or lifestyle. However, detailed information is collected about the treatments they receive in the ordinary course of living their lives, their sociodemographic characteristics, lifestyles and behaviours, and their outcomes, using questionnaires, interviews, physical examinations, and investigations such as blood tests, as well as taking information from existing sources such as medical records. Cohort studies can therefore be prospective, involving the collection of new data, or retrospective, looking back in time and using existing data

such as medical records. Retrospective studies are cheaper and faster to conduct because the data have already been collected and stored. However, because it was not collected for research purposes, information in patients' medical records may be incomplete, inaccurate, and inconsistently recorded.

Cohort studies can also vary significantly in terms of their size and complexity. At one extreme, a large population may be studied over decades. For example, the Millennium Cohort Study (MCS) follows the lives of around 19,000 young people born across England, Scotland, Wales, and Northern Ireland in 2000–2001. Families were first assessed when children were 9 months old and have been followed up at regular intervals. Interviews are carried out with the main parent and the child, and a range of information is collected, including employment and income, housing, family structure, ethnicity, parenting activities such as reading to the child, developmental indicators, and both parental and child mental health. At the other extreme, some cohort studies follow up in relatively small groups for a few days or weeks.

Cohort studies are particularly valuable in situations where RCTs cannot be performed because randomization is either unethical or impractical. They also allow rare or long-term outcomes to be assessed: for example, in the UK, the National Joint Registry collects information on a nationwide cohort of patients who have undergone joint replacement operations. Implants that fail earlier than expected can be identified and the information acted on. External validity (also referred to as generalizability) will be higher in large cohort studies such as this because a broad range of patients, from different settings, are included, without the restrictive inclusion criteria that are often applied in RCTs.

A limitation of the cohort method is losses of follow-up. Cohort members may decline to continue participation in the study or fail to maintain contact. Loss to follow-up may (or may not) be related to the exposure, the outcome, or both. For example, a study looking at the effect of high-impact exercise on risk factors for osteoporosis had a greater loss of follow-up in the group assigned to exercise. Females who dropped out may have had inherently lower bone strength, which made the exercise less tolerable for them. If so, this would have produced selective loss of females more likely to develop osteoporosis from the active treatment group, and this would cause a biased estimate: specifically, an overestimate of the benefit of exercise.

Case-control studies

In a case-control study, patients who have received a healthcare intervention or developed a disease are identified. Their past history is then compared with that of individuals—called controls—who did not receive the intervention or do not have the disease. For example, participants with a disease of interest (e.g., lung cancer) may be tracked back in time, sometimes for many years, to determine whether they were exposed to a risk factor of interest (e.g., secondhand tobacco smoke). In another example, a study used primary healthcare records to look at service use in the previous year, in order to identify factors associated with frequent attendance at hospital emergency departments.

Researchers try to ensure that the cases and controls are as similar as possible with respect to other factors. This can be done either on an individual basis (e.g., by pairing each case with a control of the same age and sex) or in groups (e.g., choosing a control group with an overall age and sex distribution similar to that of the cases). Unknown differences may still exist between the cases and controls, however, which could influence the results.

A particular limitation of case-control studies is the risk of recall bias. Case participants are usually keen to find out what caused their illness or may ascribe their positive and adverse outcomes to an intervention. As a result, they may be better motivated to remember details of their past than controls with no special interest in the study question. Assessing the exposure of controls may therefore be challenging, especially if this relies only on their personal recall. This risk can be lessened by using data from electronic medical records.

Since the assessment is retrospective and the risk of recall bias is high, case-control studies lie towards the bottom of the evidence hierarchy. However, they can provide interesting insights. For example, case-control studies start with people who are known to have a disease of interest, rather than looking at a population free of disease and waiting to see who develops it. This approach therefore lends itself to studying groups of patients with rare conditions. For example, in 1955, Richard Doll published a case-control study demonstrating a significant excess of the rare lung cancer mesothelioma among asbestos workers. This study was the first to identify asbestos as one of the world's most dangerous industrial carcinogens.

Cross-sectional surveys

These observational studies provide a snapshot of a particular group of patients or people at a single point in time and can be used to assess the prevalence of illness or a risk factor, or the range of perceived symptoms in a given population. Examples include the proportion of patients with bladder cancer who smoke, or patient-reported outcomes on discharge from hospital.

As cross-sectional data describe only what is happening at a single moment in time, they cannot be used to answer questions about the causes of disease or the impact of treatment. They also do not provide information about changes over time, unless they are repeated. Consequently, they are ranked low in the hierarchy of evidence.

Nevertheless, cross-sectional surveys can provide key information, for example, in planning services and ensuring that they are equipped to meet the needs of the local population. They may also be used to draw inferences about possible relationships between factors or outcomes, or to gather preliminary data to support further research and experimentation.

Case series or case reports

These describe a group of patients or a single individual with an outcome of interest, or novel and unusual features. Due to their descriptive nature, they provide little evidence of causality. Their small size and lack of representativeness also mean that they cannot be assumed to have relevance for the wider population of patients with the same condition. Therefore, they are regarded as the weakest form of evidence in the hierarchy.

Nevertheless, they can still provide useful information, particularly in the very early stages of research into a disease area. For example, in 1985, the Centers for Disease Control and Prevention in the United States published a report describing five cases of *Pneumocystis* pneumonia in previously healthy, gay males in Los Angeles. This was the first official reporting of what became known as the AIDS epidemic. However, case reports can also be misused: an important example is the now discredited case series published in *The Lancet* in 1998 (retracted in 2010), which falsely claimed to provide evidence for a link between the measles, mumps, and rubella vaccine; colitis; and autism.

Problems with the 'hierarchy of evidence'

Although the evidence hierarchy has become widely accepted, concerns have been raised about ranking evidence in this way. As

has been outlined, study designs traditionally ranked lower down the hierarchy may be more appropriate if, for example, outcomes of interest are rare or occur far into the future. Some observational studies also produce results with higher external validity than some RCTs, because inclusion and exclusion criteria are usually less restrictive. Additional concerns, including the extent to which the findings of RCTs can be applied to patients in other settings, are explored in the final section of this chapter.

The hierarchy focuses only on research addressing the question of the effectiveness of interventions. Other questions are equally salient, however, such as:

- prognosis (what is likely to happen?)
- harm (is this substance likely to have adverse effects?)
- experience (what is it like to receive this intervention?)
- implications for services (staff workload, the need for additional monitoring, and so on).

Different types of questions require different study designs (**Box 4.3**):

- Determining the number of patients living with a condition at a given point in time (a ***prevalence question***) may be best achieved using a cross-sectional survey. However, because it provides only a snapshot of that single time point, it cannot estimate the likely duration between diagnosis and death.
- The design needed for a ***prognosis question*** is a cohort study—a study that follows up on recently diagnosed patients and records what happens to them.
- To answer ***experience questions*** (what it is like to live with a particular condition, or whether a certain intervention is acceptable to patients), qualitative approaches, such as interviews or focus groups, are often also needed.

Combined approaches

The most comprehensive, generalizable, and useful research is that which combines quantitative and qualitative data. For example, qualitative approaches can helpfully be used within an RCT to explore how an intervention was implemented and what was actually delivered, compared with what was intended. This is known as a ***process evaluation***, as it documents the process that took place. Process evaluation can explain discrepancies between expected and observed outcomes, help us to understand how local context influences outcomes, and provide insights to aid implementation in other settings.

Further reading

Vanderbeek A, Bliss J, Yin Z et al. Implementation of platform trials in the COVID-19 pandemic: A rapid review. *Contemp Clin Trials*. 2022; 112.
The Adaptive Platform Trials Coalition. Adaptive platform trials: definition, design, conduct and reporting considerations. *Nat Rev Drug Discov*. 2019; 18: 797–807.

Critical appraisal

Critical appraisal aims to:

- identify the methodological strengths and weaknesses of a piece of research
- assess whether the researchers used the right design for their study
- make informed decisions about the quality of the evidence being appraised
- establish how useful the piece of research is for wider decision-making.

A range of tools and checklists exist to help assess the quality of each design (see 'Further reading'). However, these checklists should be regarded as an aid to careful thought, rather than a substitute for it. Regardless of the study type, critical appraisal involves some overarching questions (**Box 4.4**); the three main steps involved are described below.

Question 1: Are the results of the study valid?

Validity refers to the soundness or methodological rigour of a study. A study is considered valid if the way it has been designed and carried out means that the results provide a trustworthy estimate of clinical effectiveness. A study that is ***sufficiently free from bias*** has internal validity.

Bias is a systematic error that causes a study to produce an invalid estimate. For example, in a case-control study, recall bias may occur because a patient with an outcome of interest (e.g., lung cancer) may report their exposure experience (e.g., smoking) differently to a control without the outcome. This would then lead to an invalid estimate of the association between smoking and lung cancer.

No study is perfect and free from all bias. A systematic check should be carried out to ensure that the researchers have done all

Box 4.3 Key messages about study design

Randomized controlled trials (RCTs)

- ***Randomization*** aims to ensure that the study groups are as similar as possible, apart from whether or not they receive the treatment being studied. Consequently, observed differences are more likely to be due to differences in the treatment alone.
- ***Blinding*** ensures that there are no differences in the way in which participants in each study arm are managed or in which their progress and outcome are measured. In a double-blind study, neither patients nor investigators know the study arm to which the participant has been allocated.
- In an ***intention-to-treat analysis***, RCT findings are analysed based on the treatment arm to which participants were initially allocated, regardless of whether they dropped out of the study.
- Ensuring that RCT participants are as similar as possible enhances the ***internal validity*** of the results (i.e., their accuracy in the study context) but may reduce their ***external validity*** (i.e., their generalizability to the diverse range of patients with the condition under study).
- RCTs are a form of ***experimental study.*** The researcher manipulates exposure to the intervention (e.g., a new drug treatment) by allocating participants to the intervention or control group.

Observational studies

- In an ***observational study***, the researcher measures exposures (e.g., new surgical procedures) that the participants are already receiving. Outcomes (e.g., postoperative complications) are then observed as they occur.
- In a ***cohort study***, a group of participants (the cohort) is followed over time. Researchers observe which individuals develop outcomes of interest (e.g., breast cancer) and collect information about factors that might be relevant to the development of the outcome (e.g., smoking, alcohol consumption). Cohort studies can provide valuable evidence when conducting an RCT is unfeasible. The findings may have greater generalizability because inclusion and exclusion criteria are usually less restrictive.
- In a ***case-control study***, patients who have an outcome of interest are identified (i.e., cases). Their past history is then compared with that of individuals—or controls—who do not. Cases and controls should be as similar as possible with respect to other factors. A particular limitation of case-control studies is the risk of recall bias because cases may be better motivated to remember details of their past than controls.

Box 4.4 Key questions in critical appraisal

Why was the study done?
- A clearly focused question should address population, intervention, and outcomes.

What type of study was done?
- The study design must match the question asked.

What are the study characteristics?
- Use the PICO question format to help answer this question.

Are the results of the study valid?
- Was the assignment of patients to treatments in a controlled trial (RCT) randomized? If another design was used, how were participants allocated?
- Were patients, health workers, and study personnel 'blind' to treatment allocation?
- Were all the patients who entered the study properly accounted for at its conclusion? Look for follow-up tables and whether patients in an RCT were analysed in the groups to which they were randomized.
- Were the groups similar at the start of the study?
- Aside from the experimental intervention, were the groups treated equally?

What were the results?
- How large was the treatment effect? How precise was the estimate of the treatment effect?
- Look for confidence limits and p-values.

Will the results help?
- Can these results be applied locally? Are these findings applicable to other patients?
- Were all relevant outcomes considered, including those relevant to patients?
- Are the benefits worth the harms and costs?

they can to minimize it. Different study designs are prone to different types of bias, but observational studies (and especially case-control studies) are particularly susceptible. More than 50 types of research bias have been identified but, for simplicity, they can be broadly grouped into two categories:

- ***Information bias.*** This results from systematic differences in the way in which data are collected from the various study groups, including data relating to either the exposure (i.e., factors that may be associated with an outcome of interest, such as weight) or the outcome (e.g., type 2 diabetes). For example, participants' weights may have been recorded when they visited a GP complaining of symptoms (which may be indicative of type 2 diabetes), or the study may rely on information collected at the time the participants registered in primary care. In the former case, information might be systematically more accurate for patients with diabetes than for patients without weight-related conditions, leading to an incorrect estimate of the association between exposure and outcome.
- ***Selection bias.*** This occurs when there is a systematic difference between either:
  - those who participate in the study and those who do not, *or*
  - those in the treatment arm of a study and those in the control group.

Patients and clinicians who have strong treatment preferences are often excluded from RCTs. While these steps increase the internal validity of an RCT, they often lead to selection bias and an overestimation of the benefits of treatment, especially for patient-centred outcomes. For example, some people with early breast cancer have a strong preference for less invasive lumpectomy, whereas others prefer the reassurance of knowing that all potentially cancerous tissue has been removed by a mastectomy. In one series of RCTs, only people without a strong preference for a particular treatment were recruited. However, as few as 10% subsequently agreed to have their treatment chosen at random, significantly reducing the extent to which the findings could be generalized to the wider population.

Similarly, telehealth interventions are becoming a prominent part of healthcare delivery, with examples including online cognitive behavioural therapy or home monitoring of health parameters such as blood pressure. There is some evidence of their effectiveness from RCTs. However, in several trials a large number of eligible individuals (over 75% in some cases) have declined to take part. As well as compromising external validity (i.e., generalizability of findings to other populations outside the setting of the trial), this may also indicate problems with the acceptability of the intervention, which may have a negative impact on uptake if it is later introduced into routine practice. In the case of telehealth, issues might include physical barriers in terms of accessing technology, as well as psychological barriers such as low confidence in using it.

Question 2: What are the results?

The second step of critical appraisal is to consider the results. In any study reporting quantitative data, these should be presented using statistical methods, including statistical measures of the uncertainty around study estimates. The particular methods employed will, and should, depend on the study design and the types of data that are being analysed. The section 'Understanding statistical results' (see below) gives a short introduction to the critical appraisal of statistical reports.

Question 3: Will the results help?

The final step of critical appraisal is considering whether the study is applicable to the decision being made for a particular patient. There are three issues to consider here.

First, are there any ***important differences*** between the participants in the study and the patient in question? These must be identified, as they could have an impact on the effectiveness of an intervention. Check the study inclusion and exclusion criteria, which are usually reported in the methods section. Would the patient in your care have been eligible for inclusion in the study? Identify any characteristics that the patient has that were not considered in the study.

Second, many interventions and processes used in everyday clinical practice have both benefits and ***adverse consequences***, and these should be weighed against each other. For example, breast cancer screening programmes in the UK currently invite females aged 50 to 70 years to undergo mammography every 3 years. Breast screening helps identify breast cancer early. The earlier the condition is found, the better the chances of surviving it, and the less likely patients are to need a mastectomy or adjuvant therapy (chemotherapy and/or radiotherapy). However, all screening can lead to overdiagnosis. Overdiagnosis refers to the detection of cancers through screening that would not have become clinically apparent in the person's lifetime. Although their cancer would not have caused them ill health, these people will undergo treatment, such as surgery, which may result in side-effects and complications, as well as psychological, social, and financial consequences. Of the 470,000 UK females aged 50 years who are invited to begin screening every year, it is estimated that just over 3 per 1000 will be overdiagnosed with cancer in the next 20 years. Although most public health bodies internationally recommend screening, there remains some debate as to the balance of benefits and harms. Finally,

cost-effectiveness is often not reported in the analysis of clinical trials. If a treatment is effective but provides only a small benefit and is very expensive, it may not be a good use of resources. Equally, even if a treatment is shown to be cost-effective, it may not be affordable. In the UK, NICE uses the cost per quality-adjusted life-year (QALY) to judge the cost-effectiveness of new technologies. This is the estimated cost of each additional year of life gained by a treatment or procedure, measured not just in terms of extra months or years of life but also in terms of the quality of that life. This is defined, for example, as freedom from or reduction in pain, or the ability to carry out the basic activities of daily living. Currently, NICE uses a threshold of £20,000–£30,000 per QALY. Treatments that come in at or below the threshold are usually approved. Those that come in above tend not to be judged cost-effective and so are not recommended for adoption by the NHS. Again, however, affordability of the healthcare system is not taken into account by this approach. For example, a treatment may result in significant quality-of-life gains but be very expensive and applicable only to a small subset of patients with a particular condition.

Further reading

Black N. Why we need observational studies to evaluate the effectiveness of health care. *BMJ* 1996; 312:1215–1218.
Straus SE, Glasziou P, Richardson WS et al. *Evidence-Based Medicine*. Edinburgh: Elsevier/Churchill Livingstone; 2018; http://www.casp-uk.net/casp-tools-checklists *Critical Appraisal Skills Programme (CASP) checklists.*

UNDERSTANDING STATISTICAL RESULTS

Statistics is the science of learning from data. It has two branches: descriptive and inferential. Descriptive statistics is the art of describing and visualizing the information contained in a given dataset. Inferential statistics is the art of drawing inferences from a particular dataset to the real-world process that is being studied. Reports of statistical results in medicine should contain both descriptive and inferential statistics.

Descriptive statistics

Descriptive statistics may be presented in the form of tables, graphs, and numbers included in a text. A description of the dataset is an important part of a statistical report. Descriptive statistics provide context and allow the reader to judge the representativeness of the data and how likely it is that biases are present. For example, descriptive statistics may show that all patients studied were aged 65 or younger. It is then clear that the study's conclusions may not be generalizable to patients who are older.

Inferential statistics

Description of a particular dataset in itself does not enable us to draw conclusions about the real-world process we wish to study. To do so, we need to employ inferential statistics. Inferential statistics are based on mathematical models that account for the influence of randomness on the data that we happen to observe in a particular study. In general, inferential statistics account for randomness only and do not account for other aspects that may affect the validity or reliability of a study's results, such as poor data quality or selection bias. For the reader of research reports, the two most important techniques of inferential statistics to understand are confidence intervals and hypothesis tests.

Confidence intervals

To understand confidence intervals, consider the following example. During the COVID-19 pandemic, the UK Office for National Statistics (ONS) conducted regular surveys to estimate the SARS-CoV-2 infection rate. In the last week of September 2021, the ONS estimated that 1.6% of the population of Scotland had a SARS-CoV-2 infection (https://www.beta.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/coronaviruscovid19infectionsurveypilot/8october2021). Since this estimate was based on testing a random sample of people, rather than testing everyone in the population, there is statistical uncertainty around this estimate.

A confidence interval is used to quantify the uncertainty around an estimate. For example, the ONS reported a confidence interval of 1.3% to 1.9% around its estimate of the SARS-CoV-2 infection rate in Scotland. Such a confidence interval is often written as [1.3%, 1.9%]. Informally, this confidence interval can be interpreted to mean that the true infection rate is probably somewhere between 1.3% and 1.9%.

A confidence interval is specified by a confidence level. Typically in medical research, 95% confidence intervals (95% CIs) are reported, although other confidence levels (such as 99% or 90%) are sometimes used. A 95% CI has a 95% probability of containing the true value that we are trying to estimate, but that we do not know. Another way to understand this statement is: if I conducted 100 studies to estimate the infection rate and calculated a 95% confidence interval from each study, on average, 95 of these 100 confidence intervals would contain the true infection rate.

The importance of confidence intervals lies in the fact that they tell us something about the uncertainty around an estimate. A wider confidence interval indicates more uncertainty. For example, consider **Box 4.5** below, which contrasts two hypothetical studies: one large and one small. Both studies estimate the SARS-CoV-2 infection rate in Scotland to be 1.6%, but their confidence intervals differ. The 95% CI from the smaller study is 0.7% to 3.1%. This confidence interval is relatively wide, reflecting a high statistical uncertainty.

Generally, a published study should report confidence intervals around all important estimates. These may be estimates of percentages or proportions (as in the example above), or of other quantities, such as the size of a treatment effect. Confidence intervals are important because they indicate the statistical uncertainty around an estimate. To say that 'this study shows that the infection rate in Scotland is 1.6%' is not justified, because the true infection rate remains unknown. A better way to interpret the study's results acknowledges the uncertainty, for example, 'From this study, we estimate that the infection rate is probably somewhere between 1.3% and 1.9%'.

Hypothesis tests and p-values

Hypothesis tests are used to formally evaluate the statistical evidence for the existence of treatment effects or other effects of interest. There are many different types of hypothesis tests, but the basic logic is the same, regardless of which particular test is used.

As an example, consider a study of the effect of dietary salt intake on systolic blood pressure (SBP). Patients with chronic kidney disease (CKD) were randomized to either of two groups: 38 patients followed a diet with normal salt intake, and 38 different patients had low salt intake for 1 week. SBP was measured at the end of the week.

Box 4.5 Estimates and confidence intervals of the SARS-CoV-2 infection rate from two hypothetical studies

Study size	Estimated percentage of infected people	Estimated number of infected people (in a population of 5.3 million)	95% confidence interval for the percentage	Confidence interval for the number of infected people	The estimate is…
Large sample	1.6%	85,000	[1.3%, 1.9%]	[69,000, 101,000]	…relatively precise
Small sample	1.6%	85,000	[0.7%, 3.1%]	[37,000, 164,000]	…relatively imprecise

The mean (average) SBP in the low salt group was 115.0 mmHg, while in the normal salt group the mean was 121.6 mmHg. A measure of the effect of a low salt diet on SBP is the difference between these means. The mean difference observed in this study is 115.0 mmHg – 121.6 mmHg = –6.6 mmHg. So the 'low salt' group, on average, had a lower SBP, by 6.6 mmHg.

This mean difference in the sample (–6.6 mmHg) is a descriptive statistic. It applies to our particular sample of 38 patients. To draw conclusions about the effect of a low salt diet in patients with CKD generally, we need to use inferential statistics.

Statistical hypothesis tests assess the compatibility of a statistic with a pre-specified hypothesis about the world outside our sample. This hypothesis is formally called the *null hypothesis*. The null hypothesis in our example may be: 'Salt intake has no effect on SBP'. Is the result from our sample (mean difference of –6.6 mmHg) compatible with the absence of a real effect of salt intake on SBP in patients generally? This is the question a hypothesis test would aim to answer in this example.

The results of a hypothesis test may be reported as something like this:

$$\text{Mean}_{\text{normal_salt}} - \text{Mean}_{\text{low_salt}} = -6.6\,\text{mmHg},\ p = 0.021$$

Other numbers relevant to the test may be reported, but to explain them in detail is beyond this short introduction. For our purposes, it is important to understand the quantity p, which is commonly called the p-value. Here, the p-value is 0.021.

The p-value is a probability and can take values between 0 and 1. It indicates the probability of seeing a difference as large as the one we did see (or a larger one), if the null hypothesis was in fact true. In this example, the p-value is the probability of seeing a difference of 6.6 mmHg (or larger) between the two groups if in reality salt intake had no effect on SBP. Here, p = 0.021, or about 2%. This means that, if in reality salt intake had no effect, and we conducted our study 100 times on different patients, on average we would see a difference of 6.6 mmHg in SBP or larger in 2 out of our 100 studies. So if the null hypothesis was true (if salt intake had no effect), our observed difference would be unlikely. This is interpreted as evidence that the null hypothesis may be false, which in our example means: the small p-value is evidence that salt intake does in fact have an effect on SBP.

The null hypothesis to which a particular p-value refers may not always be explicitly stated in a publication. Usually, the null hypothesis has the form 'the treatment has no effect' or 'the exposure has no effect', although this is not necessarily so. When trying to understand the meaning of a p-value in a published study, ask yourself which null hypothesis it assesses, as this must be clearly understood in order to interpret the p-value correctly.

Statistical significance

A customary threshold of $p<0.05$ is often used to make a binary decision whether the evidence against the null hypothesis is strong enough to 'reject' it, or not. According to this method, if $p<0.05$, the result is called 'statistically significant' and is taken as evidence against the null hypothesis. For example, when a research report states that 'the treatment group had a lower blood pressure than the control group, and this difference was statistically significant', this is usually a way of saying that $p<0.05$ for the null hypothesis of 'no difference' in blood pressure between the groups.

The threshold '$p<0.05$' is arbitrary, and there is no statistical or scientific reason why it should be preferred over other possible thresholds. It is an unfortunate fact that in much medical research today, complex statistical results are often reduced to a binary judgement regarding whether some result is 'statistically significant' or not. Good publications should report precise p-values (e.g., p = 0.021, as in the example above), rather than merely stating if a result is 'statistically significant' or not. However, p-values smaller than 0.001 are often legitimately reported as '$p<0.001$' in research publications.

Many misconceptions exist about the concept of 'statistical significance' among researchers, and these misconceptions are often reflected in erroneous interpretations of study results. Three points are especially worth emphasizing:

Statistical significance does not mean strong scientific evidence. If $p<0.05$, this is often misinterpreted to mean that a finding is scientifically robust. In fact, all aspects of the study need to be considered to judge the scientific robustness of a finding, including the study design, the quality of the data, the likelihood of biases, the plausibility of the causal mechanism hypothesized to explain the finding, etc. None of these other aspects are measured by the p-value.

Statistical significance does not mean clinical relevance. If $p<0.05$, this is often misinterpreted to mean that a finding is clinically important, or important to patients. In fact, a treatment may have a very small effect (e.g., it may very slightly reduce blood pressure) that is 'statistically significant' but clinically irrelevant (e.g., the small reduction in blood pressure may not be sufficient to help prevent heart disease).

Absence of evidence is not evidence of absence. If $p>0.05$, this is often misinterpreted to mean that the null hypothesis is true. In fact, a large p-value means that this particular study has not produced strong evidence *against* the null hypothesis. But a large p-value does *not* constitute evidence in favour of the null hypothesis. For example, you may read a sentence like: 'There was no difference between the treatment and control groups (p = 0.38).' This is a misleading statement. A better report should state what the descriptive difference between treatment and control group actually was, should give the confidence interval for the estimated treatment effect, and might interpret the large p-value as indicating 'weak statistical evidence for an effect of the treatment' or 'insufficient statistical evidence to conclude that the treatment has an effect'. Just because a particular study did not find strong evidence for an effect does not mean that the study has established that no effect exists in reality. In hypothesis testing, absence of evidence for an effect is not evidence of absence of an effect.

Worked examples

We will now turn to worked examples that illustrate the use of descriptive statistics, confidence intervals, and hypothesis tests in real research studies. The four examples differ by study design and the type of outcome that is being measured. We consider, in turn:

- Treatment effects on numeric measurements
- Treatment effects on dichotomous (binary) outcomes
- Treatment effects on the time it takes for an event to occur
- Meta-analyses of studies of treatment effects.

Numeric outcomes: differences between means

When the outcome is a numeric measurement, such as blood pressure, body weight, or serum creatinine, the measure of the treatment effect is often defined to be the difference in means between the treatment group and the control group. Consider again the hypothetical study of the effect of dietary salt intake on systolic blood pressure (SBP). We saw that the mean difference in the sample was –6.6 mmHg, such that the low salt group had lower SBP on average, and that this may be reported as:

$$\text{Mean}_{\text{normal_salt}} - \text{Mean}_{\text{low_salt}} = -6.6\text{mmHg},\ p = 0.021.$$

The small p-value indicates statistical evidence that a low salt diet results in lower average SBP than a normal salt diet. Conversely, if the p-value had been large (e.g., p = 0.32), this would be taken to indicate that there is weak statistical evidence from this study for an effect of salt intake on blood pressure.

A confidence interval (CI) of the treatment effect should always be considered alongside the p-value. Here, the 95% CI for the mean difference is [–1.0 mmHg, –12.2 mmHg]. This indicates that, according to this study, a low salt diet reduces SBP by probably somewhere between 1.0 and 12.2 mmHg. It is thus possible, judging from this study alone, that the effect of the low salt diet is very small: a reduction of only 1 mmHg is unlikely to be clinically relevant. On the other hand, this study's data are also compatible with a rather large effect (a reduction of 12.2 mmHg). Larger studies, or a meta-analysis of small studies, may be needed to obtain a more precise estimate of the effect of a low salt diet on SBP.

There is a mathematical correspondence between confidence intervals and p-values. When a 95% CI contains the test value of the null hypothesis (here: 0 mmHg difference due to low salt intake compared to normal salt intake), then the p-value for that null hypothesis will always be 0.05 or larger. Conversely, when a 95% CI does *not* contain the test value, the p-value will always be smaller than 0.05.

(This example is inspired by *Konishi Y et al. Sodium sensitivity of blood pressure appearing before hypertension and related to histological damage in immunoglobulin A nephropathy.* Hypertension *2001 38; 1:81–85*, although Konishi et al. used a different study design and hence different tests.)

Dichotomous (binary) outcomes

An outcome is dichotomous when there are exactly two possibilities: either an event occurs or it does not. Examples of dichotomous outcomes are patient death (death or survival), recovery from illness (recovered or not), and hospital readmission after surgery (readmitted within 12 months or not readmitted).

Consider the following example. A randomized placebo-controlled trial was conducted in South Africa and Malawi to evaluate the efficacy of a rotavirus vaccine in preventing severe gastroenteritis in infants; 4417 infants were included in the analysis, and of these, 2974 were randomly allocated to receive the vaccine, while 1443 received a placebo. The primary outcome was gastroenteritis caused by rotavirus in the period from 2 weeks to 1 year after the last dose of the vaccine (or placebo) was given. In the vaccine group, rotavirus infection was detected in 56 infants (1.88%); the same was true for 70 infants in the placebo group (4.85%); *Madhi SA et al. Effect of human rotavirus vaccine on severe diarrhoea in African infants.* New England Journal of Medicine *2010; 362:289–298.*

A good paper should report all these descriptive statistics. They are summarized in **Box 4.6**. The results might then be further analysed in several ways.

Odds and the odds ratio

The odds ratio is a measure of the relative effect of a treatment on the odds of an effect of interest. The odds of an event are calculated as the number of people who experience the event divided by the number who do not experience the event. In our example, the odds of disease for the vaccine (intervention) group are:

$$Odds_{\text{int}} = D_{\text{int}} / H_{\text{int}} = 56/2918 = 0.0192$$

By the analogous calculation, the odds of disease in the control group are $Odds_{cont} = 0.0510$.

The odds ratio (*OR*) is calculated as the odds in the intervention group divided by the odds in the control group:

$$OR = \frac{Odds_{\text{int}}}{Odds_{cont}} = \frac{0.0192}{0.0510} = 0.38$$

This result means that the vaccine group had lower odds than the placebo group by a factor of 0.38. The same result is sometimes expressed by stating that *the vaccine group had 62% lower odds than the placebo group*, since (1 – 0.38) × 100% = 62%.

Generally, an odds ratio of 1 would indicate that there is no difference between the two groups. In our example, if *OR* = 1, this would indicate that the vaccine had no effect on the odds of disease. An odds ratio smaller than 1 indicates that the treatment group has lower odds of the event compared to the control group, while an odds ratio larger than 1 indicates that the treatment group has higher odds of the event.

Turning now to inferential statistics, the odds ratio observed in the sample (here: 0.38) is taken as an estimate of the odds ratio in the population (here: the reduction of disease odds that would result if all infants were vaccinated). A good publication should report a confidence interval around the odds ratio estimate. In the rotavirus

Box 4.6 Results from an RCT of rotavirus vaccine in Malawi and South Africa

	Event: Rotavirus gastroenteritis	Non-event: No rotavirus detected	Total
Intervention Arm: Vaccine	D_{int} 56	H_{int} 2918	N_{int} 2974
Control Arm: Placebo	D_{cont} 70	H_{cont} 1373	N_{cont} 1443

D denotes patients experiencing gastroenteritis (***D***iseased), H patients who did not (***H***ealthy), and N the total ***N***umber of patients studied. The 'int' suffix denotes patients in the '***int***ervention' (vaccine) group and 'cont' patients in the ***cont***rol group.

vaccine study, the 95% CI for the odds ratio is [0.23, 0.54]. So the data from this study suggest that the vaccine reduces the odds of severe rotavirus gastroenteritis by probably somewhere between a factor of 0.23 and 0.54. We can see that there is considerable uncertainty around the estimate—the confidence interval indicates that the data are compatible with a wide range of effect sizes, which include a halving of the odds (*OR* = 0.5) as well as a fourfold reduction of the odds (*OR* = 0.25).

Risk and the risk ratio

The risk ratio is another measure of the relative effect of two treatments. The risk of an event is estimated as the proportion of people who experience the event (i.e., the number of people who experience the event divided by the total number of people under investigation). The estimated risk of disease in the vaccine group is:

$$Risk_{int} = D_{int} / N_{int} = 56/2974 = 0.0188$$

By the analogous calculation, the estimated risk of disease in the control group is $Risk_{cont}$ = 0.0485. Risks are also sometimes expressed as percentages. Here, the estimated risk is 1.88% in the treatment group and 4.85% in the control group.

The risk ratio (*RR*) is calculated as the risk in the intervention group divided by the risk in the control group:

$$RR = \frac{Risk_{int}}{Risk_{cont}} = \frac{0.0188}{0.0485} = 0.39$$

This result means that the vaccine group had a lower proportion of diseased infants than the placebo group, by a factor of 0.39. The same result is sometimes expressed by stating that the vaccine group had a 61% lower estimated risk than the placebo group.

Generally, a risk ratio of 1 would indicate that there is no difference between the two groups. A risk ratio smaller than 1 indicates that the treatment group has a lower risk of the event compared to the control group, while a risk ratio larger than 1 indicates that the treatment group has a higher risk of the event.

The risk ratio calculated from the study data is taken as an estimate of the risk ratio in the population. A 95% confidence interval for the risk ratio is [0.27, 0.55]. So the data from this study allow us to estimate that the vaccine reduces the risk of severe rotavirus gastroenteritis by probably somewhere between a factor of 0.27 and 0.55 (approximately quartering to halving the risk). This indicates considerable statistical uncertainty.

In the case of vaccination studies, the quantity (1 – *RR*) × 100% is called the vaccine efficacy. Here, the vaccine efficacy is estimated as (1 – 0.39) × 100% = 61%. This means that a vaccinated person has a 61% lower estimated risk of contracting the disease compared to an unvaccinated person. A 95% confidence interval for the vaccine efficacy should also be reported (in our example this is [45%; 73%]).

More generally, when a treatment or exposure reduces the risk of the outcome event, a relative risk reduction (RRR) can be calculated, such that *RRR* = 1 – *RR*. In our example, the relative risk reduction is as follows: *RRR* = 1 – 0.39 = 0.61, or 61%. Conversely, if a treatment or exposure increases the risk of the outcome event, such that the risk ratio is larger than 1, a relative risk increase (RRI) can be calculated as *RRI* = *RR* – 1.

Absolute risk reduction

The *risk difference* associated with a treatment is calculated by subtracting the risk of disease in the treated group from the risk in the control group. Since usually a treatment is hypothesized to reduce the risk of the outcome event, the risk difference is often referred to as the *absolute risk reduction* (*ARR*). Should a treatment or exposure increase the risk of an outcome event, the term *absolute risk increase* would be more appropriate.

Using the estimates from the vaccination study, the *ARR* is calculated thus:

$$ARR = Risk_{cont} - Risk_{int} = 0.0485 - 0.0188 = 0.0297$$

$$= 2.97 \textit{ percentage points}$$

This result means that the proportion of diseased infants in the vaccine group was 2.97 percentage points lower than in the control group.

A 95% confidence interval for the *ARR* is [0.018, 0.042]. So the data from this study allow us to estimate that the vaccine reduces the risk of severe rotavirus gastroenteritis by probably somewhere between 1.8 and 4.2 percentage points.

About odds ratios, risk ratios, and absolute risk reduction

The odds ratio and the risk ratio are common effect size measures in clinical trials as well as observational studies. They are useful because they can be compared across studies with different baseline odds, or baseline risks. For example, suppose the risk of death from some disease is 10% in Country A, and a study finds that a new drug has the effect of halving that risk (*RR* = 0.5) so that the risk of death among patients who take the drug is 5%. A different study may investigate the same drug in a different country, Country B, where the risk of death from the disease may be only 4%. If the effect of the drug is the same in both countries (*RR* = 0.5), the risk of death among patients in Country B who take the drug is 2%. Note that, despite the risk ratio being the same in both studies, the ARR would be 5 percentage points in the first study, but 2 percentage points in the second study. So the drug, given to 100 patients in each country, would, on average, save 5 lives in Country A, but only 2 lives in Country B.

The risk ratio and the odds ratio can encourage misleading interpretations, if the baseline risk is small. For example, an epidemiological study may find that taking a certain food supplement is associated with halving the risk of some disease (*RR* = 0.5). This sounds like a large effect: 50% less risk! However, suppose the baseline risk is 0.001% (1 in 100,000 people get the disease without the food supplement). Then *RR* = 0.5 means that the risk with the food supplement is 0.0005% (1 in 200,000 people). So the absolute risk reduction is 0.0005 percentage points, which does not sound quite as impressive as a 50% risk ratio. A good research publication will report both a measure of relative effect and a measure of absolute effect. If only a relative effect measure is reported, it is often useful, as a reader, to use the descriptive statistics that have been reported to calculate an absolute effect measure oneself, in order to get a fuller picture of the size of the effect.

Number needed to treat

The number needed to treat is another measure of absolute effect size, which is popular due to its intuitive interpretation: the number of people who need to receive the treatment in order for one person to benefit. The number needed to treat (*NNT*) is calculated thus (again using the rotavirus vaccine study as an example):

$$NNT = \frac{1}{ARR} = \frac{1}{0.0297} = 33.7$$

This means that 33.7 infants would need to be vaccinated for 1 infant not to get rotavirus gastroenteritis, who would have got it without the vaccine. A 95% CI for the *NNT* is [23.9, 56.9]. So the data from this study allow us to estimate that the number needed to treat with the vaccine to avoid one case of the disease is probably somewhere between 23.9 and 56.9.

Hypothesis tests for dichotomous outcomes

The primary effect size measure for dichotomous outcomes is usually either the risk ratio or the odds ratio. To gauge the evidence for an effect of the treatment, a hypothesis test is usually conducted. The null hypothesis is then usually that *RR* = 1, or that *OR* = 1, which would indicate the absence of an effect of the treatment. In the rotavirus vaccine study, a test for the null hypothesis '*RR* = 1' yields p = 0.00000008. This very small p-value is strong statistical evidence that the vaccine has an effect. Conversely, a large p-value would be taken to indicate that there is weak statistical evidence for an effect of the vaccine from this study.

Time-to-event data (survival analysis)

Some studies follow participants over long periods of time and record the occurrence of events during that time. For example, the Scandinavian Simvastatin Survival Study (4S study) was a randomized controlled trial of 4444 patients with angina pectoris or previous myocardial infarction, who also had high serum cholesterol. Patients were randomized to receive either simvastatin or a placebo. One of the outcomes studied was death from any cause. Patients were followed up for 5 to 6 years from randomization. (See: *Scandinavian Simvastatin Survival Study Group. Randomised trial of cholesterol lowering in 4444 patients with coronary heart disease: the Scandinavian Simvastatin Survival Study [4S].* The Lancet *1994 344; 8934:1383–1389.*)

Survival analysis considers the outcome to be the time until a specific event occurs (here: the time until the participant dies, called 'time to death'). When analysing time-to-event data, it is important to consider unequal follow-up times. These come about because patients are usually not all randomized at the same time and also because some patients drop out of the study or are lost to follow-up before they experience the event. For example, if a patient moved out of the country 2 years after randomization, the investigators know that the patient was alive for the first 2 years of the study duration, but no information is available for the remainder of the intended follow-up time. This patient's follow-up is said to be 'censored at 2 years'. Survival analytic methods can account for censored data and obtain valid estimates of the survival chances of patients even in the presence of unequal follow-up times, while using all available data.

In the 4S study, death was recorded among 256 out of 2223 patients (11.5%) in the placebo group and 182 out of 2221 patients (8.2%) in the simvastatin group. This is important information, but since follow-up times varied between patients, these percentages cannot straightforwardly be used to compare the effects of treatment and placebo on the risk of death.

Kaplan-Meier plot

Time-to-event data are often visualized using a so-called Kaplan-Meier survival plot (K-M plot, **Fig. 4.2**). A K-M plot estimates the proportion of patients who have not yet experienced the event ('survived') over time and takes into account censoring. In this example, the K-M plot indicates that deaths occurred at a slower rate in the simvastatin group compared to the placebo group.

Cox regression and the hazard ratio

Several statistical methods may be used to make inferences about treatment effects from survival curves. A common method is Cox regression (also called 'Cox model' or 'proportional hazards model'). Cox regression can be used to model the effect of a treatment on the *hazard* of the outcome event (e.g., death) from time-to-event data. The hazard is the rate of events at one moment in time. The hazard may change over time, but the Cox model assumes that the ratio of the hazards of the two treatment groups remains constant over time. This is known as the 'proportional hazards assumption'.

The measure of the treatment effect provided by Cox regression is the hazard ratio (*HR*). A hazard ratio of 1 indicates that the treatments confer the same hazard of the outcome event. Hazard ratios that differ from 1 (are smaller or larger) indicate differences between the treatments.

Using the hypothetical data displayed in **Fig. 4.2**, Cox regression estimates *HR* = 0.78, for simvastatin versus placebo, indicating that the simvastatin group is estimated to have a 22% lower hazard of death than the placebo group.

A 95% confidence interval for the hazard ratio is [0.64, 0.95], and a p-value for the hypothesis of no difference between the groups is p = 0.014. Thus there would be strong statistical evidence from these hypothetical data that simvastatin reduces the hazard of mortality, and the confidence interval suggests that simvastatin reduces the hazard of death by probably somewhere between 5% and 36%, compared to a placebo.

Meta-analysis

Suppose that you would like to know if exercise can reduce cardiovascular deaths in patients with coronary heart disease. Suppose, further, that you find two high-quality studies on this topic, and that one study reports 'statistically significant' evidence for a benefit of exercise, while the other reports a 'non-significant' result. It would be tempting to conclude from this that the published evidence is inconclusive. To do so is a form of what's called *vote counting*, the logic of which goes: one study shows a benefit of exercise, one study does not, so no side wins. However, vote counting is not the best way to combine information from several studies and can, in fact, lead to erroneous conclusions, which ultimately may mean that patients miss out on the best treatments or advice. A better way to make sense of data from multiple studies is meta-analysis.

Meta-analysis pools information from several studies that assess the same research question. The studies should investigate the same treatment or exposure (e.g., exercise-based rehabilitation), the same control (e.g., passive control), and the same outcome (e.g., death from cardiovascular causes). The population studied (e.g., patients with cardiovascular disease) should also be the same or similar across studies, and reports of meta-analyses should carefully document variations in populations (e.g., if studies differ in the age distribution of included patients or differ in their inclusion criteria).

Meta-analysis is typically conducted based on published information only, without access to the original data. The effect sizes and their statistical variability, which were either reported in the original studies or can be inferred from the study reports, are collated and combined statistically. The measure of effect size (such as risk ratio, hazard ratio, mean difference, etc.) should be the same in all studies. If it is not, the meta-analyst can sometimes use the information from the study report to convert one effect size into another, for example, to convert an odds ratio into a risk ratio.

In simple terms, meta-analysis works as follows:

- Take the effect size estimate from each study.
- Weight each study by the precision with which it has measured the effect, with more precise studies getting more weight (a more precise study has a smaller confidence interval).

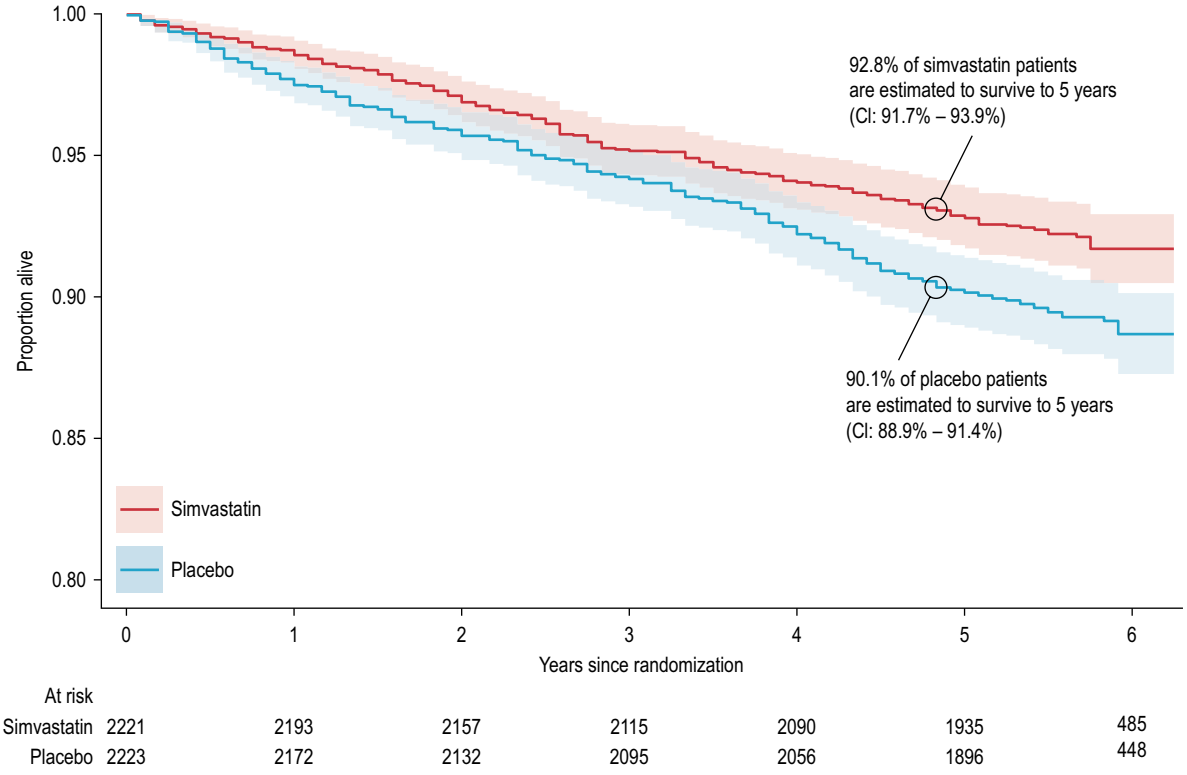


Fig. 4.2 Kaplan-Meier survival plot: Simvastatin vs placebo (hypothetical data). This plot is based on hypothetical data designed to be similar to the data from the 4S study but does not report the original data from that study. CI: 95% confidence interval (shaded areas around the survival curves). At Risk: Number of patients still alive and in the study at each time point. This is an indication of the size of the sample on which survival estimates from that point onwards are based. The number 'at risk' reduces over time for two reasons: because some patients die, and because some patients' follow-up is censored while they are still alive. *(Based on Spinks A, Glasziou PP, Del Mar CB. Antibiotics for sore throat. Cochrane Database of Systematic Reviews 2013; 11. Art. No.: CD000023.* https://doi.org/10.1002/14651858.*CD000023.pub4.)*

- Calculate a pooled effect size estimate as the weighted average of the estimates from all studies.

The details of these calculations are sometimes more complicated, but the above is the general logic. Generally, studies with larger sample sizes tend to produce more precise estimates and hence receive more weight, although factors other than the sample size can also influence the precision of a study. The results of a meta-analysis are typically displayed in a forest plot (**Fig. 4.3**), which displays a summary of the results from all studies included in the meta-analysis, as well as the result of the meta-analysis itself. The bottom row ('Common effect model') shows a risk ratio estimate of 0.56 (95% CI: 0.42, 0.76) from the meta-analysis, suggesting strong statistical evidence that exercise-based rehabilitation reduces the risk of cardiovascular death. In this special case, all eight study confidence intervals include the value 1 for the RR, which indicates no effect (lines for 95% CI cross the centre line). Therefore naïve *vote counting* would have led to the conclusion that no study has produced evidence for a benefit of exercise—but the meta-analysis shows that properly combined, the eight studies provide strong statistical evidence in favour of a benefit of exercise.

Publication bias

A good meta-analysis should assess publication bias. Publication bias comes about when the result of a study influences the likelihood of its publication. In many fields, studies with 'positive' results (e.g., that show evidence for an effect of a treatment) are more likely to be published than studies with 'negative' results (i.e., with statistically 'non-significant' findings). A naïve meta-analysis, which fails to account for publication bias, would then overestimate the efficacy of a treatment and indeed may conclude that a treatment is effective if, in reality, it is not. Therefore, many meta-analytic studies include a funnel plot that charts the study estimates around their pooled average and allows the meta-analyst to infer the presence or absence of bias (**Fig. 4.4**). Statistical methods exist to detect publication bias and adjust the analysis to remove the bias from the pooled estimate.

Summary and outlook

This section has discussed the concepts of descriptive and inferential statistics and has introduced two inferential procedures: confidence intervals and hypothesis tests. We have then illustrated how to apply these concepts in two-group randomized controlled trials with three types of outcomes (numeric, binary, and time-to-event), as well as in meta-analysis.

Of course, many more statistical procedures are found in published medical studies. Different methods and different tests are appropriate for different research designs and different types of data. A good introductory statistics textbook will introduce the reader to the technical vocabulary that is necessary to understand the rationale underlying statistical techniques, as well as to some of

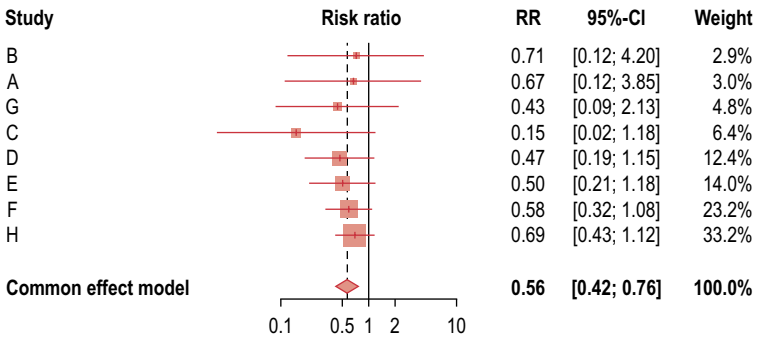


Fig. 4.3 Forest plot: risk ratios of eight studies on cardiovascular mortality: Exercise-based rehabilitation versus control. This forest plot, including some of the results, is based on Analysis 1.2.3 (p. 221) in: Dibben G et al. Exercise-based cardiac rehabilitation for coronary heart disease. Cochrane Database of Systematic Reviews 11 2021. Studies with narrower confidence intervals, and thus greater precision, are given larger weight in the analysis, represented by the size of the square. The location of the square indicates the study's effect size estimate, while lines extending to the left and right of the square indicate the 95% CI. The bottom row (diamond) is the result of the meta-analysis, with the diamond's centre indicating the pooled effect size estimate and its width the 95% CI. *CI,* Confidence interval; *RR,* risk ratio estimate.

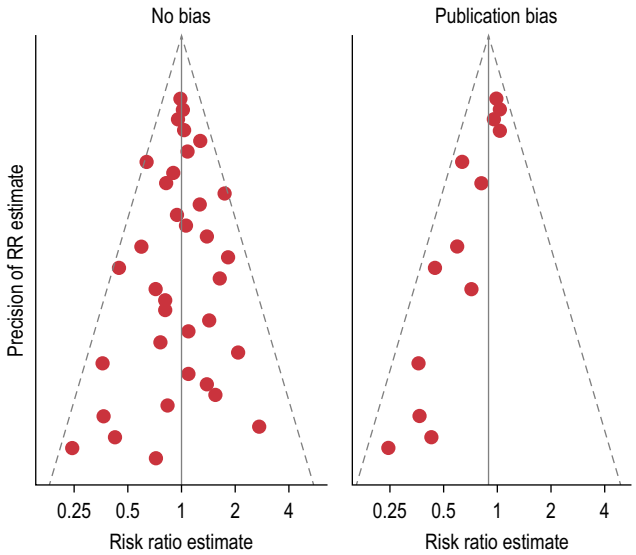


Fig. 4.4 Funnel plots illustrating two hypothetical meta-analyses. Each dot represents a study. The vertical axis represents the precision of the study, with more precise studies located towards the top. If there is no publication bias, a funnel plot should look something like the left panel: a symmetrical distribution scattered around the true effect size (in this example RR = 1, i.e., no effect), with more variation in estimates as study precision decreases towards the bottom of the plot. In the presence of publication bias, a funnel plot might look something like the right panel: the absence of symmetry suggests that small studies with 'non-significant' results or whose results indicate that the treatment may do harm (RR > 1) have been withheld from publication, while studies showing a benefit of the treatment (RR < 1, and 'statistically significant') were more likely to be published.

the most common analytic methods and when they are appropriate to use.

Further reading

Bland, M. *An introduction to medical statistics*. Oxford University Press, 2015.
Kirkwood, BR & Sterne, JAC. *Essential medical statistics*. John Wiley & Sons, 2010.

APPLYING THE EVIDENCE

The medical literature is continuously expanding, with new information published every day that could help to improve care for patients and population health. However, there are often major challenges involved in implementing the findings of research in clinical practice, with some estimates suggesting that the process takes an average of 17 years. The gap leads not only to the under-use of effective interventions but also to the incorrect use of others, as well as the over-use of unhelpful or unproven interventions. All of this means negative outcomes for patients and for population health. It is also a waste of health service resources and of costly and time-consuming research. Consequently, considerable effort has been put into improving the routine use of research findings.

Challenges of implementing evidence

Trial design

RCTs are designed to minimize the risk of bias, so differences between randomized groups can be interpreted as estimates of the causal effect of the treatment. RCTs may achieve this internal validity at the expense of external validity or generalizability. There are three reasons why RCTs in many areas of healthcare may have low external validity:

- ***The provider.*** The healthcare professionals who participate may be unrepresentative. They may have a particular interest in the topic or be enthusiasts and innovators.
- ***The setting.*** This also may be atypical: a teaching hospital, for example.
- ***The patients.*** All trials exclude certain categories of patients: for example, older patients or those with more than one illness. Often the exclusion criteria are so restrictive that the patients who are eligible for inclusion represent only a small proportion of the patients being treated in normal practice.

The outcome of drug treatment is likely to be unaffected by the characteristics of the prescribing doctor or the setting in which it is administered. On the other hand, the outcome of activities such as surgery and physiotherapy may be highly dependent on the characteristics of the provider, setting, and patients. Thus findings from

RCTs often need to be interpreted as indications of the *efficacy* of an intervention rather than its actual *effectiveness* in everyday practice. For successful implementation, however, both external validity and effectiveness are required if the findings are to be generalized to different patients and different settings.

Pragmatic RCTs

In a conventional RCT, the principal aim is to establish whether a treatment works. As outlined, because of the emphasis placed on internal validity, these 'explanatory' RCTs generally provide evidence of efficacy or what can be achieved in ideal circumstances.

However, the ultimate goal is to improve the health of real people in real settings, so it is crucial also to know whether an intervention is going to be effective in the real world. In contrast to explanatory trials, pragmatic RCTs place the emphasis on effectiveness rather than efficacy and seek to measure the benefit that the treatment produces in routine clinical practice; they therefore provide better evidence for decision-making.

There are some key differences between explanatory and pragmatic trials:

- Explanatory trials recruit a population that is as homogeneous as possible. In contrast, the design of a pragmatic trial reflects variations between patients that occur in real clinical practice.
- It may not be possible to blind participants and investigators to the treatment that is being given. Such biases are not necessarily viewed as detrimental in a pragmatic trial.
- In pragmatic trials, outcome measures should represent the full range of health gains relevant to patients and clinicians, for example, both a reduction in mortality after stroke and an improvement in quality of life.
- The two approaches to trial design will sometimes arrive at different conclusions about the benefit of a treatment, because a treatment that works in an ideal setting may not work in real life. The difference could also stem from the fact that improvement in a biomedical end-point does not produce the expected health gain. For example, although fluoride treatment for osteoporosis increases a commonly measured biomedical end-point (non-vertebral bone density), it actually raises fracture rates (an important clinical end-point) because it increases skeletal fragility. Consequently, it is not an effective treatment for postmenopausal osteoporosis.

The Pragmatic RCT design demonstrated its value early in the COVID-19 pandemic when the 'RECOVERY' trial was established in the UK to rapidly examine treatment effectiveness (see Pessoa-Amorim G et al., in 'Further Reading'). Key characteristics of the pragmatic RECOVERY design were as follows: simple recruitment of hospitalized patients using national routine datasets, allowing all hospital admissions to be quickly considered for recruitment, rapid randomization, and high follow-up rate; wide inclusion criteria allowing a diverse, representative population to be included; a platform design which allowed treatment arms to be added and removed as evidence emerged; and integration into the healthcare system allowing rapid translation into clinical practice. Within 100 days, RECOVERY demonstrated the survival benefit of dexamethasone in severe disease and enabled the swift cessation of other commonly used treatments which were ineffective, potentially harmful, and costly. Dexamethasone was promptly implemented worldwide, saving hundreds of thousands of lives.

Patient-centred outcome measures

Research only has external validity if it reflects the wider priorities of patients and, for population health research, the public. For this reason, the outcomes of most relevance to patients must be included in trials. In one study, for example, patients with multiple sclerosis and their doctors were independently asked to select the three aspects of the disease that had the greatest effect on quality of life. Doctors focused mainly on the physical effects of the disease, whereas patients were more concerned about mental health, emotional wellbeing, general health, and vitality, which are often not measured in RCTs. Similarly, in RCTs of anticonvulsants, individuals with epilepsy are much more interested in the proportion of patients rendered free of seizures than they are in changes in mean seizure frequency (a commonly used measure of benefit in RCTs).

Barriers and facilitators to change

Almost all healthcare interventions are complex and involve several interacting components. For example, a new drug therapy may superficially seem simple to implement, but the process might actually require additional patient monitoring or allow management to move from an inpatient to a community setting, requiring a change in care pathways. Other new interventions may necessitate amendments to referral processes (e.g., by allowing direct access to hospital care) or involve care provision by different staff (e.g., allied health professionals rather than doctors). These changes may involve modification of working practices, equipment, and care setting and entail additional unforeseen costs. It is often not clear which of the specific components of these complex interventions represent the 'active ingredient' for bringing about the intervention's effect.

If interventions are to be rolled out in other settings, careful consideration should be given to ensuring that interventions are designed in such a way as to be easily and successfully replicated. This involves trying to identify not only the key 'ingredients' that must be included but also whether the intervention might be sensitive to particular features of the local setting.

Intervention context

Barriers to change can arise at different levels in the healthcare system: the patient, their carer and family, the individual care professional, the healthcare team, the healthcare organization (e.g., its resources, capabilities, structure, culture, and politics), and wider society (e.g., the economic, social and political environment). Each of these components makes up the context of the intervention (**Fig. 4.5**). When a new intervention is being implemented, consideration should be given to what will happen when these individual contextual factors interact. Because of the complexity and diversity of each of these factors, the effect is likely to differ in different settings.

For example, reduction in hospital-acquired infections is a healthcare priority in many countries. Handwashing plays a key role. However, concordance on the part of healthcare workers in general, and doctors in particular, is known to be poor. Doctors are generally aware of the evidence but overestimate their own handwashing practices. Most countries and many hospitals have guidelines on the prevention of infections but they are frequently not followed. This means that, despite well-established evidence that has been properly disseminated, performance remains poor.

What is the problem? Research has shown that a range of competing demands reduce the likelihood of doctors washing their hands. These include a high workload, organization of the ward and access to facilities, concerns about skin irritation, lack of knowledge of the evidence, a lack of leadership to encourage change and mutual accountability, and an absence of institutional policies. Even if doctors are aware of the evidence and are willing to change, altering well-established patterns of care is difficult if the clinical environment is not conducive to change. These factors will be present to

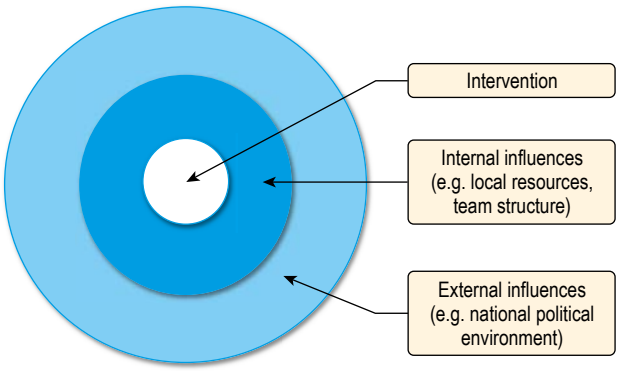


Fig. 4.5 Contextual factors that influence the implementation of evidence.

differing extents in different settings. The key ones should be identified in each location to ensure that change comes about.

Leadership

Local leadership has been identified as key to ensuring that evidence is used in practice. Leadership includes the actions taken by formal leaders in an organization to influence change. The attributes and qualities of leaders that appear to facilitate the use of research evidence to change practice include:

- valuing research
- being knowledgeable about research
- role-modelling evidence-based practice
- involving other staff in decision-making
- supporting changes in practice
- ensuring that policy and procedures are evidence based and up to date.

Resistant leadership that fails to support change reduces the likelihood that evidence will be applied in practice.

Improving the uptake of evidence

There is a growing body of evidence relating to the effectiveness of different interventions to improve the uptake of evidence and change clinical and public health practice, which includes synthesizing and disseminating evidence in more accessible formats. For example, information graphics (or 'infographics') utilize images and data visualizations (pie charts, bar graphs, line graphs) to present research in an engaging way. Infographics add value by enhancing understanding and the reach of research. Information is more likely to be retained if it has been learned from an infographic than from text alone. However, although this may help ensure the uptake of some simple changes, it is insufficient for most innovations. Instead, multifaceted approaches are required, including a combination of the following:

- ***Educational strategies.*** Large conferences and courses have limited effects; small group interactive education sessions with active participation have more positive results. Educational outreach by experts may be especially effective for improving prescribing behaviour.
- ***Audit and feedback.*** These seem to be most effective when targeting diagnostic test ordering and prevention interventions. The effect size depends on the type of feedback, its source and format, and the frequency or intensity of presentation. Feedback is recommended in combination with education, outreach visits, or reminders.
- ***Use of reminders and computers.*** Computer prompts, for example, may be one of the most effective interventions. They have been shown to be particularly effective in increasing the use of prevention interventions, For example, vaccination and cancer screening reminders.

All of these interventions have the potential to help facilitate the effective transfer of evidence to practice, although the evidence base for them is still emerging. More than one approach may be needed. Further research is ongoing to increase understanding of the processes and elements behind successful change.

Further reading

Pessoa-Amorim G, Campbell M, Fletcher L. Making trials part of good clinical care: lessons from the RECOVERY trial. *Future Healthc J*. 2021;8(2):e243–e250.
Bauer M, Damschroder L, Hagedorn H et al. An introduction to implementation science for the non-specialist. *BMC Psychol* 2015; 3:32.
Grol R, Grimshaw J. From best evidence to best practice: effective implementation of change in patients' care. *Lancet* 2003; 362:1225–1230.
Wensing M. Implementation science in healthcare: introduction and perspective. *J Evid Training Qual Health Care* 2015; 109:97–102.

Online material at Elsevier eBooks+ includes a significant amount of additional content, including practical exercises based on published clinical evidence, to help readers put into practice the principles described in this chapter.

Questions for this chapter can be found on Elsevier eBooks+.

SECTION 1: ASKING ANSWERABLE QUESTIONS AND SEARCHING FOR THE EVIDENCE

Q1: Sarah is a 37-year-old female who asks for your advice about managing her mild depression. She is reluctant to take medication and wonders if regular exercise might improve her mental health.

For the vignette above, please set out:

a. An answerable clinical question
b. PICO terms

Question 1b

Patient Problem or Population	Intervention or Exposure	Comparison or Control	Outcome Measure

c. Search terms for use in an online database (e.g., Medline)

Model answer:

a. Answerable clinical question: Is exercise as effective as drug therapy in reducing symptoms of mild depression?
b. PICO terms

Answer 1b

Patient Problem or Population	Intervention or Exposure	Comparison or Control	Outcome Measure
Mild depression	Exercise	Drug therapy	Reducing symptoms of depression

c. Search terms for use in an online database (e.g., Medline)
depression AND exercise
depression AND exercise AND drug AND symptoms

Q2: Elizabeth is 62 years old and suffers from chronic obstructive pulmonary disease, for which she takes regular medication. You wonder whether she might also benefit from pulmonary rehabilitation.

For the vignette above, please set out:

a. An answerable clinical question
b. PICO terms

Question 2b

Patient Problem or Population	Intervention or Exposure	Comparison or Control	Outcome Measure

c. Search terms for use in an online database (e.g., Medline)

Model answer:

a. Answerable clinical question: In patients under the age of 65 years who suffer from chronic obstructive pulmonary disease, does pulmonary rehabilitation in conjunction with pharmacological treatment significantly improve quality of life?
b. PICO terms

Answer 2b

Patient Problem or Population	Intervention or Exposure	Comparison or Control	Outcome Measure
Patients <65 years old suffering from COPD	Pulmonary rehabilitation in conjunction with current pharmacological treatment	Standard pharmacological treatment without pulmonary rehabilitation	Improved quality of life

c. Search terms for use in an online database (e.g., Medline)
COPD AND "pulmonary rehabilitation"
COPD AND "pulmonary rehabilitation" AND "quality of life"

Q3: David is 67 years old and has had four episodes of cellulitis affecting his lower leg. You wonder whether he might benefit from prophylactic antibiotics.

For the vignette above, please set out:

a. An answerable clinical question
b. PICO terms

Question 3b

Patient Problem or Population	Intervention or Exposure	Comparison or Control	Outcome Measure

c. Search terms for use in an online database (e.g., Medline)

Model answer:

a. Answerable clinical question: In patients with recurrent cellulitis, do prophylactic antibiotics reduce the risk of the infection recurring, compared with no treatment?
b. PICO terms

Answer 3b

Patient Problem or Population	Intervention or Exposure	Comparison or Control	Outcome Measure
Patients with recurrent cellulitis	Prophylactic antibiotics	No treatment	Repeat infection

c. Search terms for use in an online database (e.g., Medline)
"recurrent cellulitis" AND "prophylactic antibiotics"
"recurrent cellulitis" AND "prophylactic antibiotics" AND infection

SECTION 2: CRITICAL APPRAISAL

In this section, you should read and critically appraise the following papers:

Lamb SE, Sheehan B, Atherton N et al. Dementia and Physical Activity (DAPA) trial of moderate to high intensity exercise training for people with dementia: randomised controlled trial. BMJ 2018;361:k1675.

Li L, Geraghty OC, Mehta Z, Rothwell PM. Age-specific risks, severity, time course, and outcome of bleeding on long-term antiplatelet treatment after vascular events: a population-based cohort study. *Lancet* 2017; 390:490–499.

CASP checklists can be downloaded at https://casp-uk.net/casp-tools-checklists/. Specimen answers are provided for your reference at https://coursewareobjects.elsevier.com/objects/elr/Kumar/clinicalmedicine10e/PDF/.

SECTION 3: INTERPRETING ODDS RATIOS, CONFIDENCE INTERVALS, AND P-VALUES

Paper 1

If tuberculosis treatment is not taken regularly, there may be risks to the patient and the public through the generation of drug resistance, relapse, and transmission of infection. Consequently, directly observed treatment (DOT) is currently recommended by the World Health Organization for patients at high risk of poor adherence (e.g., those with a history of homelessness, imprisonment, or drug use). Although DOT can be administered in clinic, community, or home settings, it still entails substantial inconvenience to patients and service providers. Developments in video technology have raised the possibility of remote video-observed treatment (VOT) as an alternative approach.

This study was carried out to test whether or not levels of treatment observation were improved with VOT. Patients who were at least 16 years old, with active pulmonary or non-pulmonary tuberculosis, and eligible for DOT were identified from 22 clinics in England. Individuals were then randomly assigned to either VOT (daily remote observation using a smartphone app) or DOT (observations done three to five times per week in the home, community, or clinic settings). The primary outcome was the completion of 80% or more scheduled treatment observations over the first 2 months following enrolment.

A total of 226 patients were randomly assigned, namely 112 to VOT and 114 to DOT; 78 (70%) of 112 patients on VOT successfully completed ≥80% of the scheduled observations during the first 2 months, compared with 35 (31%) of 114 on DOT. Odds ratio (OR) 5.18; 95% CI 2.94–9.12; p<0.0001.

Adapted from Story A et al. Smartphone-enabled video-observed versus directly observed treatment for tuberculosis: a multicentre, analyst-blinded, randomised, controlled superiority trial. Lancet *2019; 393(10177):1216–1224.*

Q1: What type of study is this?

Answer

A randomized controlled trial. Participants were randomly assigned to receive either the intervention, VOT (daily remote observation using a smartphone app), or the usual treatment/control, DOT (observations done three to five times per week in the home, community, or clinic settings).

Q2: What is the estimated odds ratio in this study?

Answer

5.18.

Q3: What does this mean in practice?

Answer

An odds ratio is a relative measure of the effectiveness of the intervention being studied, which allows the comparison of the intervention group of a study to the comparison or placebo group.

The trial above compared participants who had received video-observed therapy (VOT, intervention arm) with participants who had received directly observed therapy (DOT, control arm). The outcome—or event—of interest was ≥80% of the scheduled observations successfully completed during the first 2 months.

An odds ratio of 5.18 means that patients in the VOT group (intervention arm) had more than 5 times higher estimated odds of the successful completion of ≥80% of the scheduled treatment observations compared with the DOT group (control arm).

Here, the outcome is a desirable one (successful completion of treatment).

The OR is >1, so the intervention is better than the control—the odds of the outcome occurring is greater in the intervention group.

If the OR were <1, the control would be better than the intervention—the odds of the outcome occurring would be lower in the intervention group.

If the OR were = 1, there would be no difference between the two study groups, as the outcome is the same in both.

Q4: What is the confidence interval?

Answer

95% CI 2.94–9.12.

Q5: What does this mean in practice?

Answer

The confidence interval indicates the amount of statistical uncertainty around the measure of the efficacy of the intervention being studied. In other words, the confidence interval indicates how precise the effect estimate is. Confidence intervals are used because a study recruits only a small sample of the overall population, so the treatment effect (odds ratio) observed in the particular study is an imprecise estimate of the true treatment effect (true odds ratio in the population). Most studies report the 95% confidence interval (95% CI).

If the 95% confidence interval for an odds ratio includes the value 1 (e.g., if 95% CI 0.9–1.1), this implies that there is weak evidence for an effect of the treatment. If it does not (e.g., 95% CI 0.7–0.8), this is equivalent to a p-value smaller than 0.05 and suggests some statistical evidence for an effect of the treatment.

In the trial comparing VOT to DOT, the confidence interval is reported as 95% CI = 2.94–9.12, or approximately 3 to 9. This can be interpreted as follows: From this study, we estimate that the treatment probably increases the odds of completing ≥80% of the scheduled observations by somewhere between 3 and 9 times.

Q6: What is the p-value?

Answer

p<0.0001.

Q7: What does this mean in practice?

Answer

The p-value in this study is <0.0001. This means that if VOT actually had no effect, compared to DOT, the probability to observe an odds ratio of 5.18, or a larger one, was smaller than 0.0001 (or smaller than 0.01%).

This very small p-value is interpreted as strong statistical evidence that the treatment does have an effect.

Q8: How would you summarize the findings of this study?

Answer

This trial was carried out to examine whether or not levels of treatment observation were improved with video-observed therapy (VOT), compared with conventional directly observed therapy

(DOT). The results were presented as an odds ratio = 5.18; 95% CI 2.94–9.12; p<0.0001. This means that the odds of patients successfully completing ≥80% of the scheduled treatment observations were more than 5 times greater in the VOT arm, compared with the DOT arm, and we can be 95% confident that the treatment effect lies between 2.94 times and 9.12 times greater.

As a consequence, the authors concluded that VOT was a more effective approach to observation of tuberculosis treatment than DOT. They note that VOT is likely to be preferable to DOT for many patients across a broad range of settings, providing a more acceptable, effective, and cheaper option for supervision of daily and multiple daily doses than DOT.

5 Human genetics

Neil Rajan and David Peter Kelsell

Clinical skills 52
Taking a genetic family history 52
Drawing a pedigree 53
Background 53
Gene editing 53
Clinical genetics 54
Approach to the patient with a suspected genetic disease 54
Ethical considerations 55
Testing for genetic disease 55
Interpreting a genetic test report 56
The cellular basis of genetics 57
DNA and the genetic code 57
Human chromosomes 57
DNA transcription 58
Control of gene expression 58
Protein synthesis and secretion 60
Genetics and the cell 61
Review of cellular genetics 61
Genetic disorders 62
Chromosomal disorders 62
Point mutations, insertions, and deletions 64
Prenatal diagnosis of congenital anomalies 65
Patterns of inheritance of single-gene diseases 66
Complex traits: multifactorial and polygenic inheritance 69
Techniques for studying the genetic code 69
Polymerase chain reaction 69
Sanger sequencing 69
Next-generation sequencing 70
Transcriptomics 70
Epigenetic techniques 70
Genome databases 70
Research techniques 71
Genomic medicine 72
Gene therapy 72
Mitochondrial replacement therapy 73
Stem cell therapy 73
Pre-implantation genetic diagnosis 73
The genetic basis of cancer 74
Inherited cancer syndromes 74
Oncogenes 76
Tumour suppressor genes 76
Genetic diagnostics for cancer 77
Targeted treatments for cancer 77
Precision (or personalized) medicine 77

CORE SKILLS AND KNOWLEDGE

The clinical practice of diagnosing and managing genetic disease is an expanding and evolving field. Increasingly, this is carried out by experts in organ-based subspecialties (e.g., neurologists and dermatologists), in partnership with clinical genetics teams. Hence, a working knowledge of skills such as documenting a family history and understanding basic genetic investigations is relevant across a range of disciplines. Typically, clinical geneticists are qualified adult or paediatric physicians who undertake postgraduate training in the specialty, often including a period of laboratory-based research. Genetics services are usually delivered in specialist centres in the outpatient setting.

Key skills in clinical genetics include:

- Obtaining a comprehensive history. A careful family history is essential.
- Drawing and interpreting a family tree.
- Estimating the risk of a future pregnancy being affected or carrying a disorder, based on common patterns of inheritance.
- Understanding the principles of genetic testing and the ethical challenges this may entail.

Exposure to genetics multidisciplinary team meetings and genetics clinics represent opportunities for medical students and training clinicians to gain insights into how to ask relevant questions when taking a history from a patient with genetic disease. This has to be coupled with thorough clinical examination, which may lead to the selection of appropriate radiological imaging tests and invasive tests such as tissue biopsies.

CLINICAL SKILLS

Taking a genetic family history

A genetic family history consists of information describing how family members are related to each other (their biological relationships) and any medical conditions they may have.

Establishing such a family history can:

- reveal patterns of inheritance
- help make or refine a diagnosis
- help assess the likelihood of genetic disease in relatives
- affect testing, treatment and management strategies
- highlight the need for referral to specialist services
- help in building rapport with patients.

Box 5.1 How to take a genetic family history

Immediate family

In adult patients, information is typically obtained first on which members of the immediate family are affected by a disorder. Begin drawing a pedigree with the patient towards the bottom of the page (allowing space below for any children).

- Ask about their spouse or partner.
- Ask about any children the patient may have.
- Enquire sensitively about stillbirths and miscarriages.

Where any individual is affected, try to determine the age at which they were diagnosed and the severity of their disease.

Patients' siblings and parents

Siblings are drawn alongside the patient and the parents are placed above, with lines indicating family relationships (which can be complex):

- Ask whether the patient's parents are alive, and about any illnesses they have at present or have been diagnosed with in the past. If the parents are deceased, record the stated cause of death and the age at which this happened.
- Ask how many siblings and half-siblings the patient has, and their ages. Enquire about their health.

Sometimes, patients may not know the family members' cause of death, which may, in some cases, reflect medical uncertainty (e.g., in conditions that cause sudden death); this uncertainty should be recorded.

Father's side of the family

Typically, these family members will be drawn above and to the left of the patient and include the father's siblings and parents. Information on a minimum of three generations should be obtained.

Mother's side of the family

Typically, these family members will be drawn above and to the right of the patient and include the mother's siblings and parents. Information on a minimum of three generations should be obtained.

Family history questionnaire

Sometimes, patients may find it difficult to recall information in a clinical consultation. A family history questionnaire can aid information-gathering, as the patient can fill this in at home, with the help of family. These questionnaires can be verified against cancer registry data, which is important as patients sometimes may not know which specific cancers affected their relatives. Increasingly, online forms are being piloted, which can be filled in electronically.

Information is usually gathered by asking a series of questions about each member of the family in a systematic order (**Box 5.1**) and recording the information pictorially as a family tree (a pedigree; see opposite). Family history can be an important risk factor for cancer and a wide range of conditions with monogenic or complex polygenic aetiology. Some national clinical guidelines therefore highlight the importance of taking and recording a genetic family history.

Drawing a pedigree

A set of agreed symbols is used when drawing a pedigree (**Fig. 5.1**). In order to show how individuals are related to each other, lines are drawn between symbols. The correct placement of these lines is a key skill to ensure that the pedigree gives an accurate picture of family relationships.

A dominantly inherited example is presented in **Fig. 5.2**. Affected males and females are filled in with black, and unaffected are left white. Males and females are equally affected, and both can transmit the disease to their offspring, who carry a 50% chance of inheriting the 'faulty' gene. More examples of patterns of inheritance are shown in the later section entitled 'Patterns of inheritance of single-gene diseases'.

BACKGROUND

The genetic code that forms the instruction manual for humans can now be sequenced in less than a day. From a clinical perspective, this brings an enormous range of new possibilities and challenges to the practice of medicine. In this chapter, we aim to provide an overview of key concepts in deoxyribonucleic acid (DNA), when we can use it to predict with confidence and the caveats that apply.

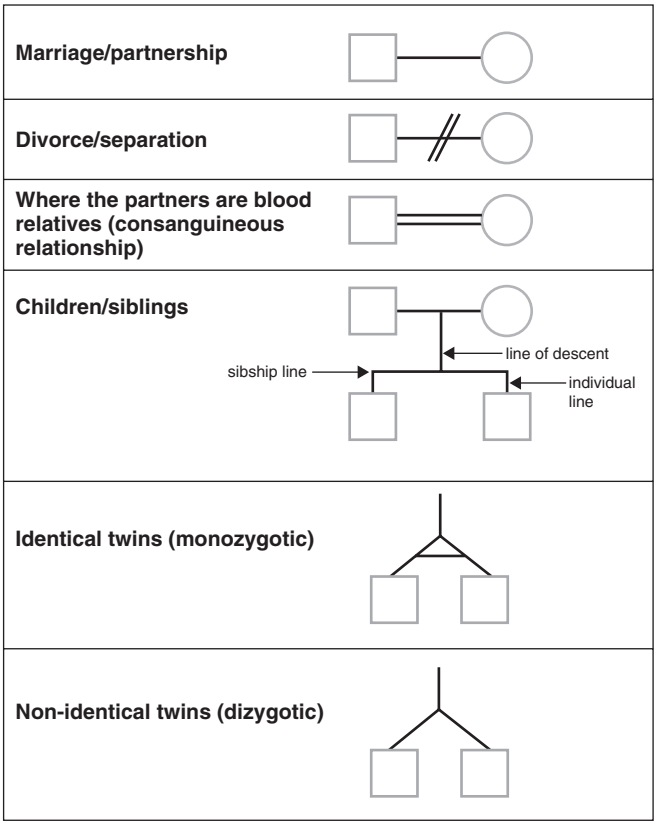
DNA is an efficient means of storing information. Three billion base-pairs (bp) sit in the nucleus of each cell and carry the instructions to make all the parts of any cell in the human body. When a cell divides, all of this information is duplicated, with systems designed to prevent mistakes from occurring during this process. Nonetheless, errors do arise, which can either be harmless or cause disease, depending on the nature and the timing of the change. The discovery that the genetic code could be 'read' or sequenced allowed changes in parts of this instruction manual to be sought as an explanation for inherited diseases. In addition the study of how the genetic code is altered in malignant cells has provided new insights into cancer biology (see Chapter 41). Progressive advances in DNA sequencing technology have meant that we are able to look at the human genome of an individual at multiple time points from conception to the grave (**Fig. 5.3**). This information can have a range of implications, from pre-implantation genetic diagnosis to predicting and stratifying disease, selecting treatments and assessing drug response and the risk of adverse side-effects. An understanding of how this new information fits into clinical medicine is immensely relevant for patient care. As the cost of DNA sequencing drops, adoption of this technology is likely to become more widespread and indeed is now implemented in the National Health Service (UK) (NHS) for rare disorders and cancer. Commercially available testing is accessible simply by supplying a saliva sample in the post.

However, while we are able to sequence DNA rapidly, much bigger challenges lie ahead. For example, it can be difficult to interpret changes in DNA found in apparently healthy individuals. Until recently, genetic changes in rare diseases were studied in the context of affected individuals and their families. A genetic change that can cause disease within such families may not have the same effect on other people. This is thought to be due to the genetic context in which the change presents. It is hoped that when careful phenotyping and long-term follow-up are combined with genomics, we can advise appropriately on the implications of a broad range of such genetic changes. Currently, in some areas there is consensus on 'actionable' incidental findings in genomics, such as when preventative mastectomy should be discussed with patients when a genetic change is found that may predict early-onset breast cancer. Work remains to be done, however, to capture how the human genotype predicts phenotype, and large consortia efforts such as the Human Phenotype Ontology project aim to achieve this across all diseases.

Gene editing

In addition to reading the human genome, we can now also edit and change the genetic code, using very precise molecular tools such as CRISPR (see **Fig. 5.19**). By 2018, reports had been verified of gene editing in humans, which can be transmitted to the next generation. There is much discussion about how gene-editing technologies should and should not be used.

Here we aim to highlight selected genetic advances that exemplify concepts of genetic disease and to describe cases where genetic information has been clinically relevant to patients, either in routine care or in the context of clinical trials. The ability to interpret and contextualize genetic information in clinical care is becoming increasingly relevant as genomics is mainstreamed across clinical specialties.



	Male	Female	Sex unknown
Individual			
Affected individual (symbol coloured in)			
Multiple individuals	5	5	5
Deceased			
Pregnancy (the unborn baby of a pregnant mother)	P	P	P
Miscarriage	male	female	
Person providing pedigree information			

Fig. 5.1 **Symbols used in drawing a genetic pedigree.**

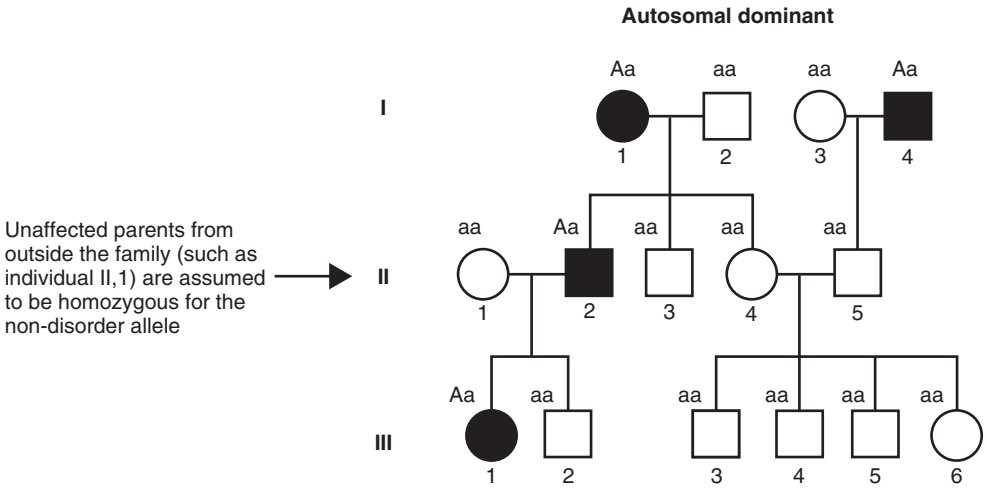


Fig. 5.2 **An autosomal dominant pattern of inheritance.** Affected males and females are shown in *black*. Capital 'A' indicates the allele associated with the disorder in question and lowercase 'a' the non-disorder allele.

CLINICAL GENETICS

Genetic disorders pose considerable health and economic problems, and early diagnosis is crucial. While individual diseases are rare, it is estimated that almost 6000 genetic disorders are recognized, and as a collective, 1 in 25 children may be affected with a genetic disorder. In any pregnancy the risk of a serious developmental abnormality is approximately 1 in 50 pregnancies, and around 15% of paediatric inpatients have a multifactorial disorder with a predominantly genetic element. In addition, 50% of clinical genetics referrals are adults with late-onset disease, including neurological, endocrine and gastrointestinal conditions, along with cancer.

Approach to the patient with a suspected genetic disease

Genetic services routinely coordinate clinical diagnosis, genetic testing, information-giving, family planning and targeted screening

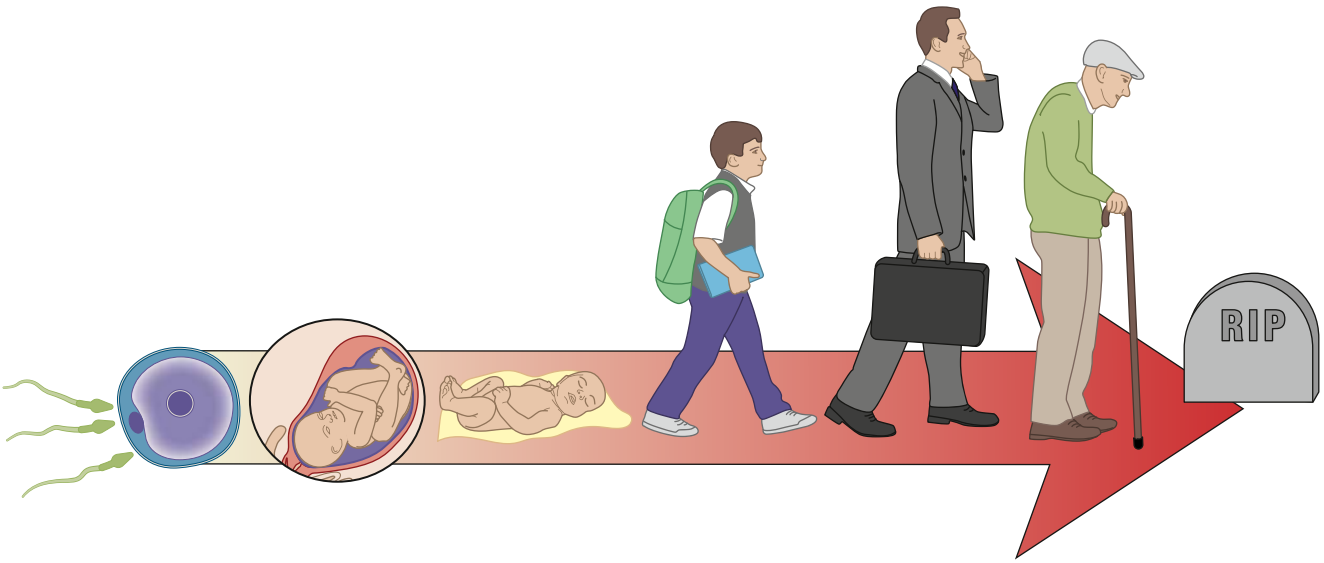


Fig. 5.3 Genotype and phenotype. Sequencing the human genome throughout our lifetime can give insights into disease susceptibility and healthcare management.

for a range of genetic disorders. Clinicians assessing a patient with a genetic disorder should keep in mind the following aims:

- ***Obtaining a full history.*** The pregnancy history, drug and alcohol ingestion during pregnancy, and maternal illnesses (e.g., diabetes) should be detailed.
- ***Establishing an accurate diagnosis.*** Examination may help in diagnosing a genetically abnormal child with characteristic features (e.g., trisomy 21) or ascertaining whether a genetically normal fetus was damaged *in utero.*
- ***Drawing a family tree.*** This is essential. Questions should be asked about abortions, stillbirths, deaths, marriages, consanguinity, and medical history of family members. Diagnoses may need verification from other hospital reports.
- ***Estimating the risk of a future pregnancy being affected or carrying a disorder.*** Estimation of risk should be based on the pattern of inheritance. Disorders caused by pathogenic variants in single genes carry a high risk; chromosomal abnormalities other than translocations typically carry a low risk. Empirical risks may be obtained from population or family studies.
- ***Information-giving.*** Information on prognosis and management should be supplied, with adequate time allowed for it to be discussed openly and freely and repeated as necessary.
- ***Continued support and follow-up.*** An explanation should be provided of the implications for other siblings and family members.
- ***Genetic screening.*** This includes prenatal diagnosis or preimplantation genetic diagnosis (*in vitro* fertilization followed by testing of embryos before implantation) and, if requested, carrier detection. A large number of molecular genetic tests are available.

Genetic counselling should be non-directive, with the couple making their own decisions on the basis of an accurate presentation of the facts and risks in a way they can understand.

Further details on taking a genetic history and drawing a family tree can be found in the 'Clinical skills' section (see earlier).

Ethical considerations

Ethical considerations must be taken into account in any discussion of clinical genetics. For example, prenatal diagnosis with the option of termination may be unacceptable on moral or religious grounds. In the case of diseases for which there is no cure and currently no treatment (e.g., Huntington's chorea), genetic tests can accurately predict which family members will be affected; many people, however, would rather not have this information. One very serious outcome of new genetic information is that disease susceptibility may be predictable—for example, in Alzheimer's disease—and so medical insurance companies can decline to issue policies for individuals at high risk. In some countries a moratorium is in place to prevent insurers from using such data against patients in this way. It remains to be decided who should have access to an individual's genetic information and to what extent privacy should be preserved.

Testing for genetic disease

Genetic testing is extremely powerful, as it can reveal changes in DNA that explain why a patient has a clinical disease. Once such information is acquired, it can, for example, guide the selection of screening tests, aid prognostication, and influence treatment. In some instances, it can elegantly replace the multiple biochemical, pathological and radiological tests that were previously required to make a diagnosis of a genetic disease. Patients with rare disease often report a 'diagnostic odyssey', which may be attenuated by genetic testing. In some situations, where the diagnosis is unknown and a strong clinical suspicion of genetic disease exists, genetic tests can provide the necessary clues to exclude certain differentials and aid the process of making and refining a diagnosis.

There will increasingly be a temptation simply to order whole-genome sequencing (see 'Techniques for studying the genetic code') for any genetic disease but this may have negative outcomes. For example, it may reveal unexpected findings that patients may not necessarily have wanted to know, such as a predisposition to early-onset dementia or cancer. Changes in disease genes may also be detected and cause anxiety, yet the genetic changes may not give rise to the expected disease in that patient. In some situations, incomplete knowledge may make any form of prognostication or reassurance impossible. The individual responses of each patient to such uncertainty can be difficult to predict, and more dialogue is needed to understand how to make better decisions for comprehensive testing in partnership with patients.

REFERRAL REASON:
Multiple monogenic benign skin tumours
Diagnostic testing for a panel of genes associated with multiple monogenic benign skin tumours John Smith who is clinically affected.

Result summary
A molecular diagnosis has not been confirmed or excluded.

Result
Next generation sequencing detected no clearly pathogenic variant John Smith see technical information below.

Further testing
Please contact the laboratory if further analysis is required.

Reported John Smith clinical scientist	Date: 12/05/23
Authorised by: John Smith principal clinical scientist	Date: 14/05/23

Technical information

Variant details
No reportable variants have been detected in the analysed genes.
Only variants that are considered likely to be relevant to this patient's clinical phenotype are reported. Variants classified as benign, likely benign or uncertain significance, where further testing would not alter classification, or monoallelic variants in recessive genes are not reported.

Panel and coverage information

Test directory indication:	R230 Multiple monogenic benign skin tumours
Panel ID:	MMBST_v1.3_R230_Green
Target bases >20x:	100.0%
Average depth across target:	1228

MMBST_v1.3_R230_Green panel:
CYLD, FLCN, LEMD3, MLH1, MSH2

For details of the sequence coverage for individual gene(s) please contact the laboratory.

Fig. 5.4 A sample clinical genetic test report.

In the UK, genetic tests accessible by NHS clinicians are now catalogued in the Genetic Test Directory and updated every 6 months. This directory has been developed as part of an agenda of 'mainstreaming' genetic testing, where clinicians across a range of specialties can order genetic tests for certain conditions. The test directory lists a range of types of tests, relevant for specific clinical indications. Based on existing data on the types of genetic changes that cause specific conditions, bespoke tests are designed that are drawn from more than 30 types of tests. These can include targeted panels, whole exome, and whole genome sequencing. Where it is more efficient, established tests that can detect single nucleotide variants, small insertions or deletions, or copy number changes are offered in place of whole genome sequencing. For the non-genetic specialists, access to such testing will need to be supported with training and education on how to counsel and choose tests.

The importance of taking a comprehensive personal and family history and performing a clinical examination is paramount. This information may identify a single gene that needs to be tested if a particular disorder is suspected. In some conditions, where a selection of genes may account for a phenotype, a panel of 10 to 20 relevant genes can be tested. Some genomic medicine centres favour this approach, in which only clinically relevant genes are analysed in a tiered manner, with further testing carried out only in cases that are negative. There will still be unsolved cases despite these approaches, and this should be one of the outcomes made aware to the patient.

Interpreting a genetic test report

Genetic test reports are written by clinical scientists who have designed and analysed specific genetic tests for specific conditions. Their knowledge of the known genetic variants linked with a clinical phenotype gives important context to the report. Genetic test reports will feed back clinically relevant changes, which are categorized according to the American College of Medical Genetics (ACMG) criteria. These utilize standardized terminology to describe any genetic variants found; 'pathogenic', 'likely pathogenic', 'uncertain significance', 'likely benign', and 'benign'. For clinical purposes, 'pathogenic and likely pathogenic' changes are useful for counselling patients. Variants of uncertain significance (VUS) may need to be investigated further across additional members of the family to determine their importance, and clinicians need to note that these can be reclassified in the future as new information becomes available in clinical variant databases and the published literature. 'Likely benign and benign' variants in relevant genes for the clinical diagnosis being tested for are unlikely to impact genetic counselling and are not routinely reported in test reports from UK Genomic Medicine Services.

In addition to stating the ACMG status of a detected variant, genetic test reports will report the change of the sequence to standardized criteria, established by the Human Genome Variant Society (HGVS). In the example of a single nucleotide variant, this will indicate the reference nucleotide and the detected nucleotide in the patient, as well as its location in the gene of interest. For example, a change in the gene causing tuberous sclerosis *(TSC1)* will be reported as 'a heterozygous pathogenic variant in TSC1 has been detected (c.1525C>T)', where the nucleotide cytosine (C) has been replaced by a thymine (T) at position 1525 within the coding sequence of the *TSC1* gene.

Genetic test reports will include details of the test type used to test the DNA sample, the genes covered, the performance of this test and the limitations of the test strategy. A sample test report is shown (Fig. 5.4).

Genetic testing is a rapidly changing field and exposure to clinical genetics clinics and multidisciplinary meetings will be crucial in training clinicians to select and implement the 'right' test.

Further reading

Wright CF, Ware JS, Lucassen AM et al. Genomic variant sharing: A position statement. *Wellcome Open Res* 2019; 4:22.
Wright CF, Campbell P, Eberhardt RY et al. Genomic diagnosis of rare pediatric disease in the United Kingdom and Ireland. *N Engl J Med* 2023: 388 (17):1559–1571.

THE CELLULAR BASIS OF GENETICS

DNA and the genetic code

Hereditary information is contained in the sequence of the building blocks of double-stranded DNA (Fig. 5.5). Each strand of DNA is made up of a deoxyribose–phosphate backbone and a series of purine (adenine [A] and guanine [G]) and pyrimidine (thymine [T] and cytosine [C]) bases. Because of the way in which the sugar–phosphate backbone is chemically coupled, each strand has a polarity, with a phosphate at one end (the 5′ end) and a hydroxyl at the other (the 3′ end). The two strands of DNA are held together by hydrogen bonds between the bases. T can pair only with A, and G can pair only with C; therefore each strand is the antiparallel complement of the other (see Fig. 5.5). This is key to DNA replication because each strand can be used as a template to synthesize the other.

The two strands twist to form a double helix with a major and a minor groove, and the large stretches of helical DNA are coiled around histone proteins to form nucleosomes (see Fig. 5.5). They can be condensed further into the chromosomes that can be visualized by light microscopy at metaphase (see Figs. 5.5 and 5.12).

Human chromosomes

Chromosomes are complex structures containing one linear molecule of DNA that is wound around histone proteins into small units called nucleosomes, and these are further wound to make up the structure of the chromosome itself.

Diploid human cells have 46 chromosomes, 23 inherited from each parent; thus there are 23 'homologous' pairs of chromosomes (22 pairs of '***autosomes***' and 2 '***sex chromosomes***'). The sex chromosomes, called X and Y, are not homologous but are different in size and shape. Males have an X and a Y chromosome; females have two X chromosomes. (Primary male sexual characteristics are determined by the *SRY* gene—sex-determining region on the Y chromosome.)

The chromosomes are classified according to their size and shape, the largest being chromosome 1. The constriction in the chromosome is the ***centromere***, which can be in the middle of the chromosome (metacentric) or at one extreme end (acrocentric). The centromere divides the chromosome into a short arm and a

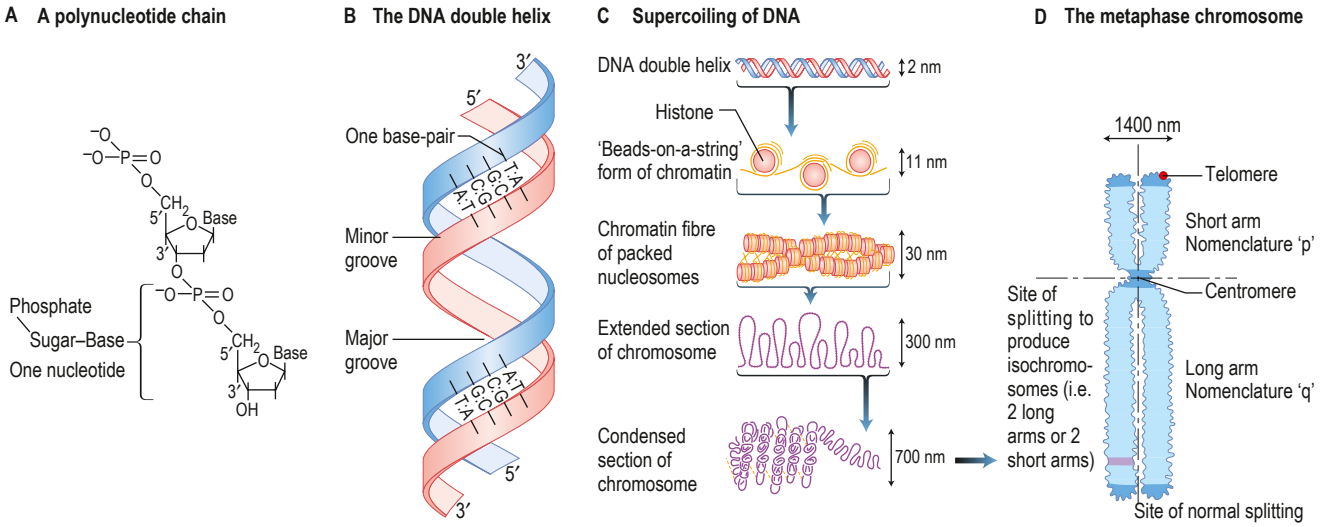


Fig. 5.5 DNA and its relationship to human chromosomes. (A) Individual nucleotides form a ***polymer*** linked via the deoxyribose sugars to form an ssDNA. The 5′ carbon of the sugar is covalently joined to phosphate. The 3′ carbon links the phosphate on the 5′ carbon of the ribose of the next nucleotide, forming the sugar–phosphate backbone of the nucleic acid. The 5′ to 3′ linkage gives orientation to a sequence of DNA. **(B)** ***Double-stranded DNA***. The two strands of DNA are held together by hydrogen bonds between the bases. T pairs with A, and G with C. The orientation of the complementary ssDNA is thus complementary and antiparallel. The helical three-dimensional structure has major and minor grooves, and a complete turn of the helix contains 12 base-pairs. The grooves are structurally important; DNA-binding proteins predominantly interact with the major grooves. **(C)** ***Supercoiling of DNA***. The large helical DNA is coiled into nucleosomes by winding around nuclear proteins (histones) and further condensed by coiling and supercoiling into the chromosomes that are visible at metaphase. **(D)** ***At the end of the metaphase, DNA replication*** results in a twin chromosome joined at the centromere. Chromosomes are assigned a number (or the letter X or Y for the sex chromosomes), plus short arm (p) or long arm (q). The region or subregion is defined by the transverse light and dark bands observed when staining with Giemsa (see Fig. 5.18) (hence G-banding) or quinacrine, and numbered from the centromere outwards. ***Chromosome constitution = chromosome number + sex chromosomes + abnormality***; for example, 46XX = normal female; 47XX + 21 = Down's syndrome (trisomy 21); 46XYt (2;19) (p21;p12) = male with a normal number of chromosomes but a translocation between chromosome 2 and 19 with breakages at short-arm bands 21 and 12 of the respective chromosomes. *DNA,* Deoxyribonucleic acid; *ssDNA,* single-strand DNA.

long arm, referred to as the ***p arm*** and the ***q arm***, respectively (see Fig. 5.5).

Chromosomes can be stained when they are in the metaphase stage of the cell cycle and are very condensed. The stain gives a different pattern of light and dark ***bands*** that is diagnostic for each chromosome. Each band is given a number, and gene-mapping techniques allow genes to be positioned within a band within an arm of a chromosome. For example, the *CFTR* gene (a pathogenic variant that gives rise to cystic fibrosis) maps to 7q21; that is, on chromosome 7 in the long arm in band 21.

During cell division (mitosis), each chromosome divides into two so that each daughter nucleus has the same number of chromosomes as its parent cell. During gametogenesis, however, the number of chromosomes is halved by meiosis so that, after conception, the number of chromosomes remains the same and is not doubled. In the female each ovum contains one or other X chromosome, but in the male the sperm bears either an X or a Y chromosome.

Chromosomes can be seen easily only in actively dividing cells. Typically, lymphocytes from the peripheral blood are stimulated to divide and are processed to allow the chromosomes to be examined. Cells from other tissues can also be used for chromosomal analysis; for example, amniotic fluid, placental cells from chorionic villus sampling, or bone marrow and skin cells (Box 5.2).

The mitochondrial chromosome

In addition to the 23 pairs of chromosomes in the nucleus of every diploid cell, the mitochondria in the cytoplasm of the cell also have their own genome. The mitochondrial chromosome is a circular DNA (mtDNA) molecule of approximately 16 500 bp, and every base-pair makes up part of the coding sequence. These genes principally encode proteins or RNA molecules involved in mitochondrial function. These proteins are components of the mitochondrial respiratory chain involved in oxidative phosphorylation, producing adenosine triphosphate (ATP). They also have a critical role in apoptotic cell death. Every cell contains several hundred mitochondria, and therefore several hundred mitochondrial chromosomes. Virtually all mitochondria are inherited from the mother, as the sperm head contains no (or very few) mitochondria. Disorders mapped to the mitochondrial chromosome are shown in Fig. 5.6.

Box 5.2 Examples of clinical indications for genetic testing

Antenatal
- Positive maternal serum screening test for aneuploid pregnancy
- Ultrasound features consistent with an aneuploid fetus or genetic syndrome (e.g., skeletal dysplasia)
- Severe fetal growth retardation
- Sexing of fetus in X-linked disorders

In the neonate
- Congenital malformations
- Suspicion of trisomy or monosomy
- Suspicion of microdeletion syndrome
- Ambiguous genitalia

In children and adolescents
- Primary amenorrhoea or failure of pubertal development
- Growth retardation
- Malignant disorders (e.g., haematological malignancies, Wilms' tumour, neuroblastoma)

In the adult
- Screening of parents of a child with a chromosomal abnormality for further genetic counselling
- Learning difficulties
- Certain malignant disorders (e.g., haematological malignancies)

DNA transcription

To express the information in the genome, cells transcribe the code into the single-stranded ribonucleic acid (RNA). RNA is similar to DNA in that it comprises four bases—A, G, and C but with uracil (U) instead of T—and a sugar–phosphate backbone with ribose instead of deoxyribose. Several types of RNA are made by the cell. Messenger RNA (mRNA) codes for proteins that are translated on ribosomes. Ribosomal RNA (rRNA) is a key catalytic component of the ribosome, and amino acids are delivered to the nascent peptide chain on transfer RNA (tRNA) molecules.

A gene is usually 20 to 40 kilobases of DNA that contains the code for a polypeptide sequence (but lengths vary—the muscle protein dystrophin is 2.4 megabases long). Three adjacent nucleotides (a codon) specify a particular amino acid, such as AGA for arginine. There are only 20 common amino acids, but 64 possible codon combinations make up the genetic code (see Fig. 5.8). This redundancy means that most amino acids are encoded by more than one triplet, and additional codons are used to signal initiation or termination of polypeptide-chain synthesis.

Human protein-coding sequences (exons) are interrupted by intervening sequences that are non-coding (introns) at multiple positions (see Fig. 5.7). These are spliced from the nascent message in the nucleus by an RNA/protein complex called a spliceosome. Differential splicing can cause exons to be spliced alongside their intervening introns. This contributes significantly to the complexity of the human transcriptome, as proteins translated from these messages lack particular domains and therefore have different activity.

There are also a variety of RNAs that regulate gene expression or RNA processing. The non-coding RNAs (ncRNAs) include a group that regulates gene expression. These include microRNA (miRNA) and small interfering RNA (siRNA), which typically bind to a subset of mRNAs and inhibit their translation and/or initiate their degradation (siRNA only initiates degradation). miRNAs and siRNAs are short ncRNAs (19–29 bp) that together regulate the expression of approximately 30% of genes by degradation of transcripts or repression of protein synthesis. A growing range of additional regulatory ncRNA classes are being identified, many of which control gene expression by epigenetic mechanisms (see below).

Control of gene expression

The genome of all cells in the body encodes the same genetic information, yet different cell types express very different subsets of proteins. Gene expression is controlled at many steps, from transcription to protein degradation. However, for many genes, transcription is the key point of regulation. This is controlled primarily by proteins, which bind to short sequences within the promoter regions that either repress or activate transcription, or to more distant sequences where proteins bind to enhance expression. These transcription factors and enhancers are often the end-points of signalling pathways that transduce extracellular signals to change gene expression. This level of regulation often involves the translocation of an activated factor from the cytoplasm to the nucleus. In the nucleus, these DNA-binding proteins recognize the shape and position of hydrogen bond acceptor and donor groups within the major and minor grooves of the double helix (i.e., the double helix

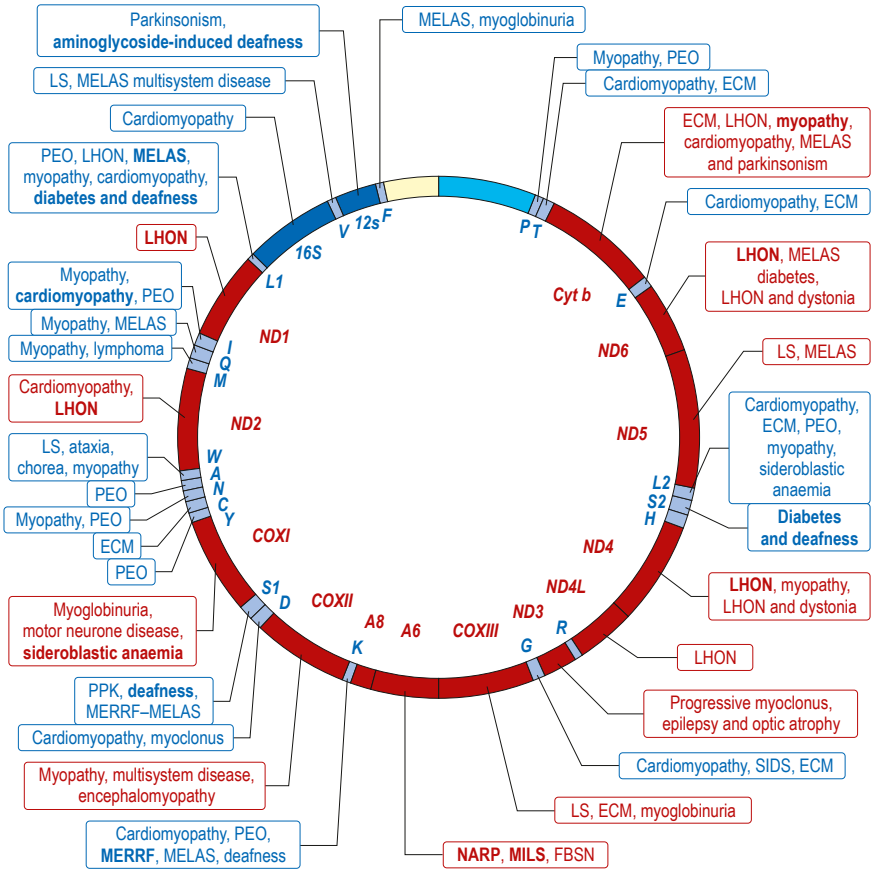


Fig. 5.6 Mitochondrial chromosome abnormalities. Disorders that are frequently or prominently associated with pathogenic variants in a particular gene are shown in bold. Diseases due to pathogenic variants that impair mitochondrial protein synthesis are shown in blue. Diseases due to pathogenic variants in protein-coding genes are shown in *red*. *ECM,* Encephalomyopathy; *FBSN,* familial bilateral striatal necrosis; *LHON,* Leber's hereditary optic neuropathy; *LS,* Leigh's syndrome; *MELAS,* mitochondrial encephalomyopathy, lactic acidosis and stroke-like episodes; *MERRF,* myoclonic epilepsy with ragged-red fibres; *MILS,* maternally inherited Leigh's syndrome; *NARP,* neuropathy, ataxia and retinitis pigmentosa; *PEO,* progressive external ophthalmoplegia; *PPK,* palmoplantar keratoderma; *SIDS,* sudden infant death syndrome.

does not need to unwind). There are several classes of DNA-binding protein that differ in the protein structural motif, which allows them to interact with the double helix. More permanent control of gene expression patterns can be achieved epigenetically.

Epigenetics

Although the importance of the genetic code has been emphasized, normal function and disease can occur, despite a particular DNA sequence being exactly the same. The term '***epigenetics***' is used to explain changes that arise due to differing gene expression but do not involve changes in the underlying DNA sequence. Despite not altering the genetic sequence the effects of epigenetic changes are stable over rounds of cell division. A number of systems initiate and sustain these changes:

- ***Modifications to DNA's surface structure*** but not its base-pair sequence: DNA methylation. Enzymes called methyltransferases modify cytosine to 5-methylcytosine. This 'tag' occurs mainly in sites where a cytosine molecule is next to a guanidine nucleotide, a CpG site. When CpG islands (groups of CpG sites) in the promoter region are methylated, gene expression is repressed.
- ***Modification of chromatin proteins*** (in particular, acetylation of histones), which not only support DNA but also bind it so tightly as to regulate gene expression. At the extreme, such binding can permanently prevent the DNA sequences from being exposed to, let alone acted on by, gene transcription (DNA-binding) proteins.

Some epigenetic modifications are also 'transmissible', meaning that a dividing liver cell, for example, can give rise to two daughter cells with the same epigenetic signals, such that they express the appropriate transcriptome for a liver cell. Epigenetic change forms the basis of genomic imprinting (see later). Epigenetic changes in gene expression also occur in cancer. An example is DNA methylation repressing tumour suppressor genes.

The X chromosome and inactivation

Although females have two X chromosomes (XX), they do not have two doses of X-linked genes (compared with just one dose for a male XY) because of the phenomenon of X inactivation or ***lyonization*** (after its discoverer, Dr Mary Lyon). In this process, one of the two X chromosomes in the cells of females is epigenetically (see earlier) silenced through the action of a regulatory ncRNA, so the cell has only one dose of the X-linked genes. Inactivation is random and can affect either X chromosome.

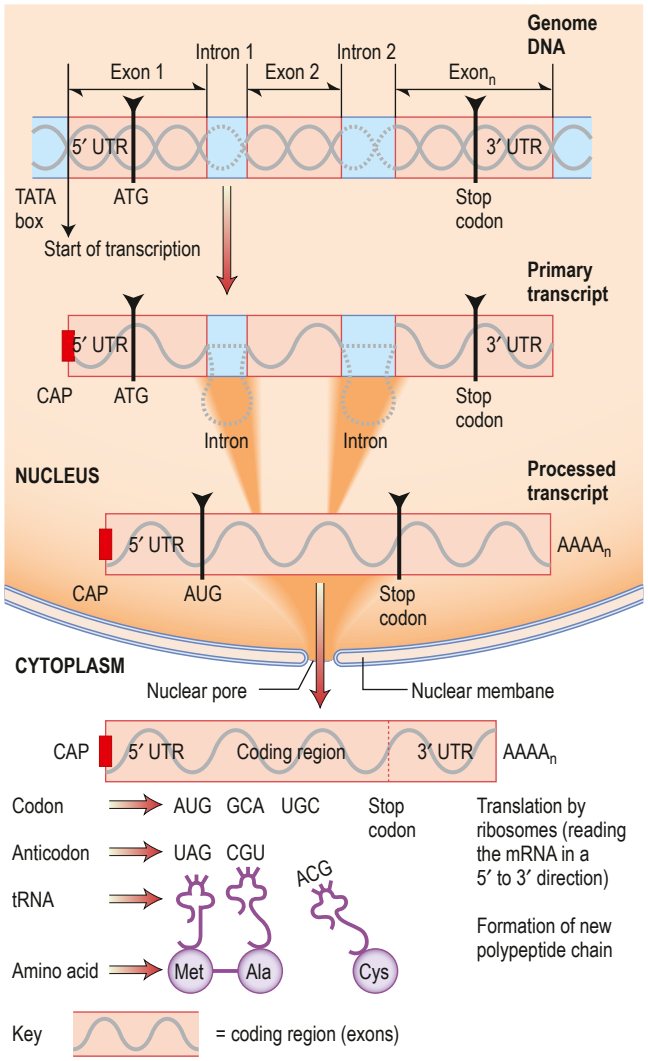
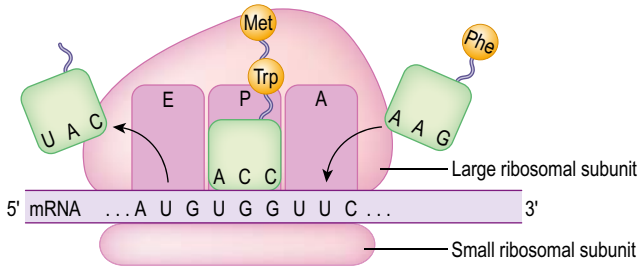


Fig. 5.7 Transcription and translation (DNA to RNA to protein). ***RNA polymerase*** transcribes an RNA copy of the DNA gene sequence. The ***transcript*** is capped at the 5′ end by the addition of an inverted guanine residue, which is then methylated to form 7-methylguanosine. At the 3′ end the sequence AAUAAA is recognized by endonuclease, and the transcript is then cleaved 20 base-pairs further downstream. ***Poly-A polymerase*** then adds adenosine residues to the 3′ end, forming a poly-A tail (polyadenylation). The ***introns*** are then spliced to produce the mature mRNA, which is trafficked to the cytosol via nuclear pores. ***Ribosomal subunits*** assemble on the 5′ end of the mRNA to translate the message into protein, for which tRNAs deliver the amino acid building blocks (see also **Fig. 5.8**). *DNA,* Deoxyribonucleic acid; *mRNA,* messenger RNA; *RNA,* ribonucleic acid; *tRNAs,* transfer RNAs; *UTR,* untranslated region.

Protein synthesis and secretion

While the instructions are essentially the same in every cell, in different tissues the regions that are transcribed are cell specific, and so are the ensuing proteins. For example, in skeletal muscle the RNA transcripts involve the generation of proteins that are needed for muscle, such as myosin. Conversely, in the cells of the cornea, specific keratins that are transparent are expressed. The actual working parts of the cell are protein, and the translation of the nucleic acid sequence to protein involves the ribosome.



A

Second letter

First letter	U	C	A	G	Third letter
U	UUU Phe, UUC Phe, UUA Leu, UUG Leu	UCU Ser, UCC Ser, UCA Ser, UCG Ser	UAU Tyr, UAC Tyr, **UAA Stop**, **UAG Stop**	UGU Cys, UGC Cys, **UGA Stop**, UGG Trp	U C A G
C	CUU Leu, CUC Leu, CUA Leu, CUG Leu	CCU Pro, CCC Pro, CCA Pro, CCG Pro	CAU His, CAC His, CAA Gln, CAG Gln	CGU Arg, CGC Arg, CGA Arg, CGG Arg	U C A G
A	AUU Ile, AUC Ile, AUA Ile, AUG Met	ACU Thr, ACC Thr, ACA Thr, ACG Thr	AAU Asn, AAC Asn, AAA Lys, AAG Lys	AGU Ser, AGC Ser, AGA Arg, AGG Arg	U C A G
G	GUU Val, GUC Val, GUA Val, GUG Val	GCU Ala, GCC Ala, GCA Ala, GCG Ala	GAU Asp, GAC Asp, GAA Glu, GAG Glu	GGU Gly, GGC Gly, GGA Gly, GGG Gly	U C A G

B

Fig. 5.8 Protein translation. **(A)** Ribosomal assembly of amino acid chains. **(B)** The genetic code and amino acids.

Protein translation

Mature mRNA is transported through the nuclear pore into the cytoplasm for translation into protein by ribosomes (**Fig. 5.8A**).

- In the cytoplasm the two ribosomal subunits interact on an mRNA molecule, usually via ribosome-binding sites encoded in the untranslated 5′ region of the message. The mRNA is then pulled through the ribosome until a translation initiation codon is encountered (usually an AUG).
- As the mRNA is pulled through the ribosome in the 5′ to 3′ direction, codons distal to the AUG are recognized by complementary sequences, or anticodons, in tRNA molecules that dock on the ribosome. Each tRNA molecule carries an amino acid specific to the anticodon (**Fig. 5.8B**). The amino acids are transferred from tRNA molecules and sequentially linked to the carboxy-terminus of the growing polypeptide by the peptidyl transferase activity of the ribosome. The poly-A tail of the mRNA is not translated (3′ untranslated region [UTR]) and is preceded by a translational stop codon; UAA, UAG, or UGA.

Protein structure

The amino acid sequence of a polypeptide chain (its ***primary structure***) ultimately determines its shape. The weak bonds (hydrogen bonds, electrostatic and van der Waals interactions) formed between the side-chains of the different amino acids and/or the peptide backbone provide the ***secondary structure*** (α-helices, β-strands, loops). These, in turn, are folded into a three-dimensional, ***tertiary structure*** to provide functional protein

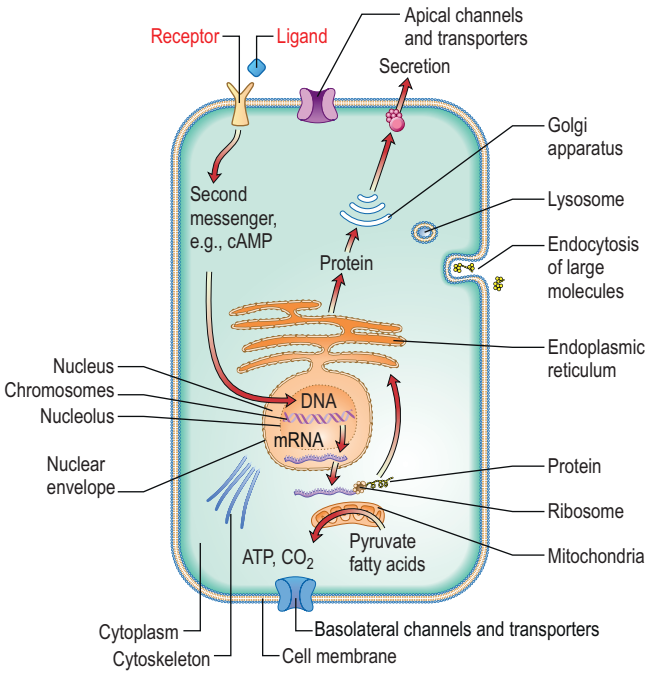


Fig. 5.9 The cell. The major organelles and receptor activation, intracellular messengers, protein formation and secretion, endocytosis of large molecules, and production of ATP are shown. *ATP,* Adenosine triphosphate; *cAMP,* cyclic adenosine monophosphate; *DNA,* deoxyribonucleic acid; *mRNA,* messenger RNA.

domains of 40 to 350 amino acids. The modular nature of domains allows their functionality to be combined in protein complexes of different proteins. This final level of organization is the ***quaternary structure***.

The folding of polypeptides into fully functional proteins is facilitated by an assortment of molecular chaperones, which bind to partially folded polypeptides and prevent the formation of inappropriate bonds.

Further reading

Beermann J, Piccoli MT, Viereck J, Thum T. Non-coding RNAs in development and disease: background, mechanisms, and therapeutic approaches. *Physiol Rev* 2016; 96(4):1297–325.
Fan K, Pfister E, Weng Z. Toward a comprehensive catalog of regulatory elements. *Hum Genet* 2023.

Genetics and the cell

The nucleus is the most prominent cellular organelle (Fig. 5.9) and has a double membrane enclosing the human genome (the outer membrane is continuous with the endoplasmic reticulum). The double membrane contains nuclear pores, through which gene regulatory proteins, transcription factors and mRNA that has been transcribed from DNA are transported. The nuclear matrix is highly organized. Microscopically dense regions of heterochromatin represent highly compacted chromosomal DNA, which tends to be transcriptionally repressed. Lighter regions of euchromatin contain extended chromosomes, which tend to be transcriptionally active. Outside the nucleus, mRNA is translated into amino acid chains in the endoplasmic reticulum.

Embryonic development and stem cells

It is critical to recognize that all of the estimated 37 trillion cells that form a person are derived from a single cell. Genetic changes that involve this first cell, whether due to natural variation or artificial manipulation via gene-editing techniques like CRISPR (clusters of regularly interspaced short palindromic repeats), are carried forward to every subsequent cell. These changes are also passed on to future generations. The gravity of intervening at this moment is captured in the ethical debates around interventions such as preimplantation genetic diagnosis and mitochondrial replacement therapy (see 'Genomic medicine').

Following fertilization the newly formed zygote and the cells it forms following the first few divisions are totipotent, meaning that they can differentiate into any cell type in the adult body (Fig. 5.10A). At the blastula stage of embryonic development, these cells undergo a primary differentiation event to become either the trophectoderm or the inner cell mass (ICM) (see Fig. 5.10A). The trophectoderm gives rise to the fetal cells of the placenta, while the ICM cells are pluripotent and give rise to all other cell types of the body (except those of the placenta); they are more commonly called embryonic stem (ES) cells. Stem cells have two properties:

- ***self-renewal:*** the ability to divide indefinitely without differentiating
- ***pluripotency*** or ***totipotency:*** the capability to differentiate, given the appropriate signals, into any cell type (except fetal placental cells).

During embryogenesis a carefully orchestrated programme ensues, with this ball of cells ultimately forming the early embryo, recognizable as human with the formation of organs such as the heart and the development of limbs (Fig. 5.10B).

As they begin to differentiate, adult progenitor cells (sometimes referred to erroneously as stem cells) retain their ability to self-renew, but their potency is reduced; with this limited ability to self-renew they can differentiate into multiple related lineages (multipotent, like haemopoietic 'progenitor cells') or single lineages (unipotent, like muscle satellite cells). The body uses these partially differentiated progenitor cells to replace or repair damaged cells and tissues, even into adulthood.

Stem cells have great therapeutic potential and can be obtained from blood from the umbilical cord, which contains embryo-like stem cells (which are not as primitive as ES cells but can differentiate into many more cell types than adult progenitor cells). Alternatively, adult cells, such as skin fibroblast cells, can be reprogrammed to regain stem-like properties (induced pluripotent stem cells, IPSCs).

Further reading

Wang JY, Doudna JA. CRISPR technology: A decade of genome editing is only the beginning. *Science* 2023; 379 (6629):eadd8643.
Doudna JA, Charpentier E. The new frontier of genome engineering with CRISPR-Cas9. *Science* 2014; 346:1258096.
Yamanaka S. Pluripotent stem cell-based cell therapy-promise and challenges. *Cell Stem Cell* 27 2020; (4):523-31.

Review of cellular genetics

According to the fundamental paradigm in genetics, information to create a human is stored in the genetic code in a cell. This code is identically replicated in every cell from conception onwards. In different cells, different parts of the code are needed, specific to that cell type, to make proteins required for the functioning of that cell; this control of gene expression is achieved through a number of

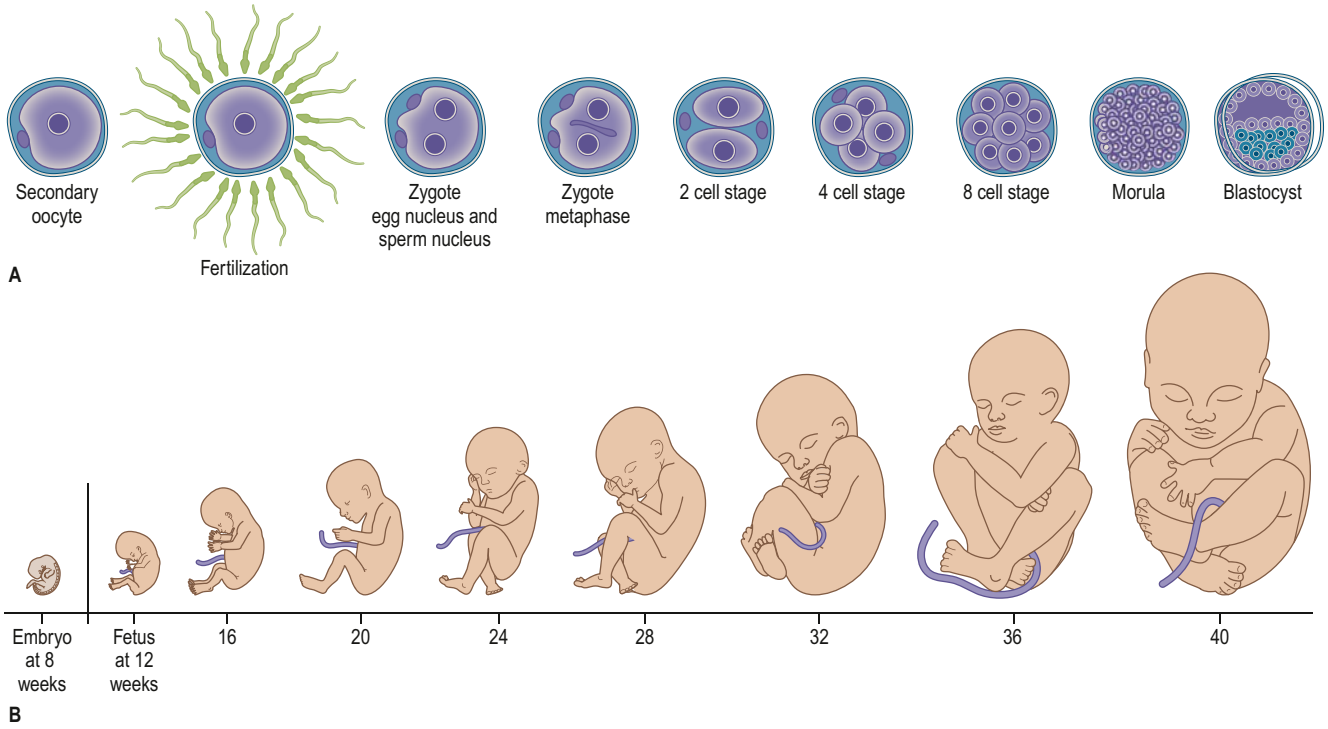


Fig. 5.10 Cell division, embryogenesis, and fetal development. **(A)** Early stages of cell division leading to the formation of the blastocyst. **(B)** Stages of fetal development.

mechanisms affecting gene transcription, processing of RNA and translation of mRNA into protein.

In the next section, we describe examples of what happens when there are specific changes in the genetic code that have been associated with the presentation of specific genetic disorders.

GENETIC DISORDERS

Inherited genetic disorders occur due to pathogenic variants in the genetic code; their prevalence is indicated in **Box 5.3**. They can arise because of new mutations in sperm or ova, or as a result of an inherited pathogenic variant passed down a family line. These disorders can stem from a single base-pair change, alteration of a region of a chromosome, or even duplication or omission of an entire chromosome (**Box 5.4**). The pattern of inheritance in a family tree depends on which chromosome the pathogenic variant occurs in, and whether both copies need to be affected to cause the disease. Somatic mutations typically occurring after birth can be responsible for benign and malignant tumours (see 'The genetic basis of cancer').

To start, we will consider conditions for which chromosomal changes are responsible. These were among the first to be studied, as the technique known as karyotyping allowed for the visualization of differences in chromosomes using a light microscope.

Chromosomal disorders

Chromosomal abnormalities are much more common than is generally appreciated (see **Box 5.3**). Over half of spontaneous abortions have chromosomal abnormalities, compared with only 4 to 6 abnormalities per 1000 live births. Specific chromosomal abnormalities can lead to well-recognized and severe clinical syndromes; autosomal aneuploidy (a differing from the normal diploid number) is

Box 5.3 Prevalence of genetic disease[a]

Genetic disease
- 0.5% of all newborns have a chromosomal abnormality
- 7% of all stillborns have a chromosomal abnormality
- 20% to 30% of all infant deaths are due to genetic disorders
- 11% of paediatric hospital admissions are for children with genetic disorders
- 12% of adult hospital admissions are due to genetic causes
- 15% of all cancers have an inherited susceptibility
- 10% of chronic diseases in the adult population (heart, diabetes, arthritis) have a significant genetic component

Congenital malformation
- 3% to 5% of all births result in congenital malformations
- 30% to 50% of post-neonatal deaths are due to congenital malformations
- 18% of paediatric hospital admissions are for children with congenital malformations

European incidences per 1000 births

Single-gene disorders: 10
- Dominant: 7.0
- Recessive: 1.66
- X-linked: 1.33

Chromosomal disorders: 3.5
- Autosomes: 1.69
- Sex chromosomes: 1.80

Congenital malformations: 31
- Genetically determined: 0.6
- Multifactorial: 30
- Non-genetic: ~0.4

[a]Although individual genetic diseases are rare, regional variation is enormous; the incidence of Down's syndrome varies from 1/1000 to 1/100 worldwide. Single-gene diseases collectively comprise over 15 500 recognized genetic disorders. The global prevalence of all single-gene diseases at birth is approximately 10/1000.

Box 5.4 Some inherited genetic conditions that exemplify different genetic mechanisms

Condition	Mechanism responsible
Down's syndrome	Trisomy 21
Cystic fibrosis	Recessive
Haemophilia	Dominant or recessive
BRCA	Dominant
Xeroderma pigmentosum	Recessive
Huntington's disease	Anticipation
Angelman's syndrome	Imprinting
Myoclonic epilepsy with ragged-red fibres (MERRFs)	Mitochondrial

usually more severe than the sex chromosome aneuploidies (Box 5.5). Abnormalities may occur in either the number or the structure of the chromosomes.

Abnormal chromosome numbers

If a chromosome or chromatids fail to separate ('non-disjunction') in either meiosis or mitosis, one daughter cell will receive two copies of that chromosome, and one daughter cell will receive no copies of the chromosome. If this non-disjunction occurs during meiosis, it can lead to an ovum or sperm having:

- ***an extra chromosome***, resulting in a fetus that is 'trisomic' and has three instead of two copies of the chromosome
- ***one less chromosome***, resulting in a fetus that is 'monosomic' and has one instead of two copies of the chromosome.

Non-disjunction can occur with autosomes or sex chromosomes. However, only individuals with trisomy 13, 18 and 21 survive to birth, and most children with trisomy 13 and trisomy 18 die in early childhood. ***Trisomy 21 (Down's syndrome)*** is observed with a frequency of 1 in 650 live births, regardless of geography or ethnic background. This rate is reduced with widespread screening and selective termination of pregnancy (see 'Prenatal diagnosis of congenital anomalies'). Full ***autosomal monosomies*** are extremely rare and generally incompatible with survival to full gestation. ***Sex chromosome trisomies*** (e.g., Klinefelter's syndrome, XXY) are relatively common. The ***sex chromosome monosomy*** in which the individual has one X chromosome only and no second X or Y chromosome (i.e., X0) is known as ***Turner's syndrome*** and is estimated to occur in 1 in 2500 live-born females.

Occasionally, non-disjunction can occur during mitosis shortly after two gametes have fused. It will then result in the formation of two cell lines, each with a different chromosome complement; the person is termed a 'mosaic' as only some of their cells carry an incorrect number of chromosomes (see later).

Very rarely the entire chromosome set will be present in more than two copies, so the individual may be triploid rather than diploid and have a chromosome number of 69. Triploidy and tetraploidy (four sets) is not usually compatible with life and results in spontaneous abortion.

Abnormal chromosome structures

Along with abnormal numbers of chromosomes, chromosomes can have abnormal structures, and the disruption to the DNA and gene sequences may give rise to a genetic disease.

- ***Deletions*** of a portion of a chromosome may give rise to a disease syndrome if two copies of the genes in the deleted region are necessary, and with just the one normal copy remaining on the non-deleted homologous chromosome the individual will not be normal. Many deletion syndromes have been well described. For example, ***Prader–Willi syndrome*** is the result of deletion of part of the long arm of chromosome 15; ***Wilms' tumour*** is characterized by deletion of part of the short arm of chromosome 11; and microdeletions in the long arm of chromosome 22 give rise to ***DiGeorge's syndrome*** (see Chapter 6).
- ***Duplications*** occur when a portion of the chromosome is present on the chromosome in two copies, so the genes in that chromosome portion are present in an extra dose. A form of neuropathy, ***Charcot–Marie–Tooth disease*** (see Chapter 28), is due to a small duplication of a region of chromosome 17.
- ***Inversions*** involve an end-to-end reversal of a segment within a chromosome, for example, 'abcdefgh' becomes 'abcfedgh', as in ***haemophilia*** (see Chapter 38).
- ***Translocations*** occur if two chromosome regions join together, when they would not do so normally. Chromosome translocations in somatic cells may be associated with tumorigenesis (see 'The genetic basis of cancer') and Fig. 5.24).

Translocations can be very complex, involving more than two chromosomes, but most are simple and fall into one of two categories:

- ***Reciprocal translocations*** occur when any two non-homologous chromosomes break simultaneously and rejoin, swapping ends. In this case the cell still has 46 chromosomes but two of them are rearranged. Someone with a balanced translocation is likely to be normal (unless a translocation breakpoint interrupts a gene); however, at meiosis, when the chromosomes separate into different daughter cells, the translocated chromosomes will enter the gametes, and any resulting fetus may inherit one abnormal chromosome and have an unbalanced translocation with physical manifestations.
- ***Robertsonian translocations*** occur when two acrocentric chromosomes join and the short arm is lost, leaving only 45 chromosomes. This translocation is balanced, and the individual is healthy. However, any offspring have a risk of inheriting an unbalanced arrangement. This risk depends on which acrocentric chromosome is involved. The 14/21 Robertsonian translocation is clinically relevant. A female with this karyotype has a 1 in 8 risk of having a baby with Down's syndrome (a male carrier has a 1 in 50 risk). However, they have a 50% risk of producing a carrier like themselves, hence the necessity for genetic family studies. Relatives should be alerted to the increased risk of Down's syndrome in their offspring and should have appropriate genetic counselling.

Mitochondrial chromosome disorders

The mitochondrial chromosome (see Fig. 5.6) carries its genetic information in a very compact form; for example, there are no introns in the genes. Therefore any mutation has a high chance of having an effect. However, as every cell contains hundreds of mitochondria, a single altered mitochondrial genome will not be noticed. As mitochondria divide, there is a statistical likelihood that there will be more mutated mitochondria, and, at some point, this will give rise to a mitochondrial disease.

Most mitochondrial diseases are myopathies and neuropathies with a maternal pattern of inheritance. Other abnormalities include retinal degeneration, diabetes mellitus, and hearing loss. ***Myopathies*** include chronic progressive external ophthalmoplegia (CPEO); encephalomyopathies include myoclonic epilepsy with ragged-red fibres (MERRFs) and mitochondrial encephalomyopathy, lactic acidosis and stroke-like episodes (MELAS) (see Chapter 28).

Box 5.5 Chromosomal abnormalities: examples of a few syndromes

Syndrome	Chromosome karyotype	Incidence and risks	Clinical features	Mortality
Autosomal abnormalities				
Trisomy 21 (Down's syndrome)	47,+21 (95%) Mosaicism translocation 5%	1:650 (overall) (risk with a 20- to 29-year-old mother 1:1000; >45-year-old mother 1:30)	Flat face, slanting eyes, epicanthic folds, small ears, single palmar crease, short and stubby fingers, hypotonia, variable learning difficulties, congenital heart disease (up to 50%)	High in first year but many survive to adulthood
Trisomy 13 (Patau's syndrome)	47,+13	1:5000	Low-set ears, cleft lip and palate, polydactyly, micro-ophthalmia, learning difficulties	Rarely survive for more than a few weeks
Trisomy 18 (Edwards' syndrome)	47,+18	1:3000	Low-set ears, micrognathia, rocker-bottom feet, learning difficulties	Rarely survive for more than a few weeks
Sex chromosome abnormalities				
Fragile X syndrome	46,XX, fra (X) 46,XY, fra (X)	1:2000	Most common inherited cause of learning difficulties, predominantly in males Macro-orchidism	Typically survive to adulthood
Female				
Turner's syndrome	45,XO	1:2500	Infantilism, primary amenorrhoea, short stature, webbed neck, cubitus valgus, normal IQ	Typically survive to adulthood
Triple X syndrome	47,XXX	1:1000	No distinctive somatic features, learning difficulties	Typically survive to adulthood
Male				
Klinefelter's syndrome	47,XXY (or XXYY)	1:1000 (more in sons of older mothers)	Decreased crown–pubis:pubis–heel ratio, eunuchoid, testicular atrophy, infertility, gynaecomastia, learning difficulties (20%; related to number of X chromosomes)	Typically survive to adulthood
Double Y syndrome	47,XYY	1:800 High incidence in tall criminals	Tall, fertile, minor mental and psychiatric illness	Typically survive to adulthood

- ***Kearns–Sayre syndrome*** includes ophthalmoplegia, heart block, cerebellar ataxia, deafness and learning difficulties due to long deletions and rearrangements.
- ***Leber's hereditary optic neuropathy*** (LHON) is the most common cause of blindness in young males, with bilateral loss of central vision and cardiac arrhythmias, and is an example of a mitochondrial disease caused by a point mutation in one gene.
- ***Multisystem disorders*** include ***Pearson's syndrome*** (sideroblastic anaemia, pancytopenia, exocrine pancreatic failure, subtotal villous atrophy, diabetes mellitus, and renal tubular dysfunction). In some families, hearing loss is the only symptom, and one of the mitochondrial genes implicated may predispose patients to aminoglycoside cytotoxicity.

Analysis of chromosomal disorders

Historically, karyotyping involved growing cells from patient samples and studying the chromosomes. The cell cycle (Fig. 5.11) can be arrested at mitosis with the use of colchicine and, following staining, the chromosomes with their characteristic banding can be seen and any abnormalities identified. This is an automated process with computer scanning software searching for metaphase spreads and then automatic binning of each chromosome to allow easy scoring of chromosome number and banding patterns.

Current approaches utilize genome-wide array-based platforms (called single nucleotide polymorphism [SNP] arrays) to identify changes in chromosome copy number and can identify very small interstitial deletions and insertions. The presence of translocations cannot be detected using SNP arrays but instead is studied in a targeted way using fluorescence *in situ* hybridization (FISH). Large, region-specific probes are labelled with fluorescently tagged nucleotides and used to allow rapid identification of metaphase chromosomes (Fig. 5.12). As whole-genome sequencing is becoming more accessible and coverage increases, it will increasingly be used to identify deletions, insertions, and translocation breakpoints.

Point mutations, insertions, and deletions

The advent of Sanger sequencing rapidly accelerated the discovery of base-pair level changes. These pathogenic variants can have various effects on the expression of the gene, and many cause a dysfunction of the protein product. In contrast to the previous section, where changes were described that affect millions of base-pairs in a cell, point mutations can involve just one critical base-pair. These are small but important changes.

Mutations

Although DNA replication is a very accurate process, mistakes occasionally occur and produce changes or mutations. These changes can also arise because of other factors, such as radiation, ultraviolet light, or chemicals. Mutations in gene sequences or in the sequences that regulate gene expression (transcription and translation) may alter the amino acid sequence in the protein encoded by that gene. In some cases, protein function will be maintained; in

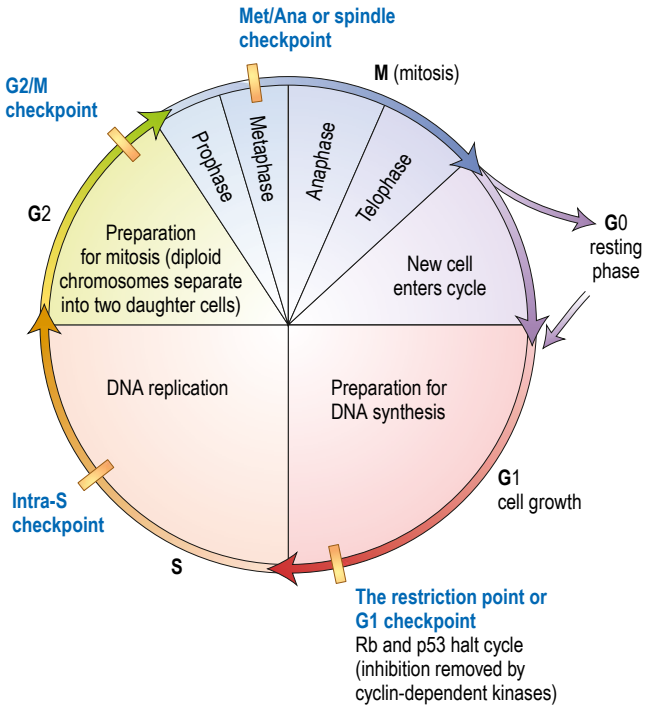


Fig. 5.11 The cell cycle. Cells are stimulated by growth factors to leave non-cycle ***G0*** to enter G1 phase. ***During G1***, transcription of the DNA synthesis molecules occurs. Rb is a 'checkpoint' (inhibition molecule) between G1 and S phases and must be removed for the cycle to continue. This is achieved by the action of the cyclin-dependent kinase produced during G1. During the ***S phase***, any DNA defects will be detected and p53 will halt the cycle. Following DNA synthesis (S phase), cells enter ***G2***, a preparation phase for cell division. Mitosis takes place in the ***M phase***. The new daughter cells can now either enter ***G0*** and differentiate into specialized cells, or re-enter the cell cycle. *DNA*, Deoxyribonucleic acid.

other cases, it will change or cease, perhaps producing a clinical disorder. Many different types of mutation occur.

Point mutation

This is the simplest type of change and involves the substitution of one nucleotide for another, thus changing the codon in a coding sequence and leading to an amino acid substitution (assuming the new sequence is non-synonymous with the former). For example, in ***sickle cell disease*** a pathogenic variant within the globin gene changes one codon from GAG to GTG so that, instead of glutamic acid, valine is incorporated into the polypeptide chain, which radically alters its properties. However, substitutions may have no effect on the function or stability of the proteins produced, as several codons code for the same amino acid (synonymous—see Fig. 5.8).

Insertion or deletion

Insertion or deletion of one or more bases is a more serious change, particularly if the inserted or deleted DNA is not a multiple of three bases, as this will cause the following sequence to be out of the reading frame and can result in a premature stop codon.

Splicing mutations

If the DNA sequences that direct the splicing of introns from mRNA are mutated, then abnormal splicing may occur. In this case the processed mRNA that is translated into protein by the ribosomes

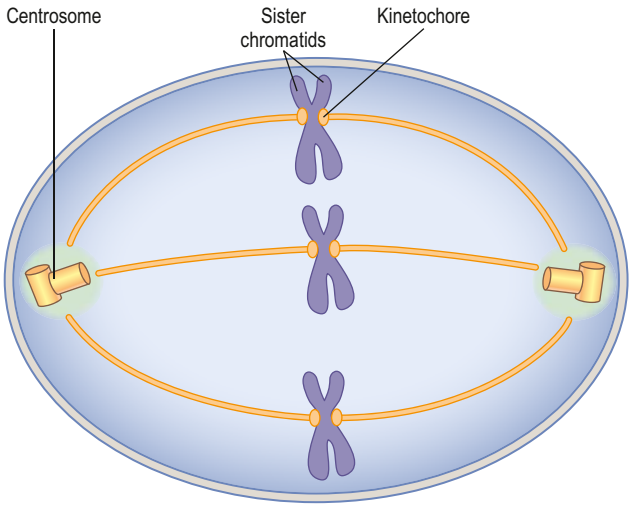


Fig. 5.12 The mitotic spindle at metaphase.

may carry intron sequences or miss exons, thus altering amino acid sequence and affecting protein function.

Nonsense mutations

A nonsense mutation is a point mutation in a sequence of DNA that results in a premature stop codon.

Prenatal diagnosis of congenital anomalies

Prenatal screening should be offered to *all* pregnant females. In England, screening for the three most common trisomies (Down's, Edwards' and Patau's) is offered universally. The risk increases disproportionately for children born to mothers older than 35 years. Infants born to mothers with a history or family history of other conditions due to chromosomal abnormalities may also be at increased risk.

Personal choice

There should be a detailed discussion with all mothers about the possible consequences of each screening test before they are offered it. In particular, they should have an understanding of the failure rates, detection rates, and false-positive and false-negative rates of each test so that they can exercise choice properly.

Screening

10–14 weeks (combined test)

A screening test for Down's syndrome, Edwards' syndrome and Patau's syndrome is available between weeks 10 and 14 of pregnancy and is called a combined test. Mothers may choose to have screening for Down's syndrome alone if they prefer.

The combined test comprises:

- ultrasound for nuchal translucency measurement
- testing of maternal serum for pregnancy-associated plasma protein-A (PAPP-A from the syncytial trophoblast) and β-human chorionic gonadotrophin.

All serum marker measurements are corrected for gestational ages, a multiple of the median value for the appropriate week of gestation. It is possible to sequence cell-free placental DNA in the maternal bloodstream to measure the copies of chromosomes that inform the risk of the fetus having a common chromosomal disorder. Known as non-invasive prenatal testing (NIPT), it is offered as a second test in females where a high-risk result (likelihood greater than 1

in 150) is returned from the combined test. In the event of an abnormal NIPT test, it may be necessary to consider a formal and more invasive diagnostic test (chorionic villus sampling or amniocentesis).

14–20 weeks (serum quadruple test)

A serum quadruple test is done if the pregnancy is too advanced for the earlier tests or if the combined test was not offered.

The quadruple test for chromosomal abnormalities consists of testing maternal serum for levels of:

- α-fetoprotein (low in Down's syndrome pregnancies).
- unconjugated oestradiol (low in Down's syndrome pregnancies)
- human chorionic gonadotrophin (raised in Down's syndrome pregnancies)
- inhibin A (raised in Down's syndrome pregnancies).

Again, a high-risk result is followed by NIPT and/or a diagnostic test.

18–20 weeks (ultrasound)

Ultrasound detects structural abnormalities (e.g., neural tube defects and congenital heart defects). Reported detection rates for congenital anomalies vary; for example, from 14% to 61% for hypoplastic ventricles to 97% to 100% for anencephaly.

Newborn screening

In the UK and other countries, there has been a longstanding program of screening normal newborns for a handful of genetic conditions where knowledge of the result will influence care of the newborn. The newborn blood spot test ('Guthrie test') relies on biochemical testing to detect nine rare but serious health conditions in neonates, including sickle cell anaemia, cystic fibrosis and phenylketonuria. Genomics England, which led the 100 000 Genomes Project, is about to embark on the whole genome screening of 100 000 apparently normal newborns in the UK. The aim is to detect a range of conditions, including inborn errors of metabolism, immunodeficiencies, and other conditions where knowledge of the genetic change within the first 2 weeks of life can make a difference to the life course of the affected child. This may be by modifications to lifestyle (e.g., avoidance of sun exposure in patients with xeroderma pigmentosum), dietary changes (patients with inborn errors), enzyme supplementation, and implanting cochlear devices in patients with some forms of congenital hearing.

Further reading

Talkowski ME, Rehm HL. Introduction of genomics into prenatal diagnostics. *Lancet* 2019; 393 (10173):719–721.
NHS. *Screening Tests in Pregnancy: Your Pregnancy and Baby Guide*. https://www.nhs.uk/conditions/pregnancy-and-baby/screening-tests-abnormality-pregnant/. https://www.gov.uk/government/publications/screening-tests-for-you-and-your-baby/downs-syndrome-edwards-syndrome-and-pataus-syndrome-combined-or-quadruple-test-taken-on-or-after-1-june-2021#the-second-screening-test-nipt.
Harrington JS, Ryter SW, Plataki M, Price DR, Choi AMK. Mitochondria in health, disease, and aging. *Physiol Rev* 2023; 103 (4):2349–2422.
Tolle I, Tiranti V, Prigione A. Modeling mitochondrial DNA diseases: from base editing to pluripotent stem-cell-derived organoids. *EMBO Rep* 2023; 24 (4):e55678.

Patterns of inheritance of single-gene diseases

Monogenetic disorders involving single genes can be inherited in a dominant, recessive or sex-linked manner.

Autosomal dominant disorders

Each diploid cell contains two copies of all the autosomes. An autosomal dominant disorder (Fig. 5.13A) occurs when one of the two copies has a pathogenic variant, and the protein produced by the normal form of the gene cannot compensate. In this case a heterozygous individual who has two different forms (or alleles) of the same gene will manifest the disease. The offspring of heterozygotes have a 50% chance of inheriting the chromosome carrying the disease allele and therefore also of having the disease. However, estimation of risk to offspring for counselling families can be difficult because of three factors:

- Some disorders have great variability in their manifestation. 'Incomplete penetrance' may occur if patients have a dominant disorder, but it does not manifest itself clinically in them. This gives the appearance of the gene having 'skipped' a generation.
- Dominant traits are extremely variable in severity (variable expression), and a mildly affected parent may have a severely affected child.
- New cases in a previously unaffected family may be the result of a new pathogenic variant. In this case the risk of a further affected child is negligible. Most cases of achondroplasia, for example, are due to new pathogenic variants.

Autosomal recessive disorders

These disorders (Fig. 5.13B) manifest themselves only when an individual is homozygous or a compound heterozygote for the disease allele; that is, both chromosomes carry the same gene pathogenic variant (homozygous) or two different pathogenic variants in the same gene (compound heterozygote). The parents are unaffected carriers (heterozygous for the disease allele). If carriers marry the offspring have a 1 in 4 chance of carrying both mutant copies of the gene and being affected, a 1 in 2 chance of being a carrier, and a 1 in 4 chance of being genetically normal. Parental relatedness increases the risk.

Sex-linked disorders

Genes carried on the X chromosome are said to be 'X-linked' and can be dominant or recessive in the same way as autosomal genes (Fig. 5.13C and D).

X-linked recessive disorders

These disorders present in males and only in homozygous females (usually rare). X-linked recessive diseases are transmitted by healthy female carriers or affected males if they survive to reproduce. An example of an X-linked recessive disorder is ***haemophilia A*** (see Chapter 38), which is caused by a pathogenic variant in the X-linked gene for factor VIII. It has been shown that in 50% of cases there is an intrachromosomal rearrangement (inversion) of the tip of the long arm of the X chromosome (one breakpoint being within intron 22 of the factor VIII gene).

Of the offspring from a carrier female and a normal male:

- 50% of the girls will be carriers, as they inherit a mutant allele from their mother and the normal allele from their father; the other 50% of the girls inherit two normal alleles and are themselves normal.
- 50% of the boys will have haemophilia, as they inherit the mutant allele from their mother (and the Y chromosome from their father); the other 50% of the boys will be normal, as they inherit the normal allele from their mother (and the Y chromosome from their father).

The male offspring of a male with haemophilia and a normal female will not have the disease, as they do not inherit his X chromosome. However, all the female offspring will be carriers, as they all inherit his X chromosome.

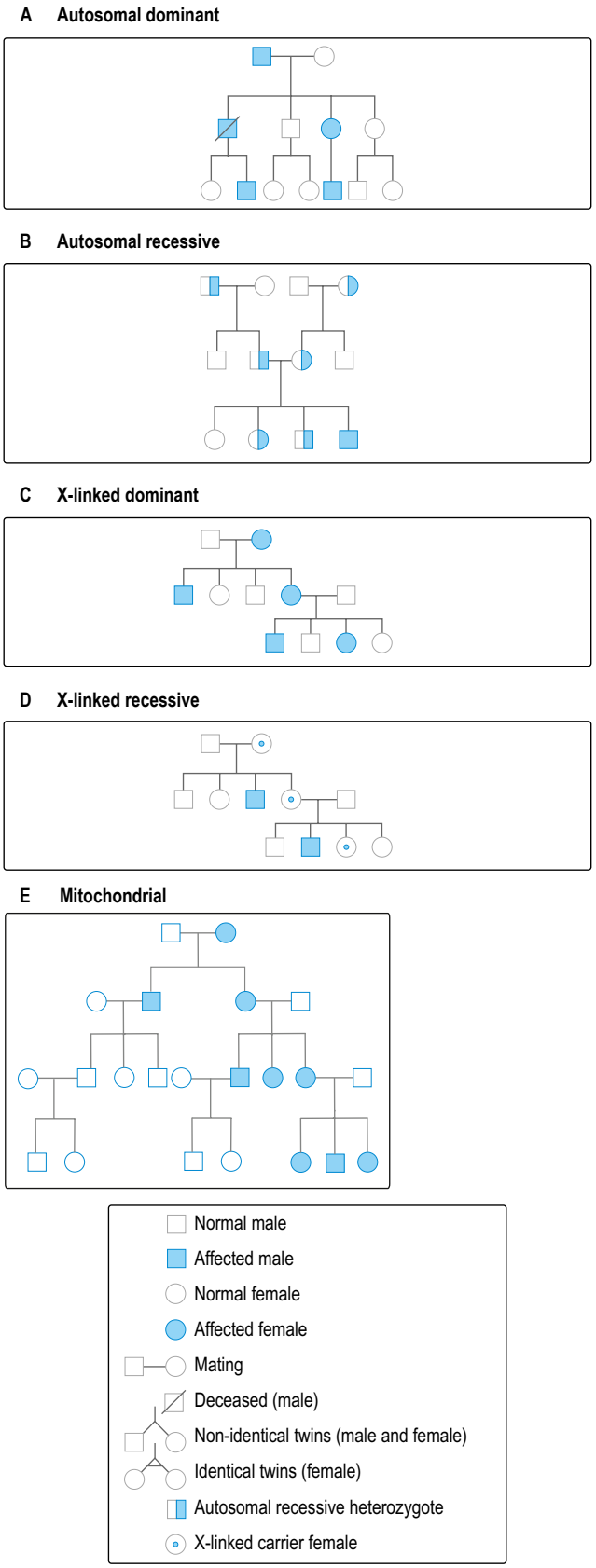


Fig. 5.13 **Modes of inheritance of simple gene disorders and mitochondrial inheritance.**

X-linked dominant disorders

These are rare. Females who are heterozygous for the mutant gene and males who have one copy of the mutant gene on their single X chromosome will manifest the disease. Half the male or female offspring of an affected mother and all the female offspring of an affected male will have the disease. Affected males tend to have the disease more severely than the heterozygous female.

Mitochondrial inheritance

The pattern of inheritance associated with alterations in the mtDNA involves both males and females, but always with the condition passed on through the female line (maternal inheritance). Since many mitochondria are passed into the egg from the cells in the ovary, all the offspring of an affected female would be expected to inherit the condition. An affected male does not pass his mitochondria on to his children, so his children will be unaffected (see **Fig. 5.13E**). Note that not all mitochondrial diseases will show this pattern of inheritance. As these conditions arise due to incorrectly functioning mitochondria, they may also be caused by an alteration in the nuclear DNA that has an impact on mitochondrial function, and not just be caused directly by a mitochondrial pathogenic variant. Hence, some mitochondrial diseases may demonstrate dominant patterns of inheritance.

Special presentations of genetic disease

These are disorders that may be due to pathogenic variants in single genes but which do not manifest as simple monogenic disorders. They can arise from a variety of mechanisms.

Mosaicism

Genetic mosaicism arises when a mutation occurs in a cell during embryonic development. This cell and its progeny are then genetically distinct from the rest of the fetus. This can give rise to a less severe form of recognized genetic disease, sometimes referred to historically as a 'forme fruste'. Equally, some mutations are not compatible with life if every cell in the fetus is affected but may be rescued if the fetus is mosaic. Schimmelpenning's syndrome is an example, where patients carry a mosaic *HRAS* gene pathogenic variant.

Triplet repeat mutations and genetic anticipation

In the gene responsible for myotonic dystrophy (see Chapter 28) the mutated allele was found to have an expanded 3′ UTR region in which three nucleotides, CTG, were repeated up to about 200 times. In families with myotonic dystrophy, people with the late-onset form of the disease had 20 to 40 copies of the repeat, but their children and grandchildren who presented with the disease from birth had vast increases in the number of repeats—up to 2000 copies. It is thought that some mechanism during meiosis causes this 'triplet repeat expansion' so that the offspring inherit an increased number of triplets. The number of triplets affects mRNA and protein function, and patients with more repeats present earlier (**Box 5.6**). The phenomenon has also been recognized with Huntington's chorea, where patients with more repeats present earlier and with progressively worse symptoms. This 'anticipation' is due to unstable mutations occurring within the disease gene. Trinucleotide repeats, such as CTG (dystrophia myotonica) and CAG (Huntington's disease), expand within the disease gene with each generation, and somatic expansion with cellular replication is also observed. This type of genetic mutation can occur within the translated region or untranslated (and presumably

regulatory) regions of the target genes. This genetic distinction has been used to subclassify a number of genetic diseases that have now been shown to be caused by trinucleotide repeat expansion and display phenotypic 'anticipation' (see Box 5.6).

Imprinting

Normal humans need a diploid number of chromosomes; that is, 46, with 23 inherited from each parent. For the majority of genes both maternally and paternally inherited alleles are functional; however, in some cases the gene from one parent is silenced. The mechanism that regulates the differential expression of two alleles of the same gene is termed ***genetic imprinting***, an epigenetic change brought about through DNA methylation.

Imprinting is relevant to human genetic disease because different phenotypes may result, depending on whether the mutant chromosome is maternally or paternally inherited. A deletion of part of the long arm of chromosome 15 (15q11–q13) will give rise to ***Prader–Willi syndrome***, if it is paternally inherited. A deletion of a similar region of the chromosome gives rise to ***Angelman's syndrome*** if it is maternally inherited. The affected gene has been identified as ubiquitin *(UBE3A)*.

Different genes that cause the same disease

Although classically divided into autosomal dominant, recessive or X-linked disorders, many syndromes show multiple forms of inheritance pattern. For example, in ***Ehlers–Danlos syndrome***, there is autosomal dominant, recessive and X-linked inheritance. This has now been explained by the finding that pathogenic variants in different genes that contribute to a common function can have the same clinical presentation. In addition, in some recessive disorders, even having just one defective allele gives a mild form of the disease (semi-dominant), while having both alleles with the pathogenic variant results in a more severe form of the syndrome.

Different diseases caused by the same gene

The following are two examples of how different pathogenic variants in a gene can cause different diseases, referred to as genotype–phenotype correlation.

CYLD

Patients with CYLD cutaneous syndrome carry germline pathogenic variants in *CYLD*, where two strikingly different phenotypes may be seen. Patients can develop tumours on the head or on the face, previously categorized as two separate diseases. It is now known that these different presentations can occur within a family where the same pathogenic variant has been documented. It is thought that the distinct phenotypes are due to yet undefined changes present in the genetic background of patients who carry *CYLD* pathogenic variants (Fig. 5.14).

Box 5.6 Examples of trinucleotide repeat genetic disorders

Syndrome inheritance pattern	Disease prevalence	Gene, product, location, and disorder	Genetic test detection rate (%)
Friedreich's ataxia—AR	2–4/100 000	*FRDA* (frataxin) 9q13—GAA trinucleotide repeat expansion disorder in intron 1 of *FRDA*	96
Fragile X syndrome—X-linked	16–25/100 000	*FMR1* (fragile X mental retardation 1 protein) Xq27.3—CGG trinucleotide repeat expansion and methylation changes in the 5′ untranslated region of *FMR1* exon 1	99
Huntington's disease—AD	3–15/100 000	*HD* (Huntingtin protein) 4p16.3—CAG trinucleotide repeat expansion within the translated protein, giving rise to long tracts of repeat glutamine residues in *HD*	98
Myotonic dystrophy—AD	1/20 000	*DMPK* (myotonin–protein kinase) 19q13.2–13.3—CTG trinucleotide repeat expansion in the 3′ untranslated region of the *DMPK* gene (dystrophia myotonica 1, DM1). Less common form—expanded. CCTG repeat in zinc finger protein 9 *(ZNF9)* gene (DM2)	100

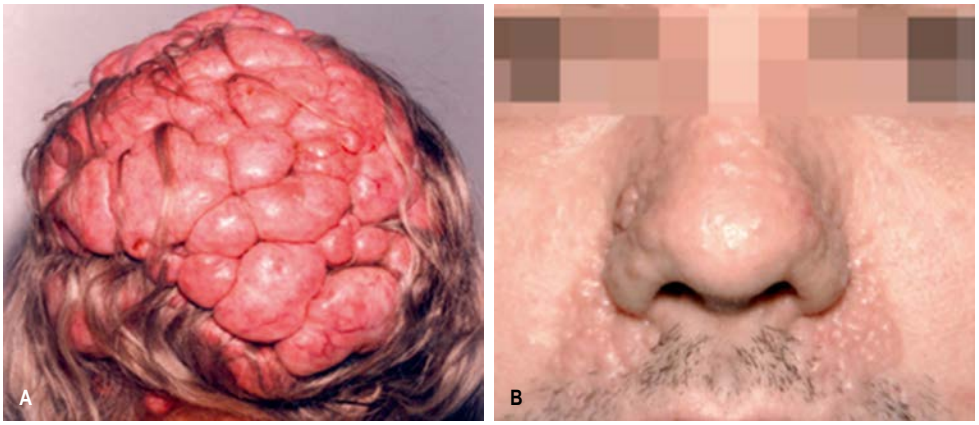


Fig. 5.14 CYLD cutaneous syndrome. Distinct phenotypes seen with identical genetic pathogenic variants reflect the impact of the genetic background on disease presentation. **(A)** Scalp cylindromas. **(B)** Trichoepitheliomas around the nasal fold.

Box 5.7 Examples of disorders that may have a polygenic inheritance

Disorder	Frequency (%)	Heritability (%)[a]
Hypertension	5	62
Asthma	4	80
Schizophrenia	1	85
Congenital heart disease	0.5	35
Neural tube defects	0.5	60
Congenital pyloric stenosis	0.3	75
Ankylosing spondylitis	0.2	70
Cleft palate	0.1	76

[a]Percentage of the total variation of a trait that can be attributed to genetic factors.

STAT3

The transcription factor signal transducer and activator of transcription 3 *(STAT3)* is a critical regulator of multiple and diverse cellular processes. Heterozygous, germline, loss-of-function pathogenic variants in *STAT3* lead to the primary immune deficiency hyper-IgE syndrome. Recently, germline, heterozygous pathogenic variants in *STAT3* that confer a gain of function have been discovered and result in early-onset, multiorgan autoimmunity.

Complex traits: multifactorial and polygenic inheritance

Characteristics resulting from a combination of genetic and environmental factors are said to be multifactorial; those involving multiple genes are said to be polygenic. There has been an explosion in genetic discoveries about these complex traits with the development of high-throughput genome-wide SNP arrays, which has allowed cost-effective and unbiased screening of large case–control cohorts (in excess of 1000 cases). These studies are known as genome-wide association studies (GWASs) and are discussed later. They have permitted unequivocal identification of SNPs associated with a variety of traits and diseases. For example, over 30 SNPs in immune-related gene loci have been associated with coeliac disease, with over half also associated with other immune-mediated or inflammatory diseases. This indicates that there are many low-risk genetic risk factors associated with complex traits, and common pathways are implicated in different diseases.

Most human diseases, such as heart disease, diabetes, and common mental disorders, are multifactorial traits (**Box 5.7**).

Further reading

Depienne C, Mandel JL. 30 Years of repeat expansion disorders: what have we learned and what are the remaining challenges? *Am J Hum Genet* 2021; 108 (5):764–785.

Gaulton KJ, Preissl S, Ren B. Interpreting non-coding sisease-associated human variants using single-cell epigenomics. *Nat Rev Genet* 2023; 24 (8):516–534.

Fabo T, Khavari P. Functional characterization of human genomic variation linked to polygenic diseases. *Trends Genet* 2023; 39 (6):462–490.

Palasantzas Vejm, Tamargo-Rubio I, Le K et al. iPSC-derived organ-on-a-chip models for personalized human genetics and pharmacogenomics studies. *Trends Genet* 2023; 39 (4):268–284.

https://www.england.nhs.uk/publication/national-genomic-test-directories/ *Genetic testing.*

https://www.ebi.ac.uk/gwas/ *A catalogue of published genome-wide association studies.*

http://www.ncbi.nlm.nih.gov/omim *Autosomal single-gene disorders.*

TECHNIQUES FOR STUDYING THE GENETIC CODE

Polymerase chain reaction

Polymerase chain reaction (PCR) is a core technique in genetics, which has revolutionized genetic research because small regions of interest in a sample of DNA—for example, the *BRCA1* gene—can be amplified over a million times within a few hours. The DNA is amplified between two short (generally 17–25 bases) single-stranded DNA fragments ('oligonucleotide primers') that are complementary to the sequences on different strands at each end of the DNA of interest (**Fig. 5.15**). These short lengths of amplified DNA can then be sequenced to obtain identification of the exact nucleotide sequence using Sanger sequencing (named after its inventor).

Sanger sequencing

This involves a chemical process known as dideoxy-sequencing. Once the sequence of interest is amplified, this is used as a template for the Sanger sequencing reaction. An oligonucleotide primer is designed to anneal adjacent to the region of interest. This

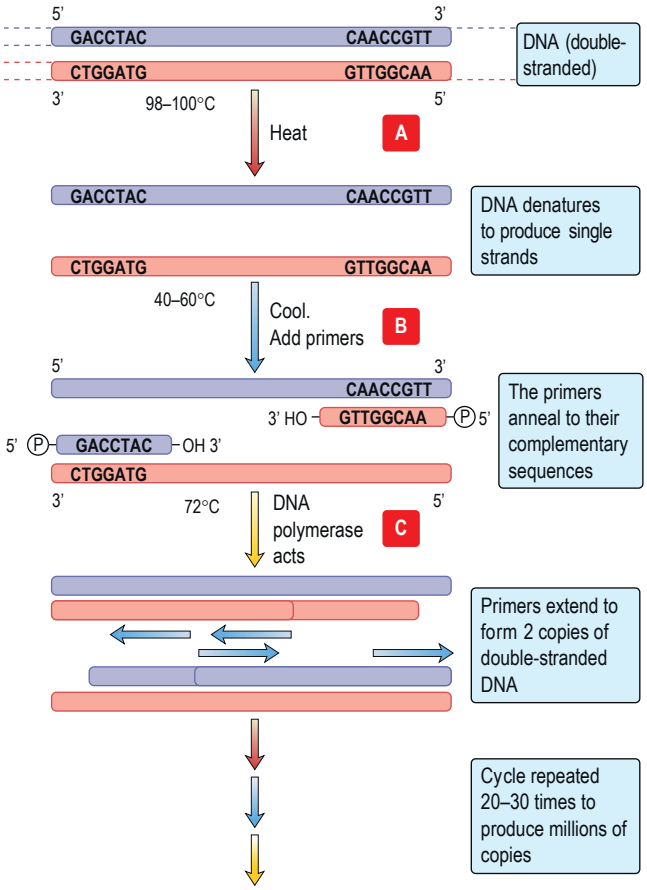


Fig. 5.15 Polymerase chain reaction. The technique is based on thermal cycling and has three basic steps. **(A)** The double-stranded genomic DNA is heat-denatured into single-stranded DNA. **(B)** The sample is cooled to favour annealing of the primers to their target DNA. **(C)** A thermostable DNA polymerase extends the primers over the target DNA. After one cycle, there are two copies of double-stranded DNA, after two cycles there are four copies, and so on. *DNA,* Deoxyribonucleic acid; *PRC,* polymerase chain reaction.

primer acts as the starting point for a DNA polymerase to build a new DNA chain that is complementary to the sequence under investigation. Chain extension is prematurely interrupted when a fluorescently labelled dideoxynucleotide becomes incorporated (this happens because it lacks the necessary 3′-hydroxyl group needed for extension). Each base dideoxynucleotide (G, C, T, A) has a different fluorochrome attached, and thus each termination base can be identified by its fluorescent colour. As each strand can be separated efficiently by capillary electrophoresis according to its size or length, simply monitoring the fluorescence as the reaction products elute from the capillary tube will give the gene sequence (Fig. 5.16). This is a widely used but relatively low-throughput method, which is increasingly being replaced with high-throughput methods.

Next-generation sequencing

Sequencing technology has developed such that it is now possible to sequence DNA molecules in a high-throughput manner ('next-generation sequencing' [NGS]), and new, faster, and cheaper methods are being developed. It is possible to sequence all the coding genes in an individual's human genome (termed the 'exome') to catalogue all possible disease-associated and non-disease-associated variants in every human gene at a low cost. If necessary, whole genomes (including non-coding regions) can also be sequenced in this way, although this involves a greater cost. The sequences are typically generated as short runs of sequences between 75 and 200 bp long, known as reads, which are aligned to the reference human genome (Fig. 5.17).

Differences from the reference genome are called 'variants' and are catalogued for each sample. The computational requirements for processing data from NGS experiments are significant and are usually performed on dedicated computer clusters. Whole-genome sequencing has the advantage over whole-exome sequencing of being able to detect intronic changes, structural variants, and changes in DNA copy number.

Transcriptomics

Along with sequencing genomic DNA, NGS technology can be used to sequence RNA (termed 'RNAseq') to assess gene expression levels accurately, in addition to determining all splice variations and allelic copy number. Analysis of all the RNA transcripts in a sample, or 'transcriptomics', allows for the comprehensive study of disease states. A limitation of this approach is that there may be a difference between levels of RNA transcripts and corresponding proteins that are encoded, warranting validation using additional techniques. It can be useful in some cases to look at affected tissue for non-coding variants such as splice site mutations. Note that, in addition to carrying out such work on whole-tissue samples, such as liver or skin, it is now possible to sequence single cells. This represents an opportunity to investigate disease states at unprecedented resolution; the Human Cell Atlas is one such project that aims to capture representative transcriptomes of human cells from all tissues. It is now possible to also study entire transcriptomes from sections of tissue, giving spatial context to transcriptomic data in health and disease.

Epigenetic techniques

NGS can also be used to study epigenetic changes. Changes in methylation can be studied using a technique that converts the unmethylated cytosines across the whole genome using a chemical process called bisulphite conversion, allowing for the absence and presence of methylation to be comprehensively detected. In addition to methylation, it is possible to determine the accessibility of chromatin using additional techniques such as assay for transposase-accessible chromatin sequencing (ATAC-seq).

Genome databases

Information arising from human genome sequencing is publicly available, providing biological information on every gene in the human genome. Information on any gene describing its protein product, function, tissue-specific expression, disease association and sequence variation can be easily obtained by searching and manipulating computer databases. In addition to genotype, it is necessary to consider databases that couple genotype with phenotype, such as DECIPHER. Some databases such as the UK BIOBANK have allowed extensive studies where genetic data can be analysed with clinical data, allowing for a range of research questions to be addressed. Phenotyping is difficult to do, although frameworks to aid this have been developed, such as the Human Phenotype Ontology (HPO).

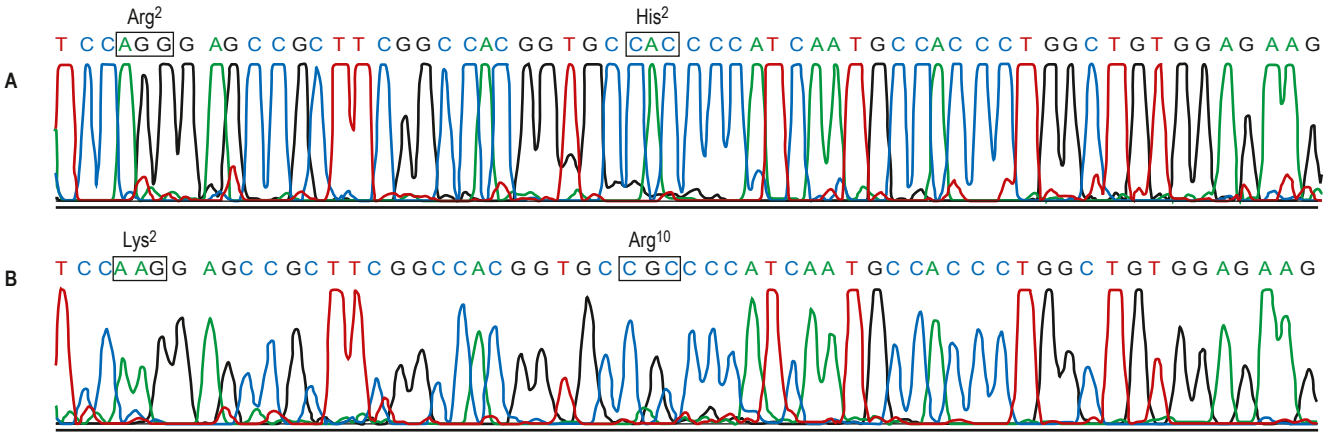


Fig. 5.16 Sanger sequencing. **(A)** Part of the normal luteinizing hormone beta chain gene. **(B)** Two polymorphic variants. The base changes and consequential alterations from arginine to lysine, and histidine to arginine amino acid residues are indicated.

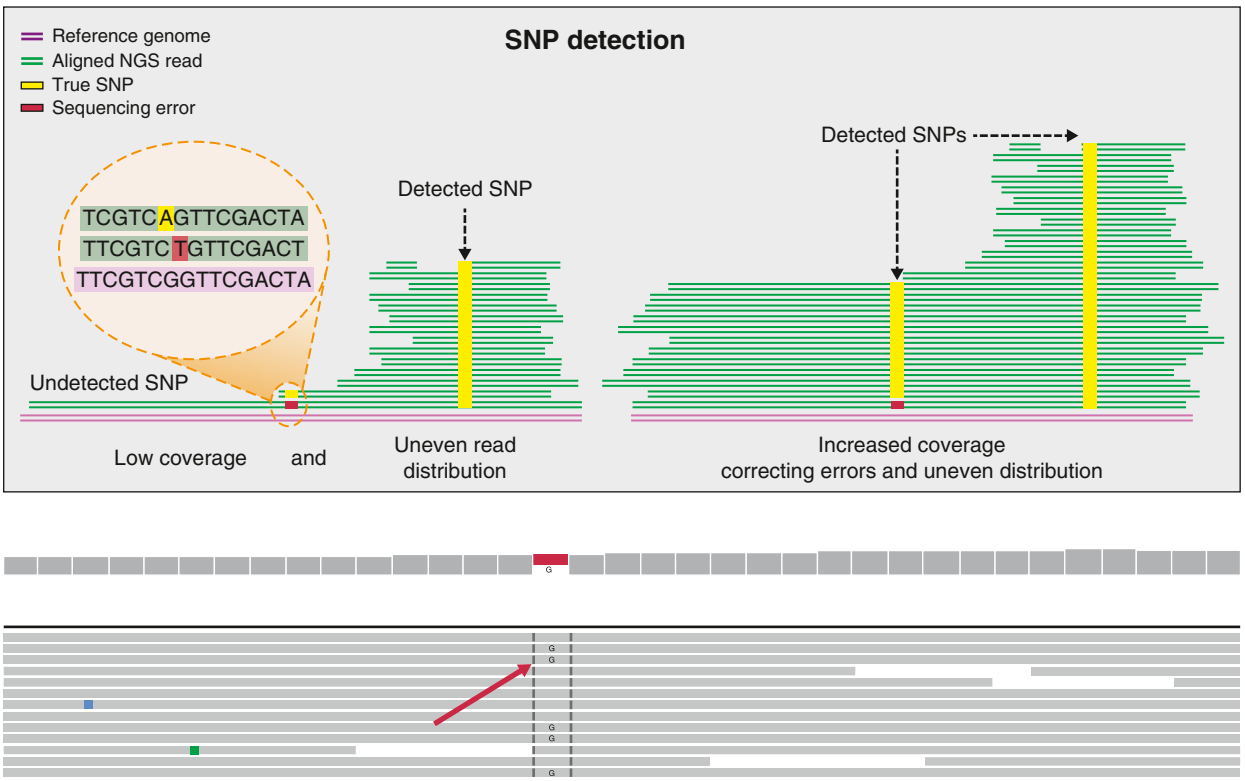


Fig. 5.17 Next-generation sequencing. The ability to detect genetic changes depends on good coverage of the region of interest, and the depth allows for true calls of variants over sequencing artefacts. *NGS,* Next-generation sequencing; *SNP,* single-nucleotide polymorphism.

Further reading

Wright CF, Campbell P, Eberhardt RY et al. Genomic diagnosis of rare pediatric disease in the United Kingdom and Ireland. *N Engl J Med* 2023; 388(17):1559–1571.
Feinberg AP. The key role of epigenetics in human disease prevention and mitigation. *N Engl J Med* 2018; 378:1323–1334.
Biesecker LG, Green RC. Diagnostic clinical genome and exome sequencing. *N Engl J Med* 2014; 370:2418–2425.

Research techniques

Discovery of disease-causing genes

Genetic polymorphisms and linkage studies

Linkage analysis has provided many breakthroughs in mapping the positions of genes that cause genetic diseases, such as the gene for cystic fibrosis, which was found to be tightly linked to a marker on chromosome 7.

These studies relied on finding markers that segregate with the disease in large families and are typically used in rare monogenic disease. Techniques that identify and quantitate genetic polymorphisms, such as SNPs (see earlier), microsatellites, and copy number variants (CNVs), are essential, as these serve as markers. Polymorphisms that are closer together are more likely to have alleles that move together in a block during meiosis than those further apart. This phenomenon is called 'linkage disequilibrium' and enables, for example, one SNP variant (tag SNP) in this block to act as a marker for the presence of other SNP variants.

Sequencing trios using whole-exome and whole-genome sequencing (WES/WGS)

The ability to study whole genomes is being used in work such as the 100,000 Genomes Project to discover new disease-causing genes. The aims of gene discovery include diagnosis, genetic counselling, and targeted screening. Typically, affected and unaffected individuals are studied, and variants that segregate with the disease are filtered. The DNA from both parents and the affected individual, known as a 'trio', is the minimum, and usually, the suspected gene in the affected individual will be found in other families with the same disease, strengthening the association. These advances have made it possible to find novel disease-causing genes in smaller pedigrees without needing to use the linkage approach.

Genome-wide association studies

Breakthroughs in complex polygenic disease are increasingly being achieved by studying large numbers (typically several thousands) of affected and unaffected patients. Many of these studies use SNP arrays, which provide informative data on hundreds of thousands of SNPs across the entire genome of an individual. Using these markers, statistical testing of association is applied, and changes that are linked with disease are typically plotted on a Manhattan plot (so called because the graph can look like the skyline of Manhattan; **Fig. 5.18**). This plot has genomic coordinates on the x-axis, and a measure of the statistical likelihood that the association detected has occurred by chance on the y-axis. SNPs that are likely to be associated with the disease have a low p-value, and as the y-axis values are plotted as negative logarithmic transformations of these p-values, a peak is displayed for such changes. These SNPs may not necessarily sit within the causative gene but may, for example, lie nearby, such as in the promoter of the gene affecting expression. GWAS can highlight regions of interest for polygenic diseases, which can then be studied in more detail, with the aim of gaining genetic insights into disease pathogenesis.

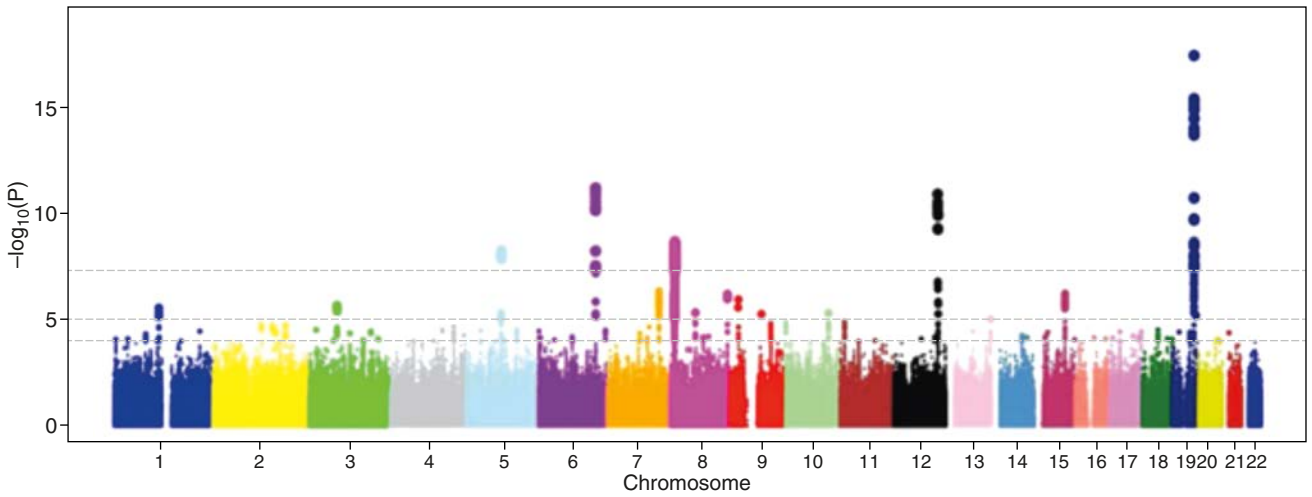


Fig. 5.18 A Manhattan plot of a genome-wide association study. SNPs at genetic loci that associate with the disease being studied are indicated; in this case, regions of interest lie in chromosomes 5, 6, 8, 12 and, most strongly, 19. Such regions can then be studied further. *SNPs,* Single-nucleotide polymorphisms.

Identification of gene function

Following the finding of genes of interest in either rare or common disease the challenge is to understand the function of these protein-coding genes. To gain insights into gene function the genetic changes found in disease can be modelled in cells or in animals. Most studies rely on the comparison of a cell's or an animal's phenotype in the presence or absence of the genetic change in question. Each approach has its respective merits and faults.

Cell culture models

DNA transfection

Human cells can be grown in cell culture flasks in the laboratory and their behaviour (growth rate, morphology, motility, gene expression profile, and biochemistry) characterized. A specific gene can then be introduced in a small plasmid (a circle of DNA from which the gene of interest can be expressed) or incorporated into a virus, and the change in cell behaviour can be assessed to provide an indication of gene function.

A recent advance is the use of genome editing, for example, CRISPR (clusters of regularly interspaced short palindromic repeats) gene editing to introduce the genetic change found in humans into cell lines (Fig. 5.19). This has the advantage of achieving physiological levels of gene expression over plasmids, where genes may be expressed at higher levels due to the presence of viral promoters in plasmid DNA. CRISPR can be used to introduce patient pathogenic variants into cells or to knock out genes completely to study their function.

RNA interference

Alternatively, if the cell line in question already expresses the gene of interest, its expression can be knocked down by RNA interference (RNAi). RNAi takes advantage of the cellular machinery that allows miRNAs encoded by the genome to regulate the expression of many genes at the level of mRNA stability and translation. This phenomenon has been exploited in the laboratory to study the function of a gene of interest or, on a much larger scale, the function of each gene in the genome. In such an RNAi screen an siRNA specific for each gene in the genome is introduced into cells grown *in vitro*, in effect knocking down expression of each gene in approximately 20 000 separate experiments. The phenotype of the cells in each experiment is then monitored to test the effect of loss of gene expression.

Animal models

The effect of a gene at the organism level can also be tested by misexpression/over-expression or knockout of a particular gene in a model animal. Nematode worms *(Caenorhabditis)*, fruit flies *(Drosophila)*, zebrafish, and rodents have all been genetically engineered to identify the function of a gene of interest. Knockout models of the higher organisms can be particularly helpful for medical research in providing a model of disease for the exploration of therapeutic intervention. Large-scale mutagenesis programmes are knocking out every gene or regulatory element in the mouse genome. However, note that the physiology of rodents and humans can differ.

GENOMIC MEDICINE

Gene therapy

A clinical aim of genetics has been to correct the disease-causing pathogenic variant and to treat the disease.

There are many technical problems to overcome in gene therapy, particularly in finding delivery systems to introduce DNA into a mammalian cell. Very careful control and supervision of gene manipulation is necessary because of its potential hazards and the attendant ethical issues.

Two major factors are involved in gene therapy:

- the introduction of the functional gene sequence into target cells
- the expression and permanent integration of the transfected gene into the host cell genome.

Haemophilia

These approaches have been gaining ground with conditions such as haemophilia, where recent trials of gene therapy have

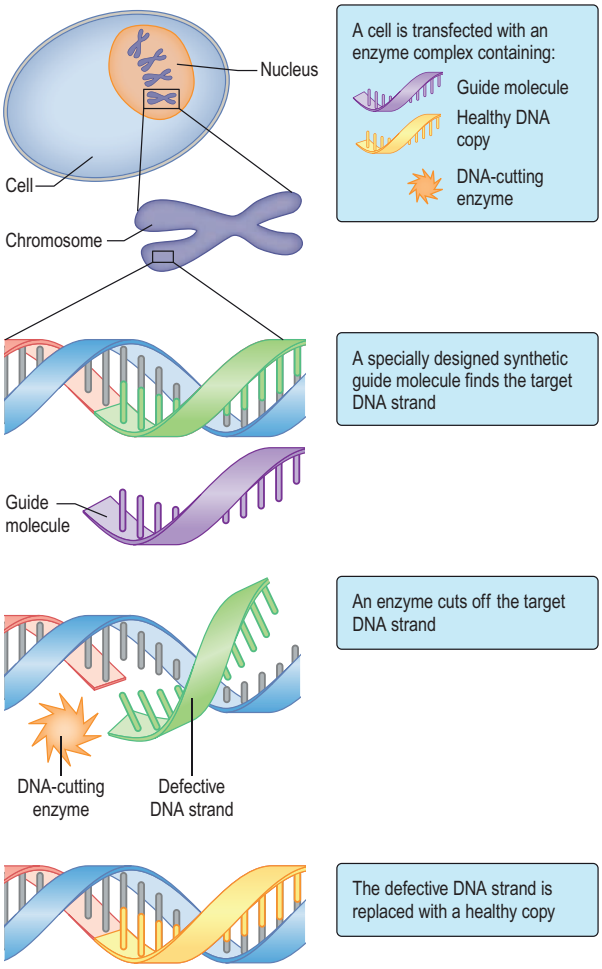


Fig. 5.19 Gene editing using technologies such as CRISPR (clusters of regularly interspaced short palindromic repeats).

meant that some patients no longer need transfusions to maintain normal clotting function. This approach involves using a virus to deliver a DNA construct with a normal sequence of factor VIII (haemophilia A) or factor IX (haemophilia B) to cells in the patient, administered as an infusion. Due to the design of this DNA construct, liver cells selectively express and produce functional clotting factors, with clinical improvement measured as reduction of bleeding episodes in the absence of regular transfusions of coagulation factors (Fig. 5.20).

Cystic fibrosis

Restoring function in diseases affecting solid organs rather than blood poses different challenges. In cystic fibrosis (CF), rather than trying to restore the altered sequence of the gene in every affected cell, a different approach is used. *CFTR*, the cystic fibrosis transmembrane regulator (see also Chapter 20), is an unusual ABC transporter in that it does not function as a primary active transporter but as a ligand-gated chloride channel (Fig. 5.21). The common CF pathogenic variant is a 3 bp deletion in exon 10 that results in the removal of a codon specifying phenylalanine (F508del). In this pathogenic variant the CFTR protein is misfolded, thereby causing ineffective biosynthesis and consequently disrupting the delivery of the protein to the cell surface. This pathogenic variant has recently been targeted with a combination of therapies (lumacaftor–ivacaftor) in phase 3 clinical trials and is now approved for clinical use. Lumacaftor brings more mutant CF channel proteins to the cell surface, and ivacaftor aids by prolonging the opening of the chloride ion channel, which restores partial function to the mutated protein. Early clinical results are encouraging, with patients demonstrating increased pulmonary function as measured by improved forced expiratory volume. More recent trials are being carried out to determine whether early intervention in childhood may modify the course of the disease.

Mitochondrial replacement therapy

In mitochondrial diseases arising due to pathogenic variants in the mitochondrial DNA, it is possible to avoid transmission of these maternally transmitted mitochondria using a technique known as pro-nuclear transfer. This involves transferring the nucleus from a fertilized cell from an affected patient with mitochondrial disease to a fertilized cell from a healthy donor, from which the nucleus has been removed. By doing this the developing embryo will have mitochondria only from the healthy donor while retaining nuclear DNA from the biological parents. In the UK, this process is underway in a small number of families (Fig. 5.22).

Stem cell therapy

Stem cell therapy has the potential to change the treatment of human disease radically. A number of adult stem cell therapies already exist, particularly bone marrow transplants for immune deficiencies. In this context a patient with a genetic immune deficiency receives a transplant from a healthy donor, and these normally functioning cells confer immunity. It is currently anticipated that technologies derived from stem cell research can be used to treat a wider variety of diseases in which replacement of destroyed specialist tissues is required, such as in Parkinson's disease, spinal cord injuries, and muscle damage.

In the era of gene editing the possibility of restoring genetically corrected cells has become a reality in some genetic blistering skin disorders. A male child with junctional epidermolysis bullosa developed blisters, with epidermal loss affecting 80% of his total body surface area. Skin cells were taken from this patient and transduced with a DNA construct that resulted in the production of the necessary protein, laminin 332. These corrected stem cells were grown in the laboratory before being transplanted back onto the patient's skin. Therapy resulted in complete coverage of the patient's skin, with genetic testing of the skin confirming the presence of the corrected cells.

These examples are indicative of the potential for using genetic information for clinical benefit and also of the length of time it takes to achieve meaningful outcomes.

Pre-implantation genetic diagnosis

Genetic diagnosis in a family can also be used to prevent transmission of a disease. This process requires knowledge of the genetic pathogenic variant that runs in the family. Early stage 8 cell embryos can have a cell removed and continue to form into a normal baby (Fig. 5.23A). It is possible to take such a cell and use it to genotype an embryo. Only embryos that do not carry the pathogenic variant are taken forward for implantation (Fig. 5.23B). This approach has been used to prevent transmission of cancer genes such as *BRCA*.

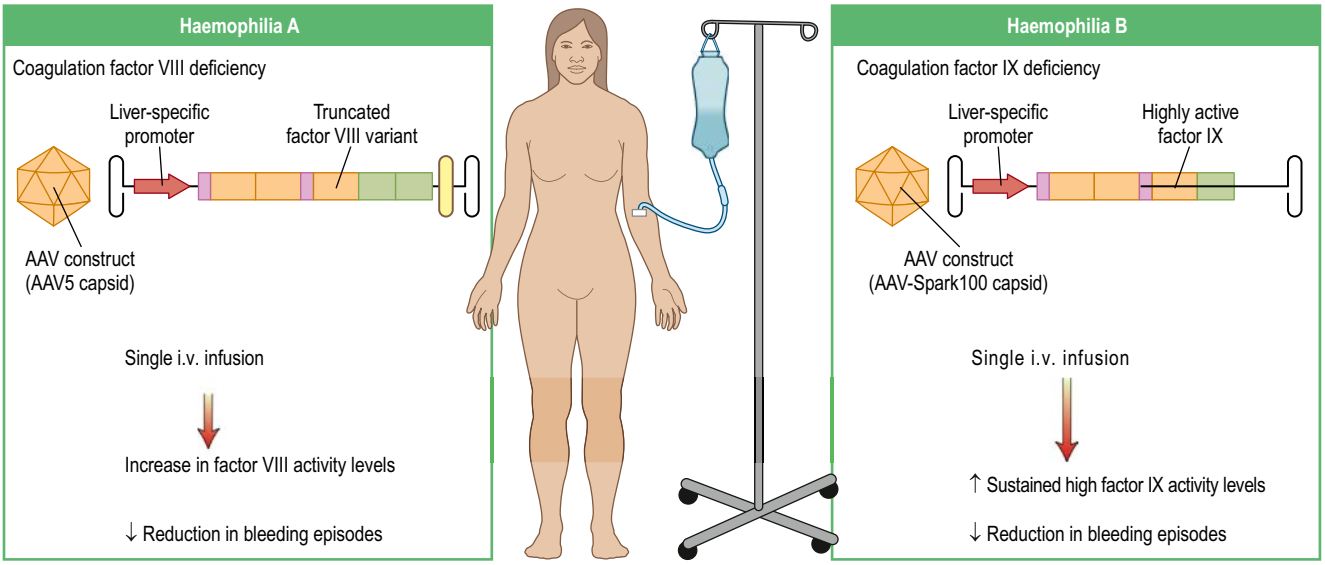


Fig. 5.20 Haemophilia targeted with gene therapy. Virally delivered DNA constructs with liver-specific promoters restore hepatic production of functioning coagulation factors in patients with haemophilia. Trial results shown are from Rangarajan S, Walsh L, Lester W et al. AA5-Factor VIII gene transfer in severe haemophilia A. *N Engl J Med* 2017; 317(26):2519–2530; George LA, Sullivan SK, Giermasz A et al. Hemophilia B gene therapy with a high-specific-activity Factor IX variant. *N Engl J Med* 2017; 377(23):2215–2227. *AAV,* Adeno-associated virus.

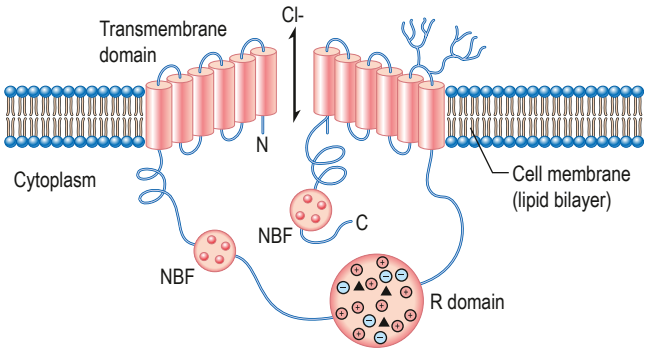


Fig. 5.21 Model of cystic fibrosis transmembrane regulator. This is an integral membrane glycoprotein, consisting of two repeated elements. The cylindrical structures represent six membrane-spanning helices in each half of the molecule. The ***NBFs*** are in the cytoplasm. The ***R*** domain links the two halves and contains charged individual amino acids and protein kinase phosphorylation sites *(black triangles)*. ***N*** and ***C*** are the amino and carboxy termini of the protein, respectively. The branched structure on the top right represents potential glycosylation sites. *CFTR,* Cystic fibrosis transmembrane regulator; *NBFs,* nucleotide-binding folds; *R,* regulatory.

Further reading

Lue NZ, Liau BB. Base editor screens for in situ mutational scanning at scale. *Mol Cell* 2023; 83(13):2167–2187.
De Rosa L, Enzo E, Palamenghi M, Sercia L, De Luca M. Stairways to advanced therapies for epidermolysis bullosa. *Cold Spring Harb Perspect Biol* 2023; 15(4).
Charo RA. Rogues and regulation of germline editing. *N Engl J Med* 2019; 380:976–980.
Hirsch T, Rothoeft T, Teig N et al. Regeneration of the entire human epidermis using transgenic stem cells. *Nature* 2017; 551:327–332.
Hyslop LA, Blakeley P, Craven L et al. Towards clinical application of pronuclear transfer to prevent mitochondrial DNA disease. *Nature* 2016; 534:383–386.
Keating D, Marigowda G, Burr L et al. VX-445–tezacaftor–ivacaftor in patients with cystic fibrosis and one or two Phe508del alleles. *N Engl J Med* 2018; 379:1612–1620.
Pickar AK, Gersbach CA. Gene therapies for hemophilia hit the mark in clinical trials. *Nature* 2018; 24:121–122.
Porteus MH. A new class of medicines through DNA editing. *N Engl J Med* 2019; 380:947–959.

THE GENETIC BASIS OF CANCER

The preceding sections have given an outline of how the genetic code gives rise to human health and disease. Cancers are genetic diseases and involve changes to the genetic code that result in abnormal function of cellular genes. The Cancer Genome Project set out to map the variety of pathogenic variants seen across human cancers, and these changes are detailed in online repositories (the Catalogue of Somatic Mutations in Cancer, COSMIC). Multiple gene pathogenic variants interact during oncogenesis and a progression of genetic defects gives a cancer cell a survival advantage over surrounding normal cells, famously termed the 'hallmarks of cancer' (see Fig. 41.2). These hallmarks include processes like increased proliferation, avoidance of control mechanisms such as apoptosis (programmed cell death) and the ability to evade immune detection. In the vast majority of cancer cases (especially those arising later in life) the multiple genetic changes that occur are somatic; that is, they occur *de novo* and are not inherited. In the next section the impact of inheriting pathogenic variants in cancer predisposition genes, such as early-onset and multiple cancers, is exemplified.

Inherited cancer syndromes

In some cancers (normally those that occur at an earlier age) an inherited single-gene defect can give rise to an almost Mendelian trend with lifetime risks of nearly 90%.

Autosomal dominant inheritance of cancer syndromes

The development of cancers at an earlier age and the occurrence of multiple tumours in these individuals highlight how genetic pathogenic variants facilitate the development of cancer. The following are examples of cancer syndromes that exhibit dominant inheritance:

- ***Retinoblastoma.*** This eye tumour is found in young children. It occurs in both hereditary (40%) and non-hereditary (60%)

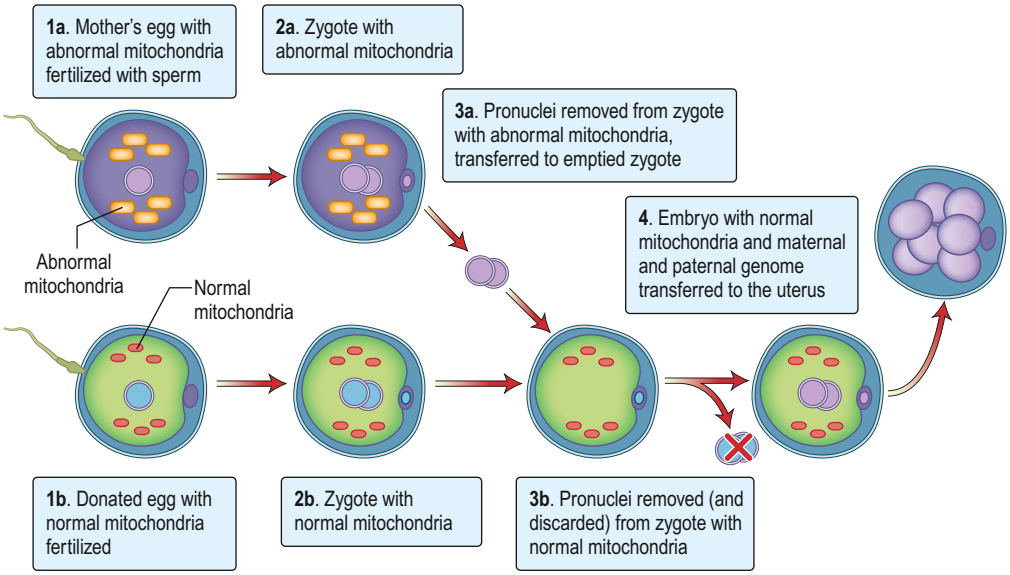


Fig. 5.22 Mitochondrial replacement therapy. Parents with transmissible mitochondrial disease provide fertilized eggs that carry abnormal mitochondria (1a, 2a). A healthy donor produces a fertilized egg, with normal mitochondria (1b, 2b). The nucleus from the affected parents' egg is transferred to the donor egg (3) and this is then transferred to the uterus to develop (4).

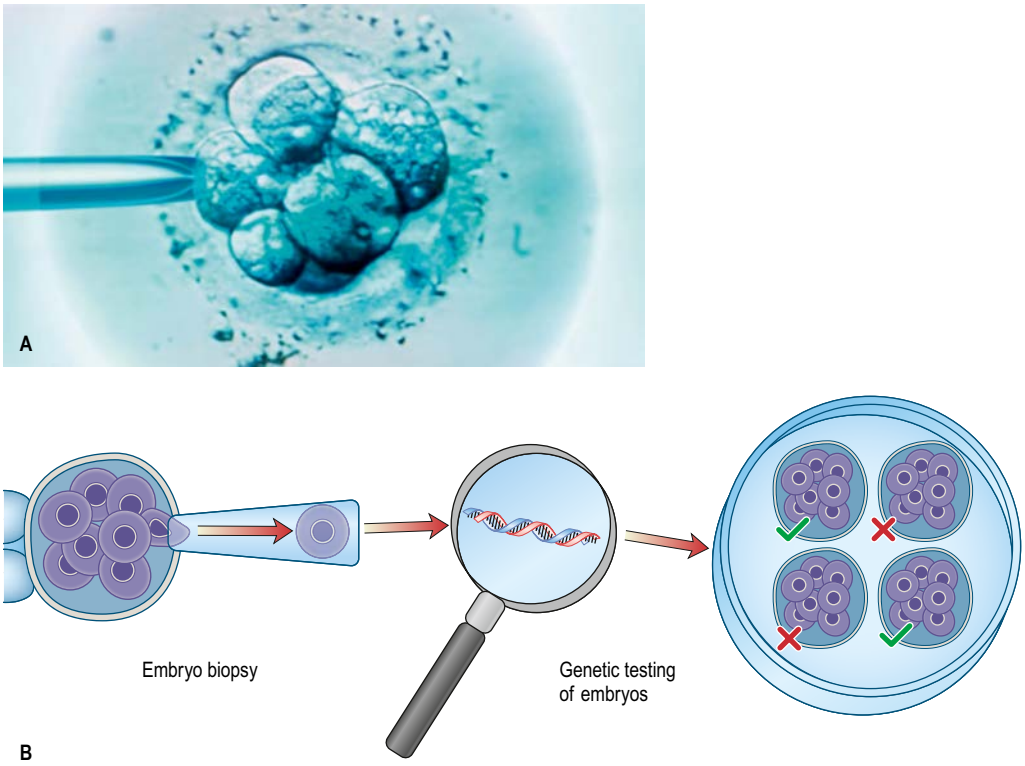


Fig. 5.23 Pre-implantation genetic diagnosis. (A) An early-stage blastocyst being sampled for genetic testing. **(B)** Each cell is tested, and only embryos without the screened pathogenic variant are selected for implantation.

forms. In the hereditary form, there is a germline pathogenic variant in the retinoblastoma gene *(RB1)* and people are also at risk of developing other tumours, particularly osteosarcoma. Knudson studied these patients, and this led to the 'two-hit hypothesis' (see later).

- ***Breast and ovarian cancer.*** Two major genes have been identified: *BRCA1* and *BRCA2*. A strong family history along with germline pathogenic variant of these genes accounts for most cases of familial breast cancer and over half of familial ovarian cancers. BRCA1 and BRCA2 proteins bind to the DNA repair

enzyme Rad51 to make it functional in repairing DNA breaks. Pathogenic variants in the *BRCA* genes will lead to the accumulation of unrepaired mutations in tumour suppressor genes and crucial oncogenes. Treatments that block the remaining repair mechanism of double-stranded breaks in these cells, a process called 'non-homologous end joining', have been used to treat *BRCA*-deficient cancers. This paradoxically increases the mutational load in these cells, resulting in cell death. The process is described as synthetic lethality and, in the case of *BRCA*-deficient tumours, is achieved by inhibitors of an enzyme called poly-ADP ribose polymerase (PARP).

- ***Neurofibromatosis.*** Inactivation of the *NF1* gene will lead to constitutive activation of Ras proteins and the development of multiple tumours in the nervous system (see Chapter 28).
- ***Lynch syndrome*** (see Chapter 22). Patients with defects in a group of genes that encode proteins involved in DNA mismatch repair are at increased risk of colorectal cancer. This syndrome has recently been highlighted as one instance where chemoprevention with aspirin can reduce the development of colorectal cancer by as much as 63%.

Autosomal recessive inheritance of cancer syndromes

Some relatively rare autosomal recessive diseases associated with abnormalities of DNA repair predispose to the development of cancer:

- ***Xeroderma pigmentosum.*** There is an inability to repair DNA damage caused by ultraviolet light and by some chemicals, leading to a high incidence of skin cancer.
- ***Ataxia telangiectasia.*** Increased sensitivity to ionizing radiation and a higher incidence of lymphoid tumours is seen.
- ***Bloom's syndrome and Fanconi's anaemia.*** An increased susceptibility to lymphoid malignancy is seen.

The majority of individuals who develop cancer do not have a genetic predisposition. In these people, somatic changes in DNA drive cancer. Studies of cancer where such changes are amplified have led to the discovery of selected genes that repeatedly cause cancer when mutated. These include oncogenes and tumour suppressor genes.

Oncogenes

The genes coding for growth factors, growth factor receptors, secondary messengers or even DNA-binding proteins would act as promoters of abnormal cell growth if mutated. This concept was verified when viruses were found to carry genes that promoted oncogenesis when integrated into the host cell. These were originally termed viral or 'v-oncogenes'; later, their normal cellular counterparts, c-oncogenes, were found. Thus oncogenes encode proteins that are known to participate in the regulation of normal cellular proliferation; for example, *erb-A* on chromosome 17q11–q12 encodes for the thyroid hormone receptor.

Activation of oncogenes

Non-activated oncogenes, which are functioning normally, have been referred to as 'proto-oncogenes' (Box 5.8). Their transformation to oncogenes can occur by mutation. A range of patterns of genetic mutations have been shown to cause cancer, and additional mechanisms such as 'chromothripsis', or chromosomal shattering, have been recently described.

Mutation

Carcinogens, such as those found in cigarette smoke, ionizing radiation and ultraviolet light, can cause point mutations in genomic DNA encoding tumour suppressor genes or oncogenes.

Box 5.8 Examples of acquired/somatic mutations and proto-oncogenes

Point mutation
- *K-RAS:* colorectal and pancreatic cancer
- *B-RAF:* melanoma, thyroid cancer
- *ALK:* lung cancer

DNA amplification
- *MYC:* neuroblastoma
- *HER2*-neu: breast cancer

Chromosome translocation
- *BCR/ABL:* chronic myeloid leukaemia, acute lymphoblastic leukaemia
- *PML/RARA:* acute promyelocytic leukaemia
- *BCL2/IGH:* follicular lymphoma
- *IGH/CCND1:* mantle cell lymphoma
- *MYC/IgH:* Burkitt lymphoma

Chromosomal translocation

If an error occurs during cell division and two chromosomes translocate so that a portion swaps over, the translocation breakpoint may occur in the middle of two genes, resulting in a mutation. If this happens, then the end of one gene is translocated onto the beginning of another gene, giving rise to a 'fusion gene'. Therefore sequences of one part of the fusion gene are inappropriately expressed because they are under the control of the other part of the gene. An example of such a fusion gene (the Philadelphia chromosome) occurs in *chronic myeloid leukaemia* (Fig. 5.24).

Viral activation

There are different ways in which viruses can cause cancer. When viral RNA is transcribed by reverse transcriptase into viral cDNA and, in turn, is spliced into the cellular DNA, the viral DNA may integrate within an oncogene and activate it. Alternatively, the virus may pick up cellular oncogene DNA and incorporate it into its own viral genome. Subsequent infection of another host cell might result in expression of this viral oncogene. For example, the Rous sarcoma virus of chickens was found to induce cancer because it carried the *Ras* oncogene.

Tumour suppressor genes

These genes restrict undue cell proliferation (in contrast to oncogenes) and induce the repair or self-destruction (apoptosis) of cells containing damaged DNA. Therefore mutations in these genes, which disable their function, lead to uncontrolled cell growth in cells with active oncogenes.

The ***RB gene*** was the first tumour suppressor gene to be described. In inherited retinoblastoma the first pathogenic variant in the *RB* gene is inherited; by chance, a second somatic mutation in the remaining normal copy of the *RB* gene then occurs, with the formation of a tumour. In the sporadic variety, mutations occur in both of the *RB* genes in a single cell. Knudson described this phenomenon as the 'two-hit' hypothesis, which accounts for the earlier and bilateral presentation of tumours in patients with germline *RB* pathogenic variants. *RB* is involved in the normal regulation of the cell cycle, and loss of control of the cell cycle in *RB*-deficient retinal cells contributes to tumour formation.

Since the finding of *RB*, other tumour suppressor genes have been described, including the gene ***p53***, known as the 'guardian of the genome'. Rarely, patients are born with an inherited change in *p53*, giving rise to Li–Fraumeni syndrome. These patients have an increased predisposition to a range of types of cancer, including

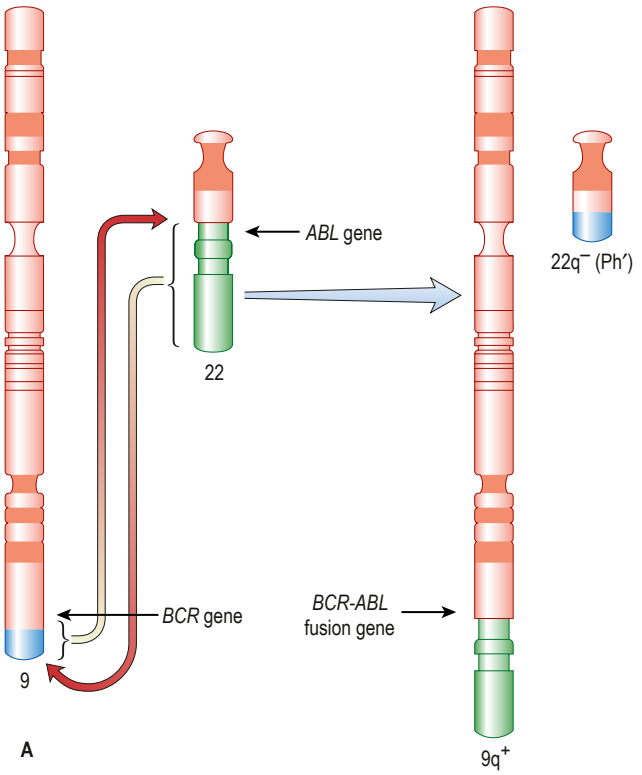


Fig. 5.24 The Philadelphia chromosome (Ph). **(A)** The long arm (q) of chromosome 22 has been shortened by a reciprocal translocation with chromosome 9. **(B)** The Philadelphia chromosome is formed by a reciprocal translocation of part of the long arm (q) of chromosome 22 to chromosome 9. It is seen in 90%–95% of patients with chronic myeloid leukaemia. The karyotype is expressed as 46XX, (9;22)(q34;q11).

breast cancer, soft tissue sarcomas, brain tumours, and leukaemia. Pathogenic variants in *p53* have also been found in almost all sporadic human cancers, suggesting that somatic mutations in *p53* are driver mutations. The protein encoded by *p53* plays a role in DNA repair and synthesis, in the control of the cell cycle, cell differentiation, and apoptosis. Four p53 protein molecules assemble into a tetramer that is functionally active.

p53 is a DNA-binding protein that activates many gene expression pathways, but it is normally only short lived. In some cancers, there is loss of *p53* from both chromosomes, accounting for loss of normal function of *p53* and ensuing tumour formation. In many tumours, mutations that disable *p53* function also prevent its cellular catabolism. Thus even a mutation in a single copy of the gene can promote tumour formation because a heterotetramer of mutated and normal *p53* subunits would still be dysfunctional.

Many other genes have been identified that function as tumour suppressor genes, including those found in Lynch syndrome, which are responsible for repairing DNA mismatches (see Chapter 22).

Genetic diagnostics for cancer

The ability to detect cancer-specific DNA sequences has led to the development of monitoring tests that can detect levels of circulating cell-free DNA in the blood. In addition, cancer cells can circulate in the bloodstream, and tests are being developed to detect these and monitor response to therapy.

Targeted treatments for cancer

The ability to sequence cancers rapidly has supported the development of targeted treatments. These aim to be selective and ideally have few side-effects compared to conventional chemotherapy. Tyrosine kinase inhibitors, such as imatinib, are examples of targeted treatments in selected haematological malignancies that carry the Philadelphia chromosome and even some solid cancers such as gastrointestinal stromal tumours. A shift heralded by cancer sequencing is the concept of orphan molecular entities. Rather than treating cancer of a tissue such as breast or lung, sequencing allows the detection of drivers in a cancer, for which a selected cocktail of treatments may apply. This may lead to the increased personalization of cancer therapy. Resistance to such treatments is well recognized, however, as cancers mutate further to become resistant to these drugs. An example is seen in metastatic melanoma, where drugs that target mutant *BRAF* in melanoma cells offer temporary but non-sustained remission for patients.

Precision (or personalized) medicine

There is a huge shift in healthcare research and delivery towards developing strategies that treat the individual patient using a combination of genetics and metabolic biomarkers with lifestyle and environmental modifications. These studies are driven by the observations that not all medical treatments including drugs will work the same or even have efficacy in every patient with a particular condition. The goal is to deliver the right drug to each patient together with appropriate lifestyle changes such as diet and exercise. Towards this ambition of 'personalized' medicine, large-scale population-based studies with ethnic diversity are being conducted to identify new drug targets and biomarkers to stratify patients into subgroups for specific disease management strategies and also to design more robust clinical trials with more defined patient recruitment. This will lead to improved individual patient diagnosis, treatment, and even disease prevention. Examples of these approaches are being employed in inflammatory conditions like psoriasis and rheumatoid arthritis and for targeted cancer therapy such as in mutant *BRAF*-positive melanoma and in mutant *BRCA*-breast cancer with the use of PARP inhibitors. In parallel, and linked to these approaches for precision medicine, there are new diagnostic tools being developed and implemented in clinical practice. These include wearable devices, for example, for monitoring heart rhythm and blood pressure, as well as the implementation of artificial intelligence (AI), for

example, in radiology and pathology. In cancer, there is also the use of liquid biopsy from blood or urine to detect cancer-associated cells, DNA, or other biomarkers to monitor cancer and metastasis.

Further reading

Xu Y, Ritchie SC, Liang Y et al. An atlas of genetic scores to predict multi-omic traits. *Nature* 2023; 616(7955):123–131.
Weeden CE, Hill W, Lim EL, Gronroos E, Swanton C. Impact of risk factors on early cancer evolution. *Cell* 2023; 186(8):1541–1563.
Burn J, Sheth H, Elliott F et al. Cancer prevention with aspirin in hereditary colorectal cancer (Lynch Syndrome), 10-year follow-up and registry-based 20-year data in the capp2 study: A double-blind, randomised, placebo-controlled trial. *Lancet* 2020; 395(10240):1855–1863.
Hanahan D. Hallmarks of cancer: new dimensions. *Cancer Discov* 2022; 12(1):31–46.
Knudson AG. Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci USA* 1971; 68:820–823.

Bibliography

Lodish H, Berk A, Kaiser CA et al. *Molecular Cell Biology*, 9th edn. Macmillan Learning; 2021.

Significant websites

http://genome.ucsc.edu *University of California at Santa Cruz Genome Bioinformatics.*
http://www.ensembl.org/index.html *Ensembl.*
http://www.ncbi.nlm.nih.gov *National Center for Biotechnology Information.*
http://www.ncbi.nlm.nih.gov/omim *McKusick's Online Mendelian Inheritance in Man (OMIM), for information on gene products and their disease association.*
https://cancer.sanger.ac.uk/cosmic *Catalogue of somatic mutations in cancer.*
https://www.deciphergenomics.org/ *DECIPHER.*

Questions for this chapter can be found on Elsevier eBooks+.

6 Immunity

Matthew Buckland and James Emmanuel Thaventhiran

Introducing the tissues, cells, and molecules of the immune system 79
Cells involved in immune responses: origin and function 81
The immune system 81
Cytokines 81
Chemokines 81
Innate immune system 81
Complement 82
Neutrophils 83
Eosinophils 83
Mast cells and basophils 84
Monocytes and macrophages 84
Dendritic cells 84
Innate lymphoid cells 85
Adaptive immune system 85
Antigen receptors on T and B lymphocytes 86
T lymphocyte development and activation 87
Natural killer cells 90
Cell migration 91
HLA molecules and antigen presentation 92
The human major histocompatibility complex 92
Antigen presentation 93
The immune system in concert 95
Acute inflammation: events and symptoms 95
Chronic inflammation: events and symptoms 96
Laboratory investigations of the immune system 97
Disorders of immunodeficiency 97
Secondary (acquired) versus primary immunodeficiency 97
Primary immunodeficiency 98
Disorders of exaggerated immune response 101
Allergic disease (immediate hypersensitivity) 101
Autoimmune disease 102
Mechanisms of tissue damage in autoimmune disease 103
Common autoimmune diseases 104
Organ rejection in clinical transplantation 105
Immune-based therapies 105
Monoclonal antibody therapy (targeted therapy) 105
Chimeric antigen receptor T-cell therapy 106
Immunosuppressive drugs 106

CORE SKILLS AND KNOWLEDGE

Aspects of immunity have an impact on all areas of medicine, from diagnosis and treatment to diseases that arise when the immune system goes wrong. Allergic disorders affect at least 1 in 10 individuals and autoimmune diseases will affect 1 in 20 individuals at some point in their life. The number of people living with secondary immune deficiency is rising as survival of malignancy and disorders treated with immune-modulating therapy improves.

The role of a clinical immunologist is to manage immune-mediated diseases and provide specialist advice and guidance to primary care physicians and hospital practitioners on diagnostic testing, interpretation, and treatment of organ-specific disorders. Clinical immunologists run clinics for severe or complex allergy and inborn errors of immunity that can lead to severe immunodeficiency and manage patients with immunodeficiency secondary to immune-modulating therapies.

Key skills in clinical immunology include:

- Recognizing the possibility of an immunodeficiency syndrome in patients presenting with recurrent or severe infections such as meningitis or pneumonia, and understanding appropriate ways to investigate and manage this.
- Assessing patients with complex allergy, including food and drug allergy, and undertaking provocation testing and desensitization including immunotherapy when appropriate.

Opportunities to learn clinical immunology in general medical contexts include seeking to apply basic immunological principles to understand patients with autoimmune conditions or those receiving immunosuppressant medications. Attending specialist immunology and allergy clinics gives insight into how patients live with chronic disease, how parenteral therapies such as immunoglobulins are administered, and how immunologically mediated conditions, including urticaria, angioedema, allergic rhinitis, and anaphylaxis, can be managed.

INTRODUCING THE TISSUES, CELLS, AND MOLECULES OF THE IMMUNE SYSTEM

Immunity can be defined as protection from infection, whether this is bacterial, viral, fungal or due to multicellular parasites. The immune system can also distinguish ***altered self*** and prevents the development of malignancy by destroying cancerous cells. The immune system is composed of cells and molecules organized into specialized tissues and acts in concert to orchestrate an immune response (Fig. 6.1).

The ***primary lymphoid organs*** (thymus and bone marrow) are where the cells originate. Cells and molecules of the immune system circulate in the blood; immune responses do not take place there but are initiated at the site of infection (typically the mucosa or skin). Immunity is then propagated and refined in the ***secondary***

Non-immune host defence mechanisms

Normal barriers	Events that compromise barrier function
Physical barriers	
Cough reflex	Suppression, e.g., by opiates, neurological disease
Skin and mucous membranes	Trauma, burns, i.v. cannulae
Mucosal function	Ciliary paralysis (e.g., smoking) Increased mucus production (e.g., asthma) Abnormally viscid secretions (e.g., cystic fibrosis) Decreased secretions (e.g., sicca syndrome)
Urine flow	Stasis (e.g., prostatic hypertrophy)
Chemical barriers	High gastric pH (gastric acid secretion inhibitors)
Resistance to pathogens provided by commensal skin and gut organisms	Changes in flora (e.g., broad-spectrum antibiotics)

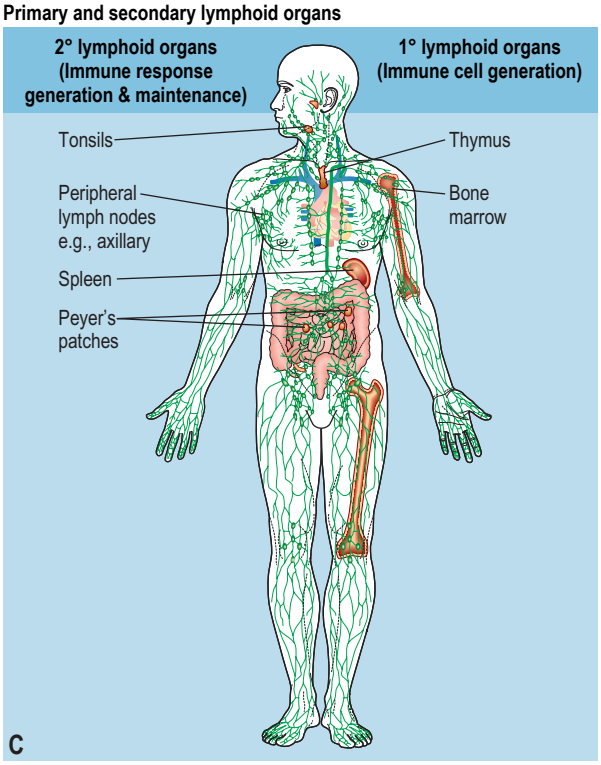
A

Cells involved in the immune response

Category	Cells	Main functions	Origin	Special features
Myeloid	Neutrophils	Immunity to bacteria and fungi	Bone marrow	Major first-line defence against pathogens
	Eosinophils, mast cells, and basophils	Immunity to parasites	Bone marrow	Role in allergy
	Monocytes and macrophages	Immunity to bacteria, fungi, parasites	Bone marrow	Specialized phagocytes; cytokine secretion
Lymphoid	Dendritic cells	Antigen presentation to T lymphocytes	Bone marrow	Key role in activating T lymphocytes
	B lymphocytes	Antibody production	Bone marrow	Have specific receptor for antigen (called 'antibody'). Mature into plasma cells that are 'antibody factories'
	T lymphocytes	Orchestration of immune response against bacteria, fungi, parasites, and viruses	Precursors come from bone marrow and undergo selection process in the thymus to avoid self-reactivity	Have specific receptor for antigen (called T-cell receptor). Two major subsets: CD4 ('helper' and 'regulatory') and CD8 ('cytotoxic')

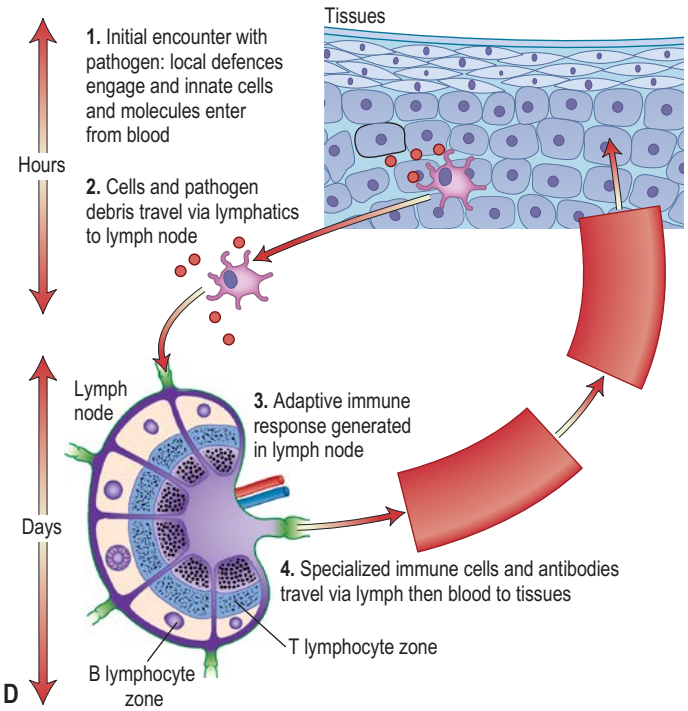
B

Primary and secondary lymphoid organs



C

Immune response: locations and dynamics



D

Fig. 6.1 Overview of the cells, molecules, and tissues of the immune system. The barrier to invasive infection includes both **(A)** non-immune and **(B)** cellular immune defences. **(C)** Lymphocytes are generated as precursors in the bone marrow and differentiate into T (thymus) or B (bone marrow) lymphocytes in the primary lymphoid tissue. Once differentiated, 98% of lymphocytes reside in the secondary lymphoid tissue, where the adaptive immune response takes place. **(D)** The immune response to infection involves a variety of local and regional responses: **(1)** There are rapid local responses that include ingress of cells (e.g., neutrophils) from the blood. **(2)** Immune cells and pathogen debris migrate via lymphatics to the local lymph node. **(3)** An adaptive immune response is generated in the regional lymph node. **(4)** Effector cells travel by lymph and then blood to the infected tissue.

lymphoid organs (e.g., lymph nodes). Following resolution of the infection, immunological memory specific for the pathogen is generated and resides in cells (lymphocytes) in the spleen and lymph nodes, as well as being widely secreted in a molecular form (antibodies). Secondary lymphoid organs include adenoids, tonsils, spleen, Peyer's patches in the small intestine, peripheral lymph nodes (cervical, axillary, inguinal), and the thoracic duct, which drains the lymph (containing immune cells) into the right subclavian vein.

Cells involved in immune responses: origin and function

All immune cells derive from pluripotent stem cells generated in the bone marrow. They have diverse functions (see **Fig. 6.1**). ***T lymphocytes*** undergo 'education' in the thymus to avoid self-recognition and populate the peripheral lymphoid tissue, where B lymphocytes also reside. Both sets of lymphocytes undergo activation in the peripheral tissue to become mature effector cells. ***B lymphocytes*** may further differentiate into antibody-secreting ***plasma cells***. Lymphoid tissue is frequently found at mucosal surfaces in non-encapsulated patches, termed ***mucosa-associated lymphoid tissue (MALT)***.

The immune system

Cells and molecules involved in immune responses are classified into innate and adaptive systems:

- The ***innate immune system*** is inborn and operates throughout life (see 'Innate immune system' section).
- The ***adaptive immune system*** changes in response to the pathogens it encounters (see 'Adaptive immune system' section).

There are also non-immunological barriers that are involved in host protection, and lowering of these may allow a pathogen to gain a foothold (see **Fig. 6.1**).

The immune system is immensely powerful, in terms of its ability to inflame, damage, and kill, and it has a capacity to recognize multiple molecular patterns in the microbial world. However, immune responses are not always beneficial. They can give rise to a range of autoimmune and inflammatory diseases, known as ***immunopathologies***. Conversely, the immune system may fail, giving rise to immune deficiency states. These conditions are grouped under the umbrella of clinical immunology.

A major feature of the immune system is the complexity of its surface-bound, intracellular, and soluble mediators. In particular, it is necessary to be aware of the clusters of differentiation classification (**Box 6.1**) and the functions of cytokines and chemokines. The clusters of differentiation classification is like a barcode for defining a cell.

Cytokines

Cytokines are small polypeptides released by a cell in order to change the function of the same or another cell. These chemical messengers are found in many organ systems but especially the immune system. Cytokines have become markers in the investigation of disease pathogenesis; they are also therapeutic agents in their own right, as well as the targets of therapeutic agents. The key features of a cytokine are:

- ***pleiotropy:*** has different effects on different cells
- ***autocrine function:*** modulates the cell secreting it
- ***paracrine function:*** modulates adjacent cells
- ***endocrine effects:*** modulates cells and organs at remote sites
- ***synergistic activity:*** acts in concert with other cytokines to achieve effects greater than the summation of their individual actions.

The main immune cytokines are the ***interferons*** (IFNs) and the ***interleukins*** (ILs). The IFNs are limited to a few major types (α, β, and γ), whereas there are 40 ILs.

i Box 6.1 The CD classification

- Immune cells are distinguished by the surface receptors and proteins that they express in order to mediate their particular range of immunological functions, for example cell–cell signalling, cell activation.
- The surfaces are covered with such proteins and indicate the cell lineage or differentiation pathway. The discovery of monoclonal antibodies (proteins tailor-made to bind to a specific target) made this feasible.
- Surface molecules defining the origin and function of selected groups of cells are known as clusters of differentiation (CD). Over 360 CD numbers exist.
- CD markers can be detected within histologic specimens of tissues by immunohistochemistry or by the use of flow cytometry.
- In clinical use CD markers are used to analyse peripheral blood lymphocytes by the use of:
  - CD3 (all T-cells)
  - CD4 (Helper T-cells)
  - CD8 (Cytotoxic T-cells)
  - CD19 (B-cells)
  - CD16/56 (NK-cells).
- These may be used to diagnose SCID ('Severe combined immunodeficiencies' section), monitor B-cell depletion therapy (see IgG4-RD, 'IgG4 disease' section), and response to HIV therapy (see Chapter 44).
- The same technique can be used to monitor or diagnose haematologic malignancy using different markers (Chapter 39).

An updated listing is available at: http://www.hcdm.org/.

Chemokines

The defining feature of chemokines is their function as chemotactic molecules: that is, they attract cells along a gradient of low to high chemical concentration, particularly from the blood into the tissues and from the tissues into lymphatics. They also have the ability to activate immune cells. All chemokines have a similar structure relating to the configuration of cysteine residues, which gives rise to four families (**Box 6.2**).

Cell-surface receptors for chemokines are denoted by 'R'. CCR5 is a co-receptor for the human immunodeficiency virus (HIV; see Chapter 39), and drugs to block this and other chemokine receptors are under active development to combat inflammation.

INNATE IMMUNE SYSTEM

Innate immunity provides immediate, first-line host defence. The key features of this system (as well as the adaptive system, 'Adaptive immune system' section) are shown in **Box 6.3**. It is present at birth and remains operative at comparable intensity into old age. Innate immunity is mediated by a variety of cells and molecules (**Box 6.4**). Activation of innate immune responses is mediated through interaction between:

- the ***pathogen side***, comprising a relatively limited array of molecules (pathogen-associated molecular patterns [PAMPs])
- the ***host side***, comprising a limited portfolio of receptors (pattern recognition receptors [PRRs]).

Activation of certain cells in the innate immune system leads, in turn, to activation of the ***adaptive immune response*** (see 'Adaptive immune system' section). The ***dendritic cell*** is especially involved in this process and forms a bridge between innate and adaptive systems.

i Box 6.2 Features of chemokine mediators and receptors

Chemokine family	Structure	Receptor family	Pathological or therapeutic target
CXC	Two cysteines (C) separated by any other amino acid residue (X)	CXCR	CXCR4—co-receptor for late HIV infection
CC	Two cysteines next to each other	CCR	CCR5—co-receptor for early HIV infection and blocked by 'entry inhibitors'
C	One cysteine	CR	XCR1—target of Kaposi's sarcoma human virus protein
CX3C	Two cysteines separated by any three amino acids	CX3CR	Possible survival factor in myeloma

R, Receptor.

i Box 6.3 Features of the innate and adaptive immune responses

Innate	Adaptive
No memory: quality and intensity of response invariant	Memory: response adapts with each exposure
Recognition of limited number of non-varying, generic molecular patterns on, or made by, pathogens	Recognition of vast array of specific antigens[a] on, or made by, pathogens
Pattern recognition mediated by a limited array of receptors	Antigen recognition mediated by a vast array of antigen-specific receptors
Response immediate on first encounter	Response on first encounter takes 1 to 2 weeks; on second encounter, 3 to 7 days

[a]Antigen is a molecular structure (protein, peptide, lipid, carbohydrate) that generates an immune response.

Complement

Complement proteins are produced in the liver and circulate in an inactive form. When triggered, complement molecules become enzymatically active and trigger several molecules of the next stage in a series. This ***complement cascade*** is initiated via three distinct pathways: alternative, classical, and lectin (Fig. 6.2). Each pathway culminates in the cleavage of C3 and C5. Cleavage of C3 has a number of biological consequences; breakdown of C5 achieves the same and, in addition, provides the triggering stimulus to the final common ('membrane attack') pathway, which provides most of the biological activity (see Fig. 6.2).

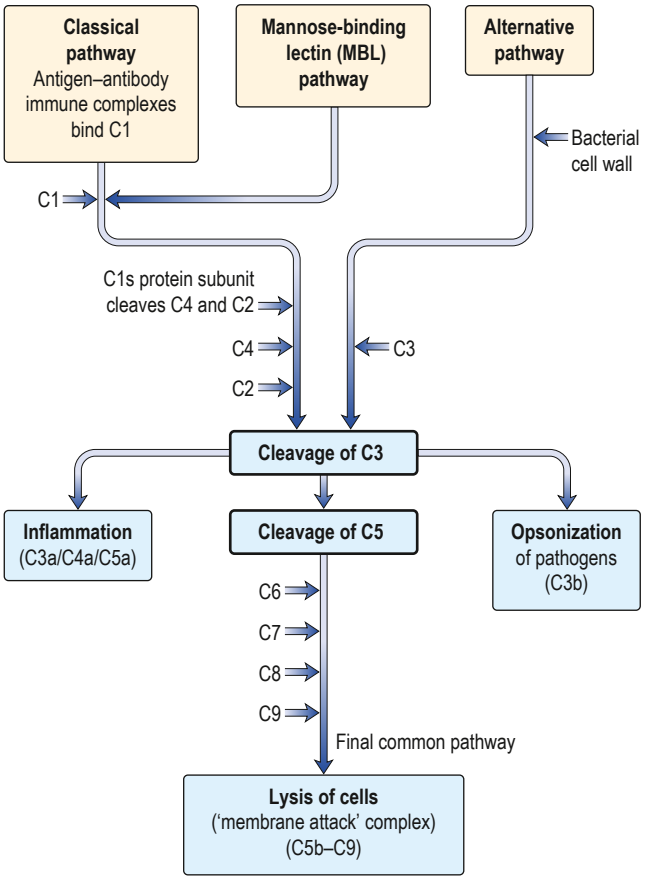


Fig. 6.2 The complement pathways and their effector functions.

i Box 6.4 Soluble components of innate immunity

Molecules	Name (examples)	Function	Provide immunity against
Complement	Cascade of >50 proteins	Lyse bacteria; opsonize[a] bacteria; promote inflammation; recruit and activate immune cells	Bacteria, viruses
Collectins	Mannose-binding lectin	Bind bacteria; activate complement	Bacteria
Pentraxins	C-reactive protein[b]	Opsonize bacteria	Bacteria
Enzymes	Lysozyme	Present in secretions; cleave bacterial cell wall	Bacteria

[a]Opsonize means to coat bacteria to enhance phagocytosis by granulocytes and monocytes/macrophages.
[b]C-reactive protein (CRP) is an acute phase protein. Blood level rises 10- to 100-fold within hours of the start of an infective or inflammatory process, making it extremely useful in monitoring infective or inflammatory diseases and their response to treatment.

The main functions of complement activation are to:

- promote inflammation (e.g., through the actions of the anaphylatoxins C3a, C4a, and C5a)
- recruit cells (e.g., through chemoattractants)
- kill targeted cells, such as bacteria
- solubilize antigen–antibody ('immune') complexes and remove them from the circulation.

During an immune response, removal of immune complexes protects unaffected tissues from the deposition of these large, insoluble composites, which could result in unwanted inflammation. Failure of this protective mechanism can result in 'immune complex' pathology, where immune complexes lodge in organs such as the joints, kidney, and eye. This can complicate infectious diseases such as infective endocarditis or result in some of the clinical manifestations of autoimmune disease such as systemic lupus erythematosus (SLE).

Neutrophils

Neutrophils phagocytose and kill microorganisms by releasing antimicrobial compounds (e.g., defensins). They are derived from the bone marrow, which can produce between 10^{11} (healthy state) and 10^{12} (during infection) new cells per day. In health, neutrophils are rarely seen in the tissues.

Neutrophil phagocytosis is activated by interaction with bacteria, either directly or after bacteria have been coated (opsonized) to make them more ingestible (Fig. 6.3). The contents of neutrophil granules are released both intracellularly (predominantly azurophilic granules) and extracellularly (specific granules) following fusion with the plasma membrane. Approximately 100 different molecules in neutrophil granules (Box 6.5) kill and digest microorganisms, for example:

- ***Myeloperoxidase and cytochrome b_{558}*** are key components of major oxygen-dependent bactericidal systems.
- ***Cathepsins, proteinase-3, and elastase*** are deadly to Gram-positive and Gram-negative organisms, as well as some *Candida* species.
- ***Defensins*** are naturally occurring cysteine-rich antibacterial and antifungal polypeptides (29–35 amino acids).
- ***Collagenase and elastase*** break down fibrous structures in the extracellular matrix, facilitating progress of the neutrophil through the tissues.

Granule release is initiated by the products of bacterial cell walls, certain complement proteins, leukotrienes (LTB_4) and chemokines (e.g., CXCL8), and cytokines such as tumour necrosis factor-alpha (TNF-α). A further defence mechanism is the formation of ***neutrophil extracellular traps*** from DNA, which bind pathogens to facilitate killing.

Eosinophils

Eosinophils release pro-inflammatory mediators to provide immunity against parasites. In contrast to neutrophils, several hundred times more eosinophils are present in the tissues than in the blood, particularly at epithelial surfaces, where they survive for several weeks. The main role of eosinophils is protection against multicellular parasites such as worms (helminths). This is achieved by the release of pro-inflammatory mediators, which are toxic, cationic proteins. In populations and societies in which such parasites are rare, eosinophils contribute mainly to allergic disease, particularly asthma (see Chapter 20). Eosinophils have two types of granules:

- ***Specific granules*** (95%) contain the cationic proteins, of which there are four main types: ***major basic protein, eosinophil***

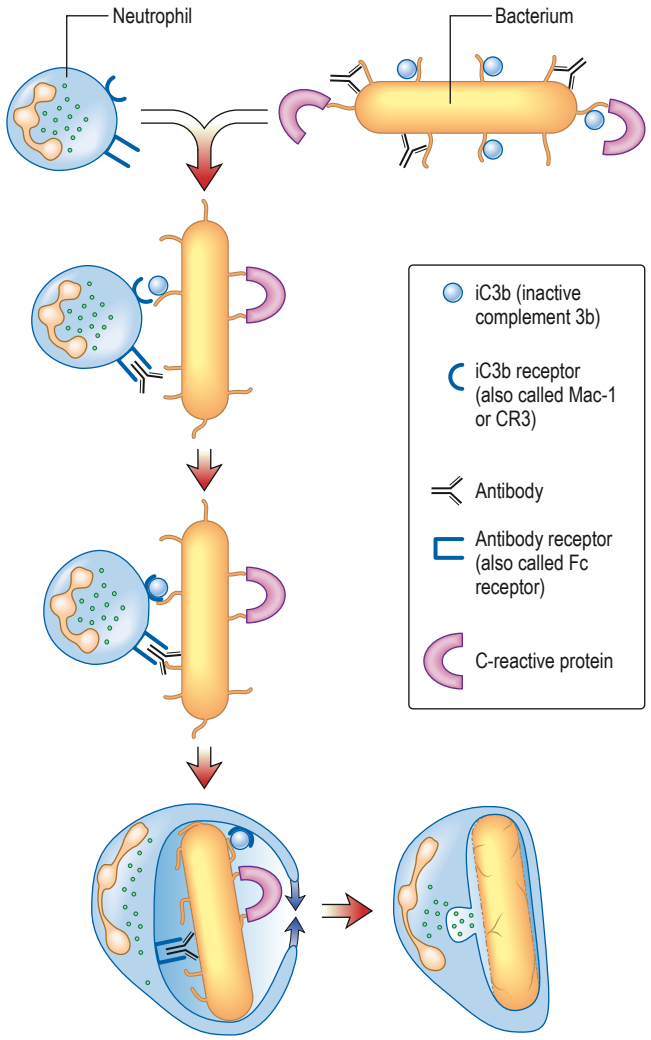


Fig. 6.3 Opsonization. Bacteria are coated with a variety of soluble factors from the innate immune system (opsonins), which enhance phagocytosis. This leads to engulfment of the bacteria into phagosomes and fusion with granules to release antibacterial agents.

Box 6.5 Contents and function of key neutrophil granules

Function	Primary or azurophilic granules	Secondary or specific granules
Antibacterial	Lysozyme Defensins Myeloperoxidase (MPO) Proteinase-3 Elastase Cathepsins Bactericidal/permeability-increasing protein (BPI)	Respiratory burst components (e.g., cytochrome b_{558}) producing reactive oxygen metabolites, such as hydrogen peroxide, hydroxyl radicals, and singlet oxygen Lysozyme Lactoferrin
Cell movement		Collagenase CD11b/CD18 (adhesion molecule) *N*-formyl-methionyl-leucylphenylalanine receptor (FMLP-R)

cationic protein, and ***eosinophil neurotoxin***, which are all potently toxic to helminths, and ***eosinophil peroxidase***, similar to neutrophil myeloperoxidase.

- ***Primary granules*** (5%) synthesize and release ***leukotrienes C_4 and D_4*** and ***platelet-activating factor***, which alter airway smooth muscle and vasculature.

Eosinophils are activated and recruited by a variety of mediators via specific surface receptors, including complement factors and LTB_4. In addition, the chemokines eotaxin-1 (CCL11) and eotaxin-2 (CCL24) are highly selective in eosinophil recruitment. There are surface receptors for the cytokines interleukin-3 (IL-3) and interleukin-5 (IL-5), which promote the development and differentiation of eosinophils.

Mast cells and basophils

Mast cells release pro-inflammatory and vasoactive mediators and have a role in allergy. Mast cells and ***basophils*** share features in common, especially in that they contain:

- histamine-containing granules
- high-affinity receptors for immunoglobulin E (IgE, an antibody type that is involved in allergic disease, see **Box 6.9**).

Mast cells are found in tissues (especially skin and mucosae) and basophils in the blood. Both mast cells and basophils release pro-inflammatory mediators, which are either pre-formed or synthesized *de novo* (**Box 6.6**).

Histamine is a low-molecular-weight amine (111 Da) with a blood half-life of less than 5 minutes; it constitutes 10% of the mast cell's weight. When injected into the skin, histamine induces the typical 'weal and flare' or 'triple' response: reddening (erythema) due to increased blood flow, swelling (weal) due to increased vascular permeability, and distal vascular changes (flare) due to effects on local axons.

The ***complement-derived anaphylatoxins*** C3a, C4a, and C5a activate basophils and mast cells, as does IgE. The mast cell also has a role in the early response to bacteria through release of TNF-α, in cell recruitment to inflammatory sites such as arthritic joints, in promotion of tumour growth by enhancing neovascularization and in allograft tolerance.

Box 6.6 Mast cell and basophil mediators

Mediators	Effects
Pre-formed	
Histamine	Vasodilation Vascular permeability ↑ Smooth muscle contraction in airways
Proteases	Digestion of basement membrane causes ↑ vascular permeability and aids migration
Proteoglycans (e.g., heparan)	Anticoagulant activity
Synthesized *de novo*	
Platelet-activating factor (PAF)	Vasodilation
Leukotriene B_4 (LTB_4), LTC_4, LTD_4	Neutrophil and eosinophil activation and chemoattraction Vascular permeability ↑ Bronchoconstriction
Prostaglandins (mainly PGD_2)	Vascular permeability ↑ Bronchoconstriction Vasodilation

Monocytes and macrophages

Monocytes (in blood) and macrophages (in tissue) ingest and kill bacteria, release pro-inflammatory molecules, present antigen to T lymphocytes, and are necessary in immunity to intracellular pathogens such as mycobacteria.

Cells of the monocyte/macrophage lineage are highly sophisticated phagocytes. ***Monocytes*** are the blood form of a cell that spends a few days in the circulation before entering into the tissues to differentiate into ***macrophages***, and some types of dendritic cell. Blood monocytes can be divided into subsets according to expression of CD14 (a receptor for lipopolysaccharide, a bacterial cell wall component) and CD16 (a receptor for IgG antibodies).

A key role of tissue macrophages is the maintenance of tissue homeostasis through clearance of cellular debris, especially following infection or inflammation. They are responsive to a range of pro-inflammatory stimuli, using their PRRs to recognize PAMPs. Once activated, they engulf and kill microorganisms, especially bacteria and fungi, and release a range of pro-inflammatory cytokines. Evidence suggests that evolutionarily conserved molecular patterns in mitochondria (organelles that were originally derived from bacteria) can also activate monocytes. These damage-associated molecular patterns could play a major role in the systemic inflammatory response that follows extensive tissue damage (e.g., following ischaemic injury).

It has been observed that some PAMPs induce the cytoplasmic assembly of large oligomeric structures of PRRs termed ***inflammasomes***. There are numerous examples: members of the Nod-like receptor family can be activated by stimuli such as viruses, bacterial toxins, and, interestingly, crystallized endogenous molecules, including urate causing inflammatory arthritis in gout. Inflammasomes have potent effects in activating caspases, leading to processing and secretion of pro-inflammatory cytokines such as IL-1β and IL-18.

Macrophages have pro-inflammatory and microbicidal capabilities similar to those of neutrophils. Under activation conditions, macrophages present antigen to T-lymphocytes (see 'T lymphocyte development and activation' section) and secrete a range of cytokines, notably TNF-α, IL-1, and interferon-gamma (IFN-γ). These are necessary for the removal of certain pathogens that live within mononuclear phagocytes (e.g., mycobacteria). Macrophages and related cells may also undergo a process termed ***autophagy***. Autophagy is a lysosomal degradative process that removes abnormal organelles, protein aggregates, and intracellular pathogens. The autophagy machinery targets an autophagosome with the relevant material to fuse with lysosomes, which then causes degradation of the contents. This self-cannibalization is used by the immune system to destroy intracellular pathogens such as *Mycobacterium tuberculosis*, which otherwise persist within cells and block normal antibacterial processes. Autophagy is also a means of enhancing antigen presentation pathways. Dysregulation of the autophagic pathway has been associated with a number of diseases, including cancer, inflammatory disease, and infections.

Tissue macrophages involved in chronic inflammatory foci may undergo terminal differentiation into multinucleated giant cells, typically found at the site of the granulomata characteristic of tuberculosis and sarcoidosis (see Chapter 20).

Dendritic cells

The function of dendritic cells (DCs) is activation of naive T lymphocytes to initiate adaptive immune responses (see 'Adaptive immune

system' section); they are the only cells capable of this. The definition of a dendritic cell is one that has:

- dendritic morphology (Fig. 6.4)
- machinery for sensing pathogens
- the ability to process and present antigens to CD4 and CD8 T lymphocytes, coupled with the ability to activate these T lymphocytes from a naive state
- the ability to dictate the T lymphocyte's future function and differentiation.

This is a powerful cell type that functions as a critical bridge between the innate and adaptive immune systems.

Types of dendritic cell

The major types are the ***conventional DC (cDC)***, the ***plasmacytoid DC***, and a variety of ***specialized DCs*** found in tissues that resemble cDCs (e.g., the Langerhans cell in the skin). DCs have several distinctive cell surface molecules, some of which have pathogen-sensing activity (e.g., the antigen uptake receptor DEC205 on cDCs), while others are involved in interaction with T lymphocytes such as Human Leucocyte Antigen (HLA) molecules. Immature cDCs and pDCs are present in the blood but at very low levels (<0.5% of lymphocyte/monocyte cells).

Pathogen sensing is a key component of the function of immature DCs, as well as monocytes/macrophages, and is achieved through expression of a limited array of specialized PRR molecules capable of binding to structures common to pathogens, aided by long cell dendrites and pinocytosis (constant ingestion of soluble material).

Pattern recognition receptors may induce opsonization (see Fig. 6.3), or act as ***signalling receptors***, such as the PRR known as Toll-like receptors (Box 6.7).

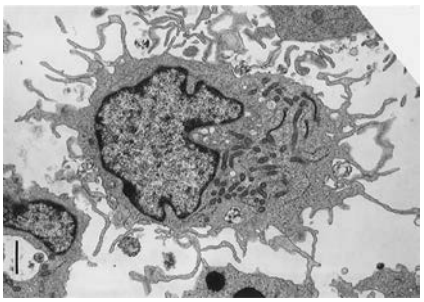


Fig. 6.4 Mature dendritic cell. The cell ingests pathogens, migrates to lymph nodes, and presents pathogen-derived antigen to T lymphocytes to enable adaptive immunity.

Box 6.7 Toll-like receptors (TLRs)

Pattern recognition receptor (PRR)	Pathogen-associated molecular pattern (PAMP)	Pathogen	PRR expressed by
TLR2	Peptidoglycan	Gram-positive bacteria	cDC
TLR3	Double-stranded RNA	Viruses	cDC
TLR4	Lipopolysaccharide	Gram-negative bacteria	cDC
TLR7	Single-stranded RNA	Viruses	pDC
TLR9	Double-stranded DNA	Viruses	pDC

cDC, Conventional dendritic cell; *pDC,* plasmacytoid dendritic cell.

The key principle is that the immune system has devised a means of identifying most types of invading microorganisms by using a limited number of PRRs recognizing common molecular patterns, or PAMPs. This recognition event has been termed a '***danger signal***'; it alerts the immune system to the presence of a pathogen.

Dendritic cells and T-cell activation

In a sequence of events that spans 1 to 2 days, immature DCs are activated by PAMPs or damage-associated molecular patterns in the tissues binding to PRRs. During maturation and migration, the cDC matures, changing its shape, gene and molecular profile, and function within a matter of hours to take on a mature form, with altered capabilities (Fig. 6.5), the most important of which is upregulating the machinery required to activate T lymphocytes. Once in the lymph node, the mature cDC interacts with naive T lymphocytes (antigen presentation), resulting in two key outcomes:

- activation of T cells with the ability to recognize peptide fragments (termed epitopes) of the pathogen
- polarization of the T-cell towards a functional phenotype (see below) that is tailored to the particular pathogen.

The mature cDC provides three major signals to naive T cells (see Fig. 6.5):

- ***signal 1:*** presentation of peptide fragments from the pathogen bound to surface HLA molecules
- ***signal 2:*** co-stimulation through CD80 and CD86 interacting with CD28 on T cells
- ***signal 3:*** secretion of cytokines, notably IL-12.

Innate lymphoid cells

Innate lymphoid cells (ILCs) are a group of immune cells that play a crucial role in the innate immune response. They are distinct from adaptive immune cells, such as T and B cells, as they lack antigen-specific receptors. ILCs are primarily involved in early immune defence against pathogens, tissue homeostasis, and regulation of immune responses at barrier surfaces, such as the skin, lungs, and intestines.

Innate lymphoid cells are classified into three major subsets based on their cytokine production profiles and transcription factor expression: ILC1, ILC2, and ILC3. Each subset has unique functions and contributes to various aspects of immune defence.

The study of ILCs is relatively recent, and ongoing research continues to uncover their diverse functions and regulatory mechanisms.

Further reading

Fucikova J, Palova-Jelinkova L, Bartunkova J, Spisek R. Induction of tolerance and immunity by dendritic cells: mechanisms and clinical applications. *Front Immunol*. 2019; 10:2393.

Nagasawa T, Kobayashi H, Aramaki M, Kiji M, Oda S, and Izumi Y. Expression of CD14, CD16 and CD45RA on monocytes from periodontitis patients. *J Periodontal Res.* 2004; 39:72–78.

Ronchetti L, Boubaker NS, Barba M, et al. Neutrophil extracellular traps in cancer: not only catching microbes. *J Exp Clin Cancer Res.* 2021; 40:231.

ADAPTIVE IMMUNE SYSTEM

The information gained by cDCs that interact with a pathogen is passed on, in the form of signals 1 to 3 (see Fig. 6.5B). These activate T lymphocytes in the adaptive immune system, which recognize the same pathogen. T lymphocytes may be involved in pathogen removal directly (e.g., by killing) or indirectly (e.g., by recruiting B lymphocytes to make specific antibody). Lymphocytes

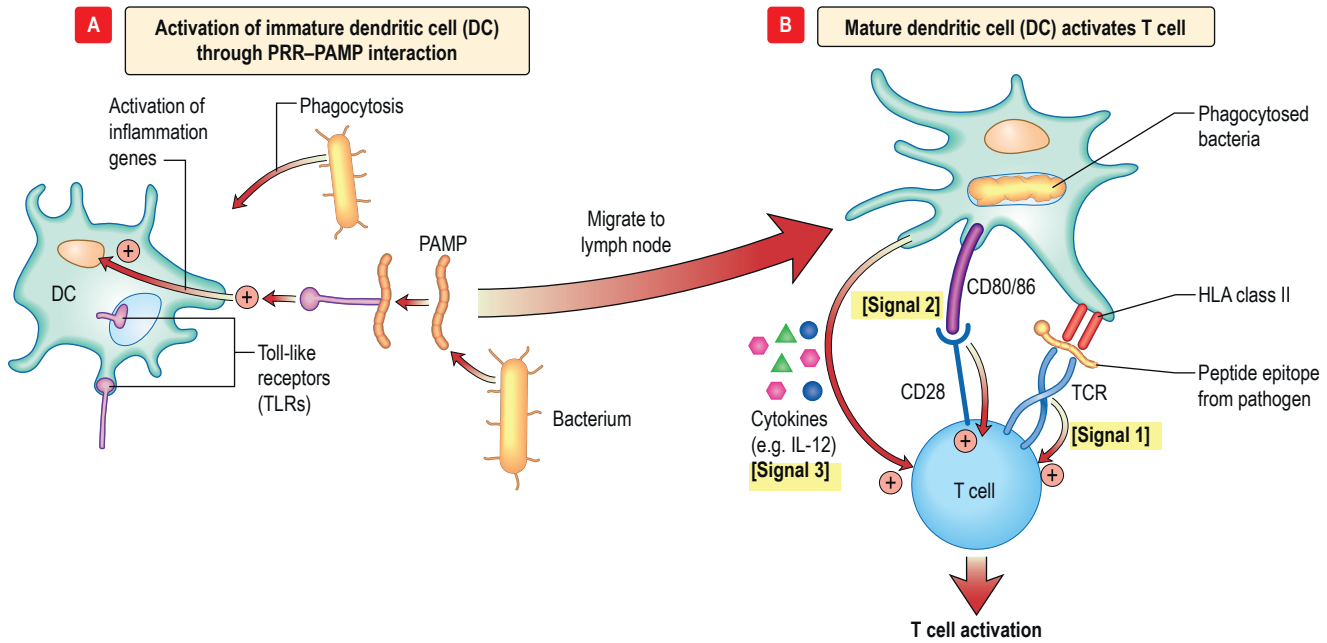


Fig. 6.5 Dendritic cell activation of T lymphocytes. **(A)** Immature dendritic cells (DCs) in the tissues are activated by pathogens through pathogen-associated molecular pattern–pattern recognition receptor (PAMP–PRR) interaction. **(B)** Multiple rapid changes in gene expression lead to migration to the lymph node as the DC takes on the mature phenotype. During migration, there is synthesis of the machinery required for activation of T lymphocytes, shown here in response to ***signals 1 to 3***. *CD,* Cluster of differentiation; *HLA,* human leucocyte antigen; *IL-12,* interleukin 12; *TCR,* T-cell receptor.

orchestrate immune responses via cell-to-cell interactions and cytokine release.

Antigen receptors on T and B lymphocytes

One of the key features of the adaptive immune system is specificity for antigen. For example, if a person is immunized against the measles virus, that person does not have immunity to hepatitis B, and vice versa. This specificity is conferred by two types of receptor: the T-cell receptor (TCR) on T lymphocytes and an equivalent on B lymphocytes, the B-cell receptor. B-cell receptors are also termed surface immunoglobulin; they differ from TCRs in also being secreted in large quantities by end-stage B lymphocytes (plasma cells) as soluble immunoglobulins, also known as antibodies.

To maintain protection against the multiplicity of pathogens in our environment, each of us generates great diversity among TCRs and antibodies. This is achieved through a mechanism shared by both TCR and antibody, in which distinct families of genes are encoded in the germline, each family (called constant, variable, diversity, and joining) contributing a sequence to part of the receptor. Recombination between randomly selected members of each family ensures diversity in the end product. Recombination frequently involves base deletions and additions, adding to the diversity. In the case of antibodies, the result is a potential capacity of more than 10^{14} different antibody molecules; for TCRs, it may be as high as 10^{18}.

Immunoglobulins

In structural terms, antibodies have four chains: two identical heavy and two identical light chains (Fig. 6.6). Each chain contains both highly variable and essentially constant regions. The variable parts of the heavy and light chains pair to form the potentially diverse part of the antibody molecule that binds antigen. The constant region of

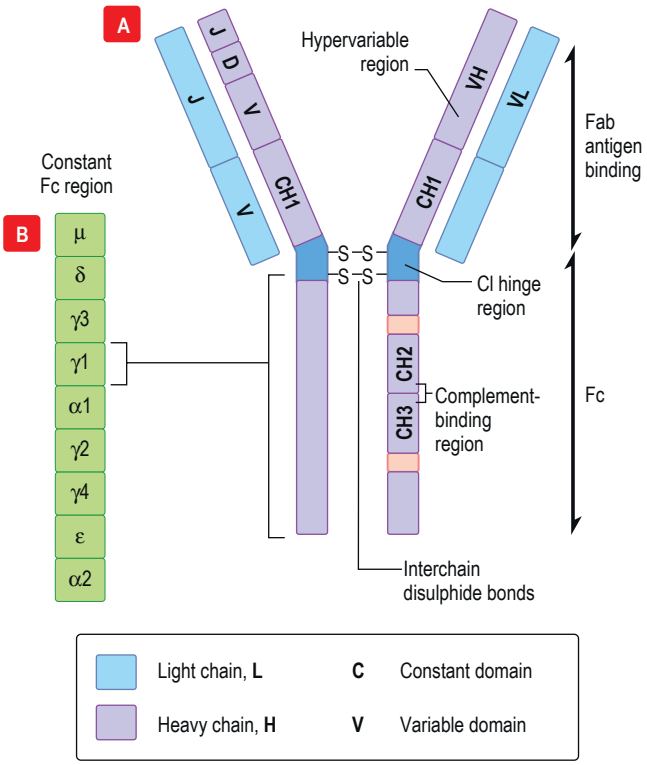


Fig. 6.6 Immunoglobulin structure. Basic subunit consisting of two heavy and two light chains **(A)**. Genes on the Fab and Fc regions of an immunoglobulin **(B)**. The chain is made up of a V (variable) gene, which is translocated to the J (joining) chain. The VJ segment is then spliced to the C (constant) gene. Heavy chains have an additional D (diversity) segment, which forms the VDJ segment that bears the antigen-binding site determinants.

the heavy chain dictates the function of the antibody and belongs to one of the classes of M, G1–4, A1–2, D, and E, giving rise to antibodies called IgM, IgG1–4, IgA1–2, IgD, and IgE, respectively. The characteristics of these different isotypes are shown in **Box 6.8**. As it matures, and under the instruction of T lymphocytes, a B lymphocyte may change the class (class switching), but never the specificity, of the antibody it makes; minor changes in antibody gene sequence can take place (somatic mutation), potentially allowing antibodies with higher affinity to arise and be selected for the effector response (affinity maturation).

Antibody function

In host defence, antibodies target, neutralize, and remove infectious organisms and toxins from the circulation and tissues, often through recruitment of innate host effector mechanisms such as complement, phagocytes, and mast cells (by binding to specific surface receptors on these cells).

In clinical medicine, specific anti-pathogen antibody levels are used in diagnosing/monitoring infectious disease and may also be administered as serum pools to provide host protection passively.

Antibodies can be raised in animals to generate monoclonal antibodies, which are commonly used in diagnostic immunology tests and increasingly employed as therapeutics (***immunotherapy***; e.g., to target cancer cells or to target immune cells and molecules to suppress or activate immune responses), often after 'humanization' (see later).

T-cell receptor genes and receptor diversity

The genomic organization of TCR genes and principles of generation of receptor diversity are similar to those of immunoglobulin genes. The TCR exists as a heterodimer, with a similar overall structure to that of the antibody molecule. There are two TCR types:

- ***α and β chains*** (αβ TCR). These are expressed on all CD4 T lymphocytes and approximately 90% of CD8 T lymphocytes; they play a role in adaptive immune responses.
- ***γ and δ chains*** (γδ TCR). These are fewer in number and are expressed mainly on intraepithelial lymphocytes; they are involved in epithelial defence.

The chains of each type of TCR are divided into variable and constant domains, each domain being encoded by separate gene pools. Like the B lymphocyte producing a single clone of immunoglobulin molecules, the T lymphocyte expresses only one form of TCR once the genes have been rearranged. Unlike antibodies, TCRs do not undergo somatic hypermutation and are not secreted.

T lymphocyte development and activation

T lymphocytes are generated from precursors in the bone marrow, which migrate to the thymus (**Fig. 6.7**). Only 1% of the cells that enter the thymus will leave it as naive T lymphocytes to populate the lymph nodes. This process (termed thymic selection) leads to a cohort of cells (**Box 6.9**) with:

- functionally rearranged genes allowing surface expression of a receptor for antigen (the TCR) alongside the CD3 accessory molecule involved in transducing the antigen-specific signal
- selection of a co-receptor, either CD4 or CD8, to stabilize the interaction between TCR and peptide-HLA:
  - CD4 T-cell responses require presentation of peptide antigens by *self* HLA class II molecules
  - CD8 T-cell responses require presentation of peptide antigens by *self* HLA class I molecules
- a reduced tendency of the selected TCR to recognize self antigens (thus avoiding autoimmunity).

Thus, during thymic education, most TCRs are rejected for further use (negative selection), either because they are unable to bind self HLA molecules or because they bind with too strong an affinity, which would run the risk of self-reactivity and autoimmune disease. The chosen TCRs (positive selection) have low/intermediate affinity for self HLA molecules. During post-thymic activation of T lymphocytes in the lymph node, TCR interaction with HLA has to be bolstered by additional signals (called co-stimulation) provided by cDCs. This ensures that T lymphocytes are activated by a mature cDC; this happens only in the presence of pathogens.

Co-stimulation of T cells is such a critical event in immune responses that it is tightly regulated. One level of control relies on molecular interactions that are termed ***immune checkpoints***. An example is cytotoxic T-lymphocyte-associated antigen 4, which is expressed shortly after T-cell activation and competes with CD80 and CD86 for CD28, at the same time providing a negative signal, reducing T-cell activation which is exploited in a rheumatoid arthritis

Box 6.8 Characteristics of the immunoglobulins (Igs)

	IgG	IgM	IgA	IgE	IgD
Description	Dominant class of antibody	Produced first in immune response	Found in mucous membrane secretions	Responsible for symptoms of allergy: used in defence against nematode parasites	Found almost solely on B lymphocyte surface membranes
Heavy chain	γ	μ	α	ε	δ
Mean adult serum levels (g/L)	IgG (total) = 8–16 G1 = 6.5 G2 = 2.5 G3 = 0.7 G4 = 0.3	0.5–2	IgA (total) = 1.4–4 A1 = 1.5 A2 = 0.2	17–450 ng/mL	0–0.4
Half-life (days)	21	10	6	2	3
Complement fixation					
Classical Alternative	++ –	+++ –	– +/–	– –	– –
Binding to mast cells	–	–	–	+	–
Crosses placenta	+	–	–	–	–

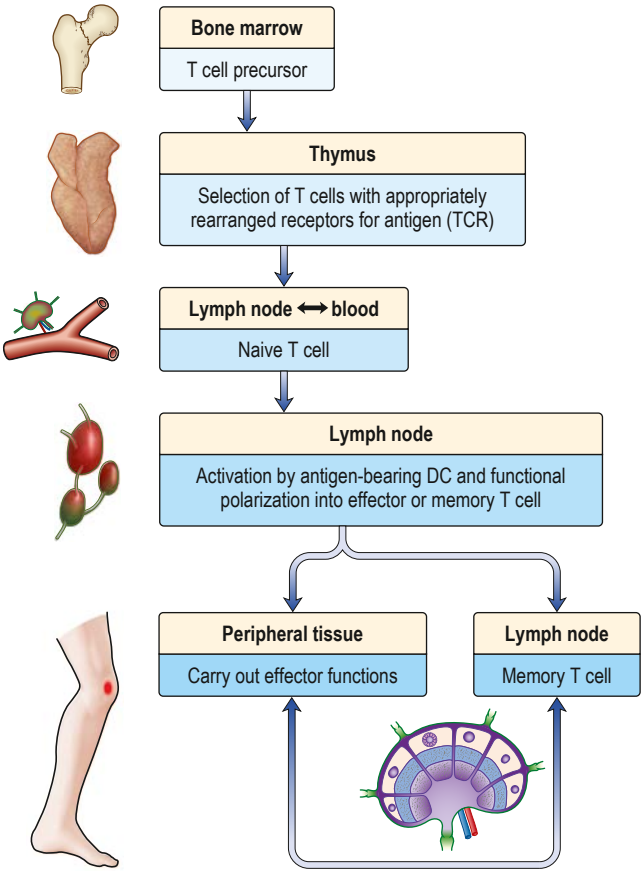


Fig. 6.7 Development of T lymphocytes in the thymus and life cycle in the secondary lymphoid tissue. *DC,* Dendritic cell; *TCR,* T-cell receptor.

Box 6.9 Identification of T lymphocytes

T-cell population	Marker	Typical percentages in blood	Additional information
T lymphocytes	T-cell receptor CD3	100% of T cells (70% of lymphocytes)	All T cells are thymus-derived
Helper T lymphocytes (Th)	CD4	66% of T cells	Th interact with antigen presented by MHC class II molecules
Cytotoxic T lymphocytes (CTL)	CD8	33% of T cells	CTL interact with antigen presented by MHC class I molecules

CD, Cluster of differentiation; *MHC,* major histocompatibility complex.

treatment with a soluble fusion protein (Fig. 6.8A). A further example is the immune checkpoint involving programmed cell death protein-1 (PD-1) and its ligands (PD-L1/L2). Like cytotoxic T-lymphocyte-associated antigen 4, PD-1 is upregulated after T lymphocyte activation, and interaction with PD-L1/L2 leads to inhibition of T lymphocyte function. PD-L1/L2 may be expressed on a variety of cells, including cDCs but also many non-immune cells, most notably tumour cells. *Immune checkpoint blockade* (see Fig. 6.8B) using monoclonal antibodies focused on PD-1/PD-L interactions releases the brakes on the immune system and is at the heart of the recent revolution in cancer immunotherapy (see Chapter 41).

Most naive T lymphocytes are resident in the lymph nodes or spleen. Naive T lymphocytes are activated for the first time in the lymph node by antigens presented to their TCRs as short peptides bound to major histocompatibility complex (MHC) molecules on the surface of DCs (see Fig. 6.5). Provision of signals 1 to 3 sets off an intracellular cascade of signalling molecule activation, leading to induction of gene transcription in T lymphocytes.

The outcome is T lymphocyte activation, cell division, and functional polarization, which is the acquired ability to promote a selected type of adaptive immune response. These processes take several days to achieve and through specific ***transcriptional regulation*** defined subsets of T cells are generated. The major functional polarities of T-cell responses (Box 6.10) are:

- $CD4^+$ pro-inflammatory T lymphocytes; T helper 1 (Th1), Th2, and Th17
- $CD8^+$ cytotoxic T lymphocytes (CTLs)
- $CD4^+$ regulatory responses (Treg).

Through cell division, a proportion of the T lymphocytes that are activated in response to a pathogen undertake these effector or regulatory functions, while a proportion is assigned to a memory pool.

CD4 T lymphocyte functions

As the pivotal cell in immune responses, the CD4 T lymphocyte influences most aspects of immunity, either through the release of cytokines or via direct cell–cell interaction, a process often termed 'licensing'. Licensing critically involves the pairing of CD40 on a specialized subset of cDC (cDC1) with CD40 ligand (CD40L, CD154) on the CD4 T lymphocyte; defects in this process result in antibody deficiency (see later).

Major functions of CD4 T lymphocytes are:

- ***licensing of cDCs*** during antigen presentation to activate CD8 T lymphocytes and generate cytotoxic cells
- ***licensing of B lymphocytes*** to initiate and mature antibody responses, leading to class switching, affinity maturation of antibodies, and generation of plasma cells or memory cells
- ***secretion of cytokines*** responsible for growth and differentiation of a range of cell types, especially other T lymphocytes, macrophages, and eosinophils
- ***regulation of immune reactions.***

T helper 1 cells

T helper 1 cells are the main effector subtype of CD4 T lymphocytes. In physiology, they drive activation of monocytes/macrophages and CTLs. In pathology, they have a key role in protection against intracellular pathogens such as viruses and mycobacteria. Th1 cells are recognized by secretion of the pro-inflammatory cytokines IFN-γ and TNF-α (Box 6.11).

T helper 2 cells

In physiology, Th2 cells drive antibody responses, especially IgE, and also promote eosinophil granulocyte functions. In pathology, they have a key role in protection from extracellular parasites (helminths) and also in the immune responses that underlie allergic disease. Th2 cells are recognized by secretion of IL-4, IL-5, and IL-13.

T helper 17 cells

In physiology, Th17 cells drive inflammatory responses, especially via recruitment of neutrophil granulocytes. In pathology, they are necessary for protection from fungal infections and are increasingly

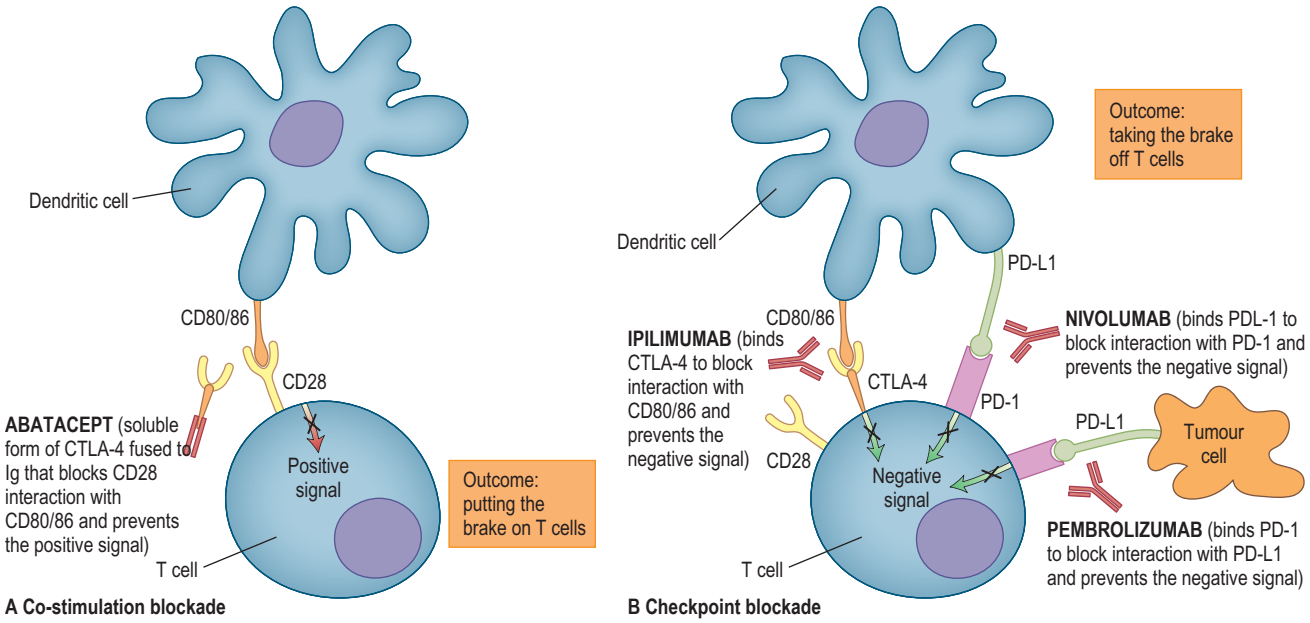


Fig. 6.8 Understanding immune checkpoints for immunotherapy. **(A)** T cells require co-stimulation, as well as recognition of antigen, to undergo full activation (see also **Fig. 6.5B**). CD80/86 on the dendritic cell binds CD28 on the T-cell receptor, providing a positive co-stimulatory signal. The drug abatacept is a soluble fusion of the cytotoxic T-lymphocyte-associated antigen 4 (CTLA-4) molecule with an immunoglobulin stalk. Since CTLA-4 has a higher affinity for CD80/86 than CD28, abatacept blockades co-stimulation and suppresses T-cell responses. **(B)** Soon after activation, T cells express CTLA-4 and programmed cell death protein-1 (PD-1), which deliver negative signals. Monoclonal antibodies aimed at blocking the interactions of these receptor/ligand pairs are thus able to take the brakes off the T-cell response. This can be critical for promoting anti-tumour responses.

Box 6.10 Identification of T lymphocyte subsets by function

T-cell type	Major CD marker	Main transcriptional regulators	Main cytokines causing polarization	Main cytokines produced	Functions	Major role in physiological immune response
T helper 1 (Th1) cells	CD4	T-bet	IL-12	IFN-γ, IL-2, TNF-α	Pro-inflammatory	Organize killing of bacteria, fungi, and viruses; activate macrophages to kill intracellular pathogens; instruct cytotoxic T-cell responses
T helper 2 (Th2) cells	CD4	GATA3	IL-4	IL-4, IL-5, IL-13	Pro-inflammatory	Organize killing of parasites by recruiting eosinophils; promote antibody responses, especially switching to IgE so play a role in allergic disease
T helper 17 (Th17) cells	CD4	RORγt	IL-6, IL-23, TGF-β	IL-17	Pro-inflammatory	Essential in preventing fungal infection, control of viral infection such as molluscum and warts, contribute to autoimmunity when dysregulated
Regulatory T cells (Treg)	CD4	FOXP3	IL-10, TGF-β	IL-10, TGF-β	Regulatory	Regulation of inflammation
Cytotoxic lymphocyte	CD8	T-bet/ Eomesodermin	IL-12, IFNa, IL-2	IFN-γ, TNF-α	Cytotoxic	Kill target cells via recognition of HLA class I+ peptide Protect from viruses

CD, Cluster of differentiation; *FOXP3,* Forkhead box P3; *GATA3,* GATA-binding protein 3; *HLA,* human leucocyte antigen; *IFN-γ,* interferon-gamma; *IgE,* immunoglobulin E; *IL,* interleukin; *RORγt,* RAR-related orphan nuclear receptor (ROR) gamma expressed in t-cells; *T-bet,* transcription box TBX21 expressed in t-cells; *TGF-β,* transforming growth factor beta; *TNF-α,* tumour necrosis factor alpha.

Box 6.11 Examples of natural killer (NK) cell receptors

Receptors on NK cells	Ligand
Inhibitory	
KIR2DL1	HLA-C molecules
KIR3DL1	HLA-B molecules
KIRDL2	HLA-A molecules
NKG2A	HLA-E molecules
Activating	
CD16 (low-affinity receptor for IgG)	IgG
NKG2D	MHC class I chain-related gene A *(MICA)*
KIR2DS1	HLA-C

CD, Cluster of differentiation; *HLA,* human leucocyte antigen; *KIR,* killer cell immunoglobulin-like receptor; *MHC,* major histocompatibility complex.

recognized as having a role in chronic inflammatory diseases such as multiple sclerosis, rheumatoid arthritis, and inflammatory bowel disease. They are recognized by secretion of IL-17.

T follicular helper cells

T follicular helper cells express CXCR5, the inducible T-cell co-stimulator molecule (ICOS), and secrete IL-21. They are responsible for licensing B lymphocytes for antibody production (see earlier).

Regulatory T lymphocytes

The generation of B and T lymphocytes provides a potentially vast array of rearranged antigen receptors. Although there are selection processes to remove lymphocytes with 'dangerous' avidity for 'self', these are not foolproof and the potential for autoreactivity remains. The fact that there is no self-destruction in the vast majority of people implies that the norm is a state of immunological self-tolerance: the controlled inability to respond to self. Several mechanisms operate to maintain this state, including CD4 T lymphocytes that respond to antigenic stimulation by suppressing ongoing immune responses. These regulatory T lymphocytes (Treg) express high levels of CD25, the receptor for IL-2; regulate other T lymphocytes by cell–cell contact; and also secrete the immune-suppressive cytokines IL-10 and TGF-β. They can be generated in the thymus or post-thymically in the periphery. Their key feature is high expression of the transcription factor Foxp3.

Evidence that Tregs are clinically relevant is provided by rare cases of genetic defects in the *Foxp3* gene, leading to IPEX (***i***mmune dysregulation, ***p***olyendocrinopathy, ***e***nteropathy, ***X***-linked) syndrome. Patients with this rare syndrome have defective Tregs and develop a range of conditions soon after birth, including organ-specific autoimmune disease such as type 1 diabetes.

CD8 T lymphocyte functions

Cytotoxic CD8 T lymphocytes are involved in defence against viruses. CTLs kill virus-infected cells following recognition of viral peptide-HLA class I complexes. CTLs must be activated first in the lymph node by a specialized cDC subset (called cDC1) that cross-presents the *same* viral peptide that is seen on infected cells and is licensed by a CD4 T lymphocyte recognizing viral peptide-HLA class II complexes. The same defence mechanism may also apply in tumour surveillance. This is a checkpoint that ensures that CTL responses, which have great destructive power, are activated only against a target for which there is also a CD4 T-cell response. CTLs kill via three mechanisms:

- cytotoxic granule proteins (cytolysins such as perforin, granzyme B)
- toxic cytokines (e.g., IFN-γ, TNF-α)
- death-inducing surface molecules (e.g., Fas ligand binds Fas on target cells mediating apoptosis via caspase activation, Fig. 6.9).

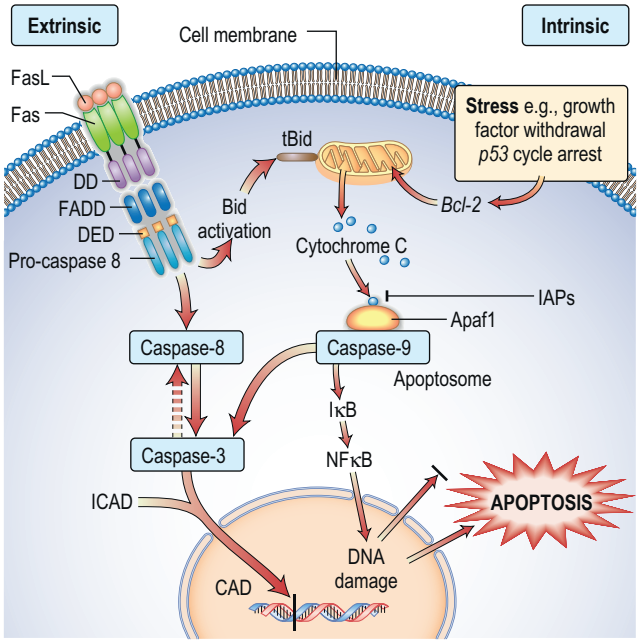


Fig. 6.9 Extrinsic and intrinsic apoptotic signalling network. The **Fas protein** and **Fas ligand (FasL)** are two proteins that interact to activate an apoptotic pathway: that is, a process of programmed cell death that removes cells that are infected or contain damaged DNA. Fas and FasL are both members of the tumour necrosis factor family. When FasL binds to Fas, it causes Fas to trimerize and brings together the death domains on the cytoplasmic tails of the protein. The **adaptor protein FADD** (Fas-associating protein with death domain) binds to these activated death domains and they bind to pro-caspase 8 through a set of death effector domains. **Pro-caspase 8s** then release caspase 8, in turn cleaving and activating other caspases, including caspase 3. **Caspase 3** cleaves ICAD, the inhibitor of caspase-activated DNase, which frees caspase-activated DNase to enter the nucleus and cleave DNA. *Apaf1,* Apoptotic protease-activating factor 1; *Bid,* family member of Bcl-2 protein; *IAPs,* inhibitor of apoptosis proteins; *IκB,* inhibitor of kappa B; *NFκB,* nuclear factor kappa B.

Natural killer cells

Natural killer cells are bone marrow derived, present in the blood and lymph nodes, and represent 5% to 10% of lymphoid cells. The name reflects two features. Unlike B and T lymphocytes, NK cells are able to:

- mediate their effector function *spontaneously* (i.e., *killing* of target cells through release of perforin, a pore-forming protein) in the absence of previous known sensitization to that target
- achieve this with a very limited repertoire of germline-encoded receptors that do not undergo somatic recombination.

For identification purposes, the main surface molecules associated with NK cells are CD16 (see later) and CD56 (note that NK cells are CD3- and TCR-negative). The role of NK cells is to kill 'abnormal' host cells, typically cells that are virus infected, or

tumour cells. Killing is achieved in similar ways to CTLs. NK cells also secrete copious amounts of IFN-γ and TNF-α, through which they can mediate cytotoxic effects and activate other components of the innate and adaptive immune systems. To become activated, NK cells integrate the signal from a potential target cell through a series of receptor–ligand pairings (**Box 6.12**). These pairings provide activating and inhibitory signals, and it is the overall balance of these that determines the outcome for the NK cell. The balance can be abnormal on a virus-infected or tumour cell, which might have altered expression of HLA molecules, for example, that mark them out for NK cytotoxicity.

Further reading

Bacchetta R, Weinberg K. Thymic origins of autoimmunity—lessons from inborn errors of immunity. *Semin Immunopathol.* 2021; 43:65–83.
Cassioli C, Baldari CT. The expanding arsenal of cytotoxic T cells. *Front Immunol.* 2022; 13:883010.
Dong C. Cytokine regulation and function in T cells. *Annu Rev Immunol.* 2021; 39(1): 51–76.
Natural killer cells and their role in immunity. *EMJ Allergy & Immunol.* 2019; 4(1):108 116.
Vivier, Eric et al. Innate lymphoid Cells: 10 years on. *Cell.* 2018; 174(5):1054–1066.

CELL MIGRATION

For the immune system to function effectively, its cells must be able to access all areas of the body. Immune cells are highly mobile, allowing recirculation through the bloodstream, tissues, and lymphatics to provide systemic pathogen surveillance. For example, migration of a DC into a regional lymph node from the tissues via the lymphatics (**Fig. 6.10**) is highly dependent on expression of the chemokine receptor CCR7 by the migrating cell and of its ligand (CCL21) by the target tissue. In addition, naive lymphocytes express L-selectin, which binds a glycoprotein cell adhesion molecule-1, found on the high endothelial venules of lymph nodes. This system can be upregulated in an inflamed lymph node, leading to an influx of naive lymphocytes and the typical symptom of a swollen node. The ability of CCR7 to retain lymphocytes in the lymph nodes is balanced by the activity of sphingosine-1-phosphatase (S1P) and its receptors. High levels of S1P allow activated immune cells to recirculate back into the tissues. Modulation of S1P can be used to retain lymphocytes in the lymph node in the treatment of multiple sclerosis (see Chapter 28).

Migration into inflamed tissue requires:

- an affected organ or tissue to signal that there is a focus of injury/infection *and*
- responding immune cells to bind and adhere specifically to that tissue.

This process is highly organized and has a similar basis for all immune cells, involving three basic steps: rolling, adhesion, and transmigration. Each of these is dependent on specialized adhesion molecules (**Fig. 6.11**).

The expression of these molecules (e.g., leucocyte function associated antigen-1) is upregulated on T lymphocytes after activation in the lymph node. Intercellular cell adhesion molecule 1 (ICAM-1) expression on tissue endothelium is sensitive to numerous pro-inflammatory molecules and allows immune cells to be guided from the blood into the tissues. Once there, cells move along a gradient of increasing concentration of mediators such as chemokines in the process of chemotaxis.

Box 6.12 HLA associations with immune-mediated and infectious diseases

Disease process	Disease	HLA type
Autoimmunity	Type 1 diabetes	Class II: *DQA1*03:01/DQB1*03:02* (susceptibility) *DQA1*05:01/DQB1*02:01* (susceptibility) *DQA1*01:02/DQB1*06:02* (protection) Class I: *HLA-A*24*, *HLA-B*18*, *HLA-B*39*
	Multiple sclerosis	*HLA-DRB1*15:01* (susceptibility)
	Myasthenia gravis	*HLA-A*01*, *HLA-B*08*, and *HLA-DRB1*03* extended haplotype (susceptibility) predominates in younger females *HLA-B*07-DRB1*15* (susceptibility) predominates in older males *HLA-DRB1*16*, *HLA-DRB1*14*, *HLA-DQB1*05* (susceptibility) for MUSK variant myasthenia
	Behçet's syndrome	*HLA-B*51:01* (susceptibility)
	Rheumatoid arthritis	*HLA-DRB1*0404* (susceptibility)
	Autoimmune hepatitis	*HLA-DRB1*03*, *HLA-DRB1*04* (susceptibility)
	Goodpasture's syndrome (anti-glomerular basement membrane disease)	*HLA-DRB1*15:01* (susceptibility)
	Pemphigus vulgaris	*HLA-DRB1*04:02*, *HLA-DQB1*05:03* (susceptibility)
Inflammatory	Coeliac disease Ankylosing spondylitis Psoriasis Idiopathic membranous nephropathy	*HLA-DQA1*05:01/DQB1*02:01* (susceptibility) *HLA-B*27* (susceptibility) *HLA-Cw*06:02* (susceptibility) *HLA-DQA1* (susceptibility)
Infectious	Human immunodeficiency virus infection	*HLA-B*27*, *HLA-B*51*, *HLA-B*57* (associated with slow progression of disease) *HLA-B*35* (associated with rapid progression)
Other	Narcolepsy	*HLA-DQA1*01:02, HLA-DQB1*06:02* (susceptibility) *HLA-DQB1*03:01* (worsens) *HLA-DQB1*05:01*, *HLA-DQB1*06:01* (protect)

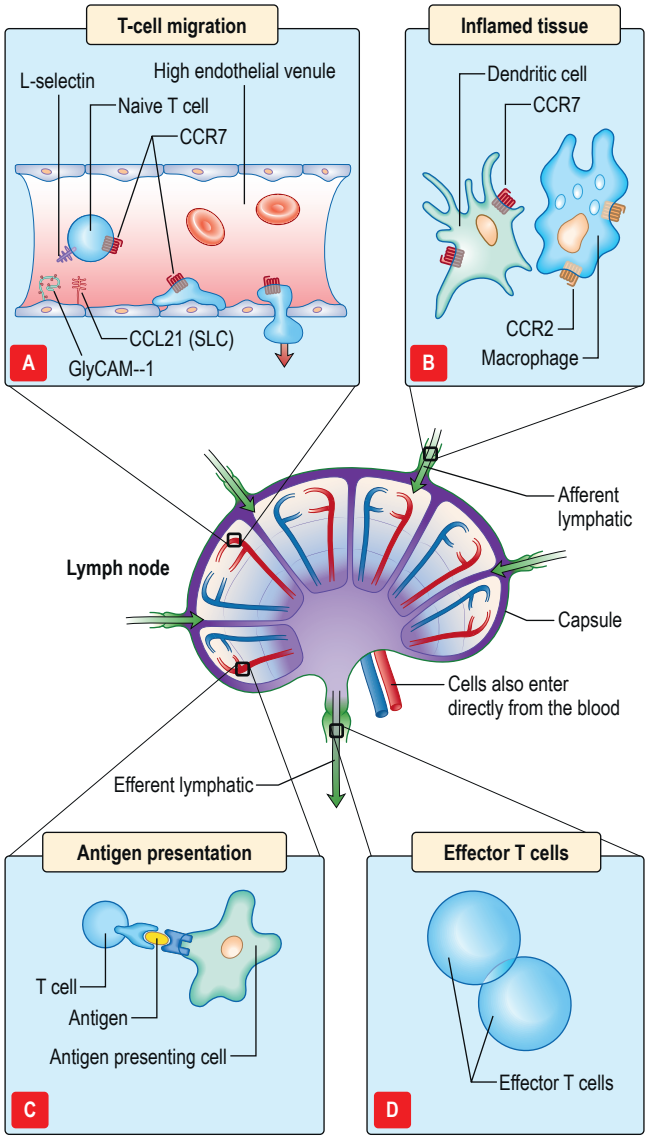


Fig. 6.10 Migration of an immune cell through a lymph node. **(A)** T lymphocytes migrate through endothelial venules, which express the chemokine CCL21 (secondary lymphoid tissue chemokine, or SLC). **(B)** ***Dendritic cells*** (expressing CCR7) and macrophages (expressing CCR2) enter lymph nodes via afferent lymphatics and localize in the paracortex. **(C)** ***Antigen presentation*** (see also Figs 6.14 and 6.15) results in activation of T lymphocytes. **(D)** ***Effector T lymphocytes*** leave the lymph node via efferent lymphatics.

HLA MOLECULES AND ANTIGEN PRESENTATION

On the short arm of chromosome 6 is a collection of genes termed the major histocompatibility complex (MHC; known as the human leucocyte antigens, or HLA, in humans), which plays a critical role in immune function. MHC genes code for proteins expressed on the surface of a variety of cell types that are involved in antigen recognition by T lymphocytes. The T lymphocyte receptor for antigen recognizes its ligand as a short antigenic peptide embedded within a physical groove at the extremity of the HLA molecule (Fig. 6.12).

Differences in HLA molecules between individuals are responsible for tissue and organ graft rejection (hence the name 'histo' [*tissue*]-compatibility). Possession of certain HLA genes is linked to susceptibility to particular diseases (Box 6.13).

The human major histocompatibility complex

The human MHC comprises three major classes (I, II, and III) of genes involved in the immune response (Fig. 6.13).

HLA classes

Classical HLA class I genes

Classical HLA class I genes (also termed Ia) are designated *HLA-A*, *HLA-B*, and *HLA-C*. Each encodes a class I α chain, which combines with a β chain to form the class I HLA molecule (see Fig. 6.12). While there are several types of α chain, there is only one type of β chain: β_2 microglobulin. The HLA class I molecule has the role of presenting short (8–10 amino acid) antigenic peptides to the TCR on the subset of T lymphocytes that bear the co-receptor CD8. As an example of HLA polymorphism, there are nearly 200 allelic forms at the *A* gene locus. Class I HLA molecules are expressed on all nucleated cells.

Non-classical HLA class I genes

Non-classical HLA class I genes are less polymorphic, have a more restricted expression on specialized cell types, and present a restricted type of peptide or none at all. These are the *HLA-E*, *F*, and *G* (Ib genes) and MHC class I–related (MIC, or class Ic) genes, *A* and *B*. The products of these genes are predominantly found on epithelial cells, signal cellular stress, and interact with lymphoid cells, especially natural killer cells (see 'Natural killer cells' section).

HLA class II genes

The class II genes have three major subregions: *DP*, *DQ*, and *DR*. In these subregions are genes encoding *A* and *B* genes that combine to form dimeric $\alpha\beta$ molecules that present short (12–15 amino acid) peptides to T lymphocytes that bear the CD4 co-receptor. Class II HLA genes (apart from *DRA*) are highly polymorphic. Other genes in this region encode proteins with key roles in antigen presentation (e.g., TAP, HLA-DM, HLA-DO, proteasome subunits; see later). Class II HLA genes are expressed on a restricted cohort of cells that go by the general term of antigen-presenting cells ([APCs]; DCs, monocyte/macrophages, B lymphocytes).

HLA class III genes

HLA class III genes encode proteins that can regulate or modify immune responses; they include tumour necrosis factor (TNF), heat shock protein, and complement protein (C2, C4).

HLA genotypes and the range of their protein products

HLA genotype is denoted first by the letters that designate the locus (e.g., *HLA-A*, *HLA-DR*, *HLA-DQ*). For class I alleles, this is followed by an asterisk and then a two- to four-digit number defining the allelic variant at that locus, often called the HLA type (e.g., *HLA-A*02* is the 02 variant of the *HLA-A* gene). The class II nomenclature is the same, except that both *A* and *B* genes are named.

Some general principles apply to the HLA genes and their protein products:

- The presence of multiple genes on each chromosome, and the fact that both maternal and paternal genes are co-dominantly

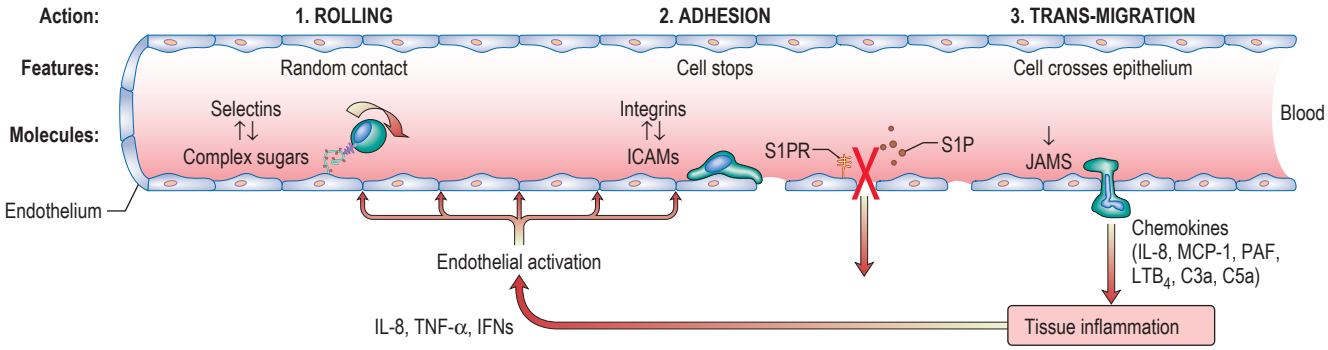


Fig. 6.11 Neutrophil migration. Neutrophils arrive at the site of the inflammation, attracted by various chemoattractants. **(1)** They roll along the blood vessel wall through interaction between L-selectin, on their surface, binding to a carbohydrate structure (e.g., sialyl-Lewis^X) on the endothelium. E-selectin on the endothelium mediates a similar effect. P-, E- and L-selectin (***p***latelet, ***e***ndothelium, and ***l***eucocyte, respectively) are named after the predominant cell type that expresses them and bind to carbohydrate moieties such as sialyl-Lewis^X (CD15s) on immune cells, inducing them to slow and roll along the vessel lumen. **(2)** ***Firm adhesion*** is then mediated via interaction between integrins on the neutrophil and intercellular and vascular cell adhesion molecules (ICAM-1 and VCAM-1) on the endothelium. The best-known integrin is the heterodimer of CD11a/CD18 (leucocyte function associated antigen-1—LFA-1), which binds ICAM-1. **(3)** ***Trans-migration*** (diapedesis) then occurs: this is a complex process that involves platelet-endothelial cell adhesion molecule-1 and junctional adhesion molecules. Following transmigration (diapedesis), mediators such as interleukin-8 (IL-8) and tumour necrosis factor alpha attract and activate the neutrophil to the infected tissue, where it phagocytoses and destroys the C3b-coated bacteria. *IFNs,* Interferons; *LTB_4,* leukotriene B_4; *MCP-1,* monocyte chemoattractant protein-1 (CCL2); *PAF,* platelet-activating factor.

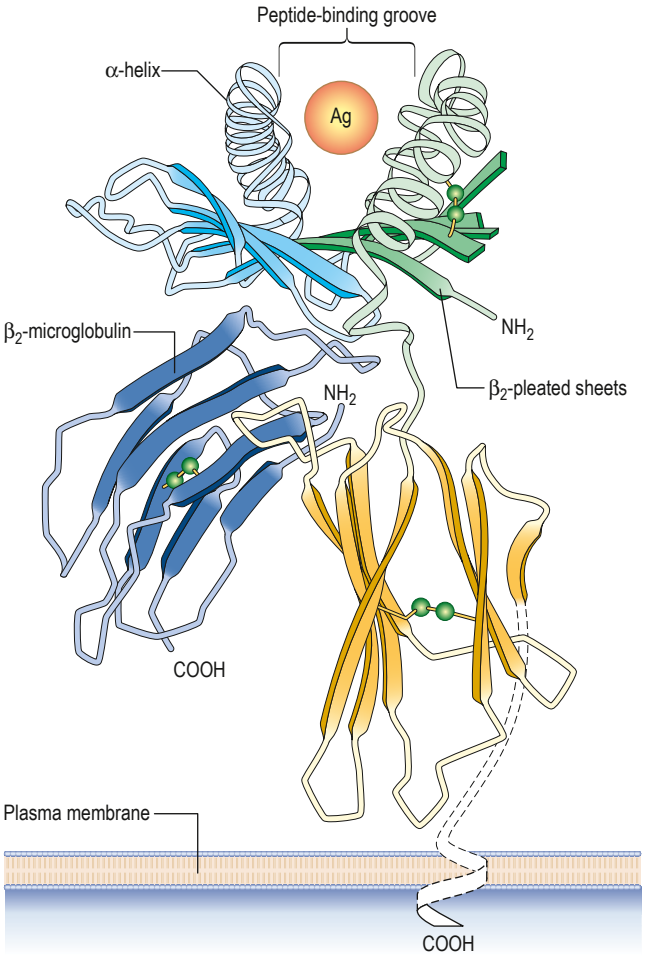


Fig. 6.12 Structure of human leucocyte antigen (HLA) class I molecule. The peptide-binding groove is used to present antigen (Ag) to T lymphocytes.

expressed, allow considerable breadth in the number of HLA molecules that an individual expresses.

- The existence of polymorphisms at each locus provides great breadth in the number of HLA molecules expressed at a population level. The polymorphic forms of HLA molecules differ predominantly in the peptide-binding groove (see Fig. 6.12).

Overall, then, each human can bind a range of peptide epitopes from pathogens to enhance individual protection; similarly, the population has an even greater range of protection to ensure population survival.

Antigen presentation

HLA molecules bind short peptide fragments that are processed ('chopped up') from larger proteins (antigens) derived from pathogens. The peptide–HLA complex is presented on APCs for recognition by TCRs on T lymphocytes. There are three major routes to antigen processing and presentation:

- *The endogenous route* (Fig. 6.14) is a property of all nucleated cells; the internal milieu is sampled to generate peptide–HLA class I complexes for display ('presentation') on the cell surface. In a healthy cell, the peptides are derived from self proteins in the cytoplasm (see Fig. 6.14) and are ignored by the immune system. In a virus-infected cell, viral proteins are processed and presented. The resulting viral peptide–HLA class I complex is presented to CD8 T lymphocytes that have cytotoxic (killer) function. In an immune response against a virus infection, CD8 T lymphocytes recognizing viral peptide–HLA complexes on the surface of an infected cell will kill it as a means to limit and eradicate infection.
- *The exogenous route* (Fig. 6.15) is a property of APCs; the external milieu is sampled. Antigens are internalized, either in the process of phagocytosis of a pathogen, through pinocytosis, or through specialized surface receptors (e.g., for antigen/antibody/complement complexes). The antigen is broken down by a combination of low pH and proteolytic enzymes for 'loading' into HLA class II molecules. At the APC surface, the pathogen peptide–HLA

Box 6.13 Examining the immune system in the clinical immunology laboratory

	Measurement	Interpretation
Proteins	C-reactive protein	Raised levels indicate infection or inflammation
	Total immunoglobulins	Low levels indicate antibody deficiency, usually a result of underlying disease or primary immunodeficiency. High levels, for example ↑ IgM, are seen in acute viral infection (e.g., hepatitis A), or may be present in plasma cell disorders such as myeloma
	IgG subclasses	Specific reductions in IgG subclasses (1–4) may indicate immune deficiency. High levels of IgG4 are observed in IgG4 disease
	Complement level and function	Low levels indicate consumption of complement in immune complex disease or primary complement deficiency, which can be confirmed with testing of the alternative or classical pathway function
	IgE	Raised levels are seen in allergy; allergen-specific IgE is useful to pinpoint the inciting stimulus (e.g., pollen, grass), component allergens may be useful in assessing anaphylaxis risk or cross-reactivity between allergens
Cells	Neutrophils	High levels are seen in bacterial infection; low levels in secondary immune deficiency
	Eosinophils	High levels are seen in allergic or parasitic disease
	CD4 T lymphocytes	Low levels are seen in HIV infection
Function (these are specialist tests and not available in routine clinical practice)	Neutrophil respiratory burst	Absent in immune deficiency chronic granulomatous disease
	T lymphocyte proliferation	Abnormally low in primary T-cell immune deficiency disease
Autoantibodies (see also Box 6.18)	Rheumatoid factor, anti-citrullinated peptide antibodies (ACPA)	Rheumatoid arthritis
	Double-stranded DNA autoantibodies	Systemic lupus erythematosus
	Acetylcholine receptor antibodies	Myasthenia gravis
	Anti-neutrophil cytoplasmic antibodies (ANCA)	Vasculitis
	Mitochondrial	Primary biliary cholangitis

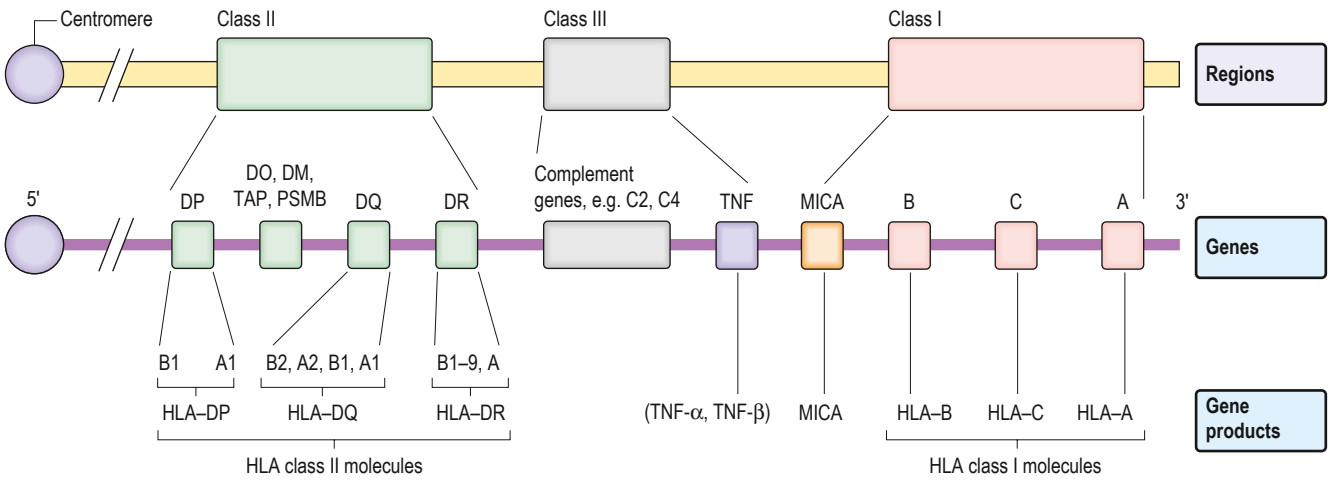


Fig. 6.13 The human leucocyte antigen (HLA) system in humans. On chromosome 6 are three major HLA regions (classes I–III), including genes that encode the HLA class I and II molecules, complement genes, the cytokine tumour necrosis factor, and other genes involved in antigen presentation (HLA-DO, HLA-DM; transporter associated with processing, TAP; proteasome subunit beta, PSMB; and the non-classical HLA class Ic molecule, MICA).

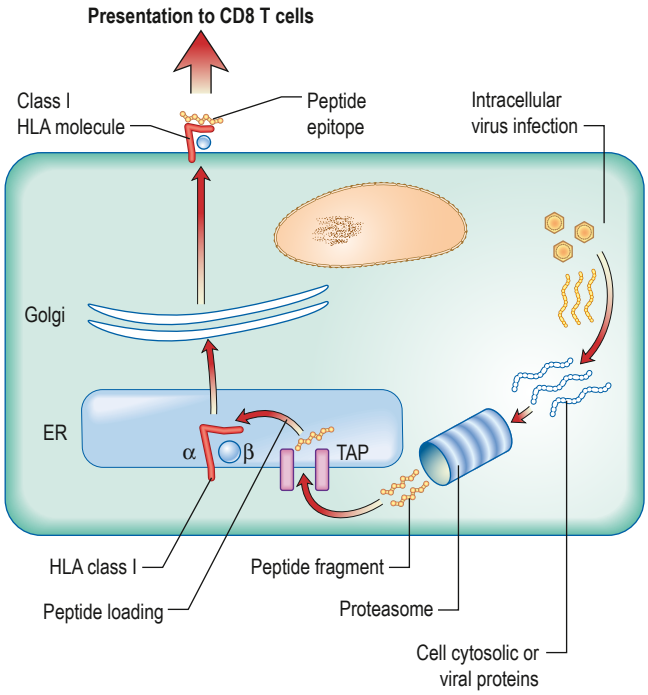


Fig. 6.14 The endogenous route of antigen presentation. Antigen is presented to CD8 T lymphocytes. Cytosolic proteins derived from self (resting cells) or viruses (virus-infected cells) are broken down by the proteasome (a sort of 'protein recycler') into fragments 2–25 amino acids long. Some of these are taken into the endoplasmic reticulum by a transporter (TAP), trimmed, and loaded into empty human leucocyte antigen (HLA) class I molecules. These are exported to the surface membrane for presentation to CD8 T lymphocytes.

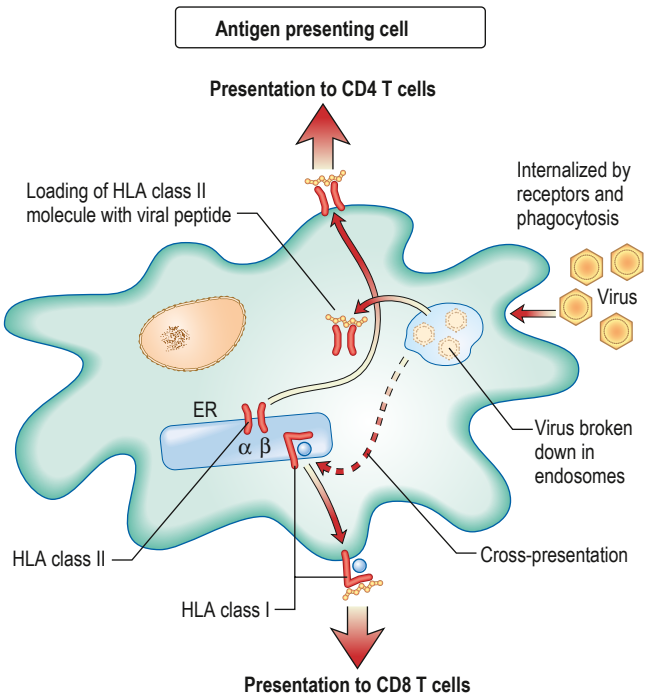


Fig. 6.15 The exogenous route of antigen presentation and cross-presentation. External material (e.g., virus particles) is taken into an antigen-presenting cell and broken down into specialized compartments by a combination of low pH and proteolysis. Peptides are then loaded into human leucocyte antigen (HLA) class II molecules for presentation to CD4 T lymphocytes. Material may also be transferred across the cell so that it is loaded on to HLA class I molecules for presentation to CD8 T lymphocytes. This process of cross-presentation is restricted to specialized antigen-presenting cells such as dendritic cells. *ER,* Endoplasmic reticulum.

class II complex is presented to, and able to interact with, CD4 T lymphocytes. Presentation by DCs can initiate an adaptive immune response by activating a naive, pathogen-specific CD4 T lymphocyte. Presentation by monocyte/macrophages and B lymphocytes can maintain and enhance this response by activating effector and memory pathogen-specific CD4 T lymphocytes.

- ***Cross-presentation*** refers to the ability of some APCs (mainly cDC1s) to internalize exogenous antigens and process them through the endogenous route (see Fig. 6.14). This is an essential component in the activation of CD8 cytotoxic T-cell responses against a virus.

Further reading

Dhatchinamoorthy K, Colbert JD, Rock KL. Cancer immune evasion through loss of MHC class I antigen presentation. *Front Immunol.* 2021; 12:636568.
Ishina IA, Zakharova MY, Kurbatskaia IN, Mamedov AE, Belogurov AA Jr, Gabibov AG. MHC class II presentation in autoimmunity. *Cells.* 2023; 12(2):314.
Liu J, Zhang X, Cheng Y, Cao X. Dendritic cell migration in inflammation and immunity. *Cell Mol Immunol.* 2021; 18(11):2461–2471.
http://hla.alleles.org/ *Information on the human MHC.*

THE IMMUNE SYSTEM IN CONCERT

Acute inflammation: events and symptoms

This is the early and rapid host response to tissue injury. To take a bacterial soft tissue infection and abscess formation as a classic example:

- Local expansion of pathogen numbers leads to direct activation of complement in the tissues, with ensuing degranulation of mast cells.
- Inflammatory mediators (from mast cells and complement) change the blood flow and attract and activate granulocytes (neutrophils).
- Concomitantly, there are the local symptoms of heat, pain, swelling, and redness, and perhaps more systemic symptoms such as fever due to the effect of circulating cytokines (IL-1, IL-6, TNF-α) on the hypothalamus. Indeed, gene mutations that lead to excessive actions of IL-1 (e.g., mutation of the *IL1RN* gene, which encodes a natural IL-1 antagonist) give rise to rare disease with just these symptoms, as well as bone erosion and skin rashes, which are treatable with IL-1 blockade using soluble IL-1 receptor antagonist (anakinra) or monoclonal anti-IL-1β antibody.
- Systemically active mediators (especially IL-6) also initiate the production of C-reactive protein in the liver.
- Bacterial lysis follows through the actions of complement and neutrophils, leading to formation of fluid in the tissue space containing dead and dying bacteria and host granulocytes ('pus').
- At the site of pathogen entry, there is often relative tissue hypoxia. The low oxygen tension has the effect of amplifying the responses of innate immune cells and suppressing the response of adaptive immune cells. This is probably an effective means of preventing excessive immune activation, which can result in collateral damage (Fig. 6.16).
- The inflammation may become organized and walled off through local fibrin deposition to protect the host.
- Antigens from the pathogen travel via the lymphatics (which may become visible as red tracks in the superficial

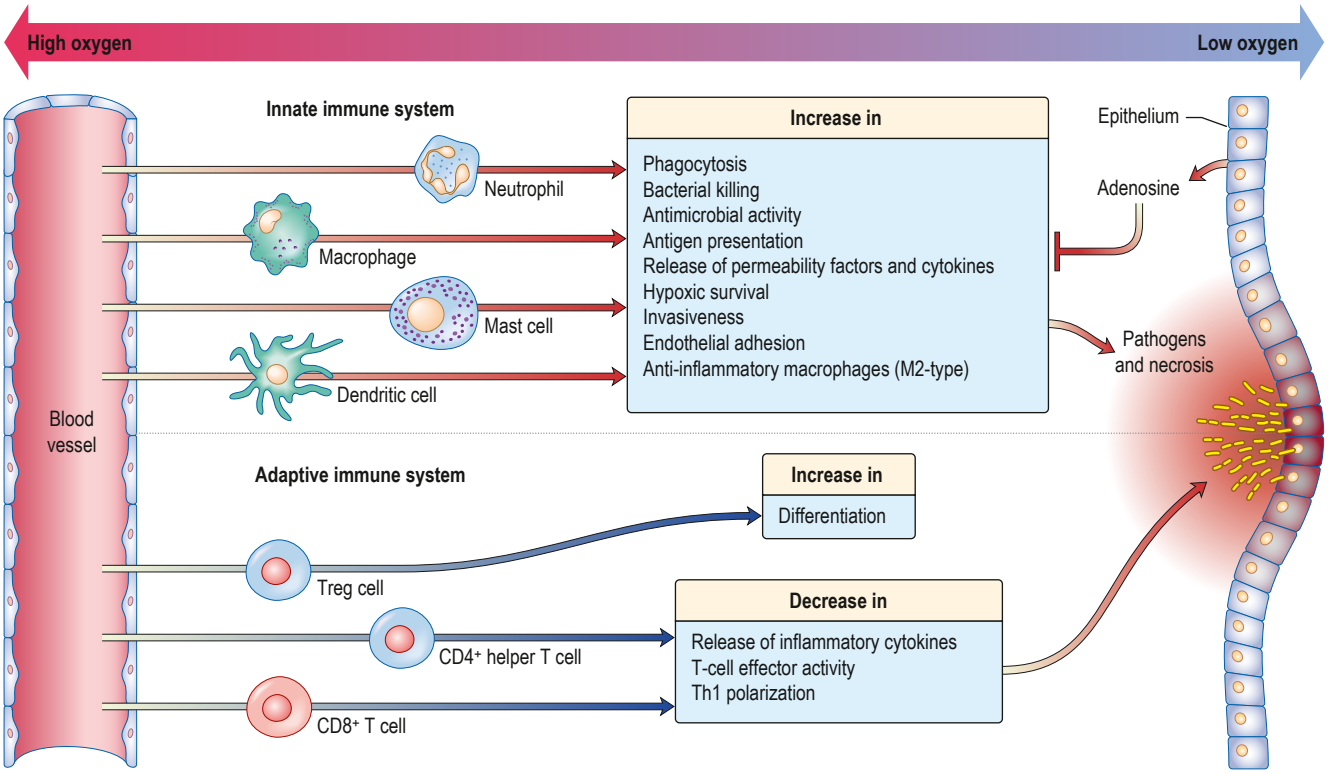


Fig. 6.16 Hypoxia and inflammation. At the site of pathogen invasion and necrosis, the oxygen tension is low *(right)*. Hypoxia acts on the inflammatory process by amplifying innate immune cell responses and dampening down the adaptive response. Overall, this benefits the host by avoiding an excessive, uncontrolled immune response.

tissues—lymphangitis) in soluble form or are carried by DCs to establish an adaptive immune response, which, at the first host–pathogen encounter, takes approximately 7 to 14 days. DCs are activated via the PRR–PAMP system (see Fig. 6.5).

- The adaptive immune response leads to activation of pathogen-specific T lymphocytes and of B lymphocytes and production of pathogen-specific antibody, initially of the IgM class and of low to moderate affinity, and subsequently of the IgG class (or IgA if the infection is mucosal) and of high affinity.
- Resolution of the infection is aided by the scavenging activity of tissue macrophages.

Chronic inflammation: events and symptoms

Inflammation arising in response to immunological insults that cannot be resolved in days or weeks gives rise to chronic inflammation. Examples include infectious agents (viruses that cause chronic infections such as hepatitis B and C or intracellular bacteria such as mycobacteria) and environmental toxins (such as asbestos and silicon). At the intracellular level, key processes of inflammasome generation and autophagy serve to enhance the chronic inflammatory process. Chronic inflammation is also a hallmark of some forms of allergic disease, autoimmune disease, and organ graft rejection.

The common feature of these pathological processes is the difficulty of removing the inciting stimulus. For example, some viruses and mycobacteria remain hidden intracellularly. In many ways, the pathology that results is inadvertent: the immune system is caught between the repercussions of not dealing with the infection or insult and the tissue damage that is caused by chronic activation of lymphoid and mononuclear cells.

Mycobacterium leprae

Leprosy, also known as Hansen's disease, is a chronic infectious disease caused by the bacterium *Mycobacterium leprae*. The immune responses that lead to different clinical manifestations of leprosy, specifically lepromatous and tuberculous leprosy, are primarily determined by the interactions between the immune system and the bacteria, so that the efficacy of cell mediated immunity determines the clinical manifestation of infection (see leprosy, Chapter 42):

- ***Lepromatous leprosy*** is characterized by a weak cellular immune response and a robust humoral immune response. In this form of the disease, the immune system fails to mount an effective defence against *M. leprae*, allowing the bacteria to multiply and spread extensively. Key features of lepromatous leprosy include:
  - Humoral Response: Antibodies, particularly IgG and IgM, are produced against *M. leprae*, leading to the formation of immune complexes. However, these immune complexes are not efficient at eliminating the bacteria.
  - Th2 Dominance: There is a predominance of Th2-type immune responses, leading to the production of cytokines such as IL-4 and IL-10. Th2 responses are associated with the production of antibodies and a weak cellular immune response.
  - Macrophage Dysfunction: Macrophages, which are immune cells responsible for engulfing and destroying pathogens, are not effectively activated in lepromatous leprosy. This allows

the mycobacteria to replicate within the macrophages and leads to the characteristic skin lesions and nodules observed in this form of the disease.

- ***Tuberculous leprosy***, also known as paucibacillary leprosy, is characterized by a stronger cellular immune response compared to lepromatous leprosy. The immune system is better equipped to control the growth of *M. leprae*, resulting in milder clinical manifestations. Key features of tuberculous leprosy include:
  - Cellular Response: There is a more effective cellular immune response, involving T cells, particularly $CD4^+$ and $CD8^+$ T cells. These T cells release cytokines such as IFN-γ, which activate macrophages to destroy the bacteria.
  - Th1 Dominance: Tuberculous leprosy is associated with Th1-type immune responses, characterized by the production of IFN-γ and other pro-inflammatory cytokines. Th1 responses are crucial for controlling intracellular pathogens.
  - Granuloma Formation: Granulomas, which are organized collections of immune cells, form around infected cells to contain the infection and limit bacterial spread. In tuberculous leprosy, granulomas are more likely to form and contribute to disease containment.

Recall responses as diagnostic markers

The cell-mediated immune response can be measured as a **recall response**. This forms the basis of diagnostic testing for mycobacterial infection and more commonly for **mycobacterium tuberculosis** (see Chapter 20):

- The measurement of the recall response can be *in vivo* as a skin test or ***Mantoux test***, where antigen is injected under the skin and primed T-cells migrate to and recognize the TB antigens causing local inflammation on the skin as a hallmark of current or previous infection.
- An alternative approach is to take blood, incubate the cells with TB antigens, and measure any interferon gamma released by previously activated T-cells. This is known as an ***interferon gamma release assay*** (IGRA).

Neither IGRA nor Mantoux tests are reliable in severe T-cell immunodeficiency diseases (e.g., advanced HIV infection), and neither can distinguish between latent or active infection. Other clinical disciplines use recall responses to determine if someone is sensitized to a specific antigen to a point that it would cause clinical symptoms; for example, in dermatology patch testing applies antigens to the skin to determine possible causes of contact dermatitis.

Crohn's disease

While in leprosy the cause of chronic inflammation is known, in several immune-mediated and autoimmune diseases (such as Crohn's disease) it is unclear what kick-starts or maintains the inflammatory process. When inflammation is sufficiently chronic, it may take on the appearance of organized lymphoid tissue resembling a lymph node germinal centre (e.g., in the joints in rheumatoid arthritis, Chapter 32). Alternatively, massive cytokine production by T lymphocytes and APCs, which contributes to local tissue damage, may give rise to the same sort of granulomata that are observed in mycobacterial disease. The granulomata 'wall off' the inciting stimulus but may also result in fibrosis and calcification. Symptoms typically relate to the site of the inflammation, so in Crohn's disease there is gut inflammation, bleeding, and fibrosis. The immune response may also be responsible for systemic effects such as fever and weight loss.

Genome-wide association studies (GWAS, Chapter 5) can help reveal the genetic basis for such chronic inflammatory disorders. Several of the polymorphisms associated with Crohn's disease reside in genes known to be involved in inflammasome induction, such as *NOD2*. Cytokine regulation has also emerged as a critical disease pathway in GWAS studies on Crohn's disease: most notably, *IL-23R* (the IL-23 receptor gene), which influences Th17 cell differentiation. Clues like these point to patients with Crohn's disease having impaired ability to control or terminate inflammasome activity, as well as a predisposition to make polarized pro-inflammatory cytokine responses.

Further reading

Zhen Y, Zhang H. NLRP3 inflammasome and inflammatory bowel disease. *Front Immunol.* 2019; 10:276.

LABORATORY INVESTIGATIONS OF THE IMMUNE SYSTEM

In the clinical immunology laboratory, proteins and cells can be measured to ascertain the status of the immune system. The results may reveal non-specific evidence of inflammatory or infectious disease (e.g., high C-reactive protein level); a state of immune deficiency (e.g., low concentration of IgG); or a state of immune pathology (e.g., the presence of autoantibodies or allergen-specific IgE).

Examples of the more common tests and their interpretation are shown in **Box 6.13**.

DISORDERS OF IMMUNODEFICIENCY

Secondary (acquired) versus primary immunodeficiency

Most forms of immunodeficiency that are encountered in clinical practice are secondary to infection (mainly HIV) or therapy (e.g., corticosteroids, monoclonal antibody [MAb] therapy, cytotoxic anticancer drugs, bone marrow ablation pre-transplant). Examples of other secondary immunodeficiencies are:

- ***Acquired neutropenias***, which are common (e.g., due to myelosuppression by disease or drugs or the increased rate of destruction in hypersplenism or autoimmune neutropenia) and carry a high risk of infection once the neutrophil count falls below 0.5×10^9/L.
- ***Acquired reductions in levels of immunoglobulins*** (hypogammaglobulinaemia), which are seen in patients with myeloma and chronic lymphocytic leukaemia or lymphoma. The use of chimeric antigen receptor (CAR)-T-cell therapy and B lymphocyte targeting monoclonal antibodies (e.g., rituximab) can cause significant depletion of B-cell numbers which has significantly added to the burden of secondary antibody deficiency.
- ***Impairment of defence against capsulated bacteria***, especially pneumococcus, following splenectomy; such patients should receive pneumococcal, meningococcal, and Hib vaccinations as a matter of course and antibiotic prophylaxis (see Box 38.17).

Primary immunodeficiencies are individually rare, but with now over 400 single gene entities recognized, collectively they are sufficiently common that most doctors will treat some of these patients.

Box 6.14 Classification of immunodeficiencies and the main diseases in each category[a]

Immune component	Examples of diseases
T lymphocyte deficiency	DiGeorge's syndrome
	Acquired immunodeficiency syndrome (AIDS)/human immunodeficiency virus (HIV) infection
	T-cell activation defects (e.g., CD3γ chain mutation)
	X-linked hyper-IgM syndrome (XHIM; CD40L deficiency)
B lymphocyte deficiency	X-linked agammaglobulinaemia (XLA)
	Common variable immunodeficiency (CVID)
	Selective IgA deficiency (IgAD)
Combined T- and B-cell defects	Severe combined immunodeficiency (SCID) (e.g., due to defects in common γ chain receptor for IL-2, 4, 7, 9, 15)
T-cell antigen–presenting cell interactions	IFN-γ receptor deficiency
	IL-12 deficiency and IL-12 receptor deficiency
Neutrophil defects	Chronic granulomatous disease (CGD)
	Leucocyte adhesion deficiency (LAD)
Deficiency of complement components	Classical pathway
	Alternative pathway
	Common pathway
	Regulatory proteins
	Mannose-binding lectin

[a]Apart from HIV infection, all the diseases shown here are primary immunodeficiencies.

Box 6.15 Immune defects and associated infections

Neutropenia and defective neutrophil function
- *Staphylococcus aureus*
- *Staphylococcus epidermidis*
- *Escherichia coli*
- *Klebsiella pneumoniae*
- *Proteus mirabilis*
- *Pseudomonas aeruginosa*
- *Serratia marcescens*
- *Bacteroides* spp.
- *Aspergillus fumigatus*
- *Candida* spp. (systemic)

Opsonin defects (antibody/complement deficiency)
- Pneumococcus
- *Haemophilus influenzae*
- Meningococcus
- *Streptococcus* spp. (capsulated)

Antibody deficiency only
- *Campylobacter* spp.
- *Mycoplasma* spp.
- *Ureaplasma* spp.
- Echovirus
- *Pseudomonas* spp.

Complement lytic pathway defects (C5–9)
- Meningococcus
- Gonococcus (disseminated)

Defect in T-cell or T-cell antigen–presenting cell responses
- *Listeria monocytogenes*
- *Legionella pneumophilae*
- *Salmonella* spp. (non-typhi)
- *Nocardia asteroides*
- *Mycobacterium tuberculosis*
- Atypical mycobacteria, especially *Mycobacterium avium-intracellulare*
- *Candida* spp. (mucocutaneous)
- *Cryptococcus neoformans*
- *Histoplasma capsulatum*
- *Pneumocystis jirovecii*
- *Toxoplasma gondii*
- Herpes simplex
- Herpes zoster
- Cytomegalovirus
- Epstein–Barr virus

They arise at birth as the congenital effect of a developmental defect or as a result of a genetic abnormality (Box 6.14). Gene defects may not manifest until later in infancy or childhood, and some forms of immunodeficiency typically present in adolescence or adulthood as a combined effect of genetics and environment.

Clinical features of immunodeficiency

The infections associated with immunodeficiency have several typical features:
- They are often chronic, severe, or recurrent.
- They resolve only partially with antibiotic therapy or return soon after cessation of therapy.
- The organisms involved are often unusual ('opportunistic' or 'atypical').

The pattern of infection, in terms of the type of organism involved, is indicative (Box 6.15):
- ***'Opportunistic' organisms*** are of low virulence but become invasive in immunodeficient states, for example atypical mycobacteria, *Pneumocystis jirovecii*, *Staphylococcus epidermidis.*
- ***Phagocyte defects*** cause deep skin infections, abscesses, and osteomyelitis.
- ***Defective antibody*** production is associated with infections with pyogenic ('pus-forming') bacteria.
- ***T lymphocyte deficiency*** causes infection with fungi, protozoa, and intracellular microorganisms or persistent viral infection.
- ***Congenital deficiencies*** of antibody production are not revealed for several months after birth, due to the 28-day half-life of maternal IgG.

The family history may reveal unexplained sibling death, the fact that only males of the family are affected (X-linked), or consanguinity; each of these makes a primary genetic syndrome more likely. In babies with severe combined immunodeficiencies (SCID) it is possible for high numbers of memory T-cells to be detected. The source of these cells is either engraftment of maternal cells ***(materno-fetal engraftment)***, an expansion of a few viable clones of the babies' own cells, known as ***Omenn syndrome***, or the inadvertent transfusion of cells as part of healthcare intervention.

Primary immunodeficiency

T lymphocyte deficiency

Complete DiGeorge's syndrome

DiGeorge's syndrome arises in 1 to 5 per 100 000 of the population due to deletions on the long arm of chromosome 22. It has a highly variable presentation, including ***c***ardiac abnormalities, ***a***bnormal

facies, ***t***hymic dysfunction, ***c***left deformities, and ***h***ypocalcaemia/hypoparathyroidism, remembered by the acronym 'CATCH-22'. In those DiGeorge individuals born without a thymus, the lack of a location for T lymphocyte development means that they have reduced or absent T lymphocyte number and proliferation responses. Apart from calcium supplementation, correction of cardiac abnormalities, and prophylactic antibiotics, cure has been reported with thymic transplantation using fetal tissue or stem cell transplant (SCT) from HLA identical siblings. Other athymic conditions include Foxn1 deficiency and CHARGE syndrome (***c***oloboma, ***h***eart defects, choanal ***a***tresia, ***r***etardation of ***g***rowth and development, ***e***ar abnormalities) caused by CHD7 mutations.

Other T lymphocyte deficiencies

Other T lymphocyte deficiencies caused by single-gene defects have been characterized and typically present from the age of 3 months with candidal infections of the mouth and skin, protracted diarrhoea, fever, and failure to thrive. These disorders are similar in presentation and management to the combined (T and B lymphocyte) immunodeficiencies.

T and B lymphocyte deficiency

Severe combined immunodeficiencies

Severe combined immunodeficiencies are a heterogeneous group of rare (1–2 live births per 100 000), genetically determined disorders resulting from impaired T, NK, and B lymphocyte immunity. The most common form of SCID is an X-linked defect in the IL-2 receptor γ chain, interfering with the function not just of IL-2 but also of IL-4, IL-7, IL-9, and IL-15, which share this component in their receptors. Clinical features are similar to those of pure T lymphocyte deficiency and, in the blood, T and NK cells are lacking while B lymphocyte numbers may be normal. The most successful treatment for all forms of SCID is SCT, preferably from HLA-identical donors such as siblings. For a few disorders, such as ADA (adenosine deaminase) and γ chain-deficient SCID, gene therapy is possible if no appropriate HLA-matched donor can be found.

Hyper-IgM syndrome

This mixed or combined immunodeficiency may not be diagnosed until later in life. It usually results from an X-linked defect in the gene encoding CD40 ligand (CD40L, also called CD154). Signalling between CD40 on B cells and CD40L on T cells is necessary for the generation of class-switched B cells bearing high-affinity immunoglobulin and for mature T-cell responses. In CD40L deficiency, B lymphocytes are unable to class-switch, producing only IgM. Levels of other types of immunoglobulin, IgG and IgA, are absent or low and there are mild defects in T-cell function. Autosomal forms are also seen with defects in CD40, activation-induced cytidine deaminase (AID), and uracil-DNA glycosylase (UNG).

Opportunistic infections occur: for example, with *Pneumocystis*, *Cryptosporidium*, *Candida*, *Cryptococcus* and herpes viral infections. Treatment is with replacement immunoglobulin, antifungals, antibiotics, and avoidance of *Cryptosporidium* by boiling water. SCT is used in some patients.

Wiskott–Aldrich syndrome

This is an X-linked defect (at Xp11–23) in a gene involved in signal transduction and cytoskeletal function. It is associated with eczema and thrombocytopenia; a mainly cell-mediated defect with falling immunoglobulins is seen, and autoimmune manifestations and lymphoreticular malignancy may develop. Activating mutations in the same gene have been described that cause severe congenital neutropenia, highlighting the key role of Wiskott–Aldrich serine protease in immune regulation.

Ataxia telangiectasia

Patients with this disease have defective DNA repair mechanisms and have cell-mediated defects with low IgA and IgG_2. Lymphoid malignancy is common and a shortened life expectancy often results from the neurological disease.

Epstein–Barr virus-associated immunodeficiencies

Apparently normal but genetically predisposed individuals develop overwhelming Epstein–Barr virus (EBV) infection, polyclonal EBV-driven lymphoproliferation, combined immunodeficiency, aplastic anaemia, and lymphoid malignancy. EBV appears to act as a trigger for the expression of a hitherto silent immunodeficiency. Many genetic susceptibilities to EBV have now been described, including X-linked lymphoproliferative disease 1 and 2 (XLP1 and XLP2), which demonstrate defects in the *SH2D1A* and *XIAP* genes, respectively, and may also present with Crohn's-like gut inflammation (see Chapter 22). Other genetic causes include mutations in *CD70*, *CD27*, *ITK*, and *CARD11*.

Antibody deficiencies

Antibody deficiency syndromes may be caused by intrinsic defects in B lymphocyte development, resulting in absence of B lymphocytes or specific subsets. Antibody deficiency due to inborn errors of immunity also occurs when normal antibody production is not well sustained. Intrinsic developmental defects tend to present 6 months after birth when transplacentally transferred maternal IgG has declined, while other syndromes may present variably throughout childhood and adult life. The clinical features of antibody deficiency are severe, recurrent sinopulmonary infection leading to bronchiectasis and persistence of gut pathogens (e.g., *Giardia*) with resultant chronic diarrhoea and malabsorption. Morbidity and mortality are high if the conditions are untreated, mainly due to chronic lung disease and central nervous system infections with enteroviruses.

X-linked agammaglobulinaemia

X-linked agammaglobulinaemia is an intrinsic defect in B lymphocyte maturation. The cause is a loss-of-function mutation of Bruton's tyrosine kinase, which is involved in cell activation and maturation. B lymphocyte development is arrested at the pre-B stage, and B lymphocytes are absent in the blood. Clinical presentation follows the fall in protective maternal IgG.

Common variable immunodeficiencies

Common variable immunodeficiencies are a heterogeneous group of disorders arising during late childhood and early adulthood, with very low IgG levels. Although rare (1/50 000), it is the most common severe primary immune deficiency and carries a significant risk of morbidity and mortality. Autoimmunity and granulomatous disease are found in up to 15% of cases. Genetic defects underlying the condition have revealed mutations in multiple genes: in Europe the most common is haploinsufficiency of *NFKB1*. In most patients, the underlying defect remains unknown.

Selective IgA deficiency

Selective IgA deficiency (IgAD; serum IgA <0.05 g/L), with normal levels of IgG and IgM, is found in approximately 1/600 Northern

Europeans. Most individuals are asymptomatic but numerically this is the most common primary immunodeficiency. The underlying defect is unknown. Patients may present at any age with recurrent infections caused by pyogenic organisms, particularly affecting mucosal sites.

Treatment of antibody deficiencies

Treatment of IgAD is not usually required; when it is, antibiotics are used. Some IgAD patients produce anti-IgA antibodies of the IgG and IgE classes, and so infusion of exogenous IgA (e.g., during a blood transfusion) could result in anaphylaxis. IgAD patients should therefore be screened for anti-IgA antibodies if there is a history of transfusion reaction and transfused with appropriate precautions if positive.

For patients with X-linked agammaglobulinaemia, common variable immunodeficiencies, or another major antibody deficiency, treatment is with replacement immunoglobulins (intravenous or subcutaneous, intravenous immunoglobulin [IVIg] or SCIg) to a level that controls infections. Trough IgG concentrations should remain well within normal limits (i.e., >6 g/L), and infections should be treated promptly and aggressively with antibiotics. Prognosis has improved in recent years as more patients survive into adulthood, but chronic lung disease and lymphomas remain life-threatening complications.

Defects in antigen-presenting cell function

A series of rare genetic defects have been uncovered in which APCs demonstrate an inability to mount protective responses to intracellular bacteria, particularly low-virulence mycobacteria and salmonella. The axis affected is the interplay between CD4 T lymphocytes and APCs that drives Th1 responses and therefore, in turn, activates mononuclear cells such as macrophages to kill and eradicate intracellular pathogens.

Neutrophil defects

Chronic granulomatous disease

Chronic granulomatous disease is a rare (1/250 000) immunodeficiency that is caused by a defect in neutrophil killing; it is characterized by deep-seated infections. The functional defect is an inability to generate antibacterial metabolites through the respiratory burst (see **Box 6.5**). X-linked and autosomal forms are described. Typical onset is at toddler age. Neutrophil numbers are normal or increased. A simple respiratory burst test of neutrophil function is diagnostic. The dysfunctional neutrophils' inability to clear infection adequately and frustrated apoptosis lead to granulomatous inflammation, and CGD may also present with a Crohn's-like gut disease (see Chapter 22).

Treatment of infections and prophylactic antibiotic and antifungal therapy are required. In many cases, SCT is necessary. For the X-linked form, gene therapy is a treatment option when a suitably matched donor cannot be found.

Leucocyte adhesion deficiency

Leucocyte adhesion deficiency results from defects in integrins (see **Fig. 6.11**). Numerous underlying defects in the genes encoding one of the component chains, *CD18*, have been described in LAD-1. LAD-1 has an autosomal recessive inheritance, presenting almost immediately after birth with delayed umbilical cord separation. In later life, there is a characteristic failure to lose primary dentition and severe gingivitis. Recurrent infections appear during the first decade of life. Blood neutrophil levels are high but cells are absent from the sites of infection, which require aggressive antimicrobial and antifungal treatment and SCT for cure.

Hyper-IgE syndrome (Job's syndrome)

This is an autosomal dominant (occasionally sporadic) immune disorder with high serum IgE levels, dermatitis, boils, pneumonias with cyst formation, and bone and dental abnormalities. Mutations in *STAT3*, *DOCK8*, and *ZNF341* have been found.

Shwachman–Diamond syndrome

This can resemble cystic fibrosis clinically, with exocrine pancreatic insufficiency and pyogenic infections. A mild neutropenia is associated with a defect of neutrophil migration.

Chédiak–Higashi syndrome

This rare recessive disorder is caused by a mutation in the lysosomal trafficking regulator gene *(LYST)* on chromosome 1q42–45. There are defects in neutrophil function with defective phagolysosome fusion, and large lysosome vesicles are seen in phagocytes. Patients have recurrent infections, neutropenia, anaemia, and hepatomegaly. As with Shwachman–Diamond syndrome, tissue infiltration with histiocytes is usually seen, including in bone marrow, and Chédiak–Higashi syndrome is one of several genetic causes of primary haemophagocytic lymphohistiocytosis.

Complement deficiency

The consequences of deficiency of complement proteins can be predicted from their functions (see **Fig. 6.2**):

- Failures in innate response components (e.g., the alternative pathway) lead to impaired non-specific immunity with an increase in bacterial infections.
- Failure of the classical pathway results in a tendency towards infection and also towards diseases in which immune complex deposition causes inflammation, such as SLE (e.g., C1 and C4 deficiency), vasculitis, and glomerulonephritis.

Neisserial infections (e.g., meningitis due to *N. meningitidis*) are often encountered in patients with complement defects of the membrane attack complex (C6–9). Lectin pathway deficiency in isolation rarely causes major ill health.

Complement regulatory proteins

Deficiency of C1 inhibitor ***(C1 esterase deficiency)*** is relatively rare. Since this enzyme is involved in regulation of several plasma enzyme systems (e.g., the kinin system) and is continuously consumed, a single parental chromosome defect resulting in 50% of normal production barely copes with the demand and fails under stress (hence has an autosomal dominant effect). As a result, uncontrolled activation of low-molecular-weight kinins may occur, leading to oedema of the deep tissues affecting the face, trunk, viscera, and airway, and explaining the alternative name of hereditary angio-oedema (HAE). Treatment is with C1 inhibitor concentrate or a selective bradykinin-2 receptor antagonist (icatibant). Steroids and antihistamines are not useful in this form of angio-oedema. Prophylaxis may be with specific kallikrein pathway inhibitors such as berotralstat (an oral kallikrein inhibitor) or lanadelumab. (MAb to Kallikrein). New genetic causes of HAE are now recognized, all with defects in the contact-kinin pathway. An acquired form of C1 esterase inhibitor deficiency (acquired angio-oedema, AAE) may be seen in lymphoproliferative disease or complicating autoimmunity. The treatment is largely for HAE.

Further reading

Bousfiha A, Moundir A, Tangye SG, et al. The 2022 update of IUIS phenotypical classification for human inborn errors of immunity. *J Clin Immunol.* 2022; 42(7):1508–1520.

Maurer M, Magerl M, Betschel S, et al. The international WAO/EAACI guideline for the management of hereditary angioedema-The 2021 revision and update. *Allergy.* 2022; 77(7):1961–1990.

DISORDERS OF EXAGGERATED IMMUNE RESPONSE

ALLERGIC DISEASE (IMMEDIATE HYPERSENSITIVITY)

Normally, host defence can cope with potentially harmful cells and molecules. Under some circumstances, a harmless molecule can initiate an immune response that can lead to tissue damage and death. Such exaggerated, inappropriate responses are termed hypersensitivity reactions or allergic disease.

In ***allergic hypersensitivity***, the binding of an antigen to specific IgE bound to its high-affinity receptor on a mast cell surface results in massive and rapid cell degranulation and the inflammatory response outlined in Chapter 11 ('Anaphylaxis' section). The antigens involved are typically inert molecules present in the environment.

The immediate effects of allergen exposure are often very florid ***(early phase response)***. Allergic disorders also have a second phase, occurring a few hours after exposure and lasting up to several days. These '***late phase responses***' are mediated by Th2 cells recognizing peptide epitopes of the allergen. Recruitment of eosinophils is often a prominent feature.

From a pathological and therapeutic viewpoint, the LPR gives rise to chronic inflammation, which is difficult to control. In ***asthma***, the LPR causes prolonged wheezing, which can be fatal. Immediate hypersensitivity is usually responsive to antihistamines but the LPR is not, requiring powerful immune modulators such as corticosteroids.

In immunopathological terms, in the LPR:

- Neutrophils and eosinophils are prominent in the first 6 to 18 hours and may persist for 2 to 3 days.
- Th2 cells accumulate around small blood vessels and persist for 1 to 2 days.
- Mediators responsible for the cellular infiltrate include platelet-activating factor and leukotrienes (**Box 6.16**).
- Th2 cytokines IL-4 and IL-5 and chemokines such as eotaxin act as growth and activation stimuli for eosinophils, which are capable of extensive tissue damage. Th2 cytokines are also responsible for the class switch of Ig production towards IgE, maintaining the cycle of immediate and late responses.

What makes an allergen so powerful?

Several allergens are proteolytic enzymes, allowing them to cross skin and mucosal barriers. They are often contained within small, aerodynamic particles (e.g., pollen grains) that gain access to nasal and bronchial mucosa.

Box 6.16 Mediators involved in the allergic response

Pre-formed mediators

- Histamine and serotonin:
  - Bronchoconstriction
  - Increased vascular permeability
- Neutrophil chemotactic factor (NCF) and eosinophil chemotactic factor (ECF):
  - Induction of inflammatory cell infiltration

Newly formed mediators (membrane derived)

- Leukotriene (LT) B_4:
  - Chemoattractant
- LTC_4, LTD_4, LTE_4 (slow-reacting substance of anaphylaxis, SRS-A):
  - Sustained bronchoconstriction and oedema
- Prostaglandins and thromboxanes:
  - Platelet-activating factor (PAF)
  - Prolonged airway hyperactivity

Why do some people react and others do not?

The tendency to develop allergic responses (known as atopy) shows strong heritability. Between 20% and 30% of the UK population is atopic, and two, one, or no atopic parents pass on the atopic trait to their children with a risk of 75%, 50%, and 15%, respectively. In developing nations, the tendency to allergy is estimated at one-tenth of the rate in industrialized countries. Among the predisposing genes are those encoding the β chain of the high-affinity receptor for IgE and IL-4, both strongly associated with Th2 pathways. The presence of Th2 cells recognizing allergens is the pathological hallmark of allergy.

What environmental factors are involved?

Early exposure to allergens (even *in utero*) may be a factor in developing atopy. Over-zealous attention to cleanliness (the hygiene hypothesis) in developed societies (use of antibiotics, reduced exposure to pathogens that might favour a Th1-like environment) may favour a reduction in Treg activity. This environmental factor is demonstrated by the rapid increase of allergy in the eastern part of Germany following reunification in 1990.

In clinical terms, approximately two-thirds of atopic individuals (who can be identified as those with circulating allergen-specific IgE) have ***clinical allergic disease*** (equating to 15%–20% of the UK population). Allergy accounts for up to one-third of school absences because of chronic illness. Allergic disorders include allergic rhinitis (hay fever), allergic eczema, bee and Wiskott–Aldrich serine protease venom allergy, some forms of food allergy, urticaria, and angio-oedema.

Diagnosis

Diagnosis of allergic disease is usually made on the history and backed up by skin-prick testing (insertion of a tiny quantity of allergen under the skin and measurement of the size of the weal) and/or measurement of serum allergen-specific IgE. Mast cell tryptase serum levels peak 1 to 2 hours after an event, remaining high for 24 hours.

Treatment

Avoidance is the first line of therapy. Other measures are as follows:

- ***Antihistamines*** are effective for many immediate hypersensitivity reactions (but have no role in the treatment of asthma).
- ***Corticosteroids*** have several well-identified modifying actions in the allergic process: production of prostaglandin and leukotriene mediators is suppressed, inflammatory cell recruitment and migration are inhibited, and vasoconstriction leads to reduced cell and fluid leakage from the vasculature.
- ***Cysteinyl leukotriene receptor antagonists*** inhibit leukotrienes (LTs) by blocking the type I receptor (e.g., montelukast, used in asthma, particularly the aspirin-induced type).
- ***Omalizumab*** is a MAb that binds IgE. It is used in severe asthma (see Chapter 20) that cannot be controlled with a corticosteroid plus a long-acting β_2 agonist. Treatment must be initiated in a specialist centre where staff have experience of treating severe, persistent asthma.
- ***Dupilumab*** is a MAb that inhibits the biological activity of both IL-4 and IL-13. It has been shown to be beneficial in severe atopic eczema and asthma.
- ***Desensitization (allergen immunotherapy)*** works on the principle that allergy can be prevented by inoculation if the allergen is given in a controlled way. Desensitization is indicated for disorders in which the hypersensitivity is IgE mediated, for example life-threatening allergy to insect stings, drug allergy, and allergic rhinitis. An induction course of subcutaneous injections of increasing doses of the allergen extract, given once every 1 to 2 weeks, is followed by maintenance injections monthly for

Box 6.17 Mechanism of action of desensitization for allergy

Probable mechanism	Effect
IgG blocking antibodies	During repeated exposure to desensitizing allergen, IgG class antibodies develop (especially IgG4 subclass); these compete with the pathogenic IgE for allergen binding and/or prevent IgE–allergen complexes from binding to mast cell high-affinity IgE receptors
Regulation	Exposure to repeated desensitizing allergen induces Treg cells, which recognize allergen but invoke regulatory immune responses, dampening down migration, infiltration, and inflammation
Immune deviation	A shift away from Th2- to Th1-producing CD4 cells results in the generation of cytokines (e.g., interferon-gamma) that are inhibitory to IgE production

2 to 3 years. A systematic review of 51 published randomized, placebo-controlled clinical trials, enrolling a total of nearly 3000 participants, showed a low risk of adverse events with consistent clinical benefit. From an immunological viewpoint, desensitization seems capable of modifying the allergic response at several levels (**Box 6.17**). Sublingual allergen immunotherapy (using grass pollen extract tablets of Phl p 5 from timothy grass) is used in hay fever that has not responded to anti-allergic drugs (the first dose is given under medical supervision) (see Chapter 21).

Anaphylaxis

This is a severe and life-threatening form of allergic reaction. The immediate management of a patient with an allergic reaction is covered in Chapter 11 ('Anaphylaxis' section).

In patients who experience anaphylaxis, the best treatment is prevention. Avoidance of triggering foods, particularly nuts and shellfish, may require almost obsessive self-discipline. Patient education is necessary and many are instructed in the self-administration of adrenaline (epinephrine) and carry pre-loaded syringes. Recent changes in UK law oblige retailers to list all allergens and ingredients in freshly prepared and pre-prepared food that they sell. Desensitization has a well-established place in the management of this disorder, particularly if exposure is unavoidable or unpredictable, as in insect stings. In peanut allergy, it has been shown that early introduction of peanuts in children modulates the response in later life, and this benefit extends to high-risk children with atopic eczema or allergy to another food. Further studies have shown that it is possible to give graded peanut exposure to those already sensitized and that oral desensitization can reduce the risk of anaphylaxis associated with inadvertent exposure in children with a history of peanut allergy. The addition of modified peanut peptides further enhances the success and tolerability of this approach.

Further reading

Du Toit G, Roberts G, Sayre PH et al. Randomised trial of peanut consumption in infants at risk for peanut allergy. *N Engl J Med.* 2015; 372:803–813.
Greenhawt M, Shaker M, Abrams EM. Peanut oral immunotherapy in very young children. *Lancet.* 2022; 399(10322):336-337.
http://www.allergen.org *Allergen nomenclature.*

AUTOIMMUNE DISEASE

Autoimmunity is when the immune response turns against self, that is recognizes 'self' antigens. The vast array of possible TCRs and antibodies that can be generated by the host make it highly probable that at least a small proportion can recognize self (i.e., are autoreactive). Moreover, a degree of autoreactivity is physiological: the TCR has evolved to interact both with the peptide epitope in the HLA molecule binding groove *and* with the HLA molecule itself.

The critical event in the development of autoimmune disease is when T and B lymphocytes bearing these receptors for 'self' become activated: this leads to loss of immunological tolerance. The major checkpoints that the immune system has in place to prevent this are:

1. removal of TCRs with very strong affinity for 'self' in the thymus
2. the presence of naturally arising regulatory T lymphocytes (Tregs)
3. the requirement for a danger signal to license DCs to activate CD4 T lymphocytes.

Failure of immune tolerance checkpoint 1: thymic education

During thymic education, TCRs with a dangerously high affinity for self are deleted. It has become apparent that this process relies on the thymic expression of self antigens. Situations that compromise the expression of a self protein would be expected to favour the development of autoimmunity. Indeed, in a rare group of patients who develop multiple autoimmune disorders affecting the adrenal and parathyroid glands (autoimmune polyglandular syndrome type 1, Chapter 34), there is a defect in the autoimmune regulator *(AIRE)* gene, which controls thymic expression of a host of self genes. When the gene malfunctions, there is reduced expression of self proteins in the thymus and autoimmune disease is a consequence.

Failure of immune tolerance checkpoint 2: regulatory T lymphocytes

An example of Treg failure is the defect in the gene encoding Foxp3, a critical transcription factor in Tregs, which leads to IPEX (see 'Regulatory T lymphocytes' section). IPEX is very rare but it serves to indicate how Treg defects can lead to autoimmune disease. Laboratory studies in this area are revealing subtle Treg defects in several autoimmune diseases (e.g., type 1 diabetes, multiple sclerosis, rheumatoid arthritis). The role of FoxP3 in regulating peripheral immune responses highlights the control of immune expansion and regulation of immunological 'space'. Another component of this process is the contraction of an immune response once the infection has been dealt with. Autoimmune lymphoproliferative syndrome is a disorder with defects in either Fas, its ligand (FasL or CD95), or the associated intracellular signalling pathway, which in healthy individuals allows immune cells to induce apoptosis in cells displaying abnormal DNA, for example, because of viral infection (**Fig. 6.9**). Individuals develop multiple cytopenias, organ-specific autoimmunity, and lymphoproliferation that may mimic lymphoma. The common problem is the inability to 'switch off' activated T cells. Treatment is with haemopoietic SCT if the disorder is severe. Many patients with multiple cytopenias, including Evans' syndrome (autoimmune thrombocytopenia and haemolytic anaemia), are revealed to have autoimmune lymphoproliferative syndrome defects when appropriately tested.

Failure of immune tolerance checkpoint 3: CD4 T lymphocyte activation against an autoantigen (or its mimic)

For an autoimmune disease to develop, there must be presentation of autoantigens to a naive, potentially autoreactive CD4 T lymphocyte by activated cDCs. Several theories exist to explain this:

- Tissue damage due to infection leads to both the release of hidden self antigens and the provision of sufficient danger signals

to activate cDCs, which in turn activate autoreactive CD4 T lymphocytes, as well as pathogen-specific ones. This is often termed 'bystander activation'.

- A pathogen mimics a self antigen. In the process of mounting an entirely appropriate immune response against the pathogen, T or B lymphocytes are generated that also have the capacity to recognize self. This is termed molecular mimicry.
- Post-translational modification. Self proteins and peptides are modified in such a way that they appear foreign. A good example is an antibody found in patients with rheumatoid arthritis, which recognizes citrulline-modified forms of self peptides.

It is unlikely that, for the common autoimmune diseases (**Box 6.18**), there is a 'single tolerance checkpoint' explanation. Rather, it is likely that multiple subtle defects, at various checkpoints, are at play.

Mechanisms of tissue damage in autoimmune disease

Fig. 6.17 illustrates potential mechanisms of immune damage in autoimmune disease.

Tissue damage via chronic inflammation

In a number of autoimmune diseases (e.g., type 1 diabetes, multiple sclerosis), autoreactive T lymphocytes are likely to be the main drivers of disease and tissue damage. Th1 cells, CTLs, and Th17 cells induce chronic inflammation and release cytokines that recruit other effector cells (macrophages) or are directly toxic (e.g., insulin-producing β cells are susceptible to the combined effects of IL-1β, IFN-γ, and IL-17 in type 1 diabetes).

Box 6.18 Some autoimmune diseases and their autoantigens

Disease	Antigens
Addison's disease	21 α-hydroxylase
Goodpasture's syndrome	Alpha-3 chain of type IV collagen
Granulomatosis with polyangiitis	Neutrophil proteinase 3
Graves' thyroiditis	Thyroid-stimulating hormone receptor
Hashimoto's thyroiditis	Thyroid peroxidase, thyroglobulin
Multiple sclerosis	Myelin basic protein Myelin oligodendrocyte glycoprotein
Myasthenia gravis	Acetylcholine receptor, muscle-specific kinase (MuSK)
Pemphigus vulgaris	Desmoglein-3
Pernicious anaemia	H^+/K^+-ATPase, intrinsic factor
Polymyositis/ dermatomyositis	Transfer RNA synthases (e.g., Jo-1, PL7, PL12)
Primary biliary cholangitis	Pyruvate dehydrogenase complex
Rheumatoid arthritis	Citrullinated cyclic peptide, IgM
Scleroderma	Topoisomerase
Sjögren's syndrome	Ro/La ribonuclear proteins
Systemic lupus erythematosus	Sm/RNP, Ro/La (SS-A/SS-B), histone and native double-stranded DNA
Type 1 diabetes	Pro-insulin, glutamic acid decarboxylase, IA-2, ZNT8
Vitiligo	Pigment cell antigens

IgG4 disease

Pathophysiology

IgG4-related disease (IgG4-RD) is characterized by chronic inflammation, fibrosis, and the infiltration of IgG4-positive plasma cells into a wide range of target tissues (**Box 6.19**). These include the pancreas, salivary glands, lacrimal glands, bile ducts, kidneys, lymph nodes, and others, with the specific clinical manifestations depending on the organs involved and varying widely between individuals. The exact cause of IgG4-RD is not fully understood; however, class-switched B-cells producing IgG4 are central to the pathogenesis and appear to respond to multiple auto-antigens.

IgG4 antibodies are a subclass of immunoglobulins with anti-inflammatory properties. However, their excessive production can lead to immune complex formation and tissue damage associated with dysregulated T-cell responses. An imbalance in cytokine signalling (with ILs 4 and 10 promoting IgG4 production and regulatory T-cell responses) promotes inflammation and fibrosis, and activation of fibroblasts can result in fibrosis or scarring of affected organs, which impairs their normal function. No unifying genetic predisposition has yet been identified.

Clinical features

IgG4-related disease is usually insidious in onset, consistent with the slow accumulation of inflammatory foci in tissues. At the time of diagnosis, there is often significant dysfunction, for example, in autoimmune pancreatitis there is loss of both exocrine and endocrine function (see 'Autoimmune chronic pancreatitis' section in Chapter 24). Clinical presentations may include pseudotumours with local anatomical disturbance (e.g., proptosis from retrobulbar lesions) to metabolic complications such as thyroiditis or pancreatitis. In some presentations, such as retroperitoneal fibrosis or pulmonary lesions, the diagnosis may only be considered following incidental imaging findings. The salivary glands and pancreas are the most commonly affected sites, with multiple organ involvement in the majority of cases.

The estimated incidence is 0.3–1 per 100 000 of population with a prevalence of 62 per million. There is equal sex distribution.

Diagnosis

The diagnosis is usually established by appropriate imaging findings of an affected site, and there are no universally accepted consensus criteria. A definite diagnosis requires:

- multiple affected sites with characteristic swelling or nodules (diagnosed clinically or radiologically)
- a significant rise in serum IgG4 (>135 mg/dL)
- the presence of classical histopathological findings with storiform fibrosis, high percentage of IgG4 positive plasma cells, or dense lymphocytic infiltrate with plasma cells and fibrosis.

Management

Glucocorticoids are usually seen as first-line management, although 40% may not achieve remission with steroids alone and their use is compounded by the high rate of pre-existing pancreatic endocrine insufficiency with pancreatic disease involvement. Remission rates are higher (and relapse rates lower) in individuals treated concomitantly with either low-dose cyclophosphamide (50–100 mg/day) or mycophenolate mofetil (1000–1500 mg daily). The use of B-cell depleting therapy (e.g., the anti-CD20 MAb rituximab) is often associated with rapid reduction in active lesions, although no therapy has any major impact on significantly fibrotic tissue. A single course of anti-CD20 therapy can induce remission for 6 to 18 months. Based

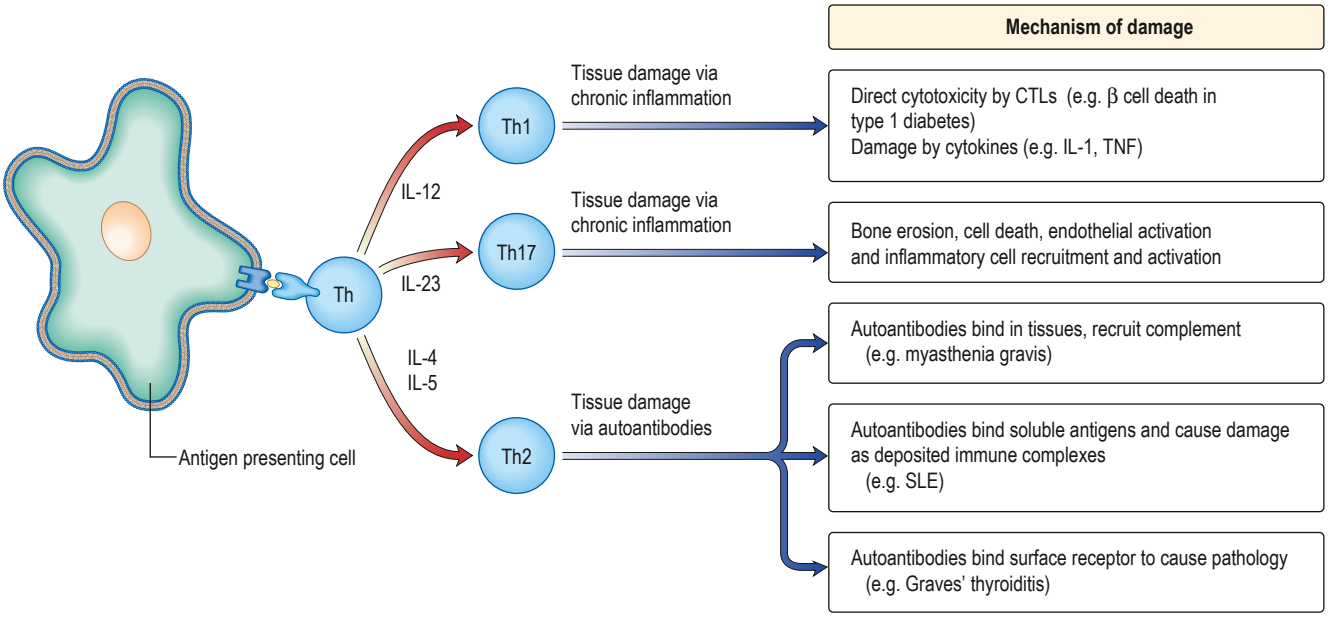


Fig. 6.17 Potential mechanisms of immune damage in autoimmune disease. *CTL,* Cytotoxic T lymphocyte; *IL,* interleukin; *SLE,* systemic lupus erythematosus; *Th,* helper T lymphocyte; *TNF,* tumour necrosis factor.

Box 6.19 IgG4-related disease

Disease	Affected Organs
Autoimmune pancreatitis	Pancreas
IgG4-related cholangitis	Bile ducts, liver
Mikulicz's syndrome	Salivary glands
IgG4-related dacryoadenitis	Lacrimal glands
Eosinophilic angiocentric fibrosis	Eye, orbit
Retroperitoneal fibrosis	Retroperitoneum
IgG4-related tubulointerstitial nephritis	Kidneys
IgG4-related lung disease including inflammatory pseudotumours	Lungs
IgG4-related lymphadenopathy	Lymph nodes
IgG4-related aortitis	Aorta and other blood vessels
Riedel's thyroiditis	Thyroid gland
Hypertrophic pachymeningitis	Brain and spinal cord meninges (rare)
IgG4-related dermatopathy	Skin (rare)
IgG4-related cardiovascular disease	Heart and blood vessels (rare)
IgG4-related prostatitis	Prostate gland (rare)

on these findings, other experimental B cell-depleting approaches are in current clinical trials.

Prognosis

The long-term prognosis in IgG4-related disease is highly dependent on the extent of the tissue fibrosis and organ dysfunction at presentation, much of which can be irreversible. Relapse is common and so maintenance therapy is usually advised until there is no clinical or radiologic evidence of relapse for 1 to 3 years. Some patients will have relapses many years after their initial presentation, so long-term follow-up is advised.

Tissue damage via autoantibodies

Examples of damage through direct binding to a target cell or structure, with recruitment of complement and other destructive processes, include:

- ***myasthenia gravis*** (autoantibodies against the acetylcholine receptor cause damage to the neuromuscular junction)
- ***autoimmune haemolytic anaemia*** (autoantibody targets red blood cell autoantigens, leading to lysis)
- ***Goodpasture's syndrome*** (anti-glomerular basement membrane autoantibodies damage glomerular integrity).

Autoantibodies can also bind their antigen in the circulation to form immune complexes. When these are deposited in the tissues, complement- and cell-mediated immune reactions are initiated. Immune complexes preferentially deposit in sites such as the kidney glomerulus, leading to chronic kidney disease. This is a feature of SLE, in which the autoantigen within the immune complexes is host-derived double-stranded DNA.

Autoantibody binding may act by binding to surface receptors. In Graves' thyroiditis, for example, the anti-thyroid stimulating hormone receptor antibody stimulates follicular cells to produce thyroid hormones, leading to hyperthyroidism.

Common autoimmune diseases

Over 80 diseases are now classified as autoimmune. The autoimmune origin of some of these diseases is very clear-cut but that is not so for others, and both major and minor factors must be defined.

Major criteria

- There is evidence of autoreactivity (e.g., activated or memory autoreactive T lymphocytes or autoantibodies).
- A clinical response to immune suppression can be demonstrated.
- Passive transfer of the putative immune effector (e.g., autoreactive T lymphocyte or autoantibody) causes the disease (this is the hardest criterion to satisfy in humans but is the most stringent).

Minor criteria

- An animal model exists that resembles the human condition, and in which there is a similar loss of immunological tolerance to self.
- There is evidence that, in the animal model, passive transfer of the putative immune effectors reproduces the disease in a naive animal.
- There is an HLA association (a frequent indicator that a disease is autoimmune).

Further reading

Moorman CD, Sohn SJ, Phee H. Emerging therapeutics for immune tolerance: tolerogenic vaccines, T cell Therapy, and IL-2 therapy. *Front Immunol.* 2021; 12:657768.
Perugino CA, Stone JH. IgG4-related disease: an update on pathophysiology and implications for clinical care. *Nat Rev Rheumatol.* 2020; 16:702–714.
Pisetsky, D.S. Pathogenesis of autoimmune disease. *Nat Rev Nephrol.* 2023; 19:509–524.
Sogkas G, Atschekzei F, Adriawan IR, Cellular and molecular mechanisms breaking immune tolerance in inborn errors of immunity. *Cell Mol Immunol.* 2021; 18:1122–1140.

ORGAN REJECTION IN CLINICAL TRANSPLANTATION

The outcome of an allograft (i.e., a graft between genetically non-identical members of the same species) in the absence of adequate immunosuppressive therapy is ***immunological rejection***. Histological analysis of rejected solid organs shows a range of immunological processes in action (**Box 6.20**). With modern tissue-matching approaches, hyperacute rejection should never occur and acute rejection can usually be prevented or treated with immunosuppression. The process of chronic rejection typically takes place over several years and is the main reason for organ graft failure.

The antigens recognized in acute and chronic graft responses are donor HLA molecules (alloantigens):

- In acute rejection, it is thought that the predominant response is against intact HLA molecules ***(direct allorecognition)*** and is mediated by $CD4^+$ and $CD8^+$ T lymphocytes.
- As the rejection process becomes more chronic, peptides from donor HLA molecules are processed and presented to T lymphocytes by host HLA molecules ***(indirect allorecognition)*.**

In haematopoietic stem cell transplantation the recipient may be recognized by the donor immune system as 'foreign' giving rise to GVHD. For GVHD to occur, there must be impaired T lymphocyte function in the recipient and the transfer of immunocompetent T lymphocytes from an HLA non-identical donor.

IMMUNE-BASED THERAPIES

Manipulation of the immune response in a therapeutic setting has seen many successes, as evidenced by the control of organ rejection in clinical transplantation through targeted immunosuppression (**Box 6.20**). Monoclonal antibodies offer the opportunity to neutralize the unwanted effects of cytokines or to direct immune responses, drugs, toxins, or irradiation against a specific target, whether it be a tumour cell or an immune cell involved in a damaging autoimmune response. Natural antiviral mediators, such as the IFNs, are already in the clinic as therapies for chronic viral infection, among other things. Genetic manipulation of T lymphocytes has generated a new class of cellular therapies.

Monoclonal antibody therapy (targeted therapy)

The combined power of MAb and recombinant DNA technology has led to a series of 'designer' drugs, which have been engineered so that they:

- are exquisitely targeted
- have optimal effector function
- do not carry antigenic segments that may incite a neutralizing response in the host.

In general, this has meant a process of 'humanizing' antibodies of mouse origin and selecting an appropriate effector function. For example, a MAb designed to remove a subpopulation of lymphocytes from the patient should have good complement-fixing ability or bind well to receptors on phagocytes, whereas a MAb designed to 'modulate' a cell without depletion should have these functions removed. Many examples of MAbs in current use are described in individual chapters. Therapeutic MAbs are potent modifiers of essential components of immune responses and therefore carry both predictable and unpredictable risks. For example, anti-TNF therapies increase the risk of invasive viral and mycobacterial infection. The new inhibitors of integrins, such as natalizumab, have been associated with JC virus reactivation in patients, leading to progressive multifocal leukoencephalopathy. Since integrins are essential for the migration of T cells into tissues, the blockade prevents egress into inflammatory sites but also reduces immune surveillance of infection.

Immunotherapy, including most monoclonal antibodies, brings unique challenges for drug safety as both efficacy and toxicity is difficult to predict for individual patients. The side-effects of immunotherapy can be life-threatening with severe immunodeficiency being a risk

Box 6.20 Classification of rejection

Description	Timing of response	Immunological mechanism	Management
Hyperacute	Minutes to hours	Pre-formed antibodies (e.g., anti-blood group, anti-HLA)	Prevention with typing to avoid mismatch, or depleting antibodies for 'high-risk' transplants
Accelerated acute	1–5 days	T lymphocytes	Combinations of immune suppression used prophylactically and as required, for example corticosteroids, ciclosporin, tacrolimus, sirolimus, polyclonal antibodies (e.g., against all T lymphocytes) and monoclonal antibodies (e.g., against activated T lymphocytes expressing CD25)
Acute	7–14 days onwards	T lymphocytes	
Chronic	Months to years	Antibodies, complement, endothelial cell changes	Chronic oral immunosuppression

HLA, Human leucocyte antigen.

of immunotherapy use in autoimmune disease, and severe autoimmunity being a risk of immune checkpoint targeting in cancer treatment.

Chimeric antigen receptor T-cell therapy

Chimeric antigen receptor T-cell therapy is an exciting new approach to the treatment of haematological malignancy. In the first licensed product of its kind (tisagenlecleucel) for B lymphocyte leukaemia and lymphoma, blood is taken from the patient and T lymphocytes are separated out. Using a retrovirus, the T lymphocytes are modified to express the CAR (which comprises a receptor for CD19 coupled to a powerful intracellular co-stimulation molecule). The CD19 target is expressed on all B lymphocytes (including lymphomas and B-cell leukaemias). The CAR-modified T lymphocytes are infused back into the patient and will then kill all target B lymphocytes. The main risk of this therapy is massive cytokine release from CAR T lymphocyte activation. Since all B lymphocytes (and not just the tumour cells) are removed, the patients develop antibody deficiency and require life-long immunoglobulin replacement. Specific co-stimulation domains in the CAR T cells have been shown to aid long-term persistence, correlating with prolonged protection against disease relapse. Further targets and approaches are now in clinical trials.

Immunosuppressive drugs

- ***Glucocorticosteroids*** (cortisone, hydrocortisone, prednisone, and prednisolone) are the most commonly used steroids and have a variety of effects on immune function, including:
  1. potent effects on monocyte production of the pro-inflammatory cytokines IL-1 and TNF-α
  2. blockade of T lymphocyte production of IL-2 and IFN-γ
  3. reduced activation and migration of a range of innate and adaptive immune cells.
- ***Ciclosporin, tacrolimus, and sirolimus (rapamycin)*** have similar effects on T lymphocyte function. Ciclosporin and tacrolimus are calcineurin inhibitors and inhibit Ca^{2+}-dependent second messenger signals in T lymphocytes following activation via TCRs. By contrast, sirolimus achieves a similar effect but acts at the level of post-activation events in the nucleus.
- ***Purine analogues*** such as azathioprine are also frequently used as anti-inflammatory drugs in conjunction with steroids and act by inhibiting DNA synthesis in dividing adaptive immune cells. Similar in mode of action, but more powerful, is mycophenolate mofetil.
- ***Alkylating agents*** that interfere with DNA synthesis, such as cyclophosphamide, are also used for immunosuppression.
- ***Tyrosine kinase inhibitors*** have profoundly improved the treatment of some malignancies with previously very poor outcomes. Tyrosine kinases are ubiquitous regulators of cell activation, promoting cell division and proliferation in response to growth factors and cytokines. Inhibition of the signalling pathways can lead to immune suppression. In immune cells, the relevant kinases include the Janus kinases (JAK) that signal through the STAT pathway. Inherited mutations in JAK-STAT pathway molecules can lead to immunodeficiency or, if the mutation causes constant (constitutive) activation, can give rise to uncontrolled cell division independent of growth factors: that is, cancer. Specific tyrosine kinase inhibitors have been developed for a range of cancers and as T- and B-cell immune suppressants.

Cytokines and anticytokines

Cytokines are pleiotropic agents with powerful pro-inflammatory and immunosuppressive effects and are attractive targets for therapies that inhibit or enhance their function. TNF-α targeting agents (cytokine modulators) have been tested in rheumatoid arthritis, as has the recombinant IL-1 receptor antagonist anakinra, although with much less success. Nonetheless, the good safety profile of anakinra has prompted its use in other diseases, and a beneficial effect has been shown in haemophagocytic lymphohistiocytosis, which may have an innate inflammatory component. IL-2 and its receptor (CD25) are also obvious candidates for immune therapies, and basiliximab which targets CD25 is used to induce immunosuppression during organ transplantation. IFNs are successfully used to boost pro-inflammatory immunity in chronic virus infection.

Restoring tolerance in autoimmune diseases and allergy

One of the goals of immune-based therapies for autoimmune disease is not simply to achieve immunosuppression but also to restore immunological tolerance against the relevant autoantigens. In animal models, an effective means of achieving this is to administer the autoantigen itself, or key peptide epitopes from it. This is known as antigen-specific or peptide immunotherapy and is under trial for several autoimmune diseases, where it appears to induce Tregs. The diseases that have the most advanced clinical data are in the field of allergy. For example, administration of cocktails of peptides of the cat allergen Fel d 1 has led to a reduction in detectable skin-prick responses and improved clinical scores.

Intravenous immunoglobulin

Intravenous immunoglobulin is a preparation of polyspecific IgG chemically purified from the plasma of large numbers (>20 000) of healthy donors. IVIg is used as a replacement therapy in patients with primary and secondary antibody deficiencies. When it is used for inflammatory conditions, however, there is also a therapeutic benefit, although randomized, placebo-controlled studies are few. In the USA, IVIg is recommended only for a small number of diseases in addition to antibody deficiency. These include immune-mediated thrombocytopenia, Kawasaki's disease, chronic inflammatory demyelinating polyneuropathy, and post-transfusion purpura.

The mechanism of action is not known, but may include:

- blockade of Fc receptors to prevent pathogenic antibodies binding to phagocytes
- inhibition of autoantibody synthesis by B lymphocytes
- modulation of dendritic cell function
- inhibition of complement activation
- inhibition of specific cytokines
- induction of T-cell regulation.

Further reading

Broderick L, Hoffman HM. IL-1 and autoinflammatory disease: biology, pathogenesis and therapeutic targeting. *Nat Rev Rheumatol*. 2022; 18:448–463.
Gubernatorova EO, Namakanova OA, Gorshkova EA, Medvedovskaya AD, Nedospasov SA, Drutskaya MS. Novel anti-cytokine strategies for prevention and treatment of respiratory allergic diseases. *Front Immunol.* 2021; 12:601842.
Hoffmann JHO, Enk AH. High-dose intravenous immunoglobulin in skin autoimmune disease. *Front Immunol.* 2019; 10:1090.
Ott N, Faletti L, Heeg M, Andreani V, Grimbacher B. JAKs and STATs from a clinical perspective: loss-of-function mutations, gain-of-function mutations, and their multidimensional consequences. *J Clin Immunol.* 2023; 43(6):1326–1359.
Sterner, R.C., Sterner, R.M. CAR-T cell therapy: current limitations and potential strategies. *Blood Cancer J.* 2021; 11:69.

Questions for this chapter can be found on Elsevier eBooks+.

7 Therapeutics and prescribing

Vikas Kapil

Background 107
Why do patients need drugs? 107
Therapeutic choice 107
The dose 108
Pharmacokinetics 109
Pharmacodynamics 110
Therapeutics in special populations 111
Pregnancy 111
Breast-feeding 112
Old age 112
Kidney disease 112
Liver disease 113
Prescribing 113
Act of prescribing 113
Electronic prescribing 113
Calculation skills 113
Ethical issues relating to therapeutics and prescribing 113
Medication review 114
Monitoring drug therapy 114
Adverse drug reactions 114
Drug interactions 115
Information sources 116

CORE SKILLS AND KNOWLEDGE

All prescribers need to learn how to use medications safely and effectively. This requires mastery of the principles in the first half of this chapter, which cover key concepts including pharmacokinetics, pharmacodynamics, drug interactions, and adverse drug reactions to apply in the specific clinical situations and related scenarios described in the second half of the chapter. Pharmacists, whether independent prescribers themselves or not, play a critical role in helping doctors make safe and effective choices when prescribing.

Key skills related to prescribing and therapeutics include:

- Developing a systematic approach to prescribing
- Using key pharmacological principles to select and administer an appropriate drug
- Learning parameters to monitor patients for both efficacy of treatment and unintended adverse reactions/toxicity
- Using information sources to assist in safe prescribing, including avoiding drug–drug interactions (DDIs).

There are many practical ways to learn prescribing skills: reviewing patients' medication charts on ward rounds, rewriting medication charts under supervision, observing annual medication reviews in primary care, reporting adverse drug reactions (e.g., using the 'Yellow Card' scheme in the UK), or working with pharmacists reviewing inpatient or outpatient prescriptions. Providing evidence of competence in these domains is increasingly required in modern healthcare. In the UK all new doctors must pass the Prescribing Safety Assessment (PSA) exam before they qualify, which covers the principles covered in this chapter.

BACKGROUND

Prescribing a medicine is the most common healthcare intervention performed by doctors and hence the correct choice of medication and the act of prescribing are key skills required by all doctors. The 1984 'Nairobi Declaration' emphasized that prescribing should be to the ***right patient*** with the ***right drug*** at the ***right dose*** and at an ***affordable cost***. Thus the prescription of a medicine needs to take into account the potential benefits and harms of that medicine and ensure that every step is taken to maximize the therapeutic benefit:harm ratio. The key principles of therapeutics and prescribing are described in **Box 7.1**.

WHY DO PATIENTS NEED DRUGS?

There are many reasons for prescribing drugs to patients:

- ***Disease treatment.*** A reliable diagnosis is usually secured before starting treatment to ensure that a patient is not exposed unnecessarily to the hazards of a particular intervention. However, a prior diagnosis may not be possible in every circumstance. For example, in sepsis the initiation of ***blind*** (empirical) antimicrobial therapy is justified, as delay would expose a patient to serious hazards (see Chapter 12).
- ***Symptom relief.*** Drugs are frequently used in the relief of symptoms, for example, in the treatment of severe pain, constipation, or pruritus. It is important to monitor for effectiveness and to consider stopping the medication if it is no longer indicated.
- ***Prevention.*** Medicines are also given to otherwise healthy individuals. In such circumstances, there must be a very clear imperative to ensure that the benefits to the individual outweigh the harm. Examples include:
  - primary cardiovascular prevention with lipid-lowering therapy (i.e., statins)
  - oral contraceptives for prevention of pregnancy

THERAPEUTIC CHOICE

The choice of the drug (**planned management** or therapeutic choice) depends on clinical factors (such as age of the patient,

Box 7.1 Key principles of prescribing

- ***Be clear about the reasons for prescribing.*** Secure, where appropriate an accurate diagnosis and have clear aims of what the benefits will be from prescribing the drug.
- ***Obtain an accurate drug history.*** Patients should be asked about their current medications, over-the-counter herbal medicines, and illicit drug usage. A past history of intolerances, including true allergic reactions, is also required.
- ***Obtain an accurate history of other factors that might affect the benefit:harm ratio of drugs.*** Take note of liver and renal impairment, age, whether patient is pregnant or breast-feeding, and co-morbidities.
- ***Establish what the patient expects from the drug, and deal with any concerns.*** Provide good information for the patient and answer any questions clearly.
- ***Select the most clinically effective, cost-effective, and safe medicine for the patient, based on an assessment of the individual.*** Optimize the benefit:harm balance by choosing the right dose and frequency and the right drug, in the best formulation, route of administration, and duration of treatment. Prescribe within the licence of the medicine, except when no alternative is available.
- ***Adhere to guidelines and local and national formularies.*** Based on the needs of the individual patient, use the most reliable information to identify the medicine most suited to the patient.
- ***Write prescriptions legibly using the correct documents, or prescribe electronically, depending on availability.*** Take care to avoid medication errors.
- ***Monitor the patient for both efficacy and safety after starting the drug.*** This is essential to optimize dose, identify adverse reactions, and determine when to stop the medicine. Report adverse reactions using spontaneous reporting schemes, if appropriate.
- ***Communicate clearly with other healthcare professionals and patients about the reasons for prescribing decisions.*** Communication is essential to optimize the benefit:harm ratio of drugs, particularly when there are multiple prescribers for a single patient.
- ***Prescribe within your competencies.*** Prescribing should be within the limits of your knowledge and experience. Do not be afraid to ask for advice and ensure that complex prescriptions are checked (e.g., when calculating doses).

Adapted from British Pharmacological Society; https://www.bps.ac.uk/education-engagement/teaching-pharmacology/ten-principles-of-good-prescribing.

concomitant disease, and concurrent therapy), pharmaceutical factors (the availability of other medicines and relative cost-effectiveness), and, increasingly, individual (host) factors.

Selecting the right drug (therapeutics) involves three elements:

- the drug's clinical efficacy for the proposed use
- the balance between the drug's efficacy and safety
- patient preference.

The most common approach to assessing a drug's efficacy is the randomized controlled trial (RCT) (**Box 7.2**), although other approaches can be informative (see Chapter 4). Although common adverse reactions are usually identified in early-phase drug development studies and RCTs, rare (often serious) adverse reactions are often only discovered after drugs are licensed by post-marketing surveillance studies (pharmacovigilance) (see **Box 7.2**). All patients should be offered the opportunity to be involved in making decisions about their medical treatment, including those about medications, by employing a ***shared decision-making*** approach (see Chapter 1). Patients' own preferences should be discussed to enable them to be equal partners in decision-making about whether, and how, they wish to be treated. This approach improves 'concordance' with treatment regimens.

Box 7.2 Evaluation of new drugs

Phase I: Healthy human subjects

- First use in humans
- Evaluation of safety and toxicity
- Pharmacokinetic assessment
- Sometimes pharmacodynamic assessment
- Approximately 100 subjects

Phase II: First assessment in patients

- Safety and toxicity evaluated
- Dose range identified
- Pharmacokinetic and pharmacodynamic monitoring
- Approximately 500 subjects

Phase III: Use in wider patient population

- Efficacy main objective
- Safety and toxicity also monitored
- Often multicentre trials
- Approximately 2000 patients involved

Phase IV: Post-marketing surveillance

- All patients prescribed drug monitored
- Efficacy, safety, and toxicity measured
- Quantification of unusual drug adverse effects
- Yellow Card and spontaneous reports of adverse reactions
- Often very large numbers of patients observed

Further reading

Maxwell SRJ. Rational prescribing: the principles of drug action. *Clin Med (Lond)* 2016; 16:459–464.

THE DOSE

The dose administered to the patient is crucial in determining both efficacy and safety of medicines. Appropriate drug dosages will usually have been determined from the results of so-called '***dose-ranging***' studies during the drug development programme (see **Box 7.2**).

In some cases, drug doses are ***fixed***, with all patients being given the same dose, for example, levonorgestrel for emergency contraception (see Chapter 47). However, these situations are unusual. Medications used for blood pressure, diabetes, pain control, and anaesthesia, to name a few examples, have an initial dose selected based on patient parameters and then the dose is titrated to response or other indicators.

More commonly, doses are ***titrated***. For many drugs, there are wide inter-individual variations in response. As a consequence, while a particular dose may, in one person, lack any therapeutic effect, the same dose in another may cause serious toxicity. This is because the concentration of drug reaching the systemic circulation ('exposure') varies between individuals even for the same dose, and it is exposure that determines the response to drugs, both in terms of both efficacy and safety. Pharmacokinetic factors (including differences in the rates of drug absorption, metabolism, and excretion) are the main determinants of such variability in exposure between individuals.

Variability in response, at an equivalent exposure, can also be due to ***pharmacodynamic factors***, where there are differences in the sensitivity of the drug target (typically due to differences in expression or biological activity of different receptors, enzymes, ion channels, transporters, DNA, or transcription factors).

Because it is not easy to measure exposure, we use the dose as a proxy. Furthermore, it is not always possible to predict

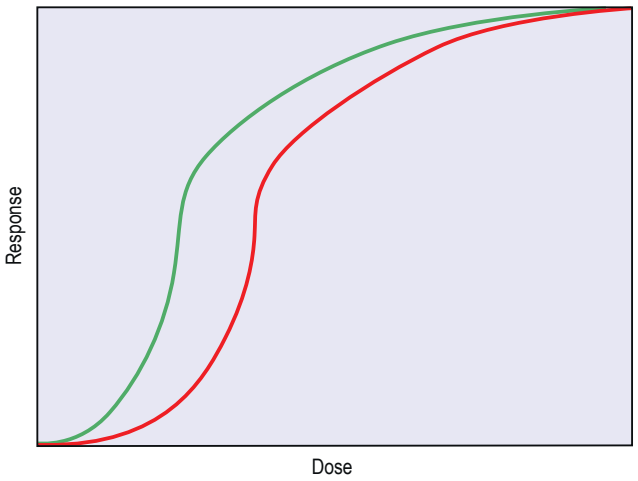


Fig. 7.1 Dose–response curve, showing the variable response of the same dose in different patients. The patient represented by the red line requires a higher dose to achieve the same desired effect than the patient represented by the green line and may therefore be at risk of experiencing more adverse effects.

pharmacodynamic variability. For these two reasons, prescribers should start most drugs at a low dose and titrate to the ***lowest effective dose***. This can help overcome the inter-individual variability in the dose–response curve, whereby there is an increase in response with an increase in dose, until a plateau effect is reached (Fig. 7.1).

Pharmacokinetics

Pharmacokinetics is the study of ***what the body does to a drug***. This can be divided into four different processes: absorption, distribution, metabolism, and excretion. Age, body weight, renal and hepatic function, concomitant drugs, co-morbidities, diet, smoking and alcohol, and genetic factors are all known to affect pharmacokinetics and so increase the inter-individual variation in required drug dose.

Absorption

Drugs can be administered by various different routes (Box 7.3) but oral administration is the most common. The main determinants of a drug's plasma concentration after oral administration are its bioavailability (Box 7.4) and its rate of systemic clearance (by hepatic metabolism or renal excretion). A drug's oral bioavailability depends on the extent to which it is:

- Destroyed in the gastrointestinal tract, especially by gastric acid.
- Unable to cross the gastrointestinal epithelium. Drugs can cross either by passive diffusion or by active uptake by gut epithelial transporters. Transporters can generally be divided into influx transporters, which increase uptake of drug (e.g., the PEPT1 influx transporter is required for the uptake of penicillins), and efflux transporters, which limit the absorption of a drug (e.g., P-glycoprotein is an efflux transporter that limits digoxin absorption).
- Metabolized by the liver before reaching the systemic circulation ('first-pass' metabolism). First-pass metabolism can be avoided by non-enteral routes.

Distribution

This is the process by which drugs are distributed from the bloodstream to organs and cells. Most drugs will circulate in the bloodstream bound to plasma proteins, most commonly albumin. Free drug, which is the active drug, is usually in dynamic equilibrium with protein-bound drug, with the latter being released as the free drug concentration decreases. The volume of distribution described the amount of body tissue into which a drug redistributes and is calculated by dividing the total amount of drug in the body (mg) by the plasma concentration (mg/L). Drugs with small volumes of

Box 7.3 Routes of drug administration

Route	Description	Example
Buccal	Rapid absorption, avoids first-pass metabolism	Midazolam for status epilepticus
Epidural	Administration of drug into epidural space around the spinal cord	Epidural anaesthetic for pain relief during labour
Inhaled	Achieves high concentrations at site of disease with limited systemic absorption	Salbutamol for acute asthma relief
Intramuscular	Easier than intravenous route but absorption is less predictable	Adrenaline (epinephrine) in anaphylaxis
Intranasal	Often used for local delivery	Intranasal corticosteroids in allergic conditions
Intrathecal	Used to achieve high local levels	Methotrexate in acute lymphoblastic leukaemia
Intravenous	100% of dose enters systemic circulation but invasive and inconvenient	Intravenous teicoplanin for treatment of invasive MRSA infections
Oral	Most common route, convenient, but subject to first-pass metabolism for certain drugs	Ramipril for hypertension
Per rectum	For local administration: for rectal/colonic disorders For systemic absorption: when oral or intravenous routes not possible	Prednisolone enema for rectal inflammatory bowel disease Diazepam for seizures
Subcutaneous	Good absorption, can be self-administered	Insulin for diabetes
Sublingual	Rapid absorption, avoids first-pass metabolism	GTN for acute relief of angina chest pain
Topical	Used for local application of cutaneous or joint diseases	Corticosteroids for eczema
Transdermal	Absorption through the skin via a patch	Nicotine replacement therapy for smoking cessation

Box 7.4 Key pharmacokinetic parameters

- ***Area under the curve:*** a plasma concentration time curve and a measure of the actual body exposure to a drug after administration of a dose of the drug; expressed in mg/L × h.
- ***Bioavailability:*** the fraction or proportion of a drug that enters the systemic circulation. Thus while intravenous formulations have 100% bioavailability, administration via other routes will have <100% bioavailability.
- ***Clearance:*** the volume of plasma cleared of the drug per unit time.
- **C_{max}:** peak plasma concentration after drug administration.
- ***Concentration:*** the amount of drug in a given volume. Even when the dose is similar, concentrations (and thus exposure) will vary between individuals.
- ***Dose:*** the amount of drug administered. This may be micrograms, milligrams, or grams. Occasionally, drugs are administered per body weight or body surface area.
- ***Loading dose:*** an initial larger dose to achieve immediate therapeutic plasma or tissue drug concentration. It is usually reserved for medical emergencies but can also be used for drugs with long half-lives, which would otherwise take a long time to reach desired therapeutic steady-state concentration.
- ***Elimination half-life:*** the time required for the drug concentration to reach half its original value. With repeat dosing, for most drugs, it takes four to five half-lives to approach steady state.
- ***Steady state:*** the situation for most drugs where drug intake is in dynamic equilibrium with drug elimination.
- **T_{max}:** time to reach C_{max}.
- **V_d:** volume of distribution—the apparent volume in which a drug is distributed. Drugs with low V_d are predominantly in the circulation, while drugs with a high V_d have diffused into the peripheral tissue compartment, such as body fat. Drugs with a low V_d are amenable to removal by haemodialysis in overdose, while drugs that have a high V_d are not.

distribution stay in the plasma, while those with larger volumes of distribution move freely into other body tissues such as adipose tissue.

Metabolism

Metabolism takes place in many different organs the liver being the most common site. Drug metabolism converts lipid-soluble drugs to water-soluble metabolites for excretion by the kidneys. Metabolism usually inactivates a drug, but in some cases, it can convert a prodrug into an active drug (e.g., enalapril to enalaprilat) or lead to the formation of metabolites that are toxic (N-acetyl-para-benzoquinoneimine in paracetamol overdose). Drug metabolism occurs in two stages:

- ***Phase I:*** oxidation, reduction, or hydrolysis. This is largely undertaken by a family of isoenzymes known as the cytochrome P450 system. Inhibition or induction of cytochrome P450 isoenzymes is a major cause of drug–drug interactions (DDIs) (see Chapter 7).
- ***Phase II:*** conjugation with glucuronate, sulphate, or other substances. This renders drugs more water soluble.

Variability in the genes that encode drug-metabolizing enzymes (**Box 7.5**) is a major determinant of the inter-individual differences in the therapeutic and adverse responses to drug treatment. One-quarter of all licensed medicines are substrates for a single P450 isoenzyme, CYP2D6. The poor metabolizer (reduced function) CYP2D6 phenotype leads to accumulation of drugs when used at conventional doses, increasing the risk of toxicity with drugs such as selective serotonin reuptake inhibitors or antipsychotics (e.g., risperidone) that undergo CYP2D6 metabolism.

Reduced conversion of a prodrug to active metabolite may also compromise efficacy in some patients; for example, poor metabolizers of CYP2D6 do not experience analgesic benefit with codeine, which is a prodrug converted to morphine by CYP2D6. Conversely, some individuals have multiple copies of the normal *CYP2D6* gene and are described as ultra-rapid metabolizers. These patients display increased drug metabolism and therapeutic failure at conventional doses for most drugs, although ultra-rapid metabolism can also lead to increased toxicity for prodrugs such as codeine, because there is increased conversion to morphine.

Understanding the role of genetic differences in drug metabolism can allow prescribing to be tailored to individual patients:

- Warfarin: variation in CYP2C9 activity and vitamin K epoxide reductase complex *(VKORC1)* genes affect warfarin pharmacodynamics. Determining the genotype before prescribing can improve anticoagulation control.

Mercaptopurine/Azathioprine: patients with reduced or absent activity of the enzyme thiopurine methyltransferase (TPMT) are at high risk of bone marrow suppression at conventional doses of thiopurine drugs such as mercaptopurine or azathioprine. Pre-treatment activity determination allows appropriate dose selection.

Excretion

Excretion of drugs and their metabolites occurs most commonly via the kidneys, although faeces, breath, and sweat represent other routes of excretion. In the kidney, free drug is filtered through the glomerulus and excreted in the urine as long as it is water soluble. Lipid-soluble drugs are usually reabsorbed into the circulation in the renal tubules. Some drugs, such as benzylpenicillin, are actively excreted by cells in the proximal convoluted tubule. Renal impairment reduces the elimination of water-soluble drugs, and thus reduction of the maintenance dose is usually undertaken to avoid toxicity.

Pharmacodynamics

Pharmacodynamics is the study of ***what the drug does to the body***. Pharmacodynamic sources of variation in drug action can be due to changes in the expression of the drug target, their affinity or their selectivity. This can be caused by genetic factors and also by disease. For example, in hepatic impairment, pharmacodynamic factors can determine the sensitivity to a drug, adding to changes in pharmacokinetics.

Genetic factors affecting drug pharmacodynamics also play a role in determining how individuals respond to drugs. For example, variation in the β_2-adrenoceptor gene can affect response to the most common bronchodilator (salbutamol) in asthma. Glucose-6-phosphate dehydrogenase deficiency the most common enzyme deficiency, can predispose patients to acute red cell haemolysis with certain drugs such as primaquine, dapsone, sulphonamides, and rasburicase.

The biggest clinical advances in this area have occurred in two therapeutic areas: targeted cancer therapies and drug safety:

- Sequencing cancer genomes has identified novel driver mutations, leading to the development of targeted therapies. These have seen remarkable success: for instance, trastuzumab is only of value in women with breast cancer whose malignant cells express the HER2 epidermal growth factor receptor.
- Some immune-mediated adverse drug reactions involving the skin and liver can now be predicted by genotyping for certain human leucocyte antigen (HLA) polymorphisms. For example, pre-prescription genotyping for *HLA-B*57:01* has largely

Box 7.5 Genetic polymorphisms affecting drug response

Organ or system involved	Associated gene/allele	Drug/drug response phenotype	Pharmacokinetic/pharmacodynamic
Blood			
Red blood cells	*G6PD* *TPMT*2*	Primaquine and others Azathioprine/6MP-induced neutropenia	Pharmacodynamic Pharmacokinetic
Neutrophils	*UGT1A1*28*	Irinotecan-induced neutropenia	Pharmacokinetic
Platelets	*CYP2C19*2*	Stent thrombosis	Pharmacokinetic
Coagulation	*CYP2C9*2, *3, VKORC1*	Warfarin dose requirement	Mixed
Brain and peripheral nervous system			
CNS depression	*CYP2D6*N*	Codeine-related sedation and respiratory depression	Pharmacokinetic
Anaesthesia	Butyrylcholinesterase	Prolonged apnoea	Pharmacokinetic
Peripheral nerves	*NAT-2*	Isoniazid-induced peripheral neuropathy	Pharmacokinetic
Drug hypersensitivity and liver injury			
	*HLA-B*57:01*	Abacavir hypersensitivity	Pharmacodynamic
	*HLA-B*15:02*	Carbamazepine-induced Stevens–Johnson syndrome (in some Asian groups)	Pharmacodynamic
	*HLA-A*31:01*	Carbamazepine-induced hypersensitivity in Caucasians and Japanese	Pharmacodynamic
	*HLA-B*58:01*	Allopurinol-induced serious cutaneous reactions	Pharmacodynamic
	*HLA-B*57:01*	Flucloxacillin hepatotoxicity	Pharmacodynamic
Malignancy			
Breast cancer	*CYP2D6*	Response to tamoxifen	Pharmacokinetic
Chronic myeloid leukaemia	*BCR-ABL*	Imatinib and other tyrosine kinase inhibitors	Pharmacodynamic
Colon cancer	*KRAS*	Cetuximab efficacy	Pharmacodynamic
Gastrointestinal stromal tumours	*c-kit*	Imatinib efficacy	Pharmacodynamic
Lung cancer	*EGFR*	Gefitinib efficacy	Pharmacodynamic
	EML4-ALK	Crizotinib efficacy	Pharmacodynamic
Malignant melanoma	*BRAF V600E*	Vemurafenib efficacy	Pharmacodynamic
Muscle			
General anaesthetics	Ryanodine receptor	Malignant hyperthermia	Pharmacodynamic
Statins	*SLCO1B1*	Myopathy/rhabdomyolysis	Pharmacokinetic

Adapted from Pirmohamed M. Pharmacogenetics: past, present and future. Drug Discov Today 2011; 16:852–861.

eliminated hypersensitivity to the anti-retroviral drug abacavir. Similarly, *HLA-B*15:02* predisposes patients from South-East Asia to Stevens–Johnson syndrome with carbamazepine.

Although not generally available yet in routine clinical practice, whole genome sequencing (see Chapter 5) may in the future be used to personalize drug therapy to improve both efficacy and safety of drugs, a key component of the drive towards ***personalized medicine***.

Further reading

Peruzzi E et al. Implementation of preemptive testing of a pharmacogenomic panel in clinical practice: where do we stand. *Br J Clin Pharmacol* 2023; epub ahead of press.

THERAPEUTICS IN SPECIAL POPULATIONS

Most medications are evaluated in clinical trials and therefore licensed for use, in adult, non-pregnant patients, usually without multimorbidity. Trial data may therefore not easily generalize to real-life patient populations, and healthcare professionals need to consider specific issues relating to prescribing in other special populations.

Pregnancy

Four in five females who are pregnant use an over-the-counter or prescribed medicine while pregnant. There is little published evidence for the safety of most drugs in pregnancy. Clinicians should be cautious about prescribing drugs to pregnant females and only essential treatments should be given, though equally necessary and effective medications should not be unduly withheld. There are significant changes in maternal physiology that have consequences for the prescribing of medications.

Pharmacokinetics

- ***Absorption:*** decreased intestinal motility and gastric emptying, increased gastric reflux and emesis.
- ***Distribution:*** increased total plasma volume (up to 40% higher), increased total body fat, decreased albumin.
- ***Metabolism:*** no major changes.

- ***Excretion:*** increased glomerular filtration rate (GFR) (up to 50% higher).

There is also significant drug transfer to the fetal circulation. Blood flow to the placenta increases 1000% over the last 6 months of gestation. Total placental surface (and therefore drug transfer) is greatest at term. Drugs that are small (<1 kDa) can cross the placenta and the main determinant of fetal drug concentration is maternal drug concentration.

Up to 1% of fetal malformations may be due to *in utero* drug exposure. Many drugs are known to cause developmental abnormality (teratogenesis) (**Box 7.6**). Furthermore, there is increased exposure to drugs that are renally excreted into the amniotic fluid as this is reswallowed and recirculated in the fetus *in utero*. Organ structure and function is most susceptible to perturbation during organogenesis between 2 and 14 weeks' gestation, which may include a period where the female is unaware of being pregnant.

When a known or potential teratogen is required during pregnancy (e.g., an anticonvulsant drug or lithium), the potential adverse effects should be discussed with the patient, preferably before conception (pre-conception counselling). The relevant drug literature should be consulted when prescribing for pregnant females.

Breast-feeding

Although most drugs can be detected in breast milk the quantity is generally small, as for most drugs, the concentration in milk is in equilibrium with plasma water (i.e., the non-protein-bound fraction). A few drugs (e.g., aspirin, carbimazole, methotrexate) may, however, cause harm to the infant if ingested in breast milk. This is especially the case for drugs that cause idiosyncratic reactions where there is little or no concentration–effect relationship (type B adverse events; see Chapter 7). The relevant drug literature should be consulted when prescribing for women who are breast-feeding.

Old age

Prescribing in old age has additional considerations. Old age alters drug pharmacokinetics:

- ***Absorption:*** diminished due to reductions in intestinal motility and blood flow.
- ***Distribution:*** reduced total body-water and increased body-fat affects the volume of distribution for water-soluble and fat-soluble drugs, respectively.
- ***Metabolism:*** decreased hepatic blood flow and enzyme function.
- ***Excretion:*** renal function declines with age at the rate of approximately 1 mL/min after the age of 40 years.

Old age also alters pharmacodynamics and exaggerated pharmacodynamic effects of drugs acting on the central nervous, cardiovascular, and gastrointestinal systems are common.

Examples of common problems encountered in the use of drugs among older people are shown in **Box 7.7**. ***Extrapolation of drug dosages*** from those appropriate in younger adults may therefore lead to toxic plasma levels in older patients.

Additional considerations include:

- ***Multimorbidity***, often associated with polypharmacy, leads to increased opportunities for disease–drug and DDIs.
- ***Concordance with treatment regimens*** diminishes as the number of prescribed drugs increases and is especially poor in the face of cognitive impairment (see Chapter 49).

Kidney disease

Prescribing drugs that require renal excretion as a major form of drug clearance can be problematic in patients with coexisting chronic kidney disease.

In chronic kidney disease there is less glomerular filtration of drugs leading to reduced renal clearance, enhanced drug half-life, and potential for drug accumulation. Patients with chronic kidney disease appear to be more susceptible and sensitive to adverse drug effects. Some medications, such as thiazide and thiazide-like diuretics, work from the luminal side of the collecting duct and therefore are ineffective in advanced chronic kidney disease.

The level of kidney function below which the dose of a drug must be reduced depends on the proportion of the drug eliminated by renal excretion and its toxicity. For many drugs with only minor or no dose-related side-effects, very precise modification of the dose regimen is unnecessary and a simple scheme for dose reduction is sufficient. For more toxic drugs with a small safety margin or in patients at extremes of weight, dose regimens based on kidney function should be used. When both efficacy and toxicity are closely related

Box 7.6 Some human teratogens

Drug or drug class	Effect
ACE inhibitors	Oligohydramnios
Retinoids, e.g., isotretinoin	Multiple abnormalities
Carbimazole	Neonatal hypothyroidism Abnormalities of bone growth
Fluconazole	Fallot's tetralogy
Antiepileptics: carbamazepine, lamotrigine, phenytoin, valproate	Cleft palate, facial abnormalities, neural tube defect, neurodevelopmental delay
NSAIDs	Delayed closure of the ductus arteriosus
Cytotoxic drugs	Most are presumed teratogens
Lithium	Ebstein's anomaly
Misoprostol	Moebius' syndrome (rare neurological condition)
Thalidomide (and possibly lenalidomide)	Phocomelia

Box 7.7 Common adverse reactions to drugs in older people

Drug or drug class	Effect
Nitrates **α-Adrenoceptor-blockers** **Loop diuretics**	Postural hypotension
Diuretics (thiazides)	Glucose intolerance, gout
Antimuscarinic drugs **Tricyclic antidepressants** **Neuroleptics** **Anticonvulsants** **Hypnotics** **Opioids**	Confusion, cognitive dysfunction
Non-steroidal anti-inflammatory drugs (NSAIDs)	Gastric erosions Small bowel and colonic lesions (less frequently than gastric) Upper gastrointestinal bleeding Perforated peptic ulcer Salt and water retention Acute kidney injury

to plasma drug concentration, dosage must be adjusted according to clinical response and plasma drug concentration.

Most drug adjustments are recommended to use estimated GFR (eGFR) or creatinine clearance (usually calculated by the Cockcroft–Gault formula). Although there is a good correlation between these two measures, at the extremes of weight there can be significant differences and so the relevant drug literature should be checked when prescribing for patients with chronic kidney disease to be sure which measure is recommended.

Various patterns of drug-induced nephrotoxicity can occur, with the commonest being:

- ***Pre-renal effects*** of diuretics (volume), angiotensin-converting enzyme (ACE) inhibitors (renal perfusion), antihypertensive medications (hypotension), and non-steroidal anti-inflammatory drugs (NSAIDs) affect autoregulation of renal perfusion.
- ***Tubulointerstitial nephritis*** caused by immune reactions to commonly used medications such as penicillin antibiotics and proton-pump inhibitors.
- ***Direct toxic effects***, typically to renal tubular cells where drugs reach high intracellular levels due to reabsorption, caused by medications including cisplatin, lithium, and aminoglycoside antibiotics.
- ***Other mechanisms***, including crystal nephropathy (e.g., some older anti-retroviral protease inhibitors), drug-induced thrombotic microangiopathy (e.g., calcineurin inhibitors), or direct glomerular effects (e.g., NSAIDs).

It is good practice to advise avoidance where possible of nephrotoxic drugs, especially over-the-counter drugs such as NSAIDs, in patients with chronic kidney disease.

Liver disease

Prescribing drugs that require hepatic metabolism as a major form of drug clearance can be problematic in patients with coexisting chronic liver disease.

In chronic liver disease, there is less first-pass metabolism and often portosystemic shunts. These combine to increase the proportion of drug that enters the systemic circulation, increasing drug exposure for a fixed-given oral dose.

However, some drugs require hepatic metabolism for efficacy (prodrugs, such as clopidogrel). For prodrugs the presence of chronic liver disease and/or portosystemic shunts reduces the proportion of orally ingested drug that undergoes hepatic activation and therefore there is relatively less circulating active drug per oral dose.

The assessment of severity of chronic liver disease that requires dose-adjustment is more nebulous than when dose adjusting in chronic kidney disease. Although there are validated scoring systems for predicting outcomes in chronic liver disease, such as Child–Pugh and MELD (see Chapter 23), these have not been routinely used to make decisions on drug-dose adjustment. Holistic clinical assessment is preferred and the relevant drug literature and specialist advice should be consulted when prescribing for patients with chronic liver disease if there is clinical uncertainty.

Drugs can cause liver injury in various patterns, for example:

- ***hepatitis:*** paracetamol, amiodarone, azathioprine
- ***fatty change:*** tetracyclines
- ***cholestasis:*** co-amoxiclav, nifedipine

Therefore in patients with chronic liver disease, avoid hepatotoxins if possible or reduce doses. For example, it is common to restrict paracetamol dosing to 2 g/day in chronic liver disease, compared with 4 g/day in patients without liver injury.

PRESCRIBING

Once the most appropriate therapeutic choice has been made taking into account of the issue above, then the medications should be safely **prescribed**. It is worth pointing out that medical gases (e.g., supplemental oxygen), intravenous fluids, and some devices (such as anti-embolic stockings) should also be prescribed routinely.

Act of prescribing

Essentials of a valid prescription include drug name (the generic unless there is a specific reason to include brand name), dose, route, frequency (or duration of administration), start date (and end date if appropriate), name and contact details of prescriber, and signature. There may be additional regulatory requirements for prescribing of certain drugs such as strong analgesics, narcotics and hypnotics (termed 'controlled drugs' in the UK), and cytotoxic chemotherapeutic agents.

Electronic prescribing

Although prescribing has been traditionally a pen-to-paper activity, there is increasing reliance on proprietary electronic prescribing systems linked to electronic health records. There are clear advantages of electronic prescribing relating to legibility, clinical audit, and linking to pharmacy dispensing and stock systems. There is additionally the possibility of building in checks for allergies, interactions, and other inappropriate therapeutic choices, though these depend on human creation of these specific rules and their integration into the systems.

Calculation skills

For certain medications that are used based on a weight or time, then basic **calculation skills** may be required to be certain of the dose. Calculation errors are common when changing units (i.e., between minutes to hours, concentration from % to mg per mL, etc.), and it is good practice to check calculations with a colleague to ensure they have been done correctly.

Ethical issues relating to therapeutics and prescribing

The use of medications in humans involves a range of considerations that are critical to both the wellbeing of patients and trust in healthcare systems. Important ethical issues include:

- ***Informed consent:*** patients should have adequate information about the medication, alternative options, the benefits, risks, and potential adverse reactions, to allow them to make an informed decision with their healthcare provider about the best course of action: this is termed ***shared decision-making*****.**
- ***Benefit/Risk:*** Use of medicines should aim to maximize benefit (beneficence) and minimize harm (non-maleficence) by selecting the most appropriate medication for the individual patient. This requires continuing professional development for prescribers to keep up to date with latest evidence—for example, if safety alerts are issued based on post-marketing surveillance of a new drug. There are important implications for end-of-life care where the shift of priorities may change from extending life to symptom management if aligned to patient wishes, so drugs with

long-term preventative effects may be discontinued in favour of those with immediate symptomatic benefits.

- ***Justice:*** medicines should be used based on equity of access and need. Affordability is included in this equation, though this is usually dealt with at formulary level rather than on an individual patient level. Confidentiality applies to medicines use as it does to any other healthcare information. Just use of medicines also relates to preserving antibiotic efficiency by preventing overuse of antibiotics and to prevent medication misuse or abuse of medicines with addictive potential.
- ***Conflict of interest:*** prescribers must disclose issues that may influence prescribing decisions, such as financial relationships with pharmaceutical producers or personal research in the area.

MEDICATION REVIEW

Timely intermittent medication review is an essential component to lifelong care of patients on medications. Reviews may occur multiple times per day for critically ill patients or once per year or opportunistically for stable patients in primary care.

The purpose of medication review is to:

- ensure all medications prescribed still have a positive benefit/risk profile with changing clinical status (there should be ***active de-prescribing*** where appropriate)
- identify and manage potential DDIs
- identify and manage (or avoid) adverse drug effects
- ensure required monitoring is in place—for instance, monitoring for toxicity to liver, kidney, or bone marrow is required for some medications.

MONITORING DRUG THERAPY

Pharmacokinetic and pharmacodynamic variability makes monitoring of the effects of many medications essential. While there are generally agreed upon principles for certain medicines (i.e., intermittent renal function monitoring for diuretics or renin-angiotensin system inhibitors), some drugs have strict recommendations in the licensing agreements for timed parameters (i.e., leucocyte and prolactin levels for clozapine).

Three approaches for monitoring are generally used:

- ***Pre-treatment dose selection.*** In patients who have known, or suspected, impaired renal or liver function, it is usually possible to predict dose requirements from their eGFR or clinical assessment of severity of chronic liver disease (see Chapter 23).
- ***Measuring plasma drug concentrations.*** For a few drugs, dosages can be effectively monitored by reference to their plasma concentrations (**Box 7.8**). This technique is useful only if both of the following criteria are fulfilled:
  1. There is a reliable and available drug assay.
  2. Plasma concentrations correlate well with both therapeutic efficacy and toxicity.

Therapeutic drug monitoring refers to the individualization of dosage by maintaining plasma or blood drug concentrations within a target range (the 'therapeutic range' or 'therapeutic window'). Using this pharmacokinetic approach, therapeutic drug monitoring is used in two major situations:

1. when drugs are used prophylactically to maintain the ***absence*** of a condition, to ensure doses are high enough to be effective but low enough to avoid toxicity. Examples include drugs to prevent seizures (many anti-epileptic drugs), depressive or manic episodes (lithium), or organ rejection (calcineurin inhibitors such as tacrolimus).
2. to avoid serious toxicity, as with the aminoglycoside antibiotics which, unlike most antibiotics, have a narrow therapeutic range.
   - Data and measuring drug effects

For many drugs, dosage adjustments are made in line with patients' responses (a pharmacodynamics-based approach). Monitoring can involve dose titration against a therapeutic endpoint or a toxic effect. Objective **data interpretation** (such as monitoring antihypertensive therapy by measuring blood pressure, glycated haemoglobin in diabetes mellitus, or cytotoxic therapy with serial white blood cell counts) is most helpful, but subjective assessments are necessary in many instances (as with antipsychotic therapy in people with schizophrenia). Pharmacodynamic tests can be used in clinical practice to target therapy or to avoid undue sensitivity. Examples are use of the INR or prothrombin time to titrate warfarin dosage to find the targeted level of anticoagulation.

Box 7.8 Some drugs for which therapeutic drug monitoring is used

Drug	Therapeutic plasma concentration range	Toxic level	Optimum post-dose sampling time (h)
Carbamazepine	20–50 μmol/L	>50 μmol/L	>8
Ciclosporin	50–200 μg/L	>200 μg/L	Pre-dose
Digoxin	1.3–2.6 nmol/L	>2.6 nmol/L	>8
Gentamicin	Trough <2 mg/L	>2 mg/L	Pre-dose
	Peak 5–10 mg/L	>12 mg/L	>1
Lithium	0.6–1.0 mmol/L	>1.5 mmol/L	>10
Phenytoin	40–80 μmol/L	>80 μmol/L	>10
Tacrolimus	4–12 μg/L	>12 μg/L	Pre-dose
Theophylline	55–110 μmol/L	>110 μmol/L	>4
Vancomycin	15–20 mg/L	>20 mg/L	Pre-dose

Further reading

Fang Z et al. Overview of therapeutic drug monitoring and clinical practice. *Talanta* 2024; 266(Pt 1):124996.

ADVERSE DRUG REACTIONS

Adverse drug reactions (ADRs), defined as 'the unwanted effects of drugs occurring under normal conditions of use', are a significant cause of morbidity and mortality. The burden of ADRs is huge in both adults and children: ADRs account for about 6.5% of hospital admissions and affect 15% of adult inpatients. In addition, about one in four patients in primary care will have an adverse reaction, but fortunately these are mild in most cases. Unwanted effects of drugs are five to six times more likely in the elderly than in young adults, and the risk of an adverse drug reaction (ADR) rises sharply with the number of drugs administered.

Classification

Two types of ADR are recognized: type A (augmented) reactions and type B (idiosyncratic) reactions (**Box 7.9**).

Box 7.9 Classification and examples of adverse drug reactions

Type A (augmented)	
• Common but only occasionally serious • Qualitatively normal, but quantitatively abnormal, manifestations of known pharmacology • Show a clear dose–response relationship • May occur after a single dose (hypotension with atenolol), or may develop after months (dependence with opioids), or years (second malignancies with anticancer drugs) • Usually reproducible in animal models	
Type of drug (or drug class)	*Type A adverse reaction*
Anticoagulants, e.g., warfarin	Gastrointestinal bleeding Intracerebral haemorrhage
Antipsychotics, e.g., haloperidol	Acute dystonia/dyskinesia Parkinsonian symptoms Tardive dyskinesia
Cytotoxic agents, e.g., 5-fluorouracil	Bone marrow dyscrasias Cancer
Diuretics, e.g., furosemide	Dehydration, renal impairment
Insulin	Hypoglycaemia
Type B (idiosyncratic)	
• Usually rare and are often serious • Qualitatively abnormal responses to the drug (not predictable from its known pharmacology) • Show a dose–response relationship that is complex and not easily discernible • Not reproducible in animal models	
Type of drug (or drug class)	*Type B adverse reaction*
Benzylpenicillin Radiological contrast media	Anaphylaxis
Broad-spectrum penicillins, e.g., co-amoxiclav	Maculopapular rash
Phenytoin Sulphonamides	Toxic epidermal necrolysis Stevens–Johnson syndrome
Diclofenac Isoniazid	Hepatotoxicity
Isotretinoin	Depression, suicidal ideation

Diagnosis

All ADRs mimic some naturally occurring disease, and drawing a distinction between an iatrogenic aetiology and an event unrelated to the drug is often difficult. The patient should be asked not only about prescription drugs but also about over-the-counter medicines, herbal medicines, and illicit drugs. Although some effects are obviously iatrogenic (e.g., acute anaphylaxis occurring a few minutes after intravenous penicillin), many are less so.

Causality assessment of a suspected ADR can be difficult, particularly in patients on multiple drugs. The temporal relationship between drug exposure and suspected adverse events should be appropriate, and other causes of the suspected adverse events should be excluded. Rechallenge with the suspected drug and dechallenge (i.e., stopping the suspected drug to see if the adverse effect disappears) may be appropriate in certain circumstances. Some ADRs are so typical, such as cough or angio-oedema following initiation of ACE inhibitors, that identification of such an ADR is likely.

Management

As a general rule, type A reactions can usually be managed by a reduction in dosage, while type B reactions almost invariably require the drug to be withdrawn (and never re-instituted). Most ADRs resolve spontaneously when the drug is withdrawn, but specific therapy is sometimes required for ADRs that cause acute patient harm. For example, bleeding with anticoagulants is actively reversed (vitamin K and prothrombin complex can be used for warfarin) and drug-induced anaphylaxis is a medical emergency.

When a patient suffers an ADR, this should be reported to the regulatory agency, for example, through the UK's ***Yellow Card reporting*** scheme to the Medicines Healthcare Products Regulatory Agency. The criteria for reporting vary across the world but, in general, all ADRs to new drugs or drugs used outside their licence should be reported, while only the serious ADRs to established drugs (i.e., those that have been on the market for usually more than 2 years) should be reported.

Further reading

Coleman JJ, Pontefract SK. Adverse drug reactions. *Clin Med* 2016; 16:481–485.

DRUG INTERACTIONS

Drugs can interact with each other DDIs, with food (drug–food interactions) and with herbal medicines (drug–herbal interactions) (**Box 7.10**). These interactions, in general, can be either pharmacokinetic (affecting the processes of absorption, distribution, metabolism, and excretion), pharmacodynamic (synergistic or antagonistic), or mixed. Approximately 1% of all ADRs that lead to hospital admission are due to DDIs. A major risk factor for DDIs is polypharmacy, which is becoming more common because of the co-morbidities associated with an ageing population (see Chapter 49). About 57% of patients above the age of 65 years are on more than five drugs, with 12% being on more than ten drugs.

Food–drug interactions are also becoming increasingly recognized. A common example is grapefruit juice that inhibits CYP3A4 in the intestinal wall, increasing the bioavailability of drugs metabolized by CYP3A4 (see **Box 7.10**), resulting in higher exposure and toxicity.

Herbal medicines are taken by almost 20% of the population but are rarely asked about when a patient's clinical history is taken. Many different herbal–drug interactions have been described the best recognized of which is with St John's Wort (SJW), which is an inducer of P450 enzymes and drug transporters, such as P-glycoprotein. SJW can lead to increased drug metabolism, reduction in drug concentration, and failure of therapeutic effect. An example is interaction with immunosuppressants such as prednisolone and ciclosporin, where the use of SJW has led to allograft rejection.

Reliable sources of information (see next section) must be used when prescribing new drugs to patients who already are taking other medications to ensure that no significant DDI exists.

Further reading

Mouly S, Lloret-Linares C, Sellier PO, Sene D, Bergmann JF. Is the clinical relevance of drug-food and drug-herb interactions limited to grapefruit juice and Saint-John's Wort? *Pharmacol Res* 2017; 118:82–92. https://bnf.nice.org.uk/drugs/grapefruit/.

Box 7.10 Some examples of drug interactions

Implicated drug(s)	Affected drug(s)	Mechanism	Clinical effect
Pharmacokinetic: absorption			
Colestyramine	Digoxin	Binding of colestyramine to digoxin	Reduced digoxin plasma concentration and reduced efficacy
Pharmacokinetic: enzyme induction			
Carbamazepine Phenytoin Rifampicin	Oral contraceptive Warfarin Ciclosporin	CYP3A4 induction CYP2C9 induction CYP3A4 and P-glycoprotein induction	Contraceptive failure Thrombosis Allograft rejection
Pharmacokinetic: enzyme inhibition			
Amiodarone Ciprofloxacin Grapefruit juice Ritonavir	Warfarin Theophylline Simvastatin Inhaled fluticasone	CYP2C9 inhibition CYP1A2 inhibition CYP3A4 inhibition CYP3A4 inhibition	Haemorrhage Arrhythmias Myopathy Cushing's syndrome
Pharmacokinetic: excretion			
Naproxen Bendroflumethiazide	Methotrexate Lithium	Inhibition of methotrexate excretion Inhibition of lithium excretion	Methotrexate toxicity Lithium toxicity
Pharmacodynamic: synergistic interaction			
Verapamil Ramipril	Atenolol Spironolactone	Atrioventricular node block Reduced excretion of potassium	Bradycardia, heart block Hyperkalaemia
Pharmacodynamic: antagonistic interaction			
Salbutamol Furosemide	Atenolol Digoxin	Opposing effects at the β-receptor Hypokalaemia	Lack of bronchodilation Digoxin toxicity

INFORMATION SOURCES

Pharmacotherapy moves at a very rapid pace and it is impossible for anyone to keep up with contemporary advances. Critical information relating to essential contraindications, interactions, and monitoring, as well as treatment pathways, can be found in reliable prescribing literature:

- The ***Summary of product characteristics (SmPCs)*** for individual drugs is produced by manufacturers and vetted by drug regulatory authorities.
- The relevant ***national formulary***. Many countries have their own formularies, for example, the UK has the *British National Formulary* (BNF).
- ***Guidance*** produced by national health technology agencies such as the UK's *National Institute for Health and Care Excellence (NICE).*

Further reading

http://bnf.nice.org.uk *British National Formulary.*
http://www.guideline.gov *National Guideline Clearinghouse clinical guidelines in the USA.*
https://www.medicines.org.uk/emc *Electronic Medicines Compendium.*
http://www.nice.org.uk *National Institute for Health and Care Excellence (NICE) clinical guidelines.*

Questions for this chapter can be found on Elsevier eBooks+.

8 Radiology

Cindy Chew, Andrew A. Plumb, and Patrick O'Dwyer

Background 117
Plain radiography/X-rays 118
Chest radiography (chest X-ray) 118
Abdominal radiography (abdominal X-ray) 122
Skeletal radiography 122
Mammography 124
Fluoroscopy 125
Barium swallow 125
Voiding micturating cystourethrogram 125
Ultrasonography 127
The kidney 127
Thyroid gland 128
Pelvic organs: uterus and ovary 128
Venous ultrasound 129
Hepatobiliary disease 130
Testis 130
Special ultrasound procedures 130
Computed tomography 131
The brain 131
Thorax 131
Abdomen and pelvis 131
CT angiography 132
CT urography 132
Special CT procedures 133
Magnetic resonance imaging 137
The brain 138
The abdomen and pelvis 138
Bones, soft tissues, and joints 138
Nuclear medicine and positron emission tomography 143
Brain—DaTScan (dopamine transporter scan) 143
Thorax—ventilation/perfusion (V/Q) scan 143
Bone scan 144
Renal function 144
Endocrine nuclear imaging 144
Positron emission tomography and computed tomography 145
Red and white cell scan 145
Gastric emptying 146
Interventional radiology 146
Surgical drainage and needle biopsies 147
Vascular procedures 147
Abdominal IR procedures 149
Interventional oncology 151
Cutting edge and future technology 151

CORE SKILLS AND KNOWLEDGE

Radiological examinations are part of every 21st-century patient's medical journey—in the outpatient clinic and community setting as well as emergency and inpatient services and across almost every clinical specialty.

From ubiquitous X-ray to complex CT and MRI, imaging is central to diagnosis, treatment planning, and therapy evaluation or simply to reassure the patient all is well. More recently, Interventional Radiology provides indispensable, often lifesaving treatment options for patients across a range of conditions.

Key learning outcomes in this chapter include:

- Learning the different imaging modalities available, the underpinning principles of physics and which modality is best suited to particular clinical questions and conditions
- Becoming familiar with the concept of radiation safety
- Learning the fundamentals of how to interpret common imaging examinations.

Opportunities for learning include spending time in radiology departments, observing how radiographers/imaging technicians acquire images and navigate concerns relating to patient comfort and safety; sitting with radiologists as they report diagnostic imaging; attending multidisciplinary imaging meetings; talking with and examining patients undergoing radiological examinations, then correlating the clinical presentations with subsequent imaging appearances and reports.

BACKGROUND

Radiology is one of the physician's most powerful investigative tools. Clinical imaging is ubiquitous in every 21st-century patient's medical journey—from disease screening and diagnosis to reassuring the patient all is well. Radiology and the radiologist are central to treatment planning, therapy evaluation, and delivering treatment through interventional radiology.

The effective use of the radiology department requires an understanding of the strengths and limitations of myriad radiological examinations to answer the clinical question (**Box 8.1**). More than mere hypothesis testing, a firm grasp of how pre-test probability impacts interpretation of imaging findings is crucial. Consideration should also be given to examination availability, patient suitability, radiation dose, and cost. Inappropriate or over-investigation harms patients, not least from issues arising from incidental imaging findings (in up to 30% of all diagnostic imaging).

Box 8.1 Categories of radiological investigation modalities

Non-ionizing radiation	Ionizing radiation
Ultrasound	Plain X-ray/radiography
Magnetic resonance imaging (MRI)	Computed tomography (CT)
	Fluoroscopy
	Nuclear medicine, e.g., bone scan
	Positron emission tomography (PET)
	Interventional radiology (IR) procedures

Box 8.2 Ionising Radiation (Medical Exposure) Regulations [IR(ME)R]—UK

The Ionising Radiation (Medical Exposure) Regulations 2017 are accessible on legislation.gov.uk. Medical ionizing radiation is used to help diagnose and treat conditions. The regulations aim to make sure that it is used safely to protect patients from the risk of harm when being exposed to ionizing radiation.

They set out the responsibilities of duty holders (the employer, referrer, practitioner, and operator) for radiation protection and the basic safety standards that duty holders must meet.

Responsibilities include:

- minimizing unintended, excessive, or incorrect medical exposures
- justifying each exposure to ensure the benefits outweigh the risks
- optimizing diagnostic doses to keep them 'as low as reasonably practicable' for their intended use.

Good communication between the clinical and radiology team is essential. While in person discussions are ideal, providing all relevant clinical information in clinical requests, most commonly via the radiology request form, helps the radiologist decide the best radiological investigations and greatly increases the likelihood of answering the clinical question and helping the patient.

Referring clinicians are bound by legal responsibilities when requesting imaging examinations (IRMER in the UK, when requesting tests using ionizing radiation, see **Box 8.2**) and to take the appropriate action(s) after reading each radiology report. Many clinical and therapeutic management decisions are made today in multidisciplinary meetings (MDTs). The radiologist, presenting each patient's imaging findings, is an indispensable presence at these discussions.

Radiology is involved in ever-increasing therapeutic procedures. These range from lifesaving haemostatic procedures (e.g., to treat bleeding aneurysms or postpartum haemorrhage) to oncologic interventions (e.g., ablating tumours) to organ-saving revascularization (e.g., thrombectomy for ischaemic limb or brain).

This chapter aims to:

- Introduce the reader to a variety of radiological examinations and interventional procedures
- Briefly describe the basic science underpinning each radiology modality
- Introduce common clinical imaging and principles of interpretation
- Discuss issues relating to the radiology patient including selection, preparation, contraindications, safety, and potential complications

PLAIN RADIOGRAPHY/X-RAYS

Plain radiographs are the most commonly performed radiological examination (**Box 8.3**). They are relatively quick to perform, well tolerated by patients, and are the investigation of choice in many conditions.

Box 8.3 Principles of plain radiography and fluoroscopy

- Image is produced when X-rays from the source pass through the body onto the detector
- Involves ionizing radiation
- Radiation dose is small (one chest X-ray = 10 days of background radiation exposure)
- Common body parts examined include chest (CXR), abdomen (AXR), spine, and limb bones
- Images produced correspond to how much X-rays are absorbed by the body: for example, bones appear 'white' or 'dense' because most of the X-rays have been absorbed. Lungs filled with air absorb little X-ray—hence they appear 'dark'
- Fluoroscopy uses the same principles but produces a continuous 'moving' X-ray image
- No special patient preparation required usually

Chest radiography (chest X-ray)

Chest X-ray (CXR) is an important first-line investigation for a broad range of symptoms including breathlessness, chest pain, and cough and is often requested and interpreted at the 'point of care' by clinicians in the emergency department and acute medical unit. Careful, systematic, and structured examination of a CXR may yield a wealth of information about a patient's health above and beyond the basic diagnostic indication for requesting the X-ray. A suggested structured approach to reporting and interpretation of the CXR can be found in the online edition of this chapter at Elsevier eBooks+ (see Box E8.1).

Many examples can be found within Chapter 17 ('Cardiology') and Chapter 20 ('Respiratory medicine'). This section will illustrate some common conditions and principles of interpretation.

Normal CXR

Fig. 8.1 shows an example of a normal CXR. Familiarity with the normal X-ray is the essential first step in confidently identifying and diagnosing pathology.

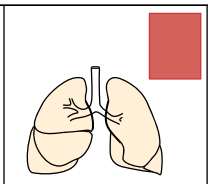
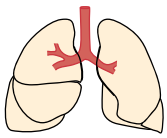
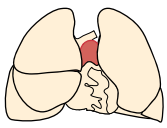
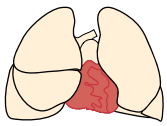
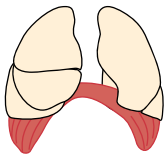
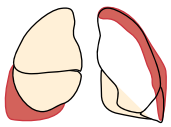
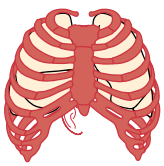
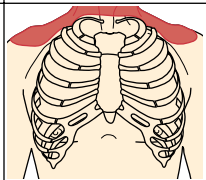
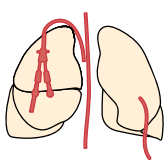
Pulmonary oedema

Pulmonary oedema results in accumulation of fluid within the alveolar, interstitial, or pleural lung spaces; common causes include left ventricular failure and renal failure. A spectrum of radiographic changes may be seen (**Fig. 8.2**): in milder cases, these may be confined to central venous prominence with engorgement of the upper lobe pulmonary veins; more severe acute pulmonary oedema leads to loss of clarity of the lung bases and air space opacification which is typically bilateral with a 'batwing' distribution. Kerley-B lines, fine horizontal lines perpendicular to the pleural surface and representing oedema of the interlobular septae, may be visible at the peripheries. Bilateral pleural effusions, often larger on the right, dominate in more chronic cases. Cardiomegaly is common and careful inspection of the X-ray may reveal pointers towards the underlying cause of cardiac disease (e.g., a mechanical heart valve); see **Fig. 8.1**.

Infection

CXR can be helpful to diagnose pneumonia (**Fig. 8.3**). Patchy or confluent air space opacities ('white areas') can be demonstrated in the normally black-appearing lungs, where secretions fill the alveolar spaces. 'Consolidation' is said to be present if aerated airways (black, branching tubes known as 'air bronchograms') can be seen against white secretion-filled alveolar spaces. The normal

Box E8.1 Stepwise approach to reporting a CXR

1. Preliminaries		• **Patient details:** - Which patient, how old, when was this image taken? • **X-ray details:** - Is it a posteroanterior or anteroposterior projection? - Is the patient rotated? - Is there adequate penetration? - Has the patient taken an adequate breath?	• Compare where available to previous radiograph images • The distance between each clavicular head and the spinal processes should be equal • In a CXR at maximal inspiration, 8-10 posterior and 5-6 anterior ribs should be visible • With modern digital imaging systems, under- or over-penetration can usually be corrected by 're-windowing' the image
2. Trachea		• The trachea should be widely patent and should be central in the chest	• The trachea may be pulled towards areas of fibrosis or collapse • It may be pushed away by masses, a goitre, lymphadenopathy, large pleural effusion or a tension pneumothorax
3. Mediastinum		• The aortic notch should be just visible • The left hilum should be slightly higher than the right • The hila should be symmetrical in size	• The mediastinum is widened by lymphadenopathy, masses such as thymoma, and aortic aneurysm • The hila may be pulled up or down by collapsed lobes, or enlarged by the presence of a tumour
4. Heart size		• Heart size should be less than 50% of the thoracic width on a PA chest film • Assess heart borders	• Cardiomegaly may be caused by hypertension, valvular disease, heart failure or cardiomyopathy • Loss of right heart border may suggest right middle lobe pathology • Loss of left heart border suggests lingular pathology
5. Diaphragm		• The right hemidiaphragm is usually higher than the left • Look for air/gas under the diaphragm and for elevation of the diaphragm on either side	• Air or gas under the diaphragm indicates a perforated abdominal viscus • A hemidiaphragm may be pulled up from above by lobar collapse, or pushed up from below by a large mass
6. Pleura		• The costophrenic angles should be clearly visible • Look for lung markings extending to the chest wall • Look for evidence of a lung edge	• Loss of the costophrenic angle is usually due to the presence of pleural effusions • Larger pleural effusions may cause a 'white out' on one side • Loss of lung markings or a visible lung edge suggests a pneumothorax
7. Lungs		• Is there consolidation of the upper lobe (stops at horizontal fissure), middle lobe/lingula (heart edge unclear) or lower lobe (diaphragm unclear)? • Inspect each part of the lung looking for round shadows • Inspect the lung parenchyma for interstitial changes • Remember to examine the apices, behind the heart and lung peripheries for subtle tumour or pneumothorax	• Areas of consolidation may often be restricted to a lobe and are usually caused by pneumonia. An air bronchogram (airways outlined) or evidence of cavitation may be seen • Rounded shadows may be caused by cancer, inflammatory lesions, such as tuberculosis or fungal disease • Interstitial changes may be caused by pulmonary oedema or fibrosis
8. Bones		• Inspect all ribs for evidence of fractures • Faint bones imply osteopenia • Are there lytic lesions?	• Think of systemic diseases: - Osteoporosis may cause osteopenia or fractures - Solid organ cancers or myeloma can cause lytic lesions
9. Soft tissue		• Observe the soft tissue around the rib cage	• Air within the skin suggests surgical emphysema
10. Miscellaneous		• Look for any externally applied devices	• Nasogastric tubes should bisect the carina and cross the diaphragm • Chest drains should terminate in the pleural space • Central venous catheters should end just above the right atrium • Pacemaker wires, and sternotomy wires imply cardiac disease

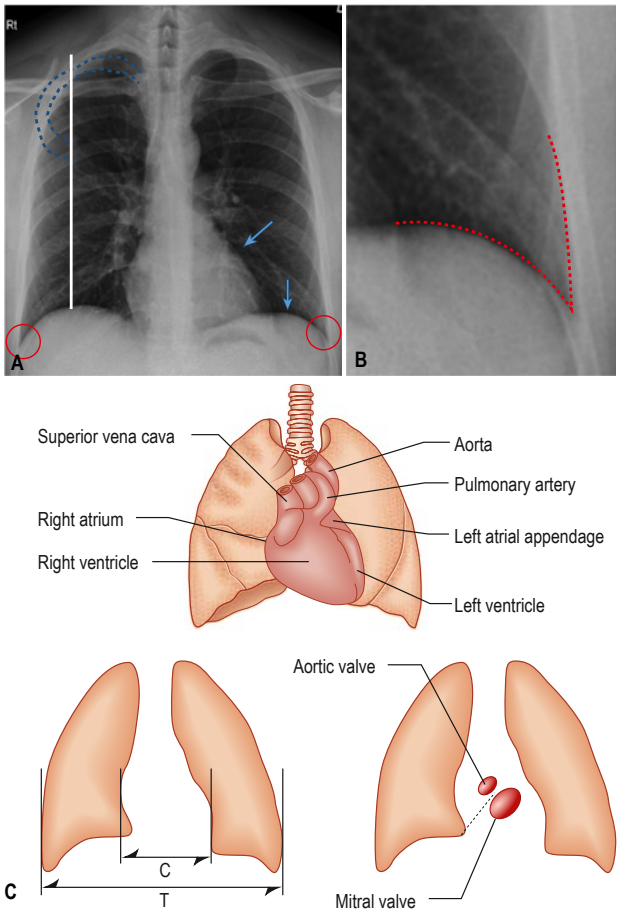


Fig. 8.1 An exemplar normal CXR. **(A)** The lungs are both well inflated—with approximately 6 anterior ribs or 9 posterior ribs *(dotted lines)* visible at the mid-clavicular line *(white)*. Notice the crisp and distinct outline of the heart and diaphragm *(arrows)*—both juxtaposed against aerated ('black') lungs. Notice the sharp costophrenic angles bilaterally *(red circles)*. **(B)** Magnified view of sharp left costophrenic angle: *red dots* outline the lateral thoracic pleural surface and left hemidiaphragm. **(C)** The cardiac silhouette as it is seen on the CXR. The heart (measurement C) should measure half the width of the thorax (T). The *dotted line* is an arbitrary line from the left hilum to the right cardiophrenic angle; a calcified aortic valve is seen above this line and mitral valve below.

sharp outlines of the heart or diaphragm may be lost in CXRs—this is a subtle clue that infection is present, also known as the 'silhoutte sign'. Consolidation may be 'lobar' (occupying an area on the CXR corresponding to an anatomical lung lobe) or 'bronchopneumonia' (usually patchy, separated by areas of normal lung, and often bilateral).

Viral pneumonitis often results in bilateral changes on CXR, and infiltrates may be either 'interstitial' (involving the connective tissues surrounding the alveoli and bronchioles, resulting in a reticulonodular pattern on X-ray) or 'alveolar' (resulting in a hazy 'ground-glass' appearance).

Pleural effusion

Pleural effusion indicates a collection of fluid within the pleural cavity and is recognized by a homogeneous 'white' opacity to lower zones of the lung on CXR, obliterating the normal sharp costophrenic angle (Fig. 8.4). The upper border is typically curved ('meniscus'). A large effusion will exert mass effect, pushing normal midline structures (trachea, heart) towards the opposite side. A careful review of other structures in the X-ray may reveal clues to the underlying diagnosis (e.g., sclerotic rib metastases secondary to a bronchogenic carcinoma; pleural plaques indicating asbestos exposure and thus risk of mesothelioma). Blunting of the costophrenic angle (the earliest radiographic evidence of effusion) implies at least 500 mL of fluid sequestered within the pleural space.

Pneumothorax

A pneumothorax is where air, often leaked from the lung, is present in the pleural cavity. On a CXR the periphery of the thoracic cavity will appear black, lacking the usual fine network of vascular markings (Fig. 8.5). Careful inspection will reveal the edge of aerated lung 'fallen away' from the chest wall; this may be difficult to spot, particularly if the pneumothorax is localized or small, and rotating the X-ray 90 degrees can sometimes be helpful.

Nasogastric tube

Nasogastric (NG) tubes allow for delivery of medications and nutrition directly to the stomach, for instance in those unable to safely swallow. Risks of a misplaced tube are significant and include inadvertent delivery of feed into the lung if mis-sited in the bronchus,

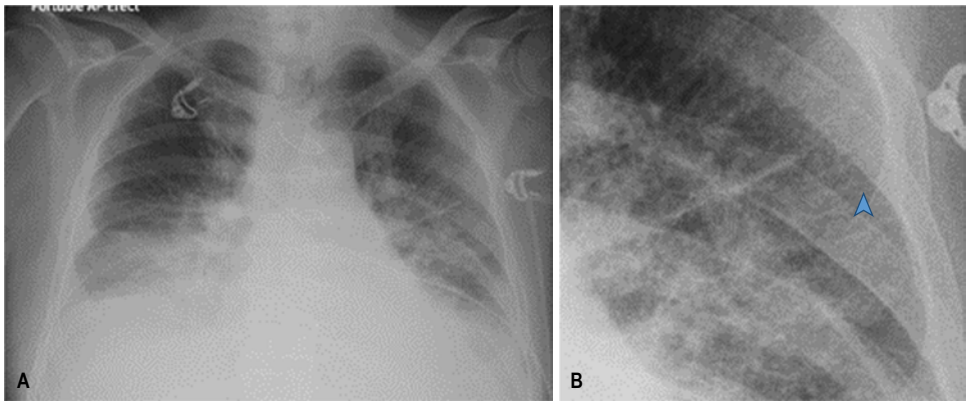


Fig. 8.2 Pulmonary oedema. **(A)** Notice the presence of median sternotomy wires which may indicate previous cardiac surgery. **(B)** Magnified image. The *blue arrowhead* points to a Kerley-B line.

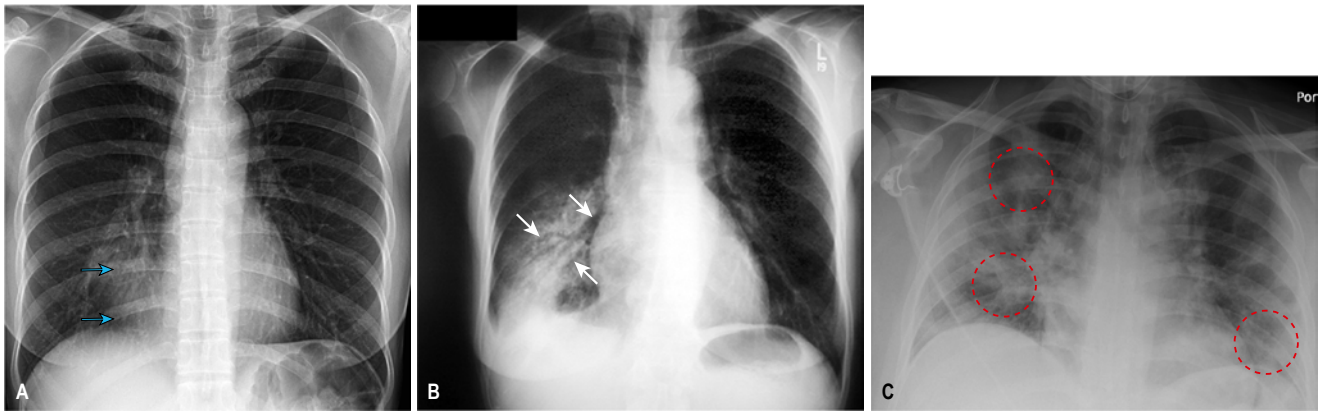


Fig. 8.3 Infection on CXR. **(A)** Right middle lobe pneumonia. The right heart border is indistinct ('silhouette sign'), localizing the consolidation to the anatomical right middle lobe *(arrows)*. **(B)** Right lower lobe pneumonia. In this instance the right diaphragmatic contour is indistinct, localizing to the anatomical right lower lob. Air bronchograms *(white arrows)* can be clearly seen coursing through the consolidated lung. **(C)** Viral pneumonitis (in this instance due to SARS-CoV-2). Notice the patchy bilateral ground-glass opacities (a few of these areas are highlighted by *red circles*). *(Panel A reproduced with permission from* ExpertDDX: Chest, *Second Edition, Elsevier; panel B reproduced from* Respiratory physiotherapy: an on call survival guide, *Elsevier.)*

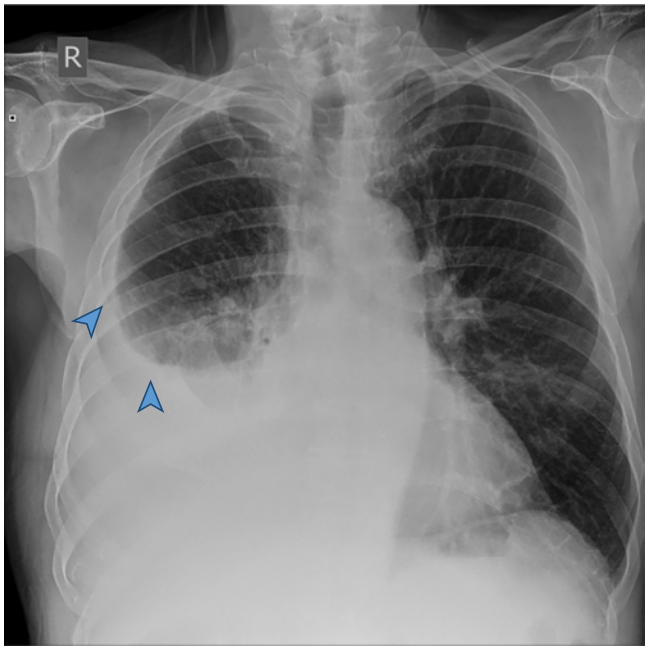


Fig. 8.4 Right pleural effusion. Notice the meniscus *(blue arrowhead)*.

which may be fatal. CXR is one means by which to confirm correct position prior to use (Fig. 8.6).

Free gas

Subdiaphragmatic free gas implies perforation of an abdominal viscus (e.g., perforated duodenal ulcer) (Fig. 8.7). Similar appearances may be seen if an X-ray is taken after recent laparoscopic surgery (where CO_2 is insufflated into the abdominal cavity to aid vision). To detect free gas, the X-ray should be taken after at least 10 minutes with patient in erect or lateral decubitus position.

Lung collapse

Lung collapse implies loss of normal aeration of a lung segment, lobe, or entire lung usually due to a blockage within the lumen of a bronchus (e.g., tumour, retained secretions, or foreign body). Collapse of a significant volume of lung tissue will result in an area of opacification (white) on CXR accompanied by signs of 'volume loss' within the corresponding hemithorax (e.g., deviation of the trachea *towards* the collapsed lung, crowding of ribs or elevated hemidiaphragm on the affected side) (Fig. 8.8).

The lung lobes collapse in characteristic directions:

- The ***lower lobes*** collapse downwards and towards the mediastinum.
- The ***left upper lobe*** collapses forwards against the anterior chest wall. This produces a 'hazy' appearance, or a 'veil-like' opacity, to the entire/much of the left hemithorax.
- The ***right upper lobe*** collapses upwards and inwards, giving the appearance of an arch over the remaining lung.
- The ***right middle lobe*** collapses anteriorly and inwards, obscuring the right heart border.
- If a ***whole lung*** collapses, that hemithorax will appear white, and the mediastinal structures will shift towards the side of the collapse.

Pulmonary fibrosis

Pulmonary fibrosis implies scarring of the interstitial scaffold of the lung tissue and may have numerous causes including connective tissue or granulomatous diseases, drug toxicity, occupational exposures, infection, or rarer idiopathic pneumonitides. Fibrosis produces a reticular pattern of shadowing on CXR accompanied by signs of volume loss in more advanced cases (Fig. 8.9). More generalized fibrosis can lead to a honeycomb appearance (see Chapter 20), seen as diffuse shadows containing multiple circular translucencies a few millimetres in diameter.

Round opacities

Lung cancer is the most common cause of large round shadows but many other aetiologies are recognized (Box 8.4).

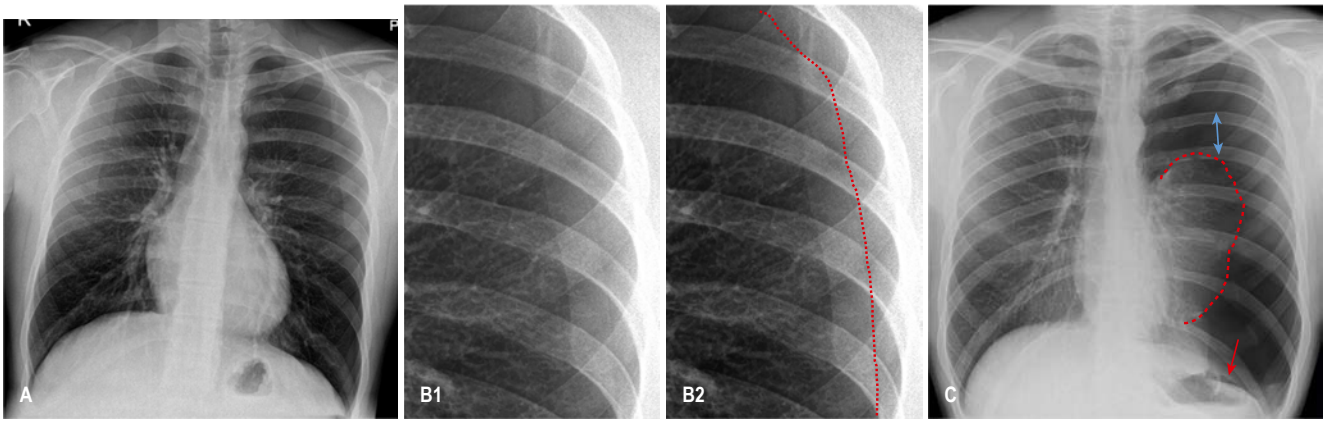


Fig. 8.5 Pneumothorax. **(A)** The left lung has 'deflated' away from the chest wall. To diagnose a pneumothorax confidently, you must see both a lung edge and an absence of lung markings peripherally. **(B)** Enlarged image with lung edge marked by *red outline*. Do not confuse the lung edge with the adjacent scapular edge. **(C)** This is a tension pneumothorax. Notice deflated lung *(red outline)*. Air is in the pleural space. Ribs are splayed apart *(double-headed arrow)*. Diaphragm is depressed downwards into the abdomen by the intrathoracic pressure *(red arrow)*. The heart is pushed towards the right. Without emergency treatment a tension pneumothorax will be fatal; ideally a tension pneumothorax is diagnosed clinically and without delaying treatment to obtain a CXR and so this image is rarely seen.

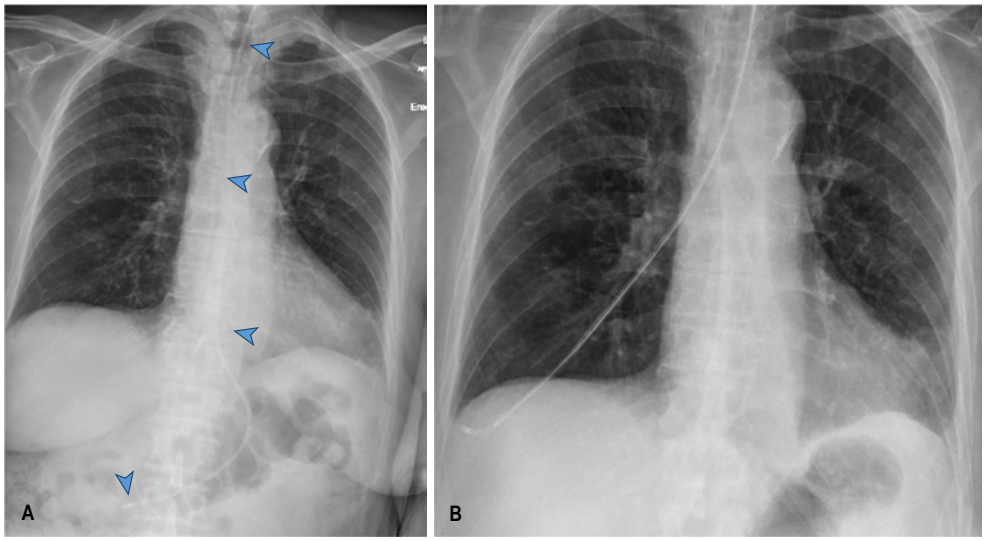


Fig. 8.6 NG tube. **(A)** The NG tube is correctly sited: (1) Tube in midline; (2) bisects carina; (3) crosses diaphragm in the midline and then deviates to the left; (4) tip is clearly seen under the diaphragm (>5 cm) *(blue arrowheads)*. **(B)** In contrast, the tube in this image is sited in the right lung.

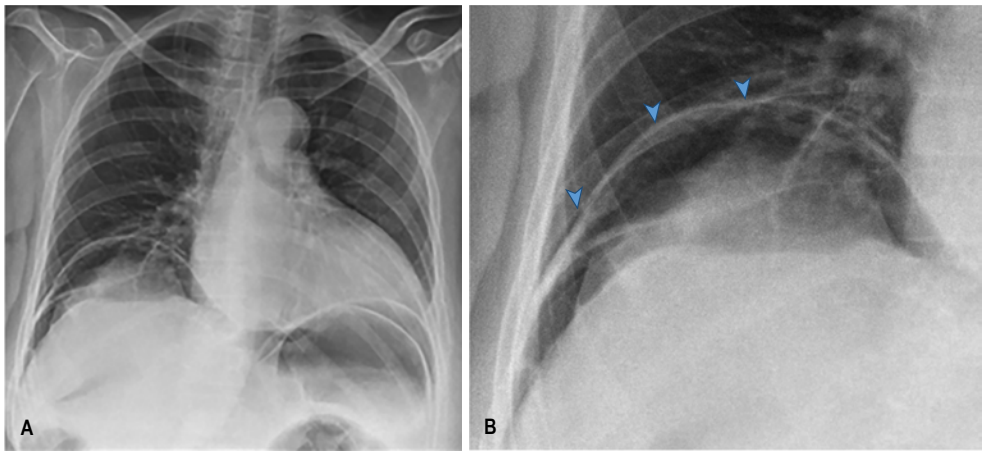


Fig. 8.7 Free gas under both diaphragms. The outline of diaphragm *(blue arrowheads)* can be clearly seen, sandwiched between air above and below it. This patient with inflammatory bowel disease came with severe generalized abdominal pain and vomiting. At operation, there was a duodenal perforation.

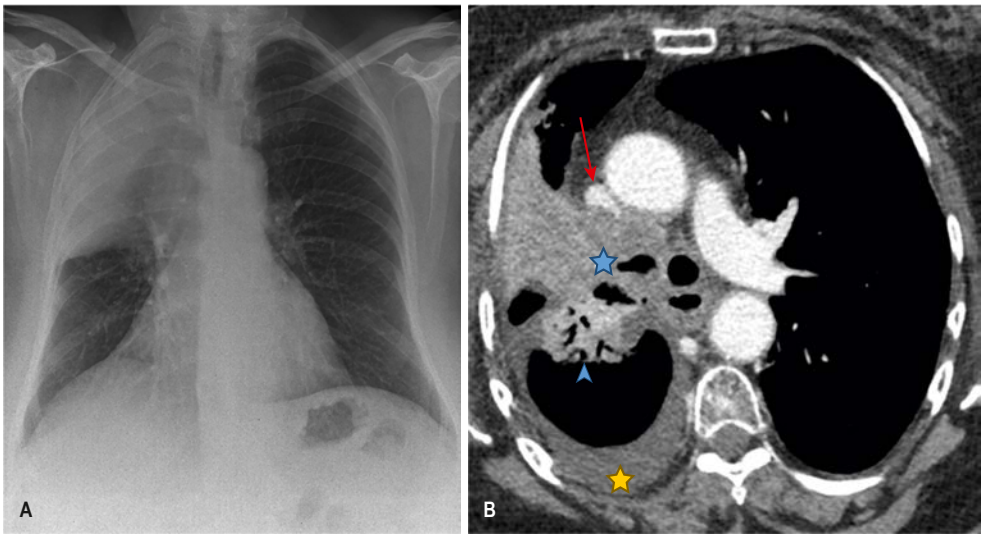


Fig. 8.8 Lobar lung collapse. This patient was a smoker and presented with lethargy and hypercalcaemia. **(A)** CXR demonstrated a collapsed right upper lobe. **(B)** CT demonstrated a large bronchial tumour *(blue star)* causing the lung collapse *(blue arrowhead)* with invasion of superior vena cava *(red arrow)* and a right pleural effusion, not visible on CXR *(yellow star)*. This is a reminder that by the time an effusion is visible on CXR, it is moderately large.

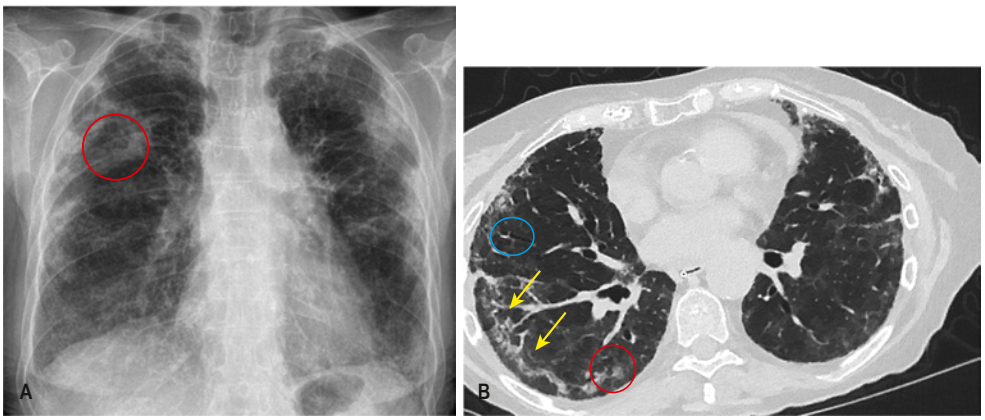


Fig. 8.9 Pulmonary fibrosis. Increased interstitial lung marking, scarring *(yellow arrows)*, ground-glass pulmonary opacity *(red circles)*, and (traction) bronchiectasis *(blue circle)* are visible on CXR **(A)** but better appreciated on CT **(B)**.

? Box 8.4 Causes of round opacities (>3 cm) in the lung

- Carcinoma
- Metastatic tumours (usually multiple shadows)
- Lung abscess (usually with fluid level)
- Encysted interlobar effusion (usually in horizontal fissure)
- Hydatid cysts (often with a fluid level)
- Arteriovenous malformations (usually adjacent to a vascular shadow)
- Aspergilloma
- Rheumatoid nodules
- Tuberculoma (may be calcification within the lesion)
- Bronchial carcinoid
- Cylindroma
- Chondroma
- Lipoma

Other shadows related to mediastinum

- Pericardium
- Oesophagus can be characterized by performing a lateral chest X-ray
- Spinal cord

Abdominal radiography (abdominal X-ray)

Abdominal radiographs are often the first examinations performed for patients with abdominal symptoms. They can be helpful in the diagnosis of ***bowel obstruction*** (Figs. 8.10 and 8.11). AXR has low sensitivity (~15%) for detection of bowel perforation (presence of 'Rigler's sign' where both sides of a bowel wall are outlined distinctly by intra- and extraluminal gas, Fig. 8.12) but may be helpful in neonates.

Skeletal radiography

Since the first X-ray image taken by Röntgen of his wife Anna's hand in December 1895, skeletal radiography continues to be the modality to investigate the skeletal system today. X-rays can diagnose ***fractures*** (Fig. 8.13), ***arthritis*** (Figs. 8.14 and 8.15), and ***bone destruction from infection*** (Fig. 8.16) or ***cancer*** (Fig. 8.17). This is because of the excellent contrast between bones (which absorb a lot of X-rays) and adjacent soft tissues (which absorb less X-rays).

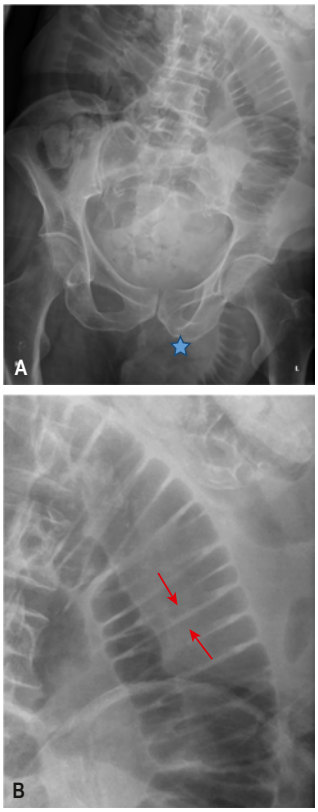


Fig. 8.10 Small bowel obstruction. This patient presented with abdominal pain, distension, and vomiting. AXR demonstrates small bowel obstruction. Notice dilated (>2.5 cm) loops of centrally positioned small bowel (recognized by valvulae conniventes which span across the whole width of the bowel, *red arrows*). The left inguinal hernia is the cause *(blue star)*.

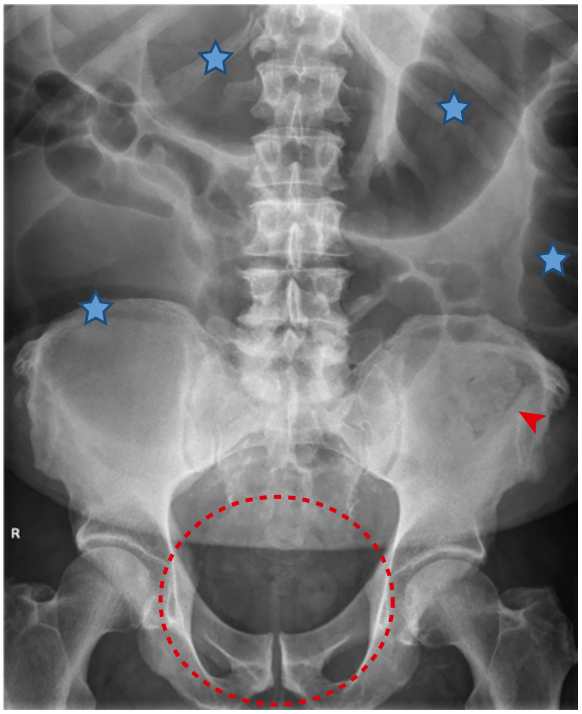


Fig. 8.11 Large bowel obstruction. This patient presented with abdominal pain and no bowel movement for 2 days. They were tender on abdominal examination. AXR demonstrated dilated peripherally placed large bowel *(blue stars)*, stopping abruptly in the descending colon *(red arrowhead)*. There is no visible rectal gas *(red outline)*. This patient had an obstructing descending colon tumour.

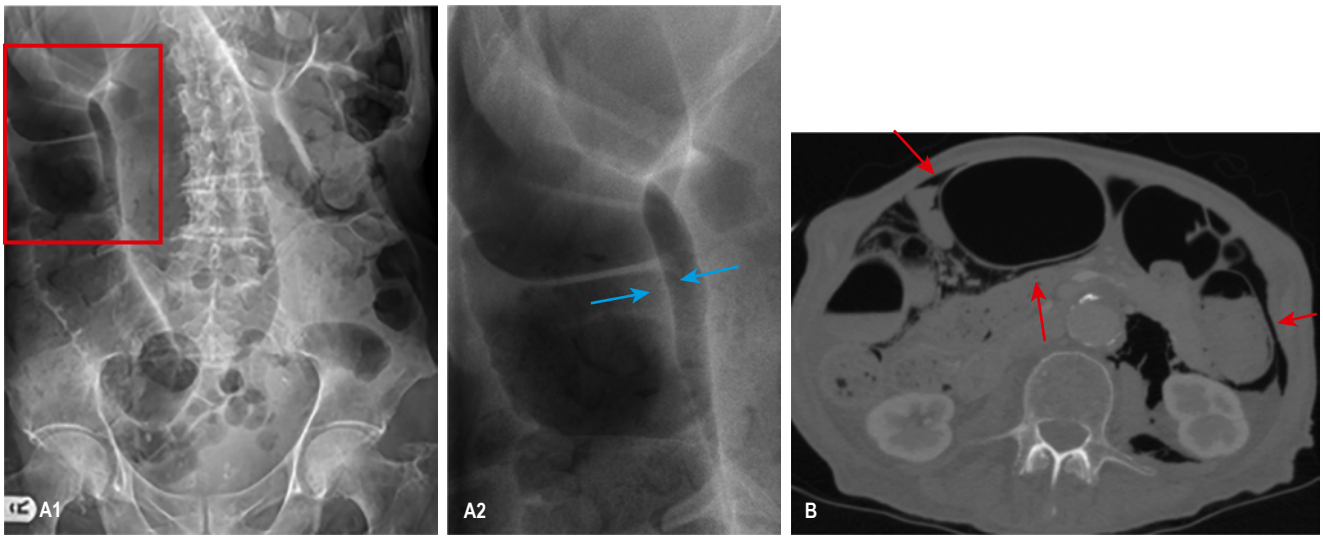


Fig. 8.12 Perforation: This patient with known sigmoid volvulus complained of severe abdominal pain. (A) AXR demonstrates the sigmoid volvulus and Rigler's sign *(blue arrows)*. **(B)** CT demonstrates large amount of free extraluminal gas *(red arrows)*, indicating perforation.

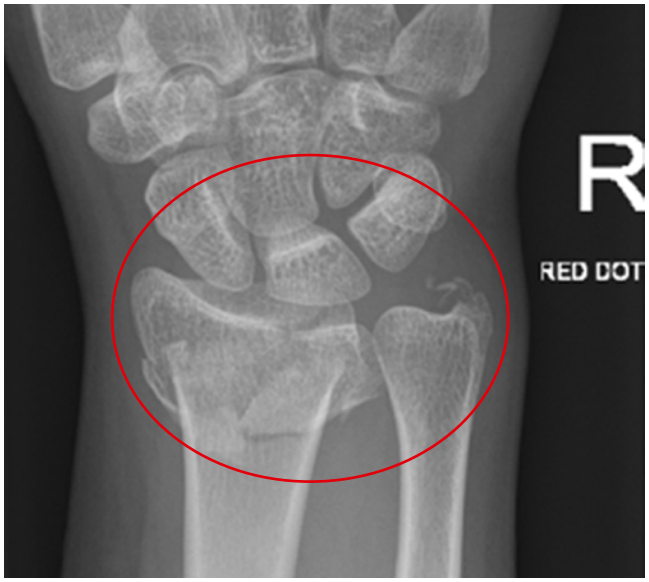


Fig. 8.13 Fractures of distal radius and ulna *(red circle)*. When reviewing an X-ray for possible fracture, trace the margin of a bone looking for disruption of the normal smooth contour and consequent loss of internal trabecular continuity. In this instance the emergency radiographer has noted an abnormality and marked the X-ray with a 'Red dot'.

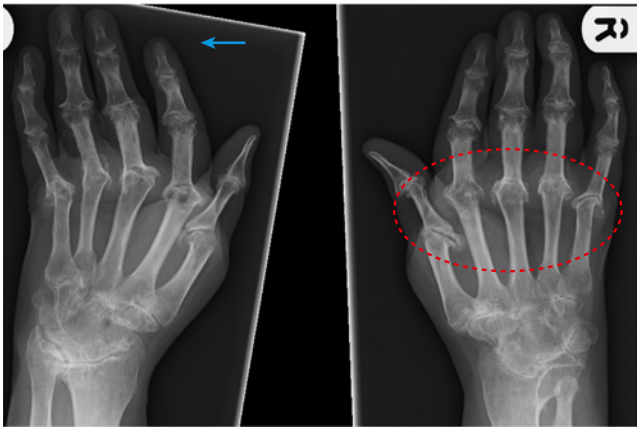


Fig. 8.14 X-rays of a patient's hands with rheumatoid arthritis. Notice the ulnar deviation (fingers deviating towards ulna: *blue arrow* demonstrating the direction) and symmetrical para-articular erosions affecting the metacarpals *(red circle)*, proximal interphalangeal joints, and wrists.

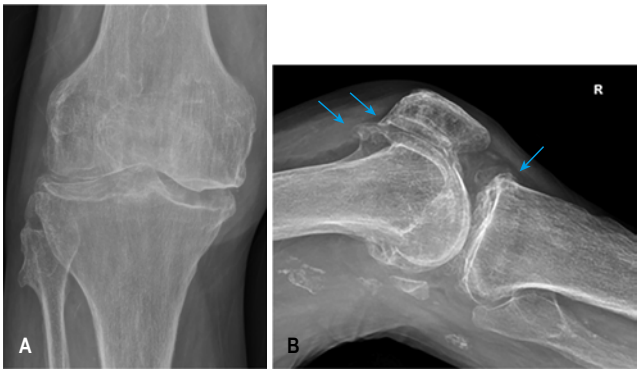


Fig. 8.15 Joint space narrowing, subchondral sclerosis, and osteophytes in a knee with marked degenerative change. *Blue arrows* pointing to visible exophytic osteophytes extending from the articulating surfaces of femur, patella, and tibia.

Fig. 8.16 Digital osteomyelitis. This patient with diabetes had a chronic ulcer on their great toe. Note the marked vascular calcification *(arrow)*. The tuft of the distal phalanx is ill-defined, in keeping with osteomyelitis. Note: normal X-rays do not exclude osteomyelitis. X-ray has a low sensitivity (5% at presentation; 33% at 1 week). If there is concern regarding osteomyelitis, bone scan or MRI should be performed.

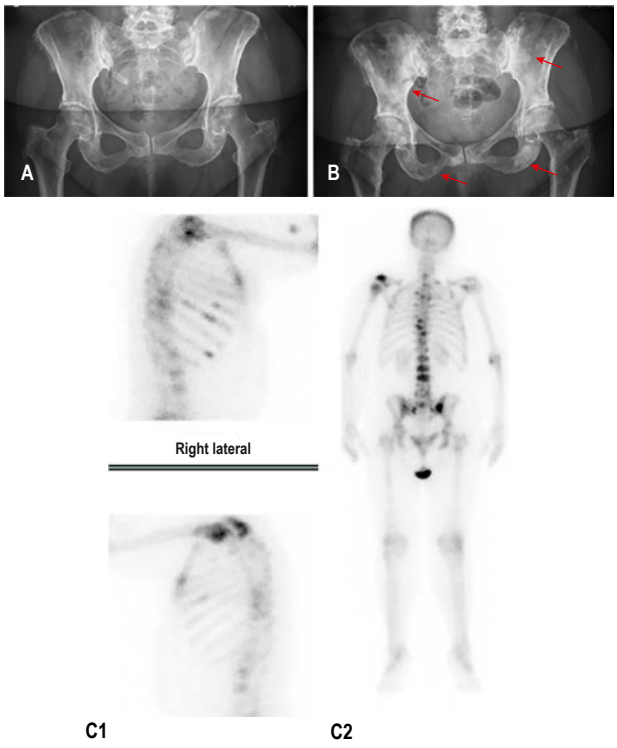


Fig. 8.17 Sclerotic bone metastases. This patient with breast cancer developed bone metastases. Notice the multiple new sclerotic bone metastases *(red arrows)* on the pelvic X-ray **(B)**, new from 8 months previous **(A)**. Bone scan (**C**, showing widespread skeletal 'hot spots' in the same patient) and MRI are much more sensitive at detecting bone metastasis and should be performed if clinical management is dependent on their identification.

Mammography

Breast imaging by mammography is used either as part of the assessment of female patients presenting with breast symptoms or as screening to detect ***early breast cancer*** (Fig. 8.18; see also Chapter 47). It uses X-rays to detect calcifications (e.g., ductal carcinoma in situ) or masses within the breast. Each breast is compressed during the procedure to acquire the best

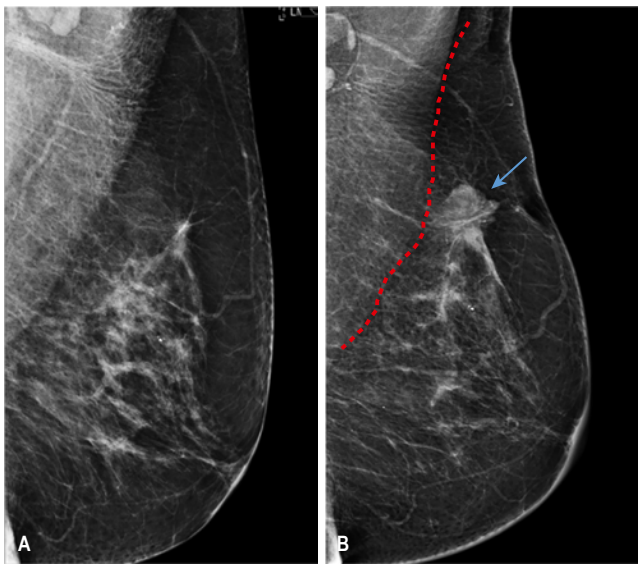


Fig. 8.18 Mammogram. This patient presented with a 2 cm palpable left breast lump. Mammography **(B)** demonstrates the mass *(blue arrow)* and tenting of the pectoralis muscle *(red dotted line)* from this tumour. These were not present on the previous examination **(A)**.

quality images. Some women may find this uncomfortable. Two (sometimes three) views are taken. Radiation-related carcinogenic risk is considered small in relation to the potential benefits of screening. Individuals who are known to be carriers of risk-increasing genetic variations and/or have an inherited predisposition to breast cancer, as well as those who suffer from hyper-susceptibility to ionizing radiation should avoid ionizing radiation as much as possible, however, and should instead be referred for ultrasound or MRI.

FLUOROSCOPY

Fluoroscopy uses X-rays to acquire images of a patient, usually during live, 'dynamic' examinations to demonstrate anatomy and interrogate function after administration of a radio-opaque contrast, for example barium or Gastrografin (an iodine-based contrast medium).

This section focuses on barium swallow and voiding micturating cystourethrogram (VMCUG) as examples of common fluoroscopic investigations. Other applications of fluoroscopy include barium defecography/proctography (imaging the patient's ability to evacuate their rectum) and hysterosalpingography (helpful in fertility investigation). Small bowel barium follow-through and intravenous pyelogram are less frequently/no longer performed in a modern radiology department and have been superseded by MRI enterography/enteroclysis and CT intravenous urogram, respectively, where superior images are obtained with similar or no (in the case of MRI) radiation.

Barium swallow

Common indications: High (cervical) dysphagia; suspicion of disordered swallowing (neurological or muscular); aspiration. (Upper GI endoscopy, rather than barium studies, is the preferred examination to investigate patients with dyspepsia and suspected gastro-oesophageal reflux or gastric ulcer disease.)

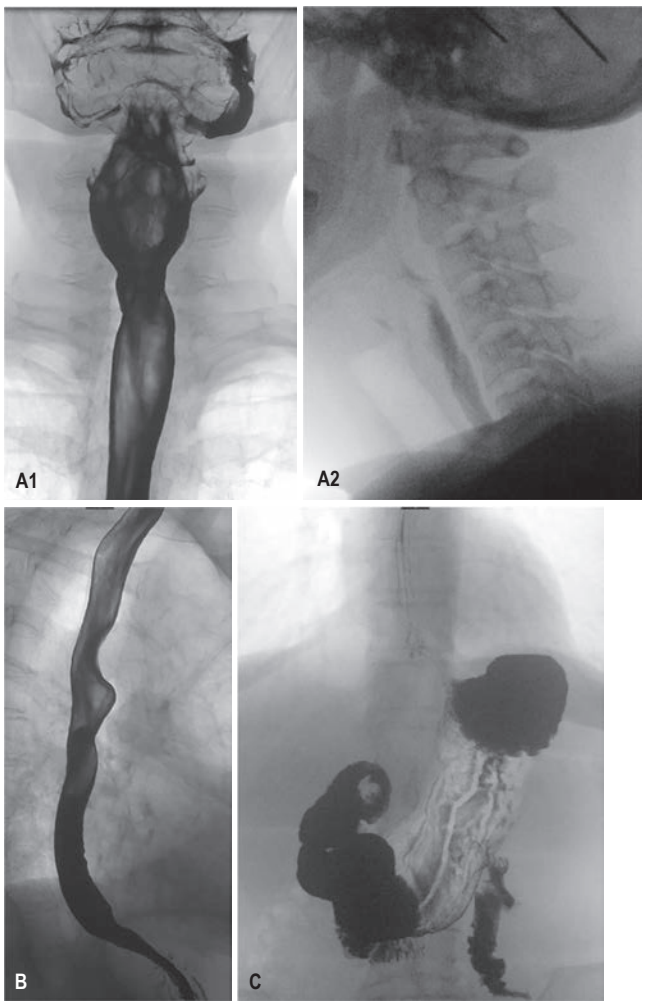


Fig. 8.19 Normal barium swallow. **(A)** Frontal and lateral view of cervical oesophagus; **(B)** oblique view of thoracic oesophagus; **(C)** frontal view of the stomach.

Patient preparation: Fasting, usually 4 to 6 hours—occasionally overnight.

Procedure: The patient is given a cup of liquid barium which can be mixed with an effervescent agent (Carbex®) to drink from. When instructed by the radiologist/fluoroscopist, the patient takes a mouthful of barium and swallows. Simultaneously the radiologist times the acquisition of 'cine'-images—much like a video—to demonstrate precisely the passage of barium from the mouth down the oesophagus.

Fig. 8.19 shows images from a normal barium swallow. Figs. 8.20 and 8.21 show findings from patients with achalasia and a pharyngeal pouch, respectively.

Voiding micturating cystourethrogram

Common indications: Urinary tract infection(s), dysuria, dysfunctional voiding, bladder outlet obstruction, trauma, neurogenic dysfunction of the bladder, congenital anomalies of the genitourinary tract, and postoperative evaluation of the urinary tract.

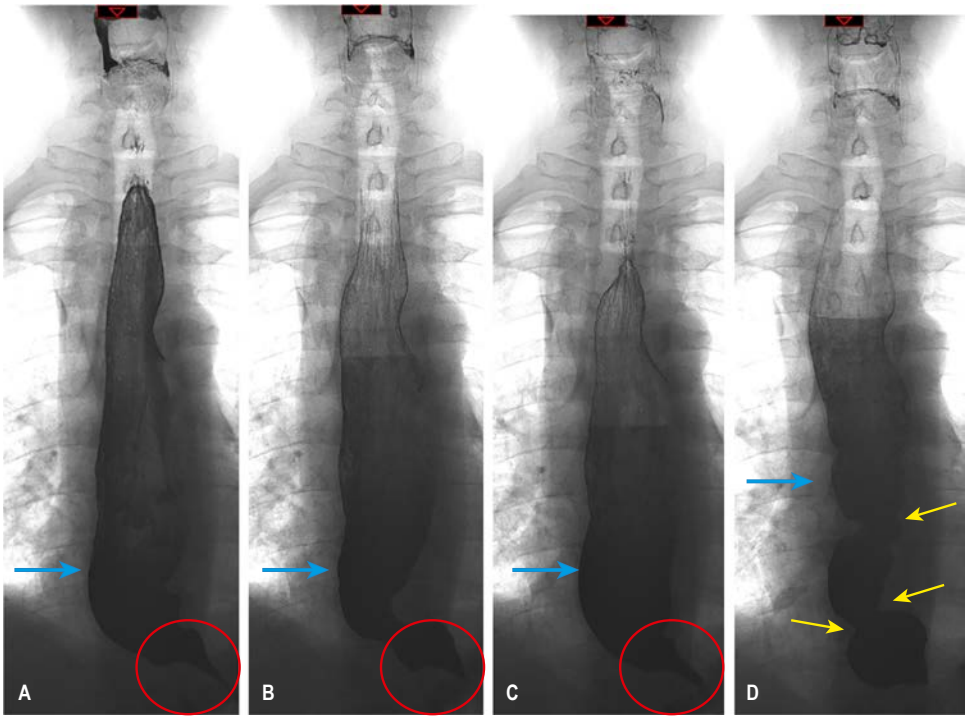


Fig. 8.20 Barium swallow: Achalasia. **(A–C)** There is a persistent narrowing at the end of the oesophagus (*red circles*, the lower oesophageal sphincter, LOS), with a dilated oesophagus *(blue arrows)* above the narrowed region. **(D)** Spastic contractions in the oesophagus above the LOS, are also visible *(amber arrows)*.

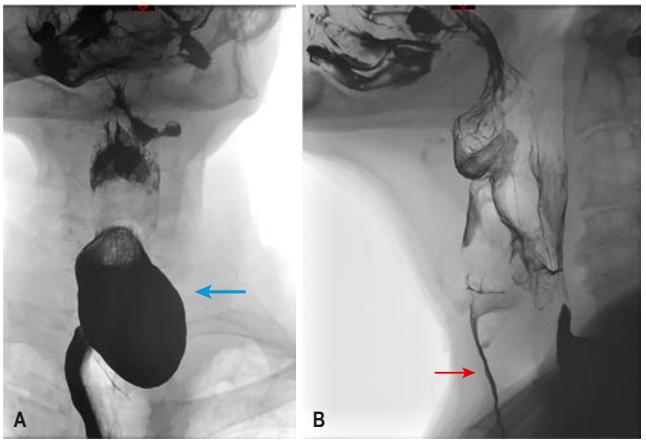


Fig. 8.21 Barium swallow: Pharyngeal pouch. This patient has a large pharyngeal pouch *(blue arrow)*. Aspiration of barium can be seen in the trachea *(red arrow)*.

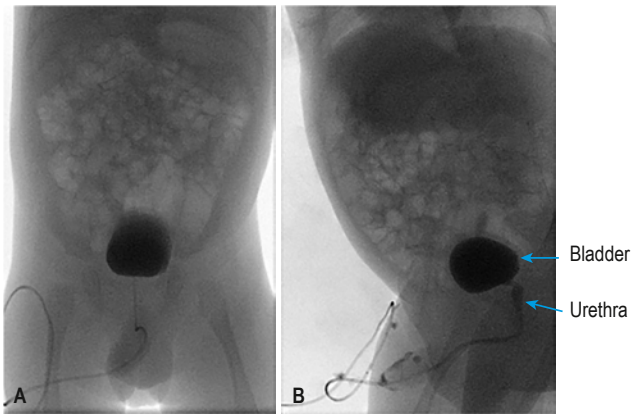


Fig. 8.22 Normal VMCUG. **(A)** Anteroposterior and **(B)** lateral views. Contrast is seen in the urinary bladder, urethra, and catheter.

Patient preparation: Patients often require antibiotics to 'cover' the procedure (this could be pre- and post-examination or doubling any usual prophylactic dosage for a few days, depending on agreed local protocols). Patients (adults or children who are able to cooperate) usually empty their bladders prior to the procedure. If the patient is an infant, an adult accompanies the infant to support them in remaining still for the examination.

Procedure: The patient having the examination lies on the fluoroscopy table. A urinary catheter is placed into the bladder through the urethra, under aseptic conditions. Contrast is then infused through the catheter to fill the bladder.

Babies and young children will naturally pass urine when the bladder is full. Older children/adults will be given a receptacle to void while lying on the table, when they feel ready. The radiologist/fluoroscopist will take cine-images/videos using X-rays while the patient voids.

A normal VMCUG is shown in Fig. 8.22 and images from a patient with reflux nephropathy in Fig. 8.23.

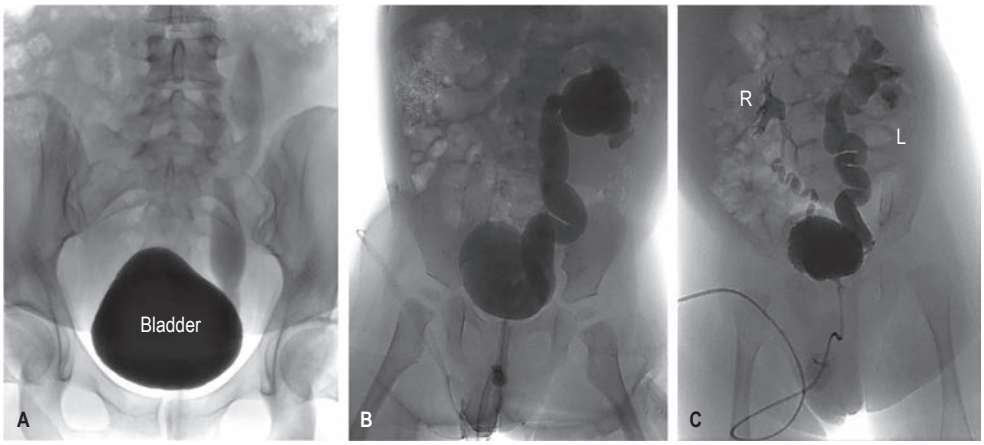


Fig. 8.23 VMCUG: Reflux nephropathy. **(A)** This study shows contrast in the urinary bladder. **(B)** Contrast can be seen refluxing into a dilated left ureter (with left hydronephrosis) and **(C)** non-dilated right ureter.

Box 8.5 Principles of ultrasound (US) imaging

- The ultrasound transducer contains piezoelectric crystals. These produce high-frequency sound waves when an electrical current is applied. They also act in reverse—producing an electric field when sound waves are applied to them. The transducer therefore both emits ultrasound waves into and detects waves reflected back ('echo') from the body. These electrical currents are converted by a computer into images by calculating the distance between the transducer and the body (using the speed of sound and the time of each echo's return).
- Minimizing the distance between the transducer and organ of interest reduces the energy lost in the body tissue, thereby improving the image quality.
- Gel applied between the transducer and body improves image quality by minimizing air pockets. Air blocks ultrasound waves and prevents images from being generated.
- Does not involve ionizing radiation.
- Relatively low cost and accessible modality; well tolerated by patients.
- Almost all parts of the body can be imaged. It is particularly useful in evaluating pelvic and intra-abdominal organs.
- Endoscopic ultrasound produces high-resolution images of many internal organs such as the oesophagus, heart (transoesophageal US), and the reproductive organs (transvaginal US).
- Ultrasound is occasionally performed with intravenous contrast agents e.g., to evaluate or treat liver or kidney tumours.
- Patient preparation is sometimes required (e.g., fasting when evaluating gallbladder or a full bladder when examining pelvic organs).

ULTRASONOGRAPHY

Ultrasound (US) is an excellent modality for evaluating many organs (**Box 8.5**). It is the modality of choice for evaluating superficially placed organs such as the thyroid gland, testes, and leg veins for venous thrombosis. US is also often used as the first-line test for the solid upper abdominal and pelvic organs with common indications including abdominal pain, deranged liver function, renal impairment, haematuria, and sepsis. US can even be used to evaluate a neonate's brain through the soft fontanelles. The use of US and Doppler to evaluate cardiac function by echocardiography is discussed in Chapter 17.

The kidney

US is frequently used as a first-line modality to image the urinary tract (**Fig. 8.24**). Common diagnoses visible on US include ***renal tract obstruction*** (**Fig. 8.25**) (resulting in hydronephrosis, a potential

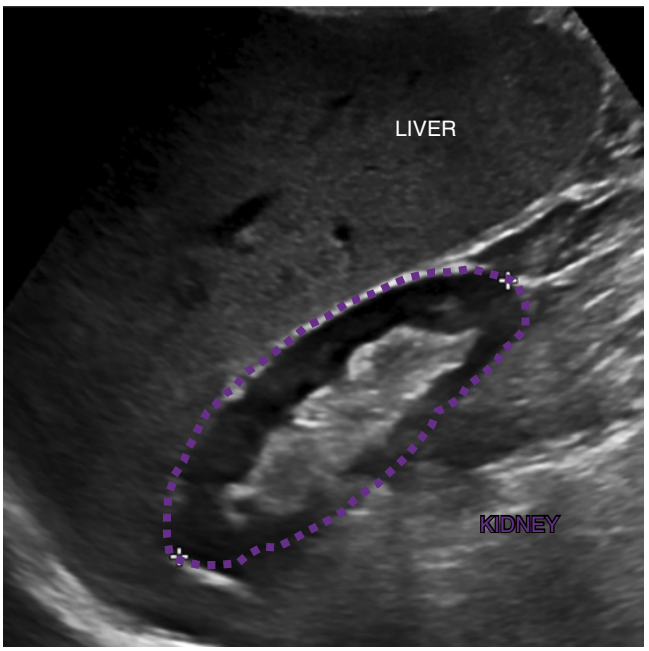


Fig. 8.24 The normal kidney on US. This US image demonstrates a normal right kidney *(purple outline)*, adjacent to a normal liver. Note normal renal cortex and medulla are distinct from the echo-bright renal sinus fat.

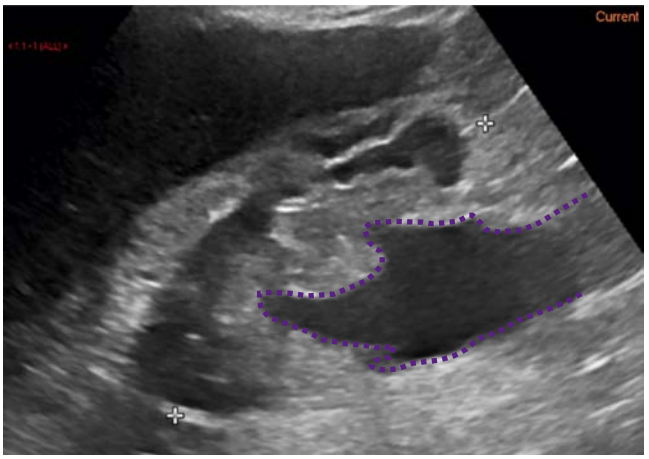


Fig. 8.25 Hydronephrosis on US. This kidney is obstructed. Notice the dilated renal pelvis and calyces *(purple outline)* extending from the renal medulla into the dilated ureter. Cortical thinning is also evident, particularly towards the lower pole, suggesting the obstruction may be long-standing.

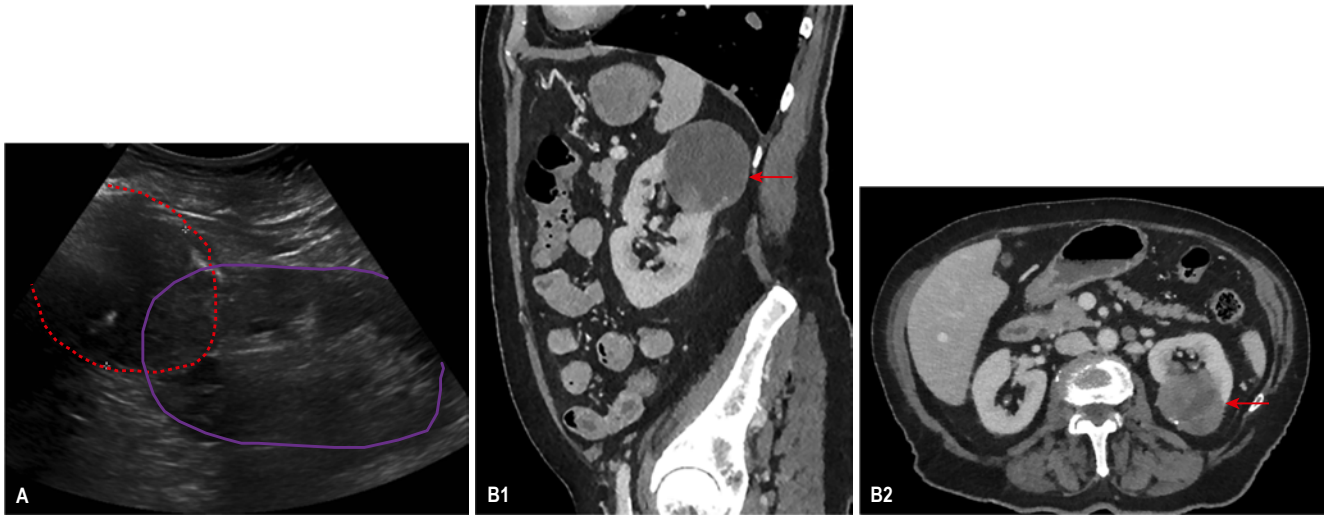


Fig. 8.26 Renal mass lesion on US. **(A)** US shows a large mass *(red outline)* at the superior pole of right kidney *(purple outline)*. **(B)** CT images of the renal tumour *(red arrow)* in the same patient.

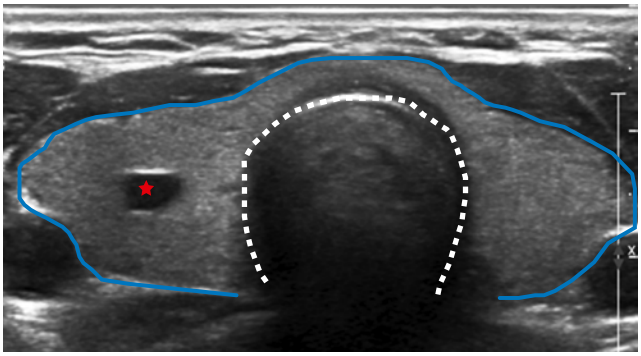


Fig. 8.27 US of normal thyroid gland. A normal thyroid *(outlined in blue)* wrapping around the air-filled trachea *(white dotted outline)*. Incidental cyst *(red star)* of no clinical significance.

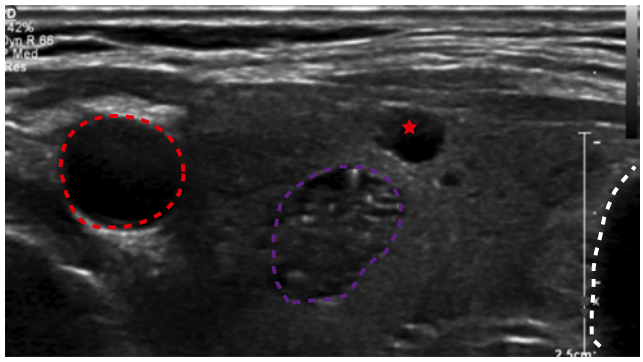


Fig. 8.28 Thyroid colloid cyst on US. This thyroid has a small colloid cyst *(purple outline)*, a common benign thyroid nodule. Trachea *(white outline)*, cyst *(red star)*, and jugular vein *(red outline)*.

cause of renal impairment or sepsis) and ***renal masses*** (Fig. 8.26; potential cause of haematuria or anaemia).

Thyroid gland

The thyroid gland lies superficially in the neck, on either side of the centrally placed trachea (Fig. 8.27). Imaging may be required if a neck swelling is palpable, to evaluate thyroid dysfunction or to further investigate abnormalities detected incidentally by other examinations such as CT or PET-CT (Figs. 8.28 and 8.29).

Pelvic organs: uterus and ovary

US is the first-line modality and investigation of choice when imaging the gynaecological pelvic organs (Fig. 8.30). These examinations require patients to have a full bladder; the urine-filled bladder allows transmission of ultrasound waves from the transducer on the abdominal wall to the pelvic organs, which would otherwise be interrupted by the gas-containing bowel, obscuring the view. Alternatively, a transducer placed internally—a transvaginal US (TVUS)—may be used, which gives much higher resolution images as the probe is now located much closer to the organs of interest (Fig. 8.31).

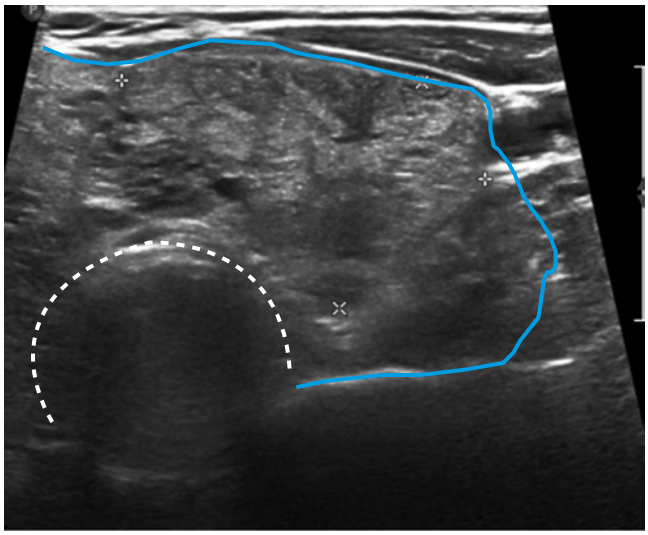


Fig. 8.29 Multinodular goitre on US. This thyroid is enlarged and contains multiple nodules—a multinodular goitre *(blue outline)*. The trachea is outlined in *white*.

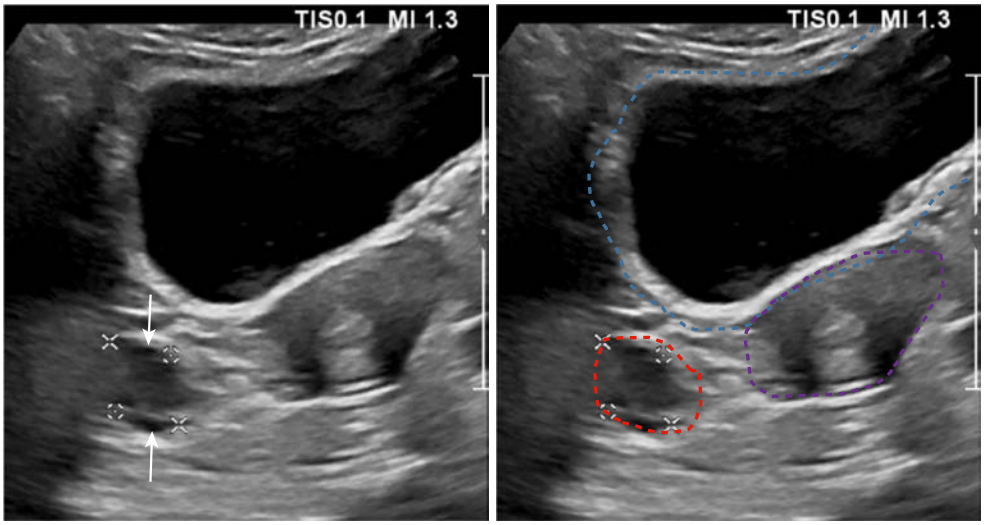


Fig. 8.30 Transabdominal US of the pelvic organs. This is a transabdominal US image of the right ovary *(red outline)* and uterus *(purple outline)*. Note the full bladder *(blue outline)* and the small cysts contained within the ovary *(white arrows)*.

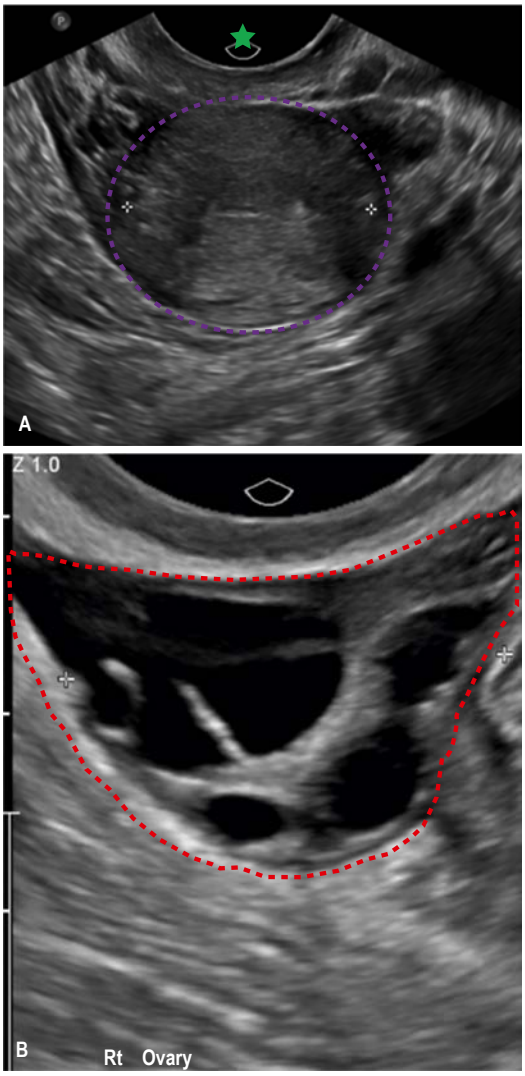


Fig. 8.31 Transvaginal US of the pelvic organs. TVUS images from the same patient as **Fig. 8.30**. The transducer *(green star)* is in much closer proximity to the uterus (**A**, *purple outline*) and right ovary (**B**, *red outline*), therefore the images obtained are of a superior quality.

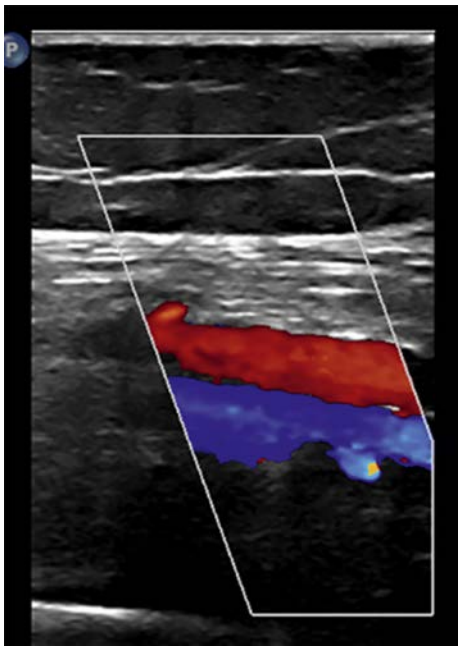


Fig. 8.32 US of femoral vein and artery. Doppler colour flow is demonstrated in the normal patent femoral vein *(blue)* and adjacent femoral artery *(red)*. Notice the colour block extends wall-to-wall within each vessel. The colour *(blue/red)* denotes the direction of flow of blood corpuscles: *blue* for blood flowing towards and *red* for blood flowing away from the transducer *(white angled box)*.

TVUS, however, may be uncomfortable for some patients, while others may find it too intrusive.

Venous ultrasound

Despite their name, the deep leg veins are readily seen on US (**Fig. 8.32**). Patent veins are easily compressible, completely anechoic ('black') internally within the lumen and when colour Doppler sound waves are applied, wall-to-wall colour can be seen inside the vessel. Venous US is commonly employed in the patient presenting with a swollen leg, where it can aid in discriminating between venous thrombosis, cellulitis, and 'simple' oedema (**Fig. 8.33**).

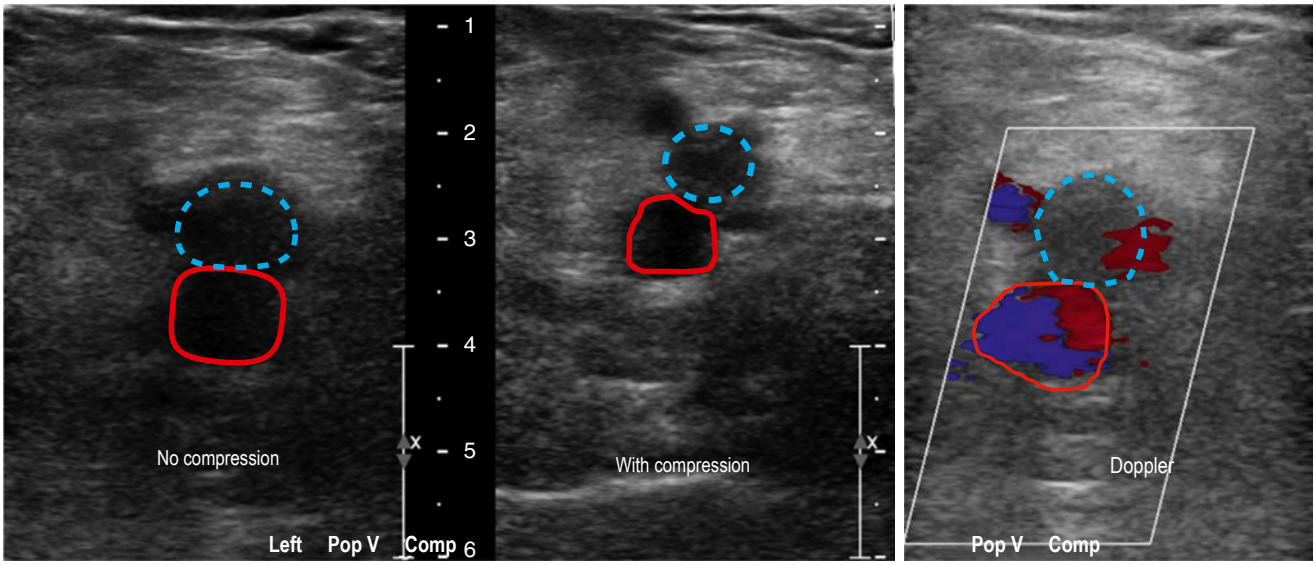


Fig. 8.33 Popliteal vein thrombosis. This patient has a thrombus in their popliteal vein. Notice the echogenic ('bright') thrombus lying within the lumen of the popliteal vein *(blue outline)*, in contrast to the patent popliteal artery *(red outline)*, which appears black. Notice too that the calibre of the popliteal vein does not change with external compression as the thrombus prevents the vein from completely flattening—as would be observed in a patent vein. No flow (colour) is seen in the vein when Doppler is applied. Wall-to-wall Doppler flow is visible within the patent artery.

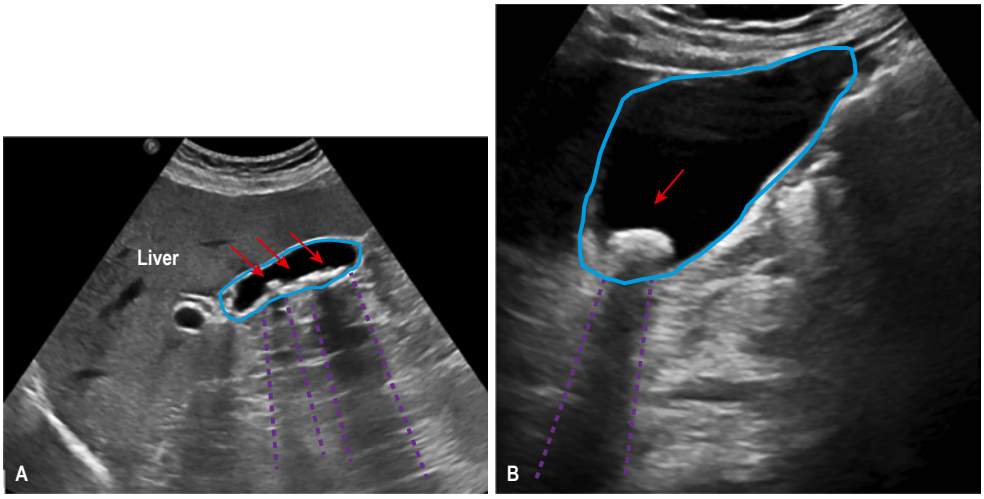


Fig. 8.34 US of liver and gallbladder. (A) There are multiple echogenic ('bright') gallstones *(red arrows)* within a normal thin-walled gallbladder *(blue outline)*. Notice the acoustic shadows cast by the gallstones *(purple lines)*. **(B)** Another example of an echogenic gallstone *(red arrow)* within a normal gallbladder *(blue outline)*, casting an acoustic shadow *(purple)*.

Hepatobiliary disease

The liver and gallbladder are easily visualized and well evaluated by US. Indeed, US is the most sensitive test for detecting the presence of ***gallstones*** (Fig. 8.34). Gallstones are best seen when the gallbladder is maximally distended—this requires the patient to be adequately fasted prior to the examination. US is also used as a first-line tool to evaluate a patient with ***deranged liver function tests***, a screening tool to detect ***hepatocellular carcinoma*** in patients with chronic liver disease and can be used to evaluate liver stiffness/fibrosis using a technique called elastography (see also Chapter 23).

Testis

Just like the thyroid gland, the testes are readily accessible to US, lying immediately under the skin (Fig. 8.35). Abnormalities within the testes are often palpable and easily visualized by US (Fig. 8.36).

Special ultrasound procedures

High-intensity focused ultrasound (HIFU): Therapeutic ultrasound which uses high levels of acoustic output to focus on specific targets for the purpose of heating, ablating, or

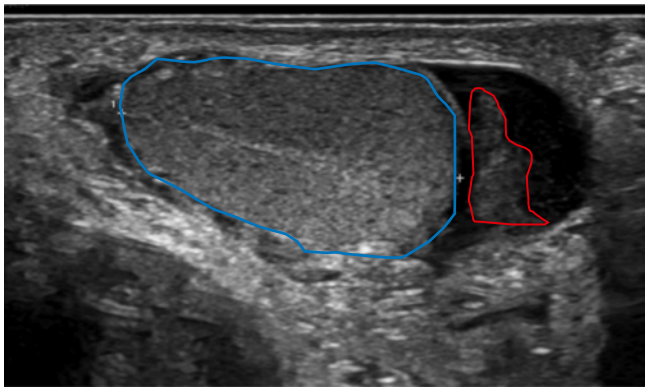


Fig. 8.35 US of the normal testis. A normal testis *(blue outline)* lying superficially under the scrotal skin. Epididymis is outlined in *red*.

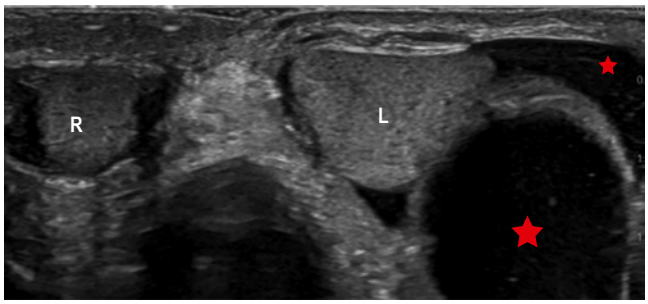


Fig. 8.36 Hydrocoele. Both testes are seen in this image (R and L). A left-sided hydrocele (dark excess fluid around the testis, *red stars*) is present.

breaking up tissue. Examples include treatment of uterine fibroids, to alleviate pain from bone metastases, and ablation of prostate tissue.

Ultrasound contrast: Intravenous injection of a highly echo-reflective substance, usually microbubbles filled with gas, allows assessment of vascularization of target structures—commonly used for focal masses in the liver or kidneys.

COMPUTED TOMOGRAPHY

Computed tomography (CT) is the 'workhorse' of the modern radiology department (Box 8.6). It is the modality of choice for evaluating many acute or emergency presentations on account of the excellent diagnostic images, ready availability, and speed of image acquisition—ideal for the unwell patient. It is also used for routine diagnostic or follow-up indications.

Examinations may be performed with or without iodinated intravenous contrast agents (Box 8.7), depending on the clinical indication. This section will discuss the common examinations under each anatomic system.

It should be emphasized that CTs are complex examinations. Interpretation of these large datasets of images requires years of training and (uninterrupted) time. By providing the reporting radiologist with as much salient information as possible, you will help improve the accuracy of the report—and patient outcome—greatly.

Tip: when viewing a CT scan in the axial plane, imagine you are viewing the image from the perspective of the patient's feet—structures on the right of the patient's body (e.g., liver) will be seen on the left side of the image.

Box 8.6 Computed tomography (CT)

- Images are produced with the same principles as X-rays.
- The scanner contains one or two X-ray sources and multiple detectors spin rapidly around the patient as they move through the scanner on the table.
- Images are produced by a computer from the information received by the detectors. Images can be reconstructed in multiple planes (axial, sagittal, and coronal) and even in three-dimension.
- Just like X-rays, images relate to how much X-ray energy is absorbed by the body: for example, bones appear 'white' or 'dense' on a CT because most of the X-rays have been absorbed. Lungs filled with air absorb little X-ray so appear 'dark'.
- Density of images is described using 'Hounsfield Units' (HU). Water is assigned as having 0 HU. Bones are dense with high HU (often >1000), air is low density (<–500 HU), and soft tissue is around 50 HU.
- Involves ionizing radiation.
- Radiation dose is higher than X-ray.
- All parts of the body can be imaged.
- Patient preparation is sometimes required (e.g., venous access may be required for intravenous contrast).

Box 8.7 Iodinated intravenous contrast: patient considerations

- Require venous access—large bore if angiogram is being performed.
- ALLERGY: Some patients may be allergic to iodinated contrast. It is important to alert the Radiology Department if the patient has had a previous contrast-related reaction.
- RENAL IMPAIRMENT: Patients with renal impairment should receive intravenous contrast if its use is deemed necessary to answer the clinical question with monitoring of renal function following the exposure.
- HYPERTHYROIDISM: Patients with a history of hyperthyroidism rarely develop thyrotoxicosis. However, iodinated contrast should be avoided in patients with acute thyroid storm or those considering/undergoing radioactive iodine therapy.

The brain

CT brain is often the first imaging examination for patients presenting with neurological impairment, confusion, headache, or trauma. It demonstrates the cerebral soft tissues, ventricles, and cerebrospinal fluid spaces as well as the cranium with exquisite clarity—allowing the diagnosis of ***infarction*** (Fig. 8.37), ***haemorrhage*** (Figs. 8.38 and 8.39), or ***tumour*** (Fig. 8.40).

Thorax

Common indications for performing CT of the thorax include further assessment of an abnormal CXR (Fig. 8.41); unexplained weight loss in a smoker (Fig. 8.42) or if there is clinical concern of an underlying pulmonary embolus (CT pulmonary angiography [CTPA], see Fig. 8.49). With the exception of pulmonary nodule follow-up and CT for diagnosing interstitial lung disease, most CT thorax examinations are performed with intravenous contrast. Low-dose CT scans of the thorax are used to screen for lung cancer in patients with elevated risks of developing lung cancer (i.e., those with significant smoking history).

Abdomen and pelvis

The patient presenting acutely with abdominal pain or trauma is frequently sent for a CT of abdomen and pelvis. Common diagnoses

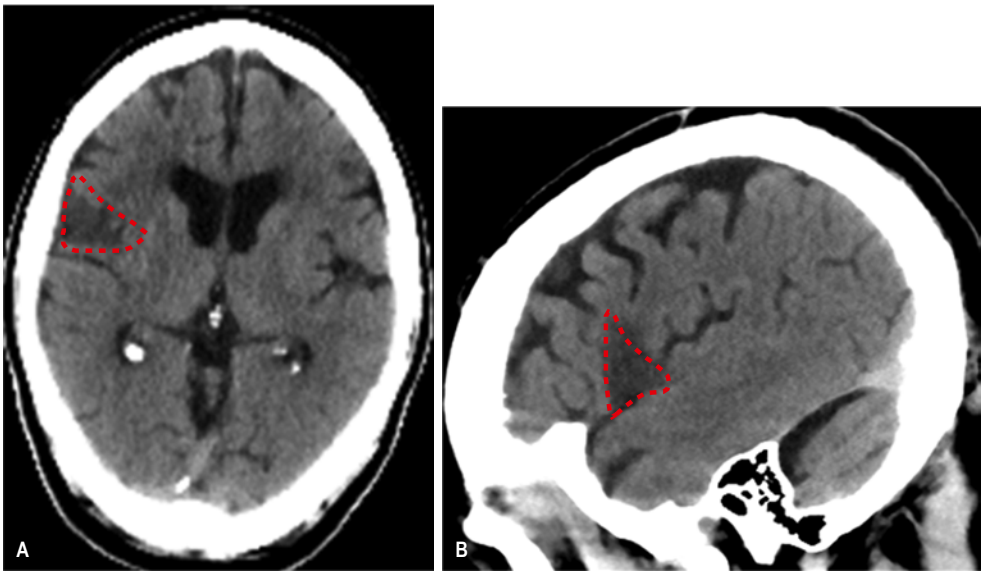


Fig. 8.37 Cerebral infarction on CT. This patient presented with a sudden onset of frontal headache, slurred speech, and facial droop. A wedge-shaped low-density zone is apparent in the right frontal lobe *(red outline)*: an area of infarction consistent with an ischaemic stroke. **(A)** Axial view; **(B)** saggital view.

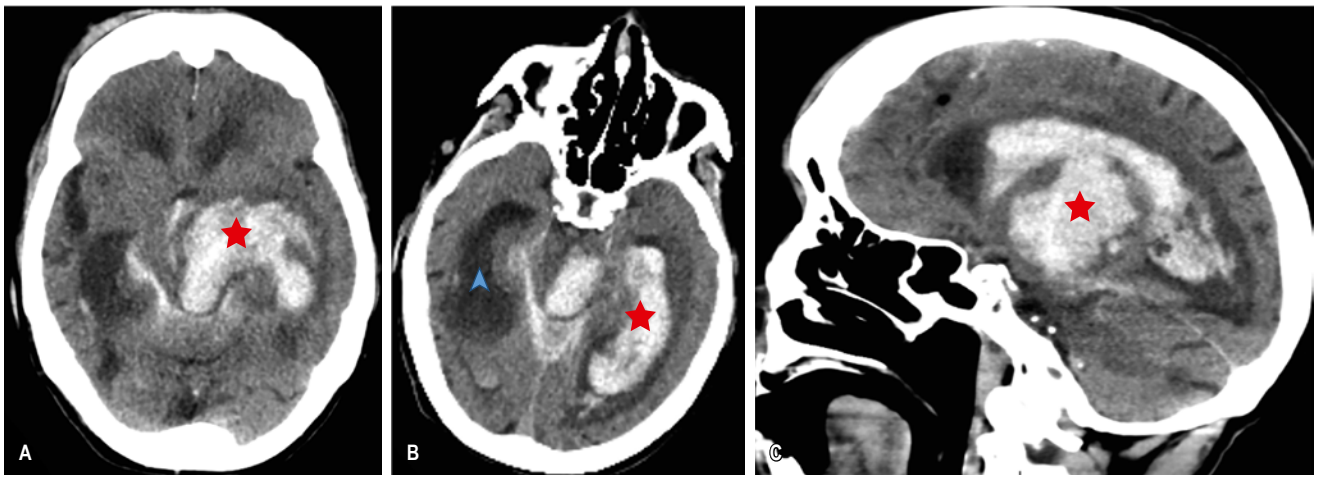


Fig. 8.38 Intraventricular haemorrhage on CT. This patient was found collapsed on the floor at home having last been seen the night before. There were signs of external head injury and GCS was 3. CT demonstrates blood in all ventricles *(red star)*, causing obstructed hydrocephalus *(blue triangle)* and left to right midline shift. **(A)** and **(B)** Axial views; **(C)** saggital view.

include ***renal calculi*** (Fig. 8.43), ***hernias*** (Fig. 8.44), ***bowel obstruction*** (Figs. 8.45 and 8.46), intra-abdominal causes of sepsis, ischaemia, haemorrhage, and complications of severe pancreatitis.

CT allows for optimized preoperative surgical planning and may prevent unnecessary exploratory laparotomies.

CT angiography

CT angiography is performed with high-flow contrast injected intravenously through a wide-bore catheter. Typically, the scan is triggered when the Hounsfield index (density) within the target vessel (aorta or pulmonary artery) reaches the predetermined level to allow maximal contrast opacification of the arterial system being studied (so-called 'bolus-tracking'). The catheter is usually placed in the arm. Specially manufactured central venous catheters able to withstand the injection pressures (without splitting or embolizing) may also be used.

Common applications of CT angiography include to diagnose ***aortic dissection*** (Fig. 8.47), investigate ***internal bleeding*** (Fig. 8.48), and to diagnose ***pulmonary embolism*** (PE; Fig. 8.49).

CT urography

CT urography is performed approximately 5 to 10 minutes after injection of intravenous iodinated contrast medium, when the iodine is excreted into and opacifies the urinary tract. A second dose of intravenous contrast is often given at the time of scan to allow better evaluation of the solid upper abdominal organs including the kidneys and liver. CT urography is most commonly requested in the investigation of ***haematuria*** (Fig. 8.50).

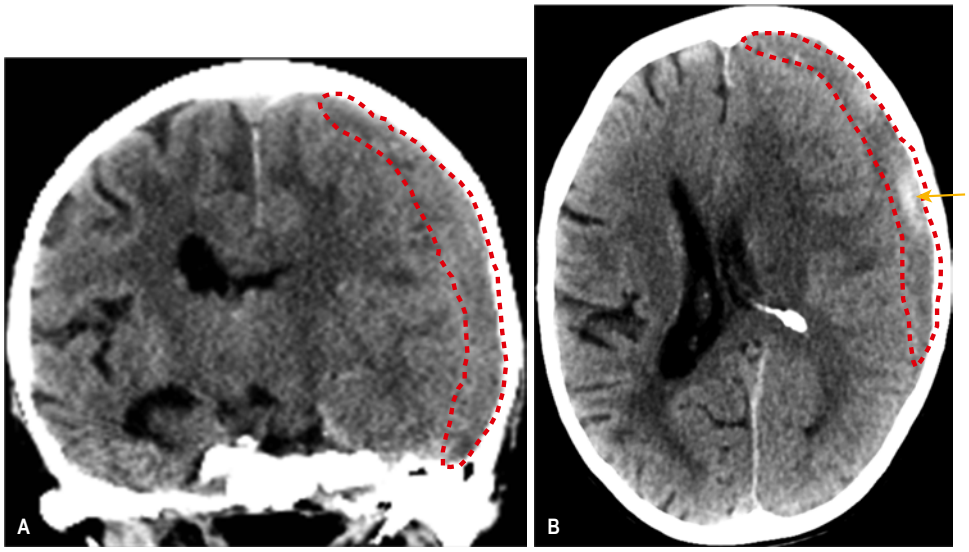


Fig. 8.39 Subdural haematoma on CT. This patient, taking regular apixaban, presented with a fall and head injury followed by nausea, vomiting, and headache. CT demonstrates a crescent-shaped left extra-axial acute-on-chronic subdural haematoma *(red outline)*. Acute bleeding appears of high density *(amber arrow)*. There is cerebral oedema (visible as low-density change in the white matter) and left-to-right midline shift due to the brain swelling and the mass effect of the haematoma itself. This patient needs an urgent neurosurgical opinion. **(A)** Coronal view; **(B)** axial view.

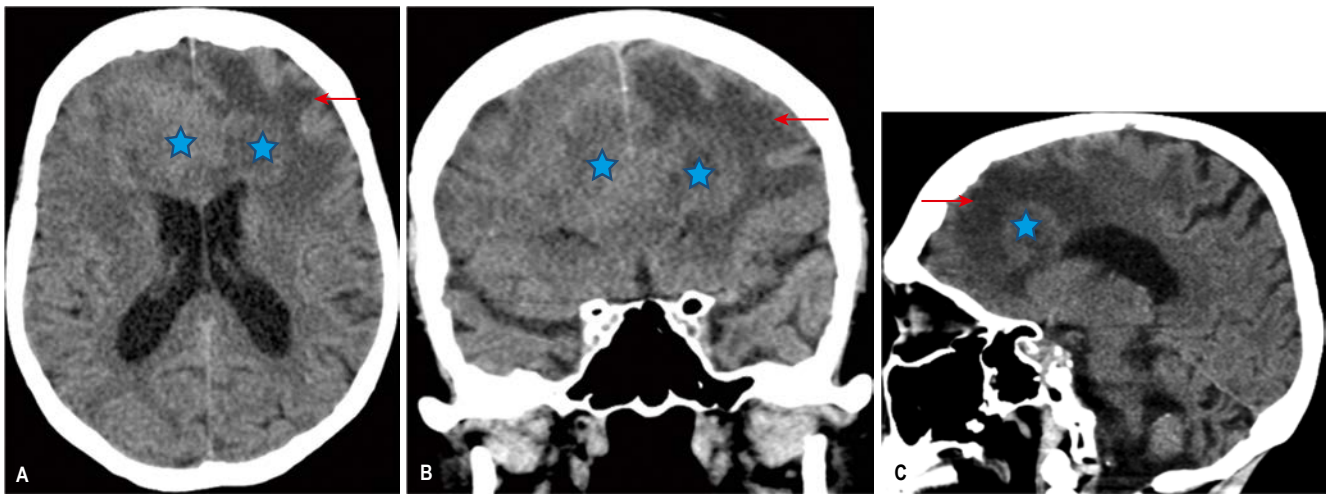


Fig. 8.40 Brain tumour on CT. This patient has a 3-month history of worsening confusion, agitation, and falls. CT demonstrates bilateral frontal masses *(blue stars)* with adjacent cerebral white matter oedema *(red arrows)*, consistent with metastatic cerebral deposits. **(A)** Axial view; **(B)** coronal view; **(C)** saggital view.

Special CT procedures

CT colonography (CTC) and CT coronary angiography (CTCA) are specialized examinations evaluated by subspecialist radiologists.

CT colonography (colloquially known as 'virtual colonoscopy') has been shown to be almost as sensitive as colonoscopy at detecting colonic cancer and larger (over 6 mm) polyps. It is a good alternative to investigate the large bowel, particularly for frailer or elderly patients. A big advantage of CTC over colonoscopy is the ability to provide information on abdominal organs outwith the bowel, which may be causing the patient's symptoms (e.g., anaemia secondary to renal tumour).

Patients are given oral laxatives to cleanse the bowel and oral iodinated contrast or barium to 'tag' residual colonic content with high-density material—meaning stool can be distinguished from colonic polyps or cancers. The colon is then inflated with carbon dioxide (via a pressure-sensitive catheter inserted per rectum) and CT of the abdomen is performed twice, in the supine and prone positions. Antispasmodic medication is usually given intravenously if there is no contraindication (e.g., significant cardiac disease such as heart failure, poorly controlled coronary heart disease, or cardiac tachyarrhythmia). It is important to inform the radiologist if the patient has these conditions. Risk of complication (e.g., perforation) is extremely low (1–2 per 10 000 examinations). CTC should be

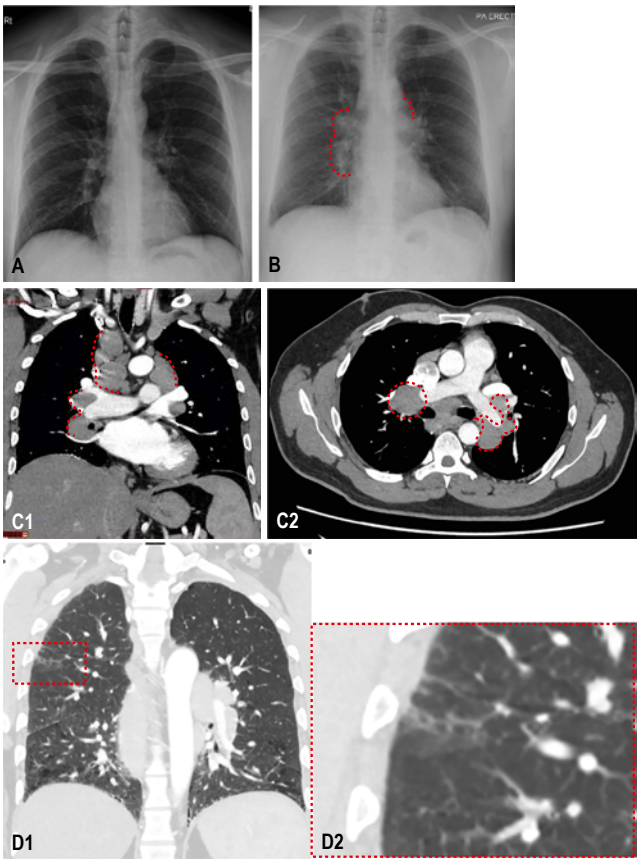


Fig. 8.41 Sarcoid lung disease. An abnormal CXR (**A** from 2 years ago; **B** current). CT demonstrates the extensive enlarged mediastinal lymph nodes *(red dotted lines)* with pulmonary ground-glass opacity and bronchovascular nodularity visible **(C, D)**.

Fig. 8.43 Renal calculi on CT. This patient presented with bilateral flank pain, haematuria, reduced renal function, a high serum calcium, and elevated PTH. Unenhanced, non-contrast CT kidneys, ureter, and bladder (KUB) demonstrate bilateral renal calculi *(red arrows)* and left renal pelvis calculi with mild left hydronephrosis (*blue arrow*; compared to normal renal collecting system on the right, *gold arrow*). The patient was subsequently diagnosed with primary hyperparathyroidism. **(A)** Coronal view; **(B)** axial view..

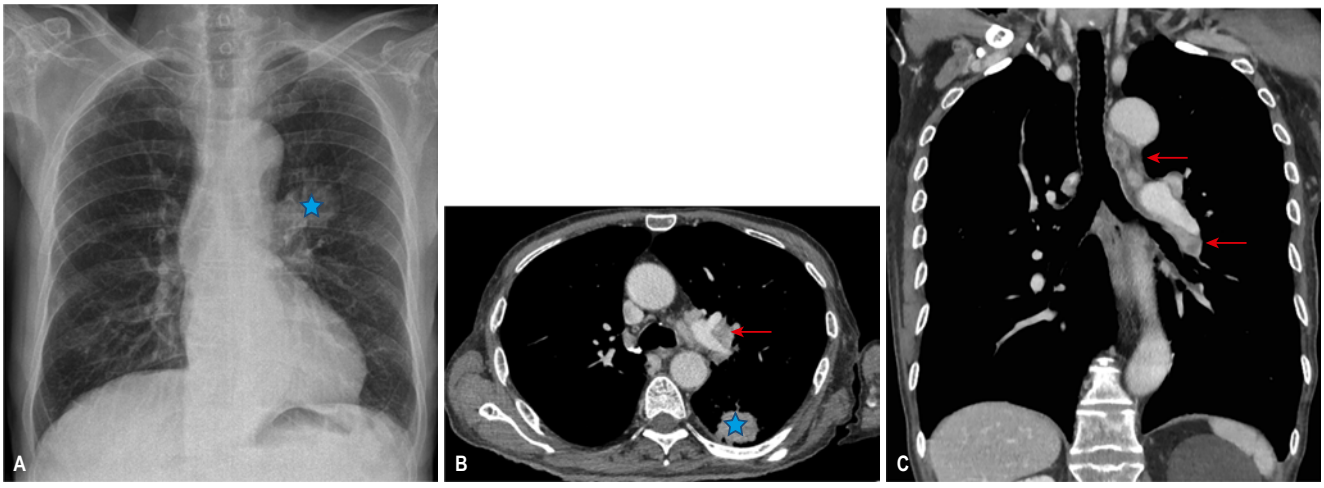


Fig. 8.42 Lung cancer. This patient was a smoker and presented with left shoulder pain and weight loss. CXR **(A)** was abnormal. CT demonstrates left pulmonary mass *(blue star)* and left hilar abnormal lymph nodes *(red arrow)*. **(B)** Axial view; **(C)** coronal view.

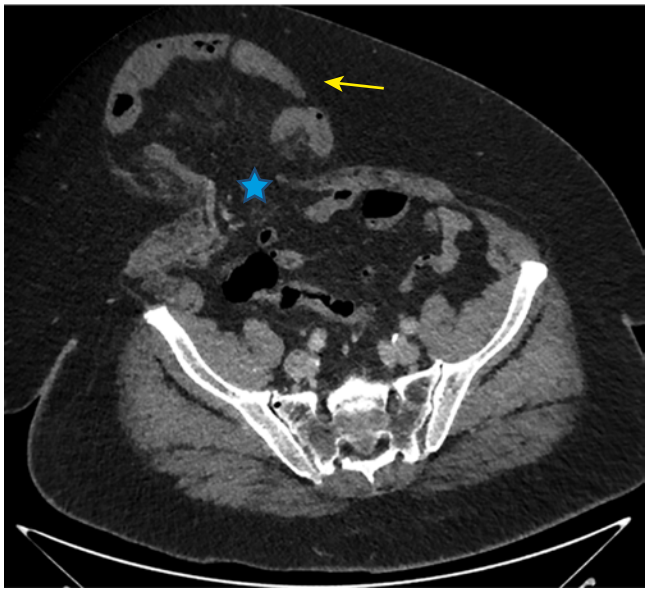


Fig. 8.44 Incarcerated umbilical hernia. This patient with a history of severe obesity described a 5-day history of central abdominal pain. On examination, there was a large 20 × 20 cm firm tender swelling in the mid-upper abdomen, consistent with an incarcerated umbilical hernia. The patient was not vomiting and was opening bowels normally, but white blood cell count was elevated. CT (axial view) demonstrates the large anterior abdominal wall hernia *(blue star)* containing small bowel *(amber arrow)*. No obstruction nor bowel ischaemia was evident on CT, but operative small bowel resection was later performed for infarcted bowel.

Fig. 8.46 Small bowel obstruction. This patient with a past history of abdominal surgery presents with abdominal pain, distension, and vomiting. CT (coronal view) demonstrates dilated loops of small bowel with a transition point at an abrupt calibre change *(green arrow)*. Appearance is in keeping with small bowel obstruction from adhesions.

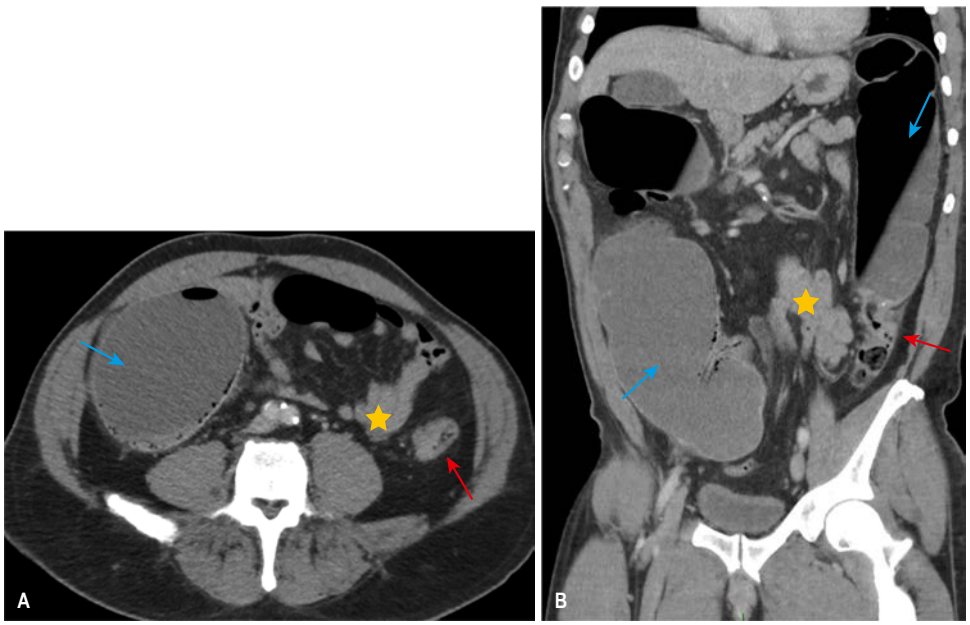


Fig. 8.45 Large bowel obstruction. This patient presented with abdominal pain, constipation, and distended abdomen. CT demonstrated large bowel obstruction with a competent ileocaecal valve (non-dilated small bowel, *amber star*; dilated right and left colon, *blue arrows*) from a descending colon tumour *(red arrow)*—a surgical emergency due to the high risk of colonic perforation. No metastases were evident on CT, so subtotal colectomy was performed. **(A)** Axial view; **(B)** coronal view.

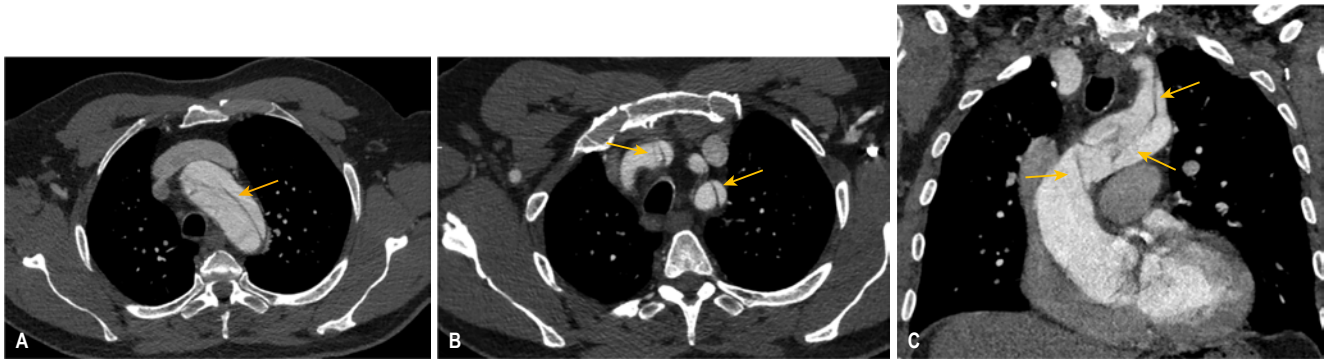


Fig. 8.47 CT angiogram showing aortic dissection. This patient presented with sudden onset neck pain and syncope. On examination, there was 50 mmHg difference in systolic blood pressure between the arms and a loud aortic regurgitant murmur. CT angiography demonstrated type A aortic dissection, extending into the carotid and brachiocephalic arteries *(amber arrows)*. Aortic arch and tissue aortic valve replacement were performed. **(A)** and **(B)** Axial views; **(C)** coronal view.

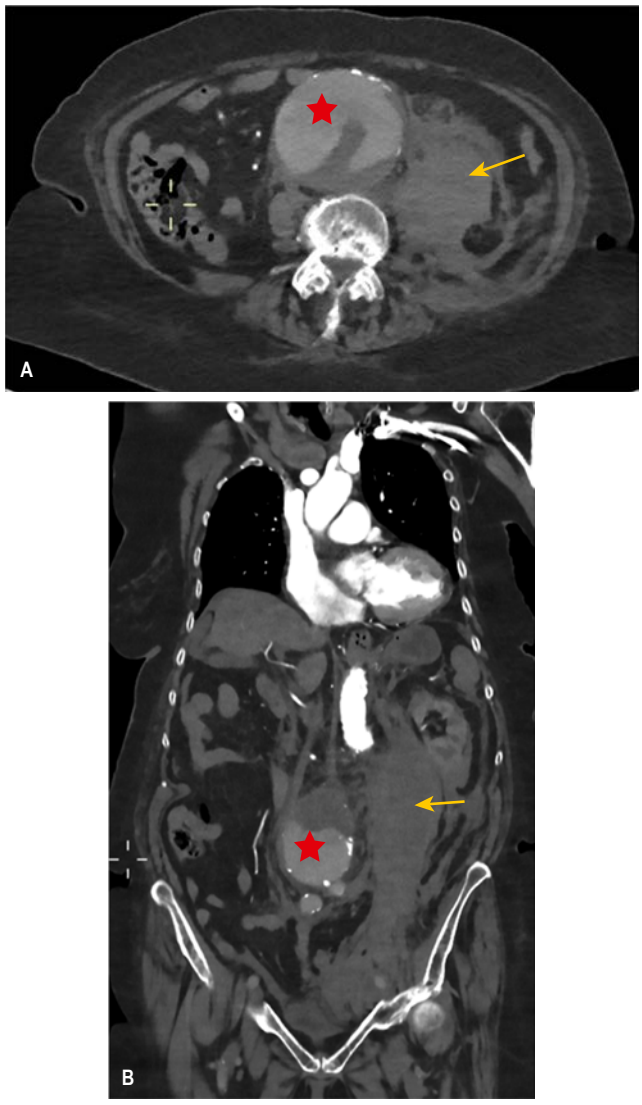


Fig. 8.48 CT angiogram showing aortic aneurysm rupture. This patient presented with severe back pain, radiating to the front of the abdomen. CT angiogram demonstrated a large aortic aneurysm *(red star)*, which was leaking into the left retroperitoneal/para-psoas space *(amber arrow)*. **(A)** Axial view; **(B)** coronal view.

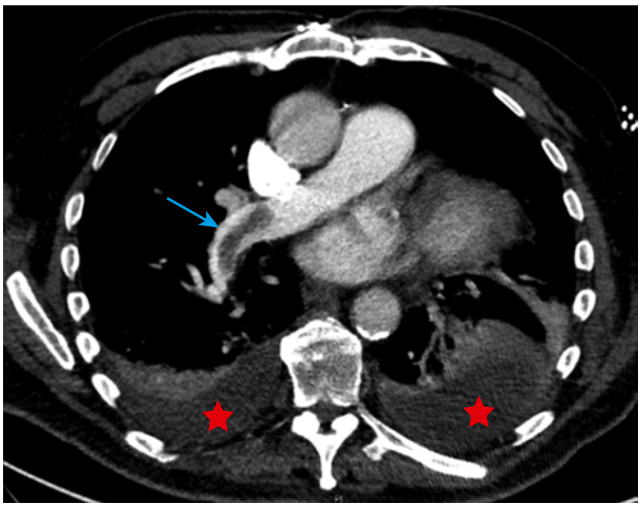


Fig. 8.49 CT pulmonary angiogram showing pulmonary embolism. This patient presented with severe sudden-onset chest pain and tachycardia. They had undergone recent surgery. CT pulmonary angiogram (axial view) demonstrates a right pulmonary artery embolus *(blue arrow)* and bilateral effusions *(red stars)*. No pulmonary trunk enlargement or ventricular septal right to left deviation (signs of right ventricular dysfunction indicative of sub-massive PE) are evident.

deferred if there has been recent large colonic biopsy. It is contraindicated in patients with large bowel obstruction or active inflammatory bowel disease.

CT coronary angiography is recommended as the first-line test for evaluation of stable coronary artery disease in chest pain pathways, allowing invasive percutaneous angiography to be reserved for therapy. CT, performed with intravenous contrast, beta-blockers to slow the heart rate and occasionally nitroglycerin produces exquisite images to identify narrowed or blocked arteries in the heart. The patient is connected to an electrocardiogram (ECG) monitor during the test, and the CT scanner is synchronized with the heartbeat to ensure imaging occurs when the heart is most likely to be still.

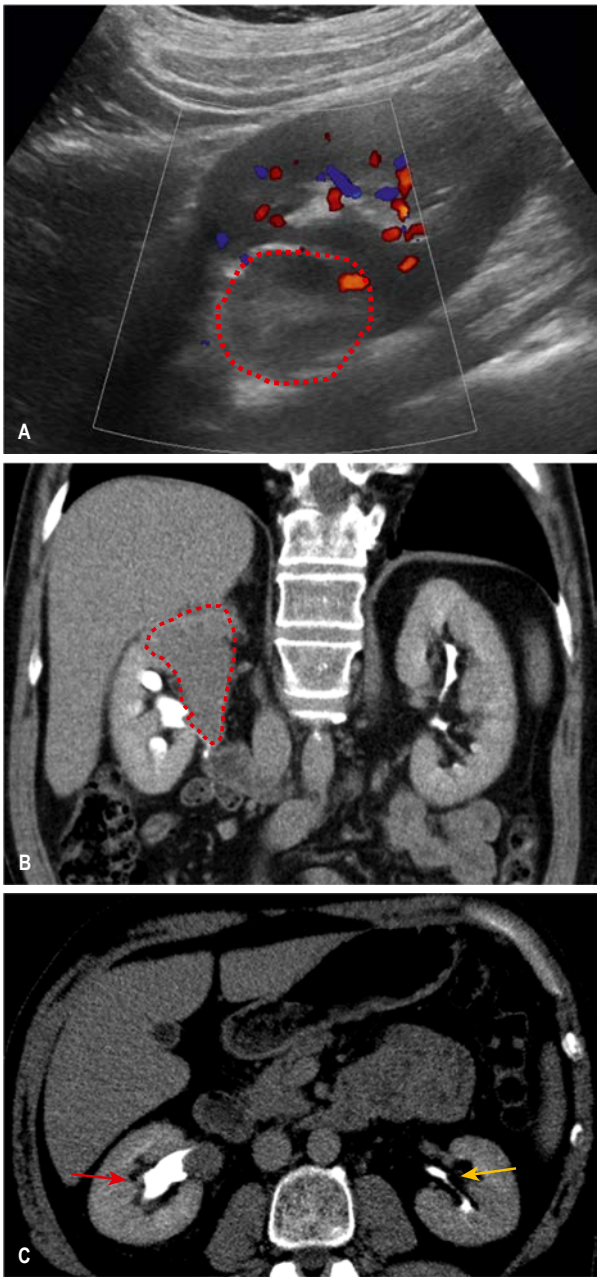


Fig. 8.50 CT urogram showing renal tumour. This patient presented with visible haematuria. **(A)** An initial US was abnormal. **(B)** CT urogram (coronal view) demonstrates a mass *(red dotted line)* arising from the upper pole of the right kidney, extending into the renal pelvis. **(C)** The mass is causing a degree of hydronephrosis, seen well on axial view (*red arrow*; compared to normal left system, *amber arrow*).

The patient should fast and avoid caffeinated beverages 12 hours before the test. Even though the actual scan is over in seconds, the whole process of preparation can take up to an hour with the patient on the scanner table.

While the risk of serious complications from CTCA is lower than percutaneous coronary angiography, subsequent percutaneous coronary angiography or cardiac surgery may still be required if CTCA demonstrates a treatable lesion.

MAGNETIC RESONANCE IMAGING

Magnetic resonance imaging (MRI) excels in delineating soft tissues, making it well suited to the investigation of musculoskeletal tissues, brain, liver, and pelvic organs (**Box 8.8**). Factors limiting its use include scanner availability and length of time required for image acquisition—not ideal for the unwell patient. Examinations may be performed with or without intravenous gadolinium-based contrast agents (**Box 8.9**). The use of powerful medical magnets has important safety implications requiring strict protocols, including the completion of detailed checklists to ensure that patients with MRI contraindications are not exposed to the magnetic field (**Fig. 8.51** and **Box 8.10**).

i Box 8.8 Magnetic resonance imaging (MRI)

- MRI scanners create powerful magnetic fields to align protons that exist in body tissues that contain water. The applied radio waves interact with these protons to produce signals that are picked up by a receiver within the MRI scanner. The signals are specially characterized using the rapidly changing magnetic fields. Computer processing converts these signals into images, which like CT images, can be viewed in any orientation.
- Intravenous contrast agents used in MRI are gadolinium based.
- Does not involve ionizing radiation.
- All parts of the body can be imaged. It is particularly helpful in evaluating soft tissues, brain, liver, and pelvic organs such as rectum, prostate, uterus, cervix, and ovary.
- Patient preparation is sometimes required (e.g., venous access may be required for intravenous contrast). Patients would routinely be dressed in a hospital gown. They should be warned they may hear a loud sound while the scanner is operating. Ear plugs/headphones may be provided for their comfort. Some examinations may require the patient to lie still for an extended period of time.
- Special considerations may be required for some patients, for example patients who experience claustrophobia, are pregnant or breast-feeding.
- MRI and patient safety are important considerations; see **Box 8.10**.

i Box 8.9 MRI contrast agent: gadolinium chelates

Gadolinium is a paramagnetic heavy metal and is toxic when a free ion; to ensure safety it is administered as a chelate, with the gadolinium safely locked away inside a larger molecule. However, post-contrast gadolinium retention in the brain as well as in extracranial tissues such as the liver and bone is documented and the long-term impact of this is unknown. Gadolinium-based contrast medium can rarely cause allergic reactions in patients. Care should be taken before administering gadolinium in patients with the following:

- Renal impairment or history of renal disease requiring dialysis
- Received a dose of gadolinium-based contrast medium in the last 24 h
- Patients who had a previous allergic or anaphylactic reaction to gadolinium-based contrast medium
- Patients who have risk factors for nephrogenic systemic fibrosis (NSF)
- Patients who are pregnant: The administration of contrast during pregnancy is predicated on a risk-benefit assessment.

Gadolinium-based contrast is excreted into breast milk in less than 0.04% of the dose injected to a mother. Also, of that tiny amount excreted into the milk, only 0.8% is absorbed by the baby. Thus, while there should be no concern about continuing breast-feeding after contrast administration, you may wish to keep the patient informed.

Fig. 8.51 MRI safety warning sign. *(Produced by the Institute of Physics and Engineering in Medicine [www.ipem.ac.uk].)*

Box 8.10 MRI safety considerations

- The magnetic field used for MRI is never switched off. The MRI room is under strong magnetic fields at all times.
- Free ferromagnetic (iron-based) material could get heated or become attracted by the strong magnetic field, becoming hazardous flying projectiles. These are not allowed into the scanner room.
- Examples of patient implants which are contraindicated:
  - Older implanted pacemakers (newer pacemakers may be conditionally safe)
  - Intracranial aneurysm clips
  - Cochlear implants
  - Implanted drug infusion pumps
  - Internal metallic objects such as bullets, shrapnel, and occupational iron embedded in the eyes
- Examples of relative contraindications:
  - Coronary stents
  - Recently placed surgical clips, pins, plates, screws, metal sutures, or wire mesh
  - Pregnancy

Box 8.11 Common MRI sequences

T1: Fluid appears dark. Useful when looking at anatomy and the soft tissues.
T2: Fluid appears bright. Useful when evaluating inflammation or oedema.
DWI (diffusion-weighted imaging): Areas where proton diffusion is restricted appear bright. It is useful for evaluating for ischaemic strokes, abscesses, and most tumours.
FLAIR (fluid-attenuated inversion recovery): Essentially T2, but fluid is suppressed, so it appears dark. It is useful for evaluating multiple sclerosis in the brain.
STIR (short tau inversion recovery): Essentially T2, but fat is suppressed so it appears dark. It is useful for evaluating oedema in soft tissues e.g., perianal abscesses.
MRA (magnetic resonance angiography): Intravenous contrast is administered, so vessels appear bright. It is useful for evaluating vascular anomalies such as arteriovenous malformations or aneurysms.

The magnetic field can be manipulated to produce differing radiofrequency pulses and gradients referred to as 'sequences', allowing the scan to be tailored to optimize assessment of different tissues and disorders (Box 8.11).

MRI, like CT, are complex examinations, for which accurate interpretation critically depends on salient clinical information. Providing relevant clinical background and potential diagnoses under consideration will greatly help to improve the quality of the radiology reports you receive.

The brain

MRI brain is usually performed when either CT has not provided the information required for patient management, or if CT is not as sensitive for the suspected clinical diagnosis. Indications include evaluation for ***demyelinating disorders*** (Fig. 8.52), diagnosis of ***stroke*** (Fig. 8.53), and characterization of ***intracranial mass lesions*** (Fig. 8.54) including pituitary tumours and lesions of the cerebellopontine angle.

The abdomen and pelvis

Magnetic resonance cholangiopancreatography

Magnetic resonance cholangiopancreatography (MRCP) utilizes heavily T2-weighted sequences to evaluate the gallbladder and biliary tree. Common indications include the work-up of obstructive (cholestatic) jaundice, or ***further investigation of dilated bile ducts*** identified on US (Fig. 8.55).

Liver MRI

MRI is very helpful in the evaluation of liver lesions. Besides standard T1, T2 weighted, and contrast-enhanced dynamic sequences, using hepatocyte-specific contrast agents and diffusion-weighted sequences can help ***distinguish benign from malignant liver lesions*** (Fig. 8.56).

Pelvic malignancies

MRI is a frontline modality for diagnosis and staging of pelvic malignancies including ***rectal cancer*** (Fig. 8.57) and ***cervical cancer*** (Fig. 8.58) and in characterization of ***ovarian masses*** (Fig. 8.59).

Bones, soft tissues, and joints

Common indications to request MRI include ***evaluation of the spinal cord and canal*** (Fig. 8.60) including in patients with possible spinal cord compression or cauda equina syndrome, ***characterization of bone lesions*** (Figs. 8.61 and 8.62) including tumour, osteomyelitis, and osteonecrosis and diagnosis of the painful or swollen joint.

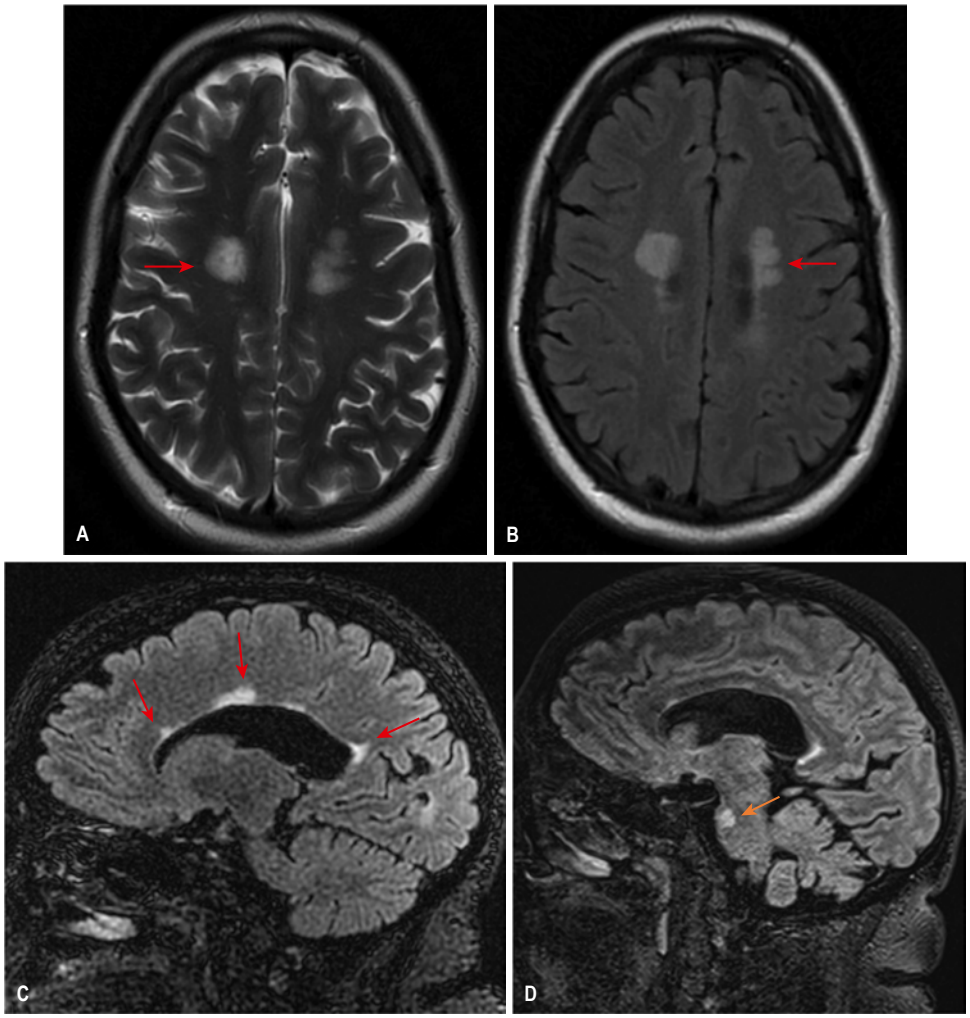


Fig. 8.52 MRI brain in a patient with multiple sclerosis. This patient has relapsing multiple sclerosis on disease-modifying medication. T2W **(A)** and FLAIR **(B–D)** MRI images demonstrate white matter changes *(red arrows)* in the periventricular white matter consistent with zones of demyelination, with a further lesion within the pons *(orange arrow)*. Occasionally, clinical decision-making may be helped by the use of contrast to determine if such plaques enhance.

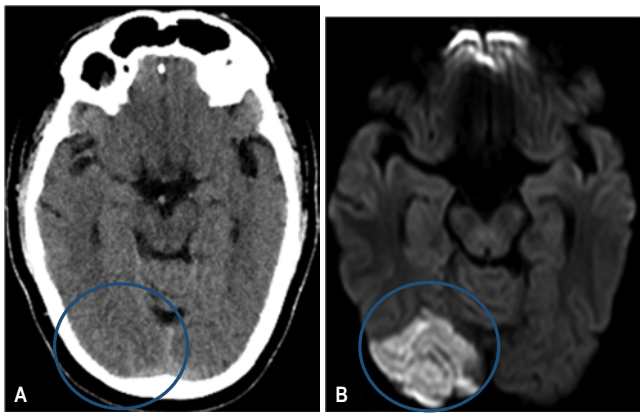


Fig. 8.53 MRI brain in a patient with ischaemic stroke. This 65-year-old male presented with sudden onset loss of coordination down his left side, unsteady on feet with left facial droop and dysaesthesia in left upper and lower limbs. CT brain **(A)** was normal. Subsequent MRI **(B)** demonstrated a large right occipital infarct *(blue circles)* visible on diffusion-weighted imaging.

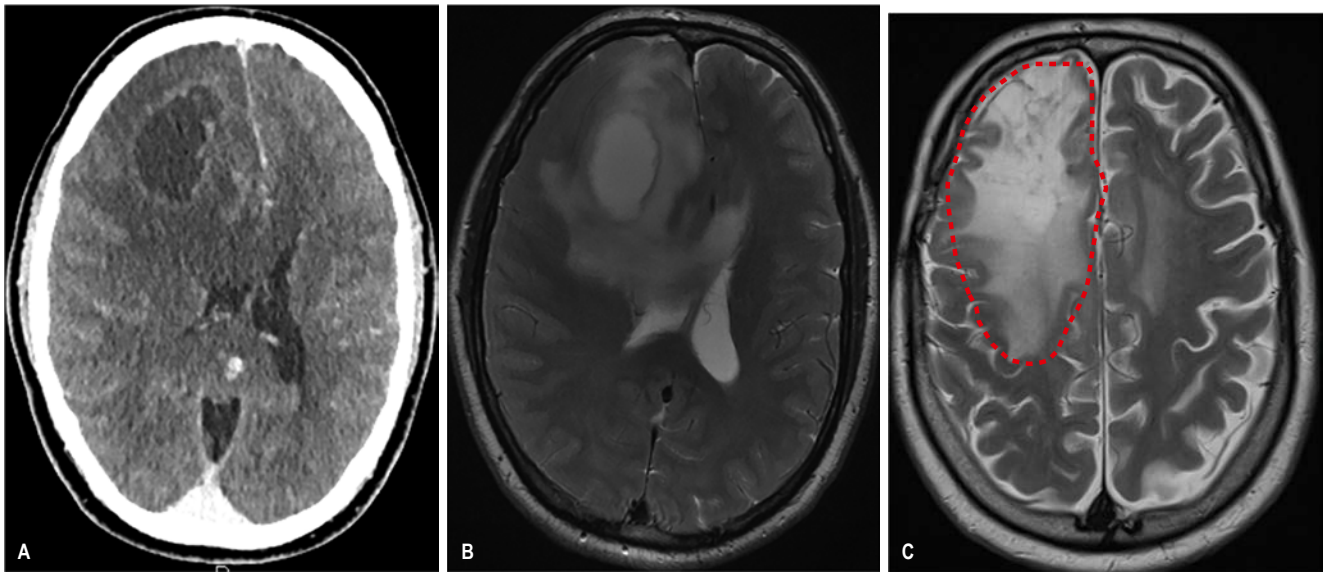


Fig. 8.54 MRI in a patient with a glioblastoma. Right frontal glioblastoma seen on CT with contrast **(A)** and MRI **(B)**. The patient underwent craniotomy with resection, followed by chemoradiotherapy. MRI 7 years later **(C)** demonstrates gliosis of the area post treatment *(red outline)*.

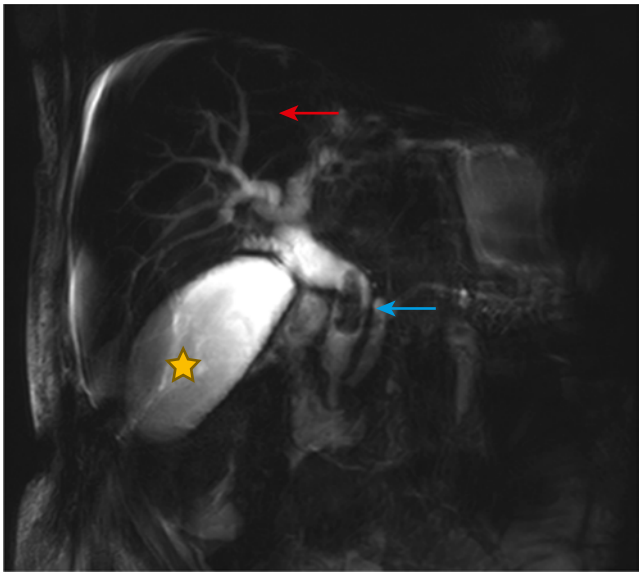


Fig. 8.55 MRCP. This MRCP performed in a patient with obstructive jaundice has identified gallstones within the dilated common bile duct *(blue arrow)*. Intrahepatic ducts are also dilated *(red arrow)*. *Star:* gallbladder.

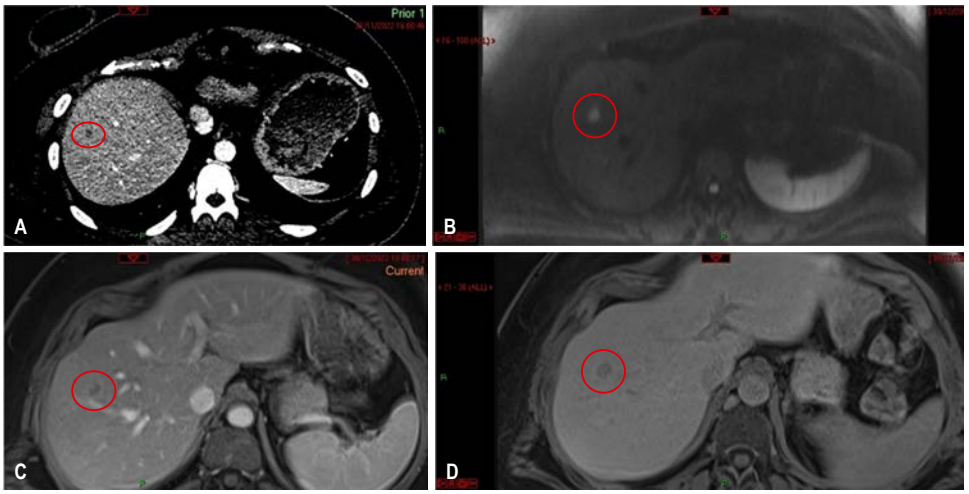


Fig. 8.56 Characterizing a liver lesion with MRI. This patient has a known gynaecologic tract malignancy with enlarged local and regional lymph nodes. Staging CT discovered an indeterminate liver lesion in segment 8 (**A**, *red circle*). MRI showed the lesion to be high signal in relation to background liver on T2W, with restricted diffusion **(B)**, rim enhancement **(C)**, and low signal to background liver on the hepatobiliary phase sequence **(D)**. Further, other lesions not seen on CT were identified on MRI. These MRI appearances are in keeping with multiple liver metastases.

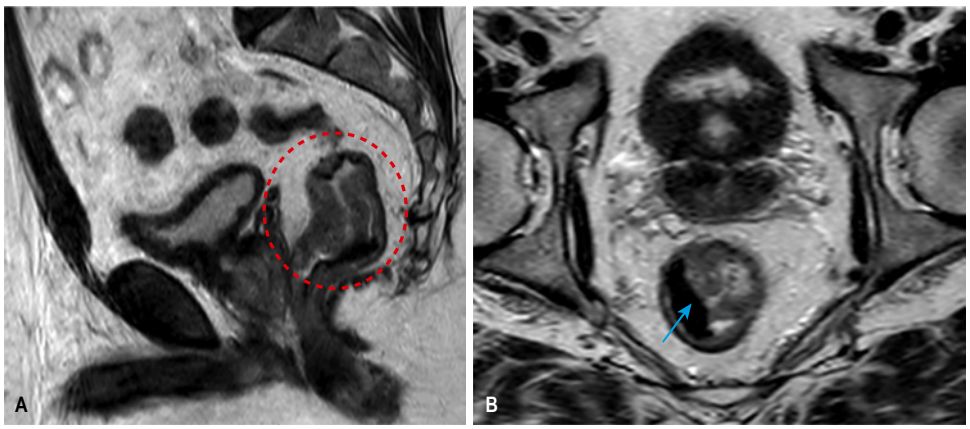


Fig. 8.57 Rectal cancer on MRI. Tumour *(circle and arrow)* seen within the lumen of rectum. The relationship between the tumour, adjacent organs, and surgically relevant operative planes is clearly visible on MRI. **(A)** Sagittal view; **(B)** axial view.

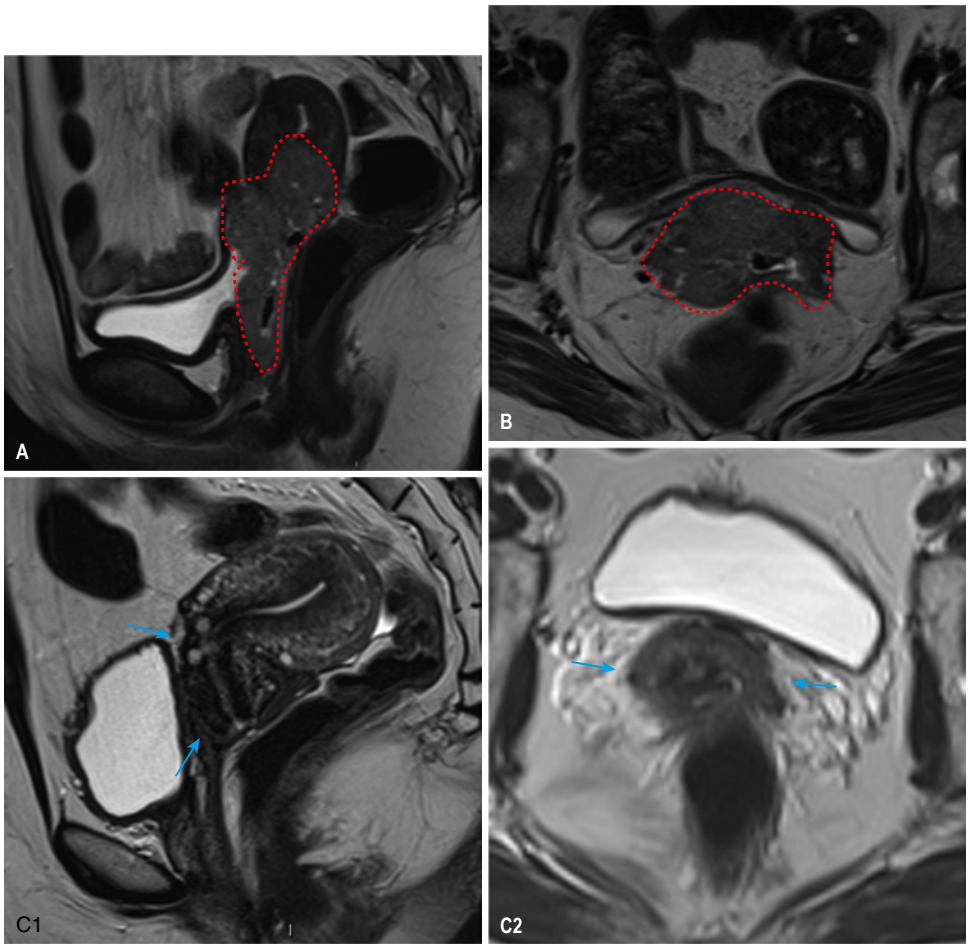


Fig. 8.58 Cervical cancer on MRI. These images (**A**, sagittal plane, **B**, axial plane) demonstrate a locally advanced cervical cancer *(red outline)* invading the bladder, in a patient with retroverted uterus. The second set of images **(C)** are of a normal cervix *(blue arrows)* in a retroverted uterus for comparison.

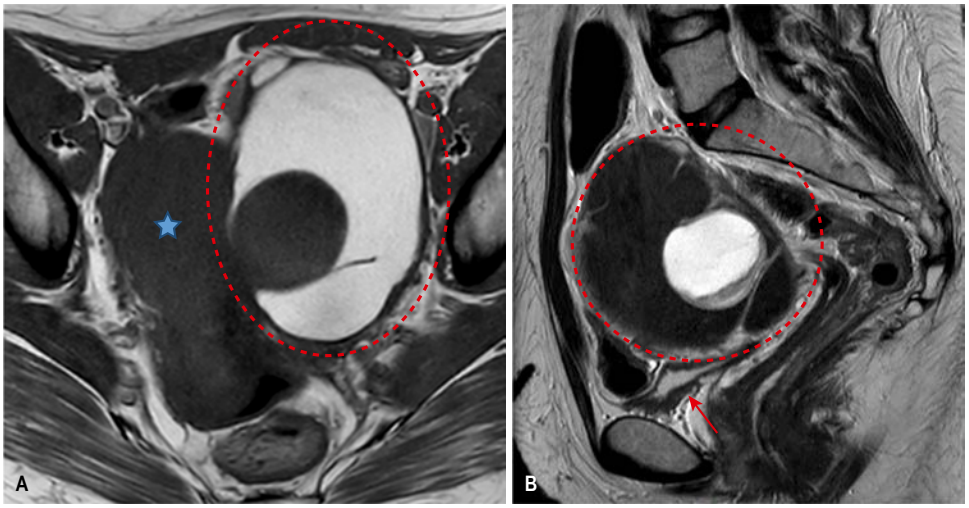


Fig. 8.59 Ovarian mass evaluated by MRI. MRI complements US in the evaluation of ovarian masses. Here is an example of a large septated left ovarian mass *(red outline)* which has high signal on T1W **(A)**, low signal on T2W **(B)** in keeping with an endometrioma. (Uterus: *blue star*; urinary bladder: *red arrow*.)

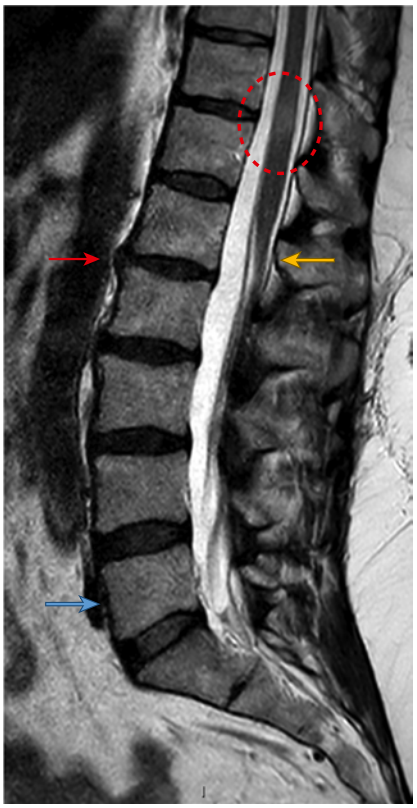


Fig. 8.60 MRI in a patient with spastic paraparesis. This patient with known multiple sclerosis presented with new bilateral leg weakness and difficulty lifting their right leg and walking, accompanied by bladder and bowel incontinence, as well as pain down the right leg. Deep tendon reflexes were brisk and plantar responses upgoing bilaterally. This image (T2W sequence) demonstrates minimal degenerative intervertebral disc changes with reasonably maintained vertebral body heights. It also demonstrates a demyelinating plaque (high-signal intensity area inside the *red dotted circle*) in the lower lumbar spinal cord which could explain the patient's symptoms. (*Amber arrow:* end of spinal cord, beginning of cauda equina; *blue arrow:* vertebral body, normal marrow signal; *red arrow:* intervertebral disc with some degenerative desiccation and loss of height.)

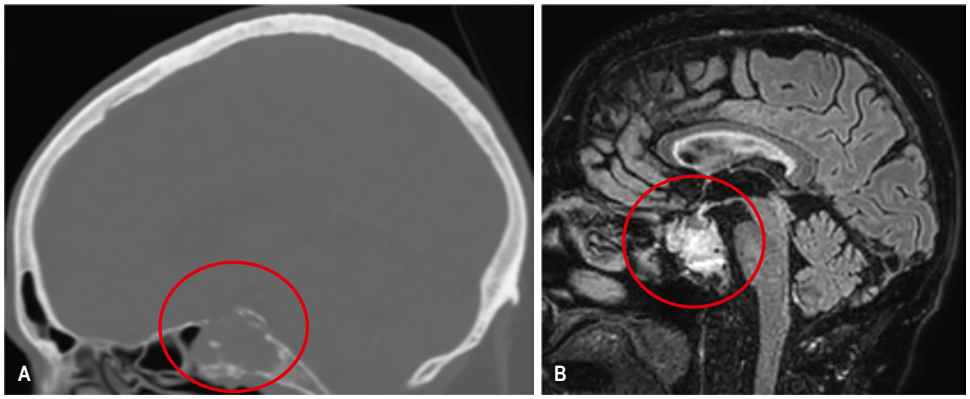


Fig. 8.61 Destructive lesion of the skull base on MRI. In this patient, bone destruction of the clivus was identified on CT **(A)**. A subsequent MRI **(B)** demonstrates the large mass causing this appearance.

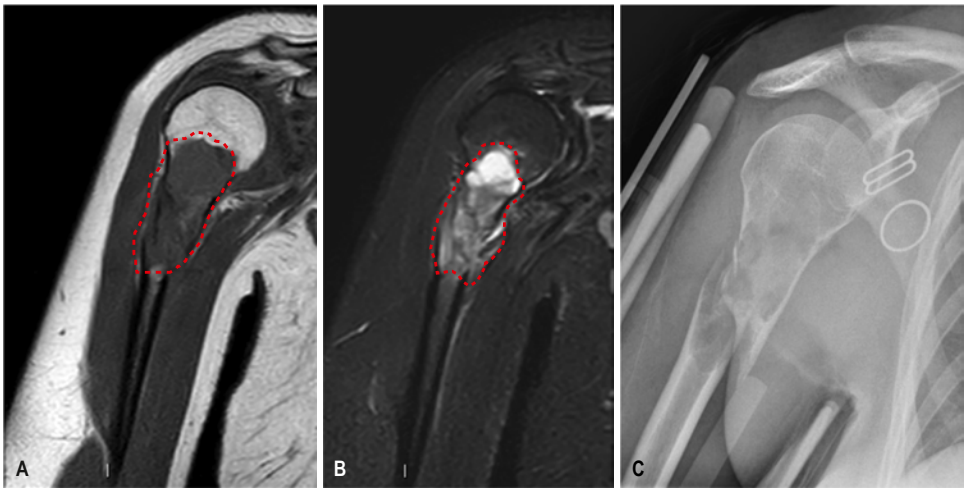


Fig. 8.62 Pathological fracture. This patient had a fracture of their right humerus through an area of abnormal bone, with extensive cystic change—a pathological fracture. MRI demonstrates the abnormal soft tissue in the marrow **(A)** T1W, **(B)** STIR sequence; **(C)** X-ray of humerus). This was a bone metastasis.

NUCLEAR MEDICINE AND POSITRON EMISSION TOMOGRAPHY

Nuclear medicine imaging uses radioactive tracers to examine organ function as well as anatomy (Box 8.12). A CT or MRI scanner may be co-located with a positron emission tomography (PET) or gamma scanner to improve anatomical correlation. This section illustrates commonly encountered system-based nuclear medicine examinations.

Brain—DaTScan (dopamine transporter scan)

DaTScan enables assessment of dopaminergic neuronal tissue in the brain and may be utilized in the evaluation of patients with tremor and dementia. Iodine-123 radioactive tracer is administered intravenously, attaches to dopamine transporters on nerve cells in the basal ganglia, and is visualized using single-photo-emission CT (SPECT) (Fig. 8.63). While DaTScan can assist in distinguishing Parkinson's disease (PD) from benign tremor, and dementia with Lewy bodies (DLB) from Alzheimer's dementia, it cannot distinguish between idiopathic PD, Parkinson-Plus syndromes (e.g., progressive supranuclear palsy or multiple system atrophy), and DLB. Patients are administered an oral dose of potassium iodide 1 hour before the tracer is injected to minimize uptake and radiation exposure to the thyroid.

Thorax—ventilation/perfusion (V/Q) scan

V/Q scanning is most commonly used in diagnosis of PE. The patient lies on a scanner bed while breathing room air mixed with minimally radioactive, tasteless, and odourless 82m krypton gas through a face mask for 10 minutes. This shows the aeration (ventilation) of the lungs. Patients are also injected with a small dose of radioactive tracer (Tc-99m-MAA) for the perfusion part of the scan. A 3D SPECT scan acquires the radiation emitted (up to 20 min) and images are processed and produced by the computer, allowing comparison and identification of ventilation/perfusion 'mismatch' (Fig. 8.64). V/Q scanning has been superseded by CTPA as a first-line diagnostic test for PE but is still employed in instances where CTPA cannot be performed (e.g., severe iodinated contrast allergy). In the absence of chronic lung

Box 8.12 Nuclear medicine imaging

- Nuclear imaging produces images by detecting radiation from different parts of the body after a radioactive tracer material is administered.
- The radioactive tracer is either injected into a vein, swallowed or inhaled as a gas and eventually accumulates in the organ or area of the body being examined. The unstable tracer degrades to a more stable isotope, releasing energy in the form of gamma rays.
- This energy is detected by a gamma camera placed close to the patient's body, then sent to a computer to generate the image. Many of these cameras are co-located with a CT scanner within the same unit to allow for better anatomical localization.
- Radiation dose is small.
- The radionuclide tracer used depends on the type of study and the body part being studied. Some examples include technetium, thallium, gallium, iodine, xenon, or fluorodeoxyglucose (FDG, for PET).
- Specific patient preparation depends on the test being performed.
- Pregnant, breast-feeding, and diabetic patients may require special consideration.

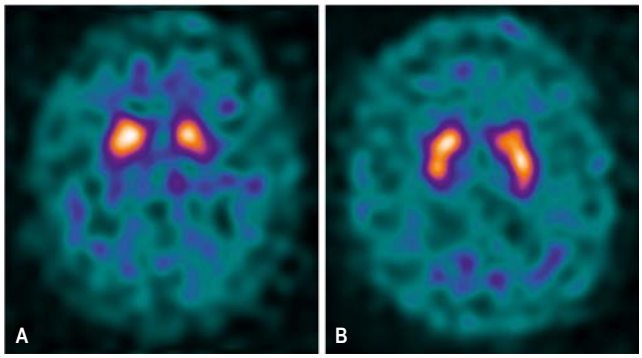


Fig. 8.63 DaTScan. This patient had a tremor in their right hand, with associated bradykinesia and rigidity (cogwheel) movement at their wrist. Scan **(A)** demonstrates markedly reduced tracer uptake in the basal ganglia, compared to a normal 'comma shaped' tracer uptake in the whole of basal ganglia in scan **(B)**, confirming a gross presynaptic dopaminergic deficit. This is consistent with a diagnosis of degenerative parkinsonism.

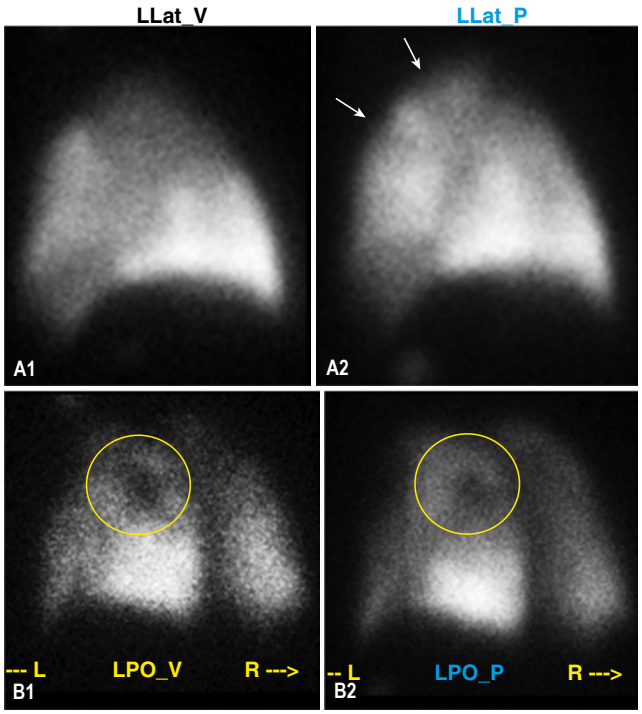


Fig. 8.64 V/Q scan. **(A)** Small wedge-shaped perfusion defects seen on left lateral lung views (*arrows*, LLat_P), mismatched to the corresponding ventilation scan which is normal (LLat_V). This is diagnostic for PE. **(B)** The perfusion defect seen on left posterior (LPO_P) view is matched to a defect on the corresponding ventilation scan (LPO_V, *circles*). This zone corresponded to severe emphysematous changes on a subsequent CT thorax.

disease and with a normal contemporaneous CXR, a perfusion-only scan may be performed, with the assumption that any perfusion defect identified would be 'unmatched', if a ventilation scan was performed.

Bone scan

Bone scan might help determine the cause of bone pain. The test is sensitive to differences in bone metabolism, which are highlighted in the body by the injected radioactive tracer (Fig. 8.65). Scanning the whole skeleton helps in diagnosing a wide range of bone disorders, including fractures, arthritis, Paget's disease of bone, cancer (primary or secondary), and infection of the bones, joints, or joint replacements.

Renal function

DMSA (dimercaptosuccinic acid, radioisotope) or **MAG-3** (mercapto acetyl triglycine) scans can both be used to estimate the individual kidneys' contribution to total renal function (so-called 'divided renal function'). DMSA images are usually acquired 2 to 3 hours after the injection (allowing sufficient time for tracer uptake into renal tubular cells; so-called 'static scanning') while MAG-3 images are obtained immediately after the injection ('dynamic scanning'). Both are bound to technetium 99m (Tc-99m) to form the radiopharmaceutical. Static DMSA scans are also used to identify zones of renal scarring (Fig. 8.66) while dynamic MAG-3 scans can be used to evaluate renal blood flow and perfusion, drainage of urine from the kidney to bladder, and assess for vesico-ureteric reflux.

Endocrine nuclear imaging

Tc-99m-Sestamibi scintigraphy is the main imaging technique for preoperative location of parathyroid adenomas (Fig. 8.67). Sestamibi is taken up by both the thyroid and parathyroid glands, but adenomatous and hyperplastic parathyroid tissue shows more avid uptake of the radiotracer and often retains the radiotracer longer than adjacent thyroid tissue. Asymmetric foci of increased radiotracer uptake on early images can be seen, representing abnormal parathyroid tissue superimposed on the normal thyroid. Delayed images, obtained approximately 2 hours after radiotracer administration, are acquired to look for foci of retained radiotracer characteristic of hyperfunctioning parathyroid gland.

Subtraction techniques have also been used to image the parathyroid glands by administering a second radiotracer taken up only by the thyroid gland. Either 123-I or Tc-99m pertechnetate can be used in conjunction with Tc-99m-Sestamibi to produce subtraction images of the parathyroid glands (although these techniques have not been clearly shown to be diagnostically superior to non-subtracted techniques).

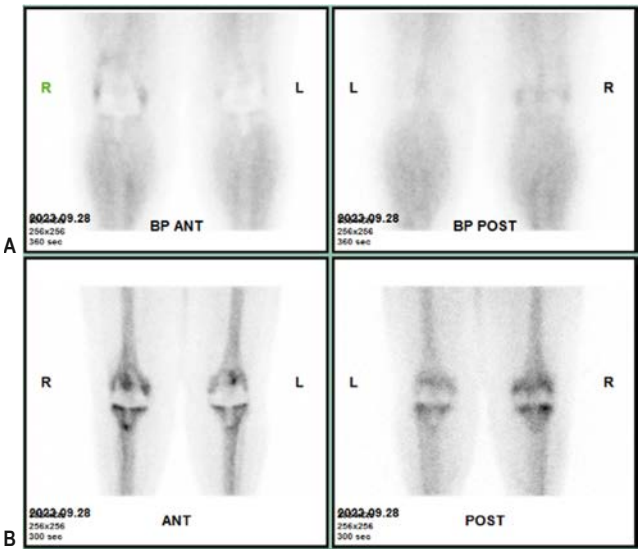


Fig. 8.65 Bone scan. This patient with bilateral knee prostheses has right knee pain. Bone scan image of the blood pool phase (**A**, anterior/posterior views) shows increased uptake in bone around the right knee prostheses. Increased uptake is also seen in the right knee more than the left during bone phase imaging **(B)**. Appearances suggest loosening and possible right knee prosthetic infection. White cell scan will distinguish between the two conditions.

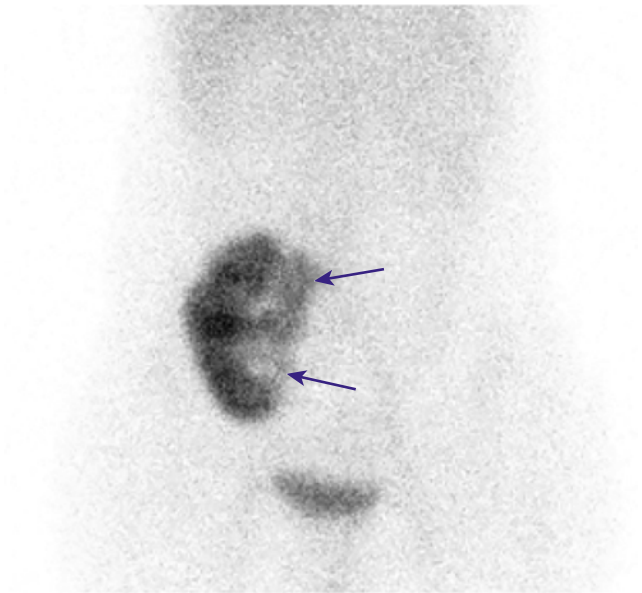


Fig. 8.66 DMSA renogram of a kidney transplant. This kidney transplant recipient had experienced repeated episodes of pyelonephritis resulting in a progressive decline in the transplant's function. In this anterior view of the transplant, defects in the usually homogeneous tracer uptake can be seen in the anteromedial portion of the upper pole and medial mid-lower pole of the transplant *(blue arrows)*.

Thyroid gland function (thyrotoxicosis, thyroid nodules, suspected ectopic thyroid tissues, and activity and/or location of the thyroid gland in neonatal hypothyroidism) can also be evaluated with nuclear medicine. Technetium 99m is taken up by the thyroid in the same way as iodine.

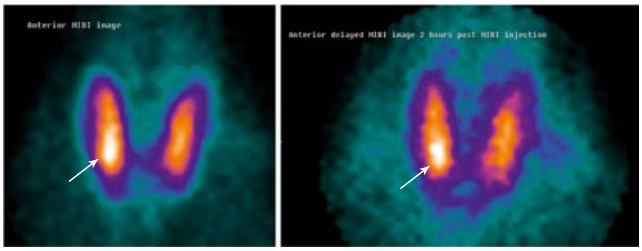


Fig. 8.67 Sestamibi scan of a parathyroid adenoma. This patient was hypercalcaemic. Increased sestamibi uptake in the parathyroid gland overlying the lower pole of right thyroid *(arrow)* is detected. Two hours after the injection, persistent uptake in the parathyroid adenoma remains visible.

Patients must refrain from ingesting iodine prior to the scan. These include certain medications or food substances such as kelp, seaweed, cough syrups, and multivitamins, as these can interfere with thyroid radiotracer uptake and the scan. Intravenous iodine-based contrast agents for CT should also be avoided.

MIBG scan involves an injection of a liquid radioactive material called iodine-123-meta-iodobenzylguanidine (MIBG), which binds to adrenergic tissue. It is typically used to identify neuroendocrine tumours such as pheochromocytomas and neuroblastomas (cancers that affect nerve tissue).

An MIBG scan takes two days—on the first day, the patient will have the isotope injection and on the second, they will have the scan. The scan is performed with a SPECT-CT scanner, which provides detailed anatomical localization of the tumour.

Octreotide scans are also used to localize neuroendocrine tumours. Octreotide binds to somatostatin receptors on cells and is therefore useful in identifying carcinoid tumours and pancreatic neuroendocrine tumours. This examination also takes 2 days.

Positron emission tomography and computed tomography

PET generates images to allow the evaluation of cellular metabolic function. Coupled to a CT scan, it allows for more precise anatomic location of the radioactivity detected.

Positron emission tomography and computed tomography (PET-CT) are routinely used to evaluate cancer, heart disease, gastrointestinal, endocrine, and neurological disorders—both to aid diagnosis as well as to evaluate treatment response.

Specific radiotracers are injected which accumulate in tumours or regions of inflammation or bind to specific proteins in the body (Box 8.13). The most common radiotracer is F-18 fluorodeoxyglucose (FDG), a molecule similar to glucose. Cancer cells are more metabolically active and may absorb glucose at a higher rate. This higher rate can be seen on PET scans (Fig. 8.68). FDG is just one of many radiotracers in use or in development.

Red and white cell scan

Occasionally the source from which a patient is bleeding is not identified by either CT or percutaneous angiography. In such cases, **red cell scans** may be helpful. A small amount of the patient's blood is extracted and mixed with radioactive technetium. The blood with

Box 8.13 PET-CT patient preparation

- No strenuous physical activity in the days (2–3) leading up to the scan.
- Limited carbohydrate diet for the previous 24 h before the examination.
- Fast (except water) for 6 h before the exam. Patients may drink as much water as they are able before and after the scan.
- Routine medications may be taken.
- Patients with diabetes are given specific instructions designed to avoid hyperglycaemia at the time of the examination (which can interfere with radiotracer uptake).
- After the radiotracer is injected, the patient will be asked to rest quietly on a reclining chair for approximately 30–60 min before being escorted into the scanner room.
- The scan will take another 30 min.
- Nursing mothers should express and delay breast-feeding for 24 h.

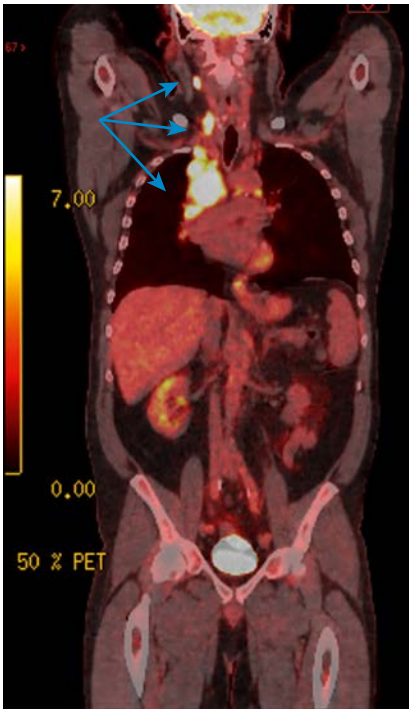


Fig. 8.68 This PET/CT image demonstrates increased FDG uptake in the cervical and upper thoracic lymph nodes *(blue arrows)*. No abnormal uptake is demonstrated below the diaphragm or spleen in this patient with lymphoma.

the technetium is then injected back into the patient's bloodstream and sequential imaging is performed to determine the site of blood loss. Red cell scan can detect bleeding at rates of 0.1 to 5 mL/minute.

White cell scan can be helpful to detect occult infection/inflammation not identified by other modalities. The process is very similar to red cell scan above, but the radioisotope used to tag the isolated white blood cells is Indium-111. Common indications include investigating prosthetic joint infections, osteomyelitis, vascular graft infections, and fever of unknown origin. The radioactive white cells are injected 1 to 2 hours after Indium-111 application. Images are taken sequentially. The whole procedure can take 1 to 4 hours or up to 30 hours depending on the clinical question to be answered.

Gastric emptying

A gastric emptying scan is performed to help diagnose motility disorders. The patient ingests a standardized meal mixed with radioisotopes (Tc-99m sulphur colloid). Sequential imaging allows accurate comparison to validated normal values for gastric emptying. This examination can take up to 4 hours.

The patient fasts for 6 hours before the scan. They should stop medicines that affect motility (e.g., metoclopramide, tegaserod, erythromycin, domperidone, and antispasmodics) for 3 days prior to the test. Patients should refrain from taking laxatives the day before or any time during the study and should not smoke on the morning of the test and during the hours of the scan.

INTERVENTIONAL RADIOLOGY

Interventional radiology (IR) is a growing field within radiology that is involved in the diagnosis and image-guided treatment of disease. With modern advancement in technology and equipment, image-guided minimally invasive procedures can be performed instead of surgery for many conditions ranging from biopsy and abscess drainage to cancer treatment (**Box 8.14**). These procedures are performed with expert technical skill and precision after years of training, and require careful patient selection, preparation, and consent in the same way as a patient undergoing surgery (**Box 8.15**). This section will consider some of the common procedures performed by the IR.

Box 8.14 Glossary of some procedures performed by interventional radiologists

- **Angiography:** Imaging of the arteries or veins to identify stenosis, occlusion, or bleeding point (fluoroscopy, MRI, CT).
- **Angioplasty:** Dilation of a (stenosed) vessel by inflating a balloon attached to a fine-bore catheter (tube) inserted percutaneously and guided into place using imaging (fluoroscopy).
- **Embolization:** Stopping or reducing the blood flow through a vessel (artery or vein) by injecting an embolic agent (coils, gelfoam, liquid or solid agents [e.g., glue, ethylene-vinyl alcohol co-polymer or cyanoacrylates] or microvascular plugs) guided precisely into place by imaging (fluoroscopy). This can be life-saving in major haemorrhage.
- **Thrombolysis:** Injecting clot-dissolving medicine (e.g., tissue plasminogen activator) directly onto the thrombus—through a percutaneously inserted fine catheter guided precisely into place through imaging.
- **Thrombectomy:** Blood clots are mechanically removed from within the vessel via a catheter inserted percutaneously and guided into place with imaging.
- **Gastrostomy tube placement:** A feeding tube is inserted (under imaging—fluoroscopy, ultrasound) from the abdominal wall directly into a stomach (percutaneously).
- **Stent placement:** A tube is precisely placed (image guided) within (i) a narrowed segment of blood vessel or bowel to 'open' up a blockage or (ii) an aneurysm to redirect flow to prevent rupture of the vessel.
- **Needle biopsy:** A sample of tissue is obtained by guiding a needle from the skin precisely into the area/organ of interest under image guidance (ultrasound, fluoroscopy, CT, or MRI).
- **Interventional oncology:** Variety of techniques (ablation, chemo-/radioembolization) are available using image guidance to treat cancer minimally invasively.
- **IVC filter insertion:** A small umbrella-shaped device placed in the inferior vena cava under image guidance to trap small thrombi, preventing embolization into the lungs.
- **Intravenous catheter insertion:** Long-term catheters are placed under image guidance to allow doctors to deliver intravenous chemotherapy, nutrition, or dialyze patients with renal failure.

Box 8.15 Patient selection, preparation, and after-care: general considerations for IR

- Patients undergoing IR procedures should be given the same considerations as for surgery.
- Informed consent should be obtained by the responsible radiologist after providing the patient with written and verbal information regarding the procedure, risks, benefits, and alternative options. The patient should be given the opportunity to ask questions and adequate time to consider their decision.
- Most procedures are performed under local anaesthesia (LA), occasionally with sedation. This is the major advantage of IR compared to surgery which usually requires general anaesthesia. Occasionally antibiotic cover may be given.
- The patient must be able to cooperate with the radiologist and tolerate the procedure by lying still on a bed. Occasionally they may be required to fast before the procedure.
- Patients should generally suspend anticoagulant therapy. Guidance from haematology services may be helpful if the patient is on long term anticoagulation for metallic heart valves or similar conditions and the radiologist should be informed if the patient is on such medications. Occasionally patients may be prescribed anticoagulation postprocedure.
- Depending on the type of procedure, patients are usually discharged home after minimal departmental observation (e.g., needle biopsy).
- Complex procedure (e.g., endovascular abdominal aneurysm repair [EVAR] or intracranial aneurysm coiling) however may require general anaesthetic and postprocedure ITU support.
- General complications for any procedure include infection, haemorrhage, unintended organ perforation, and even potentially death. However, IR procedures are less invasive than open surgery.

Surgical drainage and needle biopsies

These are commonly requested procedures for the septic patient with for example, abdominal abscess requiring drainage (Fig. 8.69) or to help facilitate diagnosis by sampling abnormal tissues in the body.

Procedure: Under image guidance (usually US or CT), a needle is guided precisely to the area inside the patient to be drained/sampled. These are minimally invasive performed under aseptic conditions with local anaesthetic (LA).

Practicalities: The area of interest should be visible on imaging, with a direct path to safely access from the skin and within the range of needles/catheters (length).

Contraindications: Area not visible on imaging for the procedure to be safely performed. Bleeding dyscrasia. Patient unable to cooperate or lie still for the duration of procedure.

Complications: These are generally safe procedures with minimal complications. (See General Considerations, Box 8.15.)

Vascular procedures

Vascular IR forms a big part of the work of the interventional radiologist. This can be life-saving in a patient with aneurysm or postpartum haemorrhage, where radiologically guided

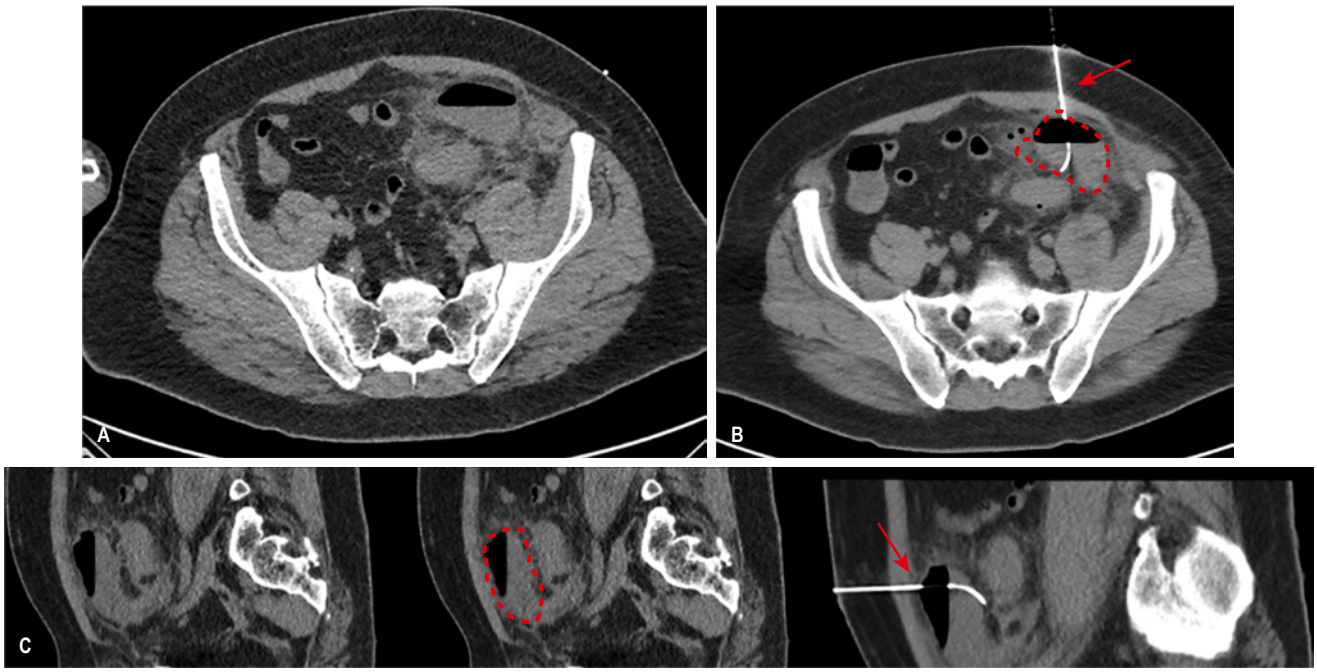


Fig. 8.69 IR drainage of an intra-abdominal collection. Drainage of the intra-abdominal collection *(red outline)* in this patient is being performed under CT guidance. A guidewire can be seen passing through the introducer needle *(red arrows)* into the collection, securing a path along which a drain can be passed. **(A), (B)** Axial views; **(C), (D)** sagittal views. *(Images courtesy of Dr Navin Mathias.)*

Box 8.16 Tumour ablation

- Common sites treated by ablation include the lung, liver, kidneys, and prostate.
- **Radiofrequency ablation (RFA)** uses a high-frequency electrical wave delivered through a probe the size of a long pencil to heat and 'burn' the tumour. A closed loop circuit is created using an RF generator, an electrode, grounding pads, and the patient. This results in ionic agitation, creating heat, coagulative necrosis, and cell death. Temperatures reach 50–100°C for approximately 4–5 min. The procedure takes around 30 min.
- **Microwave ablation (MWA)** uses microwaves, delivered via a small antenna inserted directly into the tumour. MWA is much faster than RFA, although both have similar results. It may be used with larger tumours.
- **Cryoablation** uses liquid nitrogen or argon gas to freeze tumour cells, which then rupture from the cold temperature and die. Cryoablation is often used to treat kidney tumours without damaging or affecting kidney function—useful in patients with a single kidney or in those with impaired renal function who might need dialysis if surgery is performed.
- **Percutaneous ethanol ablation (PEI)** or alcohol ablation uses absolute ethanol injected directly into the tumour. Alcohol is a potent toxin to cells and the blood vessels, making it effective in treating small tumours. PEI is used less as it often causes more discomfort than the other tumour ablation options.
- **Vertebroplasty.** Bone metastasis can also be treated by ablation, often in conjunction with cement injection to stabilize the vertebra and provide pain relief. It has been shown to reduce opioid use, improving patient function and quality of life more rapidly than radiotherapy.

interventions (e.g., stenting or embolizing blood vessels) can prevent or stop major haemorrhage (Fig. 8.70). Vascular IR can also improve blood flow to organs (e.g., limbs or the brain) through angioplasty or stent placement in stenosed or occluded vessels. When performed for the lower limbs (Fig. 8.71), this could improve the patient's function, quality of life, improve healing, and be limb saving.

Procedure: Under image guidance (usually fluoroscopy or CT), a needle then a guidewire, followed by a catheter are sequentially and expertly guided precisely to the stenosed, occluded, or bleeding vessel. Vascular access is often achieved via the superficially located femoral artery or vein at the groin. The index vessel of interest can be dilated, stented, or embolized endovascularly. These are minimally invasive performed under aseptic conditions with LA.

Practicalities: The area of interest should be visible on imaging.

Contraindications: Uncontrolled coagulopathy. No safe percutaneous access. Patient unable to cooperate or lie still for the duration of procedure.

Complications: These are generally safe procedures with minimal complications. However, vessel occlusion, haemorrhage, organ infarction, or death could ensue (particularly for more complex procedures such as intracranial or abdominal arterial coiling and aneurysm stenting). (See General Considerations, Box 8.15.)

Fig. 8.70 Coil embolization of a renal artery. This patient was bleeding from his right kidney **(A)** CT showing haematoma, *red outline*; contrast extravasation/bleeding point, *red arrow*). Angiogram showed the contrast extravasation **(B)**. A coil *(blue arrow)* was deployed and successfully stopped the bleeding **(C)**. *(Images courtesy of Dr Navin Mathias.)*

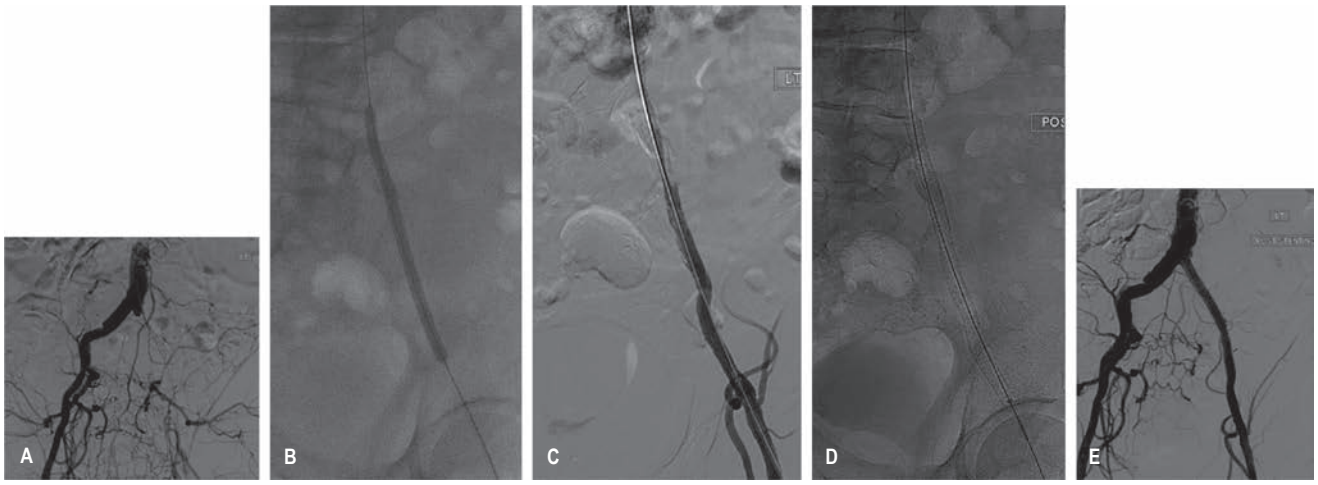


Fig. 8.71 Angioplasty of the left common iliac artery. This patient had an occluded left common iliac artery **(A)**. This was crossed with a guidewire and angioplasty performed **(B)**. Post-angioplasty angiogram showed persistent significant stenosis **(C)**. A stent was placed **(D)**. Good vascular perfusion seen with the stent in place **(E)**. *(Images courtesy of Dr Navin Mathias.)*

Abdominal IR procedures

These procedures alleviate obstructed or stenosed lumina within the abdomen. Under careful image guidance, a tube or stent can be placed to relieve an obstructed and infected urinary system (nephrostomy), palliate the cancer patient to allow swallowing/bowel movement (oesophageal or colonic stent), facilitate enteral feeding (radiologically inserted gastrostomy [RIG]) or improve transplant-free survival in patients with cirrhosis related portal hypertension and refractory ascites (transjugular intrahepatic portosystemic shunt—TIPS).

Oesophageal stenting

Indication: Palliate obstruction and dysphagia from oesophageal tumour. Treat tracheo-oesophageal fistula, post-oesophagectomy leak, or oesophageal stricture.

Procedure: A slender catheter supported by a guidewire is fed down the oesophagus past the tumour or fistula, under image guidance (fluoroscopy). The length of obstruction is measured and an appropriately sized stent is placed over the stiff guidewire at the site of luminal narrowing, creating space for food and liquid to pass. Occasionally balloon expansion of the stent may be required to create a large enough neo-lumen.

Practicalities: The patient must be able to cooperate with the radiologist for the duration of the procedure. Occasionally, an endoscopist may be required to guide the wire past a large obstructing tumour. The procedure may be performed under conscious sedation.

Contraindications: Tumour located too close to the upper oesophageal sphincter, presence of tracheal stent, and oesophageal perforation measurement greater than 50% circumference of the oesophagus or greater than 5 cm in length (absolute contraindications). Hemodynamically unstable patients, unresolved coagulopathy, tumour located too close to the gastroesophageal junction (GOJ) (relative contraindications).

Complications: Chest pain, bleeding, perforation, aspiration, stent migration, stent blockage, intractable reflux, tracheoesophageal fistula. Late complication rate ~53% to 65%, re-intervention rate up to 50%. Death 0.5% to 2%.

Colonic stenting (Fig. 8.72)

Indication: Alleviate bowel obstruction (tumour)—palliation or as a bridge to surgery.

Procedure: Guidewire is guided past the tumour per rectum (fluoroscopy). Remainder of procedure as for an oesophageal stent. Occasionally an endoscopist is required to help direct the wire through a very large tumour obstructing the lumen.

Practicalities: As for oesophageal stent above.

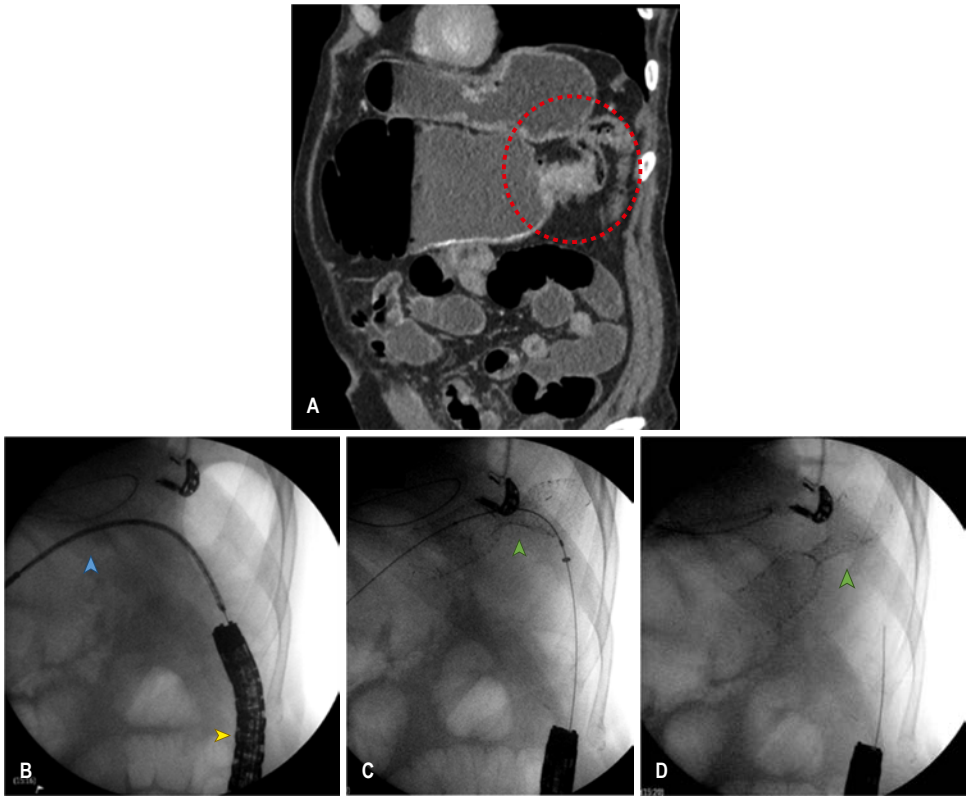


Fig. 8.72 IR colonic stent deployment for an obstructing colonic cancer. This patient has an obstructing left colonic tumour (*red circle* on CT, **[A]**). At fluoroscopy, colonoscope *(yellow arrowhead)* can be seen assisting the placement of a guidewire *(blue arrowhead)* **(B)**. The stent *(green arrowhead)* is delivered and deployed over the guidewire **(C)**. Self-expanding colonic stent can be seen in place **(D)**. *(Images courtesy of Dr Navin Mathias.)*

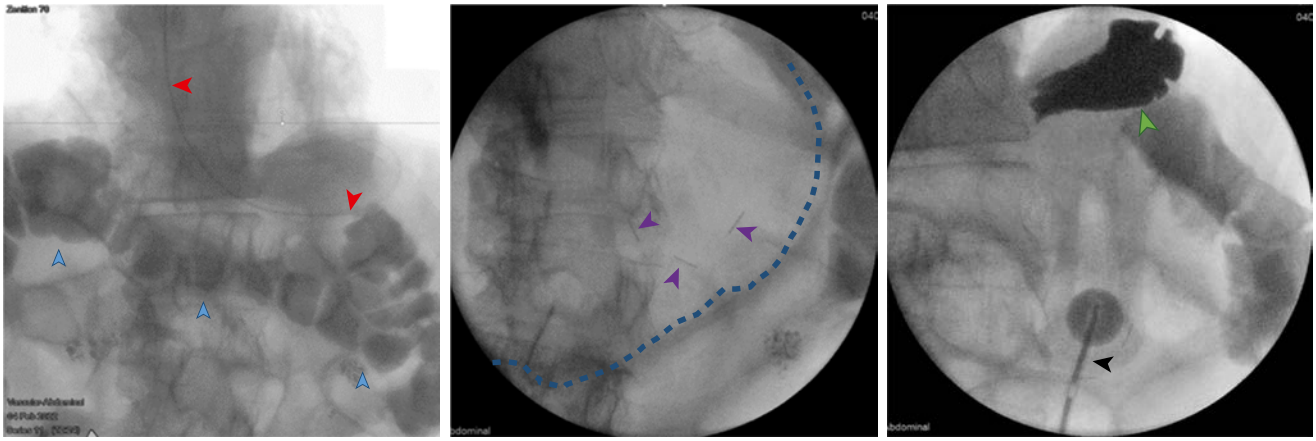


Fig. 8.73 Radiologically inserted gastrostomy (RIG) insertion. There is barium outlining the patient's colon *(blue arrowheads)* and nasogastric tube is in place *(red arrowheads)*. The stomach *(blue dotted line)* is distended with air via the nasogastric tube. Stomach is secured to the skin with clips *(purple arrowheads)*. Gastrostomy tube *(black arrowhead)* inserted percutaneously—avoiding colon—into the stomach and position checked with barium *(green arrowhead)*. *(Images courtesy of Dr Navin Mathias.)*

Contraindications: Perforated tumour. Tumour located <5 cm from anal verge (due to the likelihood of tenesmus and a higher risk of stent migration).

Complications: Perforation (associated with mortality of 16%), stent migration, tumour regrowth and re-obstruction, pain, and bleeding.

Radiologically inserted gastrostomy (Fig. 8.73)

Indication: To facilitate enteral feeding in patients with oesophageal obstruction or neurological impairment preventing safe swallow.

Procedure: A nasogastric tube (NGT) is inserted prior to the gastrostomy. The stomach is inflated with air via the NGT and visualized on fluoroscopy. Under aseptic conditions, with conscious sedation, a needle is inserted through the skin directly into the patient's stomach, avoiding the colon. A guidewire followed by the gastrostomy (feeding) tube is placed in the needle tract. The tube is stitched into place on the abdominal wall.

Practicalities: Patients may be given oral barium 1 to 2 days before the procedure to outline the colon. Oxygen is given via face mask. Procedure could take between 30 and 90 minutes. Bed rest for 4 hours. Nil by mouth for 12 hours, flushing the tube every 2 hours. Clear water per gastrostomy for 12 hours before commencing feeding per gastrostomy.

Contraindications: Severe coagulopathy. Large hiatus hernia. Interposed colon.

Complications: Leak around the gastrostomy. Infection. Bleeding.

Transjugular intrahepatic portosystemic shunt

Indication: Symptomatic portal hypertension, for example refractory ascites or variceal bleeding. Creation of a shunt between the portal and systemic circulations reduces portal pressure, thereby reducing the risk of variceal bleeding.

Procedure: Under aseptic conditions and using LA, access to the liver is achieved by puncturing the jugular vein and feeding a guidewire and catheter into the hepatic vein. Using image guidance (fluoroscopy), a needle and catheter are then passed from a hepatic vein into a branch of the portal vein. The communication is secured with a vascular stent.

Practicalities: Patient must be able to cooperate, lie still on the fluoroscopy table for the procedure (up to an hour). Patient should be fasted. Model for End-Stage Liver Disease (MELD) score predicts 90-day mortality after a TIPS procedure.

Contraindications: A high bilirubin level; severe intrinsic kidney disease (stage 4 chronic kidney disease or above), and a MELD score greater than 20 (relative contraindications). Absolute contraindications include severe refractory HE, severe pulmonary hypertension, heart failure, and significant right ventricular dysfunction (could be exacerbated by the procedure).

Complications: Hepatic encephalopathy (~24% at one year). Bleeding, infection, abscess, the aforementioned cardiac complications, bile leak or injury, acute liver failure, or stent thrombosis or migration (~1%–2%).

Nephrostomy (Fig. 8.74)

Indications: Drain an infected (pyonephrosis) or obstructed (hydronephrosis) urinary system; to divert urine from a downstream leak (e.g., ureteric injury).

Procedure: Under LA, and using US and fluoroscopic guidance, a tube is inserted from the patient's back into the renal pelvis of the kidney.

Practicalities: Patient must be able to lie still on a narrow bed for the duration of the procedure (up to an hour).

Contraindication: Uncontrolled coagulopathy. Patient unable to cooperate or tolerate the procedure.

Complications: Bleeding. Infection.

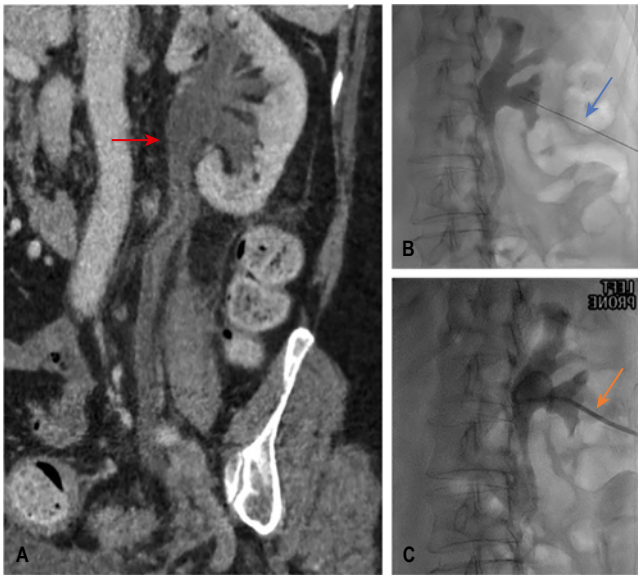


Fig. 8.74 IR nephrostomy insertion. This patient had left hydronephrosis (*red arrow*, panel **A**) secondary to a pelvic tumour. Panel **B** demonstrates guidewire *(blue arrow)* being inserted percutaneously through the introducer needle securing the tract. Contrast delineates the left renal pelvis and ureter (fluoroscopy). A nephrostomy tube *(amber arrow)* was inserted over the guidewire **(C)**. *(Images courtesy of Dr Navin Mathias.)*

Biliary drainage

Indication: Infected or obstructed biliary system, for example from bile duct or pancreatic tumour. Palliation or to allow treatment by relieving jaundice; to divert bile from a downstream leak, for example bile duct injury after cholecystectomy.

Procedure: Under LA, and using US and fluoroscopic guidance, a needle is inserted from the skin into a dilated bile duct. A guidewire is fed down the needle, the needle is removed, and a catheter is passed into the bile duct over the guidewire. If necessary, a guidewire can be navigated into the duodenum and a stent placed into the bile duct to keep it open.

Practicalities: Patient must be able to lie still on a narrow bed for the duration of the procedure (up to an hour).

Contraindication: Ascites. Uncontrolled coagulopathy. Patient unable to cooperate or tolerate the procedure.

Complications: Biliary sepsis. Bleeding. Bile leakage. (~1%–5%).

Interventional oncology

Modern multidisciplinary oncologic care is delivered by specialists in Medical, Surgical, Radiation, and Interventional Oncology (IO). IO is the branch of interventional radiology, dedicated to the diagnosis, treatment, and palliation of cancer and cancer-related problems using minimally invasive procedures performed under image guidance.

While surgical excision of tumours is generally accepted to offer the best long-term solution, occasionally this may not be possible due to the size, number, or location of the tumour. IO therapies may be applied to shrink the tumour, making surgery possible or help increase the efficacy of other treatments.

Patients can greatly benefit from IO treatments. The minimally invasive nature of the treatments means they cause less pain, fewer side-effects, and shorter recovery times. Many IO procedures can be performed on an outpatient basis, freeing up hospital beds and reducing costs. Serious complications occur in approximately 1% of patients.

Examples of IO techniques include administering chemotherapeutic drugs intra-arterially (increasing their local tumour-related potency and reducing the side-effects from system wide delivery e.g., TACE for hepatocellular carcinoma) or ablation of tumour by direct application of heat or cold energy (**Box 8.16**).

Radiofrequency ablation: liver metastasis

Indication: As a standalone therapy, in combination with surgery (curative) or for irresectable disease secondary to patient or tumour factors (see above).

Procedure: Under ultrasound and/or CT guidance, a needle probe is inserted from the skin into the liver and tumour. Heat (RFA/MWA) is transmitted from a special machine through the needle into the tumour (see **Box 8.16**).

Practicalities: Procedure is usually performed under general anaesthesia for patient comfort and control of diaphragmatic and patient movement. Recovery time is between a few hours and an overnight stay in the hospital.

Contraindications: Bile duct or major vessel invasion. Significant extrahepatic disease. Child-Pugh class C cirrhosis or active infection.

Complications: Serious complication (e.g., infection, liver failure, death) is low. Overall reported complication rate is 2.2% to 9.5%.

CUTTING EDGE AND FUTURE TECHNOLOGY

Dual/Multienergy CT is a promising new technology whereby images are acquired using 2 (or more) different spectra of X-ray energy. Harnessing variation in absorption of different X-ray energy at the molecular level, dual/multienergy CT can help identify the composition of internal body constituents, with quantification also possible. It has been shown to reduce artefacts generated from metallic prostheses and render free, active contrast extravasation/endo-leak more conspicuous. The computer processing power that is required for this technology means it is not yet widely adopted currently.

Optical imaging. Using light and special properties of photons (e.g., visible, ultraviolet, and infrared light), detailed images of organs, tissues, and cells at the cellular to subcellular level can be obtained, offering a minimally/non-invasive method to evaluate the body—without ionizing radiation. While this technology is already in use in some specialties (e.g., endoscopy and retinal imaging), clinical use in radiology is some time away.

High-intensity focused ultrasound (HIFU). Guided by MRI, radiologists are able to accurately target and use HIFU energy to ablate the area of the brain that helps regulate movement in patients with essential tremor. The procedure is done with the patient fully awake and effect is immediate. While NICE and FDA approval have been granted for essential tremor, their role in the treatment of brain tumours, epilepsy, and dementia is being investigated.

Questions for this chapter can be found on Elsevier eBooks+.

9 Public health

Sarah Elizabeth Winchester, Simona Baracaia, and Ian Basnett

Background 153
Public health and global health 153
Key concepts in public health 155
Epidemiology and causality 155
Health needs assessments 155
Life expectancy and quality of life 155
Health inequalities 155
Politics and advocacy 157
Pillars of public health 157
Health improvement 158
Health protection 159
Healthcare public health 163

CORE SKILLS AND KNOWLEDGE

Factors that influence health stretch from climate change and poverty to health behaviours and health care. A multiprofessional approach, incorporating the epidemiological study of disease alongside practical health interventions, is required to address these complex and high-level challenges. When John Snow analysed cases of cholera in London in 1854, he not only mapped the distribution of cases and the possible link to a water pump but also dealt with the source of the outbreak by removing the pump.

The core workforce includes doctors within public health medicine, working alongside epidemiologists, analysts, commissioners, nurses, health visitors, and environmental health officers. In many parts of the world public health is led by community workers. Professionals across all medical specialties often take a population approach too, especially in primary care.

Key to understanding public health is recognizing that many patients have had their health adversely affected by social, environmental, or behavioural factors (called the 'wider determinants of health') which may contribute to health inequalities. The discipline of public health strives to improve outcomes while reducing these inequalities through three core pillars:

- ***Health improvement:*** implementing interventions to prevent ill-health and improve population health
- ***Health protection:*** preventing, controlling, and monitoring infectious, environmental, chemical, and radiological hazards to population health
- ***Healthcare public health:*** maximizing the benefits of healthcare interventions and reducing inequalities using tools such as screening and health economics.

Skills required in public health can be developed by critically appraising relevant articles in journals, by undertaking local quality improvement projects, and by seeking out opportunities to work with national or local groups focusing on health or sustainability.

BACKGROUND

When viewed on a population level, health is a broad concept. Social determinants of health are numerous and varied, stretching from climate change through to health behaviours (**Fig. 9.1**). The Alma Ata (World Health Organization [WHO] conference) declaration of 1978 stated that 'health, which is a state of complete physical, mental and social wellbeing, and not merely the absence of disease or infirmity, is a fundamental human right and that the attainment of the highest possible level of health is a most important worldwide social goal whose realisation requires the action of many other social and economic sectors in addition to the health sector'. It goes on to assert the key importance of addressing inequalities. Those working in public health strive to bring about change at a scale that will maximize benefit, while reducing inequalities in health.

For much of human history communities have been concerned about communicable disease, the availability of food and clean water, and care for the sick. The origins of modern public health in the UK lie in the big cities of industrialized nations in the mid-19th century, where pioneers such as John Snow (**Box 9.1**) began studying patterns of transmissible disease and intervening to prevent the spread of infection. The development of germ theory, the discovery of means to prevent communicable diseases (such as vaccination and sanitation), and epidemiological insights into key causes of non-communicable diseases (such as smoking) allowed public health teams, consisting of qualified individuals drawn from a variety of backgrounds, to attain a crucial position in improving the health of the whole population (**Box 9.2**). Latterly, healthcare public health has brought academic rigour to the study of aspects of healthcare delivery and to ensuring that maximum population benefit is derived from healthcare services. Globally, there has been an increasing move towards a population health approach, aiming to improve physical and mental health outcomes across populations, as well as address inequalities.

Public health and global health

This chapter focuses primarily on public health in higher-income countries. Public health issues more commonly associated with low- to middle-income countries, such as communicable and nutritional diseases, are covered in Chapter 10. However, non-communicable

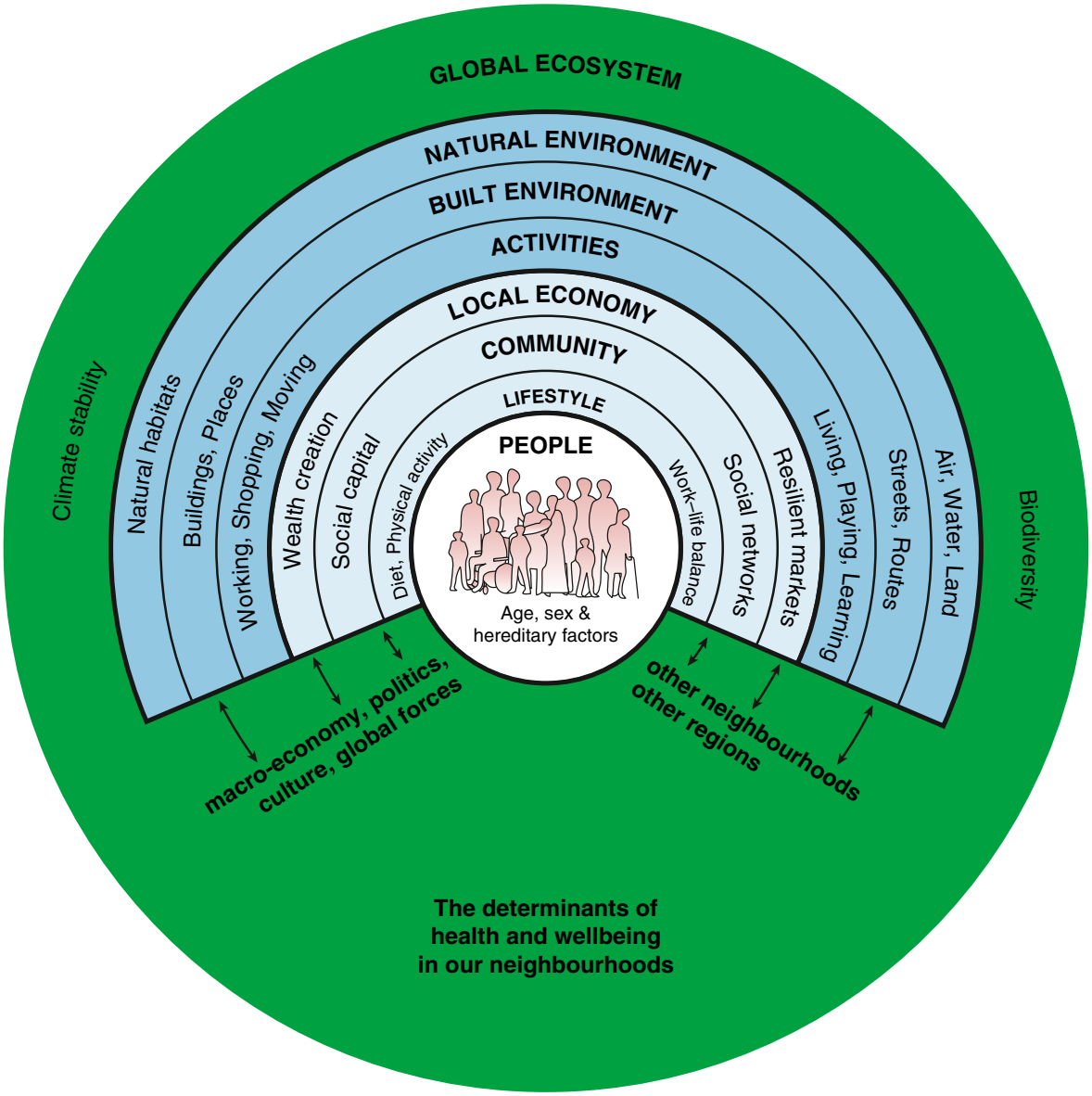


Fig. 9.1 The health map. *(From Barton H, Grant M. A health map for the local human habitat.* J Roy Soc Promot Health *2006; 126:252–253. ISSN 1466-4240 developed from the model by Dahlgren G, Whitehead M. 'The main determinants of health' model, 1991, version accessible in: Dahlgren G, Whitehead M.* European Strategies for Tackling Social Inequities in Health: Levelling up, Part 2. *Copenhagen: WHO Regional Office for Europe, 2007; http://www.euro.who.int/__data/assets/pdf_file/0018/103824/E89384.pdf.)*

Box 9.1 John Snow and the Broad Street pump

In 1854 an outbreak of cholera struck Soho in central London. Dr John Snow was a local physician who had written a paper theorizing about the transmission of cholera 5 years previously. Snow began to investigate the outbreak, reviewing local hospital records, talking to local residents, and mapping the cases he identified. He noted that 'within 250 yards of the spot where Cambridge Street joins Broad Street there were upwards of 500 fatal attacks of cholera in 10 days […] I suspected some contamination of the water of the much-frequented street-pump in Broad Street'. Snow proceeded to remove the handle of the water pump to prevent future cases, translating epidemiology into the practice of public health.

diseases, such as cardiovascular disease, cancer, and chronic respiratory disease, are now the leading cause of death globally, meaning that the contents of this chapter are relevant for all settings. Furthermore, resource prioritization and understanding population needs are integral to developing healthcare systems in low- and middle-income as well as in high-income countries.

Further reading

Institute for Health Metrics and Evaluation. *Global Burden of Disease Compare*. 2019; Global Burden of Disease (GBD) (healthdata.org).
International Conference on Primary Health Care. *Declaration of Alma-Ata*. 1978; https://www.who.int/publications/almaata_declaration_en.pdf?ua=1.

Box 9.2 Public health teams

Public health is increasingly a multidisciplinary specialty, accessible to applicants who are not doctors but who can demonstrate suitable academic attainment and experience working in the field. Public health specialty registrar training in the UK has been multidisciplinary since 2003 and around half of public health trainees are doctors, the rest being graduates from backgrounds that include epidemiology, health policy, biostatistics, or behavioural science. Public health teams exist at local, regional, and national levels; depending on the context, they may work closely with local or national government, the emergency services, education professionals, policy experts and advocates, healthcare academics, and other local medical services.

Box 9.3 Bradford Hill criteria

Causality is more likely if:

- the effect size is large (strength)
- findings are reproducible across different samples (consistency)
- the association is demonstrable in only a specific population (specificity)
- the effect occurs after the cause (temporality)
- greater exposure leads to greater incidence of the effect (biological gradient)
- there is a feasible mechanism for the association (plausibility)
- the proposed causal relationship is not at odds with previous research (coherence)
- the effect varies measurably following a change to the causal factor (experiment)
- the effects could be proposed to be similar to other situations (analogy).

KEY CONCEPTS IN PUBLIC HEALTH

Epidemiology and causality

Epidemiology is the study of how often diseases occur in different groups of people, and why. It is central to the discipline and practice of public health medicine. Understanding the patterns of disease distribution in a population can reveal factors that may be associated with that disease. For example, the demonstration of a link between smoking and lung cancer by Sir Richard Doll and Sir Austin Bradford Hill in 1950 is a seminal piece of epidemiological research. If epidemiological associations are causal, then working to reduce or mitigate identified causes may provide an opportunity to prevent future disease. Bradford Hill laid out nine criteria to be considered when judging whether an association is likely to be causal (Box 9.3).

Health needs assessments

Health needs assessments (HNAs) are key tools used in public health medicine to drive improvement in health within a population. HNAs aim to determine healthcare needs systematically and to compare them with current service provision to identify gaps. A health need (whenever individuals are not completely well) and a health*care* need (if there is an ability for them to benefit from care) should be differentiated: health*care* needs exist only if an evidence-based intervention has been shown to improve the health of the individuals with the condition in question, as not all diseases are amenable to treatment. Real healthcare needs must be distinguished from demand (what individuals may wish to buy or use) and supply (current provision) (Fig. 9.2). Health equity audits, a process to examine how health determinants, access to services, and related outcomes are distributed across groups, may be performed in relation to needs within an HNA.

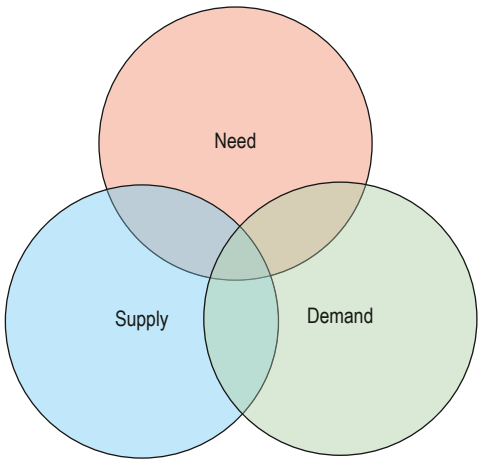


Fig. 9.2 Demand, supply, and need.

Three techniques for producing HNAs have been described (Stevens and Raftery 1997):

- ***Epidemiological***—summarizing data on healthcare needs from a range of sources by time, place, and person
- ***Comparative***—comparing needs, as well as service provision and service usage, between different populations
- ***Corporate***—consulting with stakeholders, such as healthcare professionals, commissioners, and patients, to identify priorities.

Life expectancy and quality of life

Life expectancy at birth is defined as the number of years that a newborn could be expected to live if current mortality rates continue. Life expectancy was generally rising across the world prior to 2020, although an Organization for Economic Co-operation and Development (OECD) report from 2019 noted that the rate of improvement had slowed in several European countries since 2011 (including France, Germany, Sweden, the Netherlands, and the UK), and life expectancy was falling in the USA. Falls in life expectancy became more widespread in 2020 with the COVID-19 pandemic, particularly in non-Scandinavian Europe and the USA. The largest decline within the OECD countries was reported for the USA (1.8 years), Spain (1.6 years), and Poland (1.5 years).

Healthy life expectancy, defined as the number of years lived in self-assessed good health, adds a quality-of-life component to estimates of life expectancy. While healthy life expectancy in many countries was generally rising prior to the COVID-19 pandemic, it did so at a lower rate than life expectancy. This means that people in most countries are experiencing more years in poor health. Healthy life expectancy is inversely associated with deprivation. Females in the most deprived areas of England not only die on average 8 years younger, they also experience an additional 11 years of poor health, resulting in a healthy life expectancy that is 19 years shorter than females in the least deprived areas (Fig. 9.3).

Health inequalities

Health inequalities are differences in health status between groups of individuals, which differ based on gender, ethnicity, disability, area

of residence, socioeconomic status, or other factors. Health differences arising from such characteristics often represent an injustice, especially if they are potentially avoidable and thus inequitable (inherently unjust and unfair). One of public health's fundamental principles is to strive to reduce such inequalities and to make health outcomes fairer for the population (**Box 9.4** and **Fig. 9.4**). Tudor Hart's ***inverse care law*** states that the availability of good healthcare often varies inversely with need, particularly where market forces are operating. These inequalities can be identified and addressed in relation to the access to, experience of, and outcomes from healthcare services. The COVID-19 pandemic has brought these inequalities to the fore and contributed to their widening.

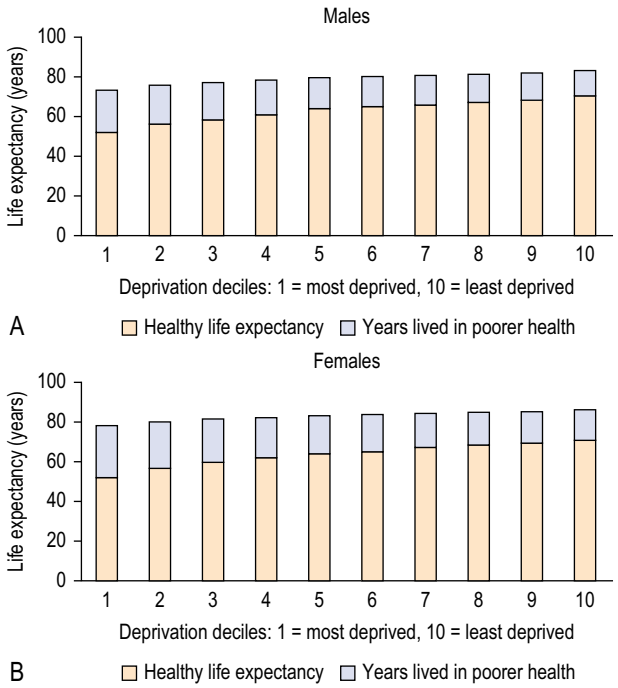


Fig. 9.3 Healthy life expectancy at birth and years lived in poorer states of health by national deprivation deciles (England, 2018–2020). (A) Males. **(B)** Females. *(Office for National Statistics. Annual Population Survey.)*

Deprivation

Examples of health inequalities that are amenable to change are those associated with deprivation. Deprivation can be measured by a basket of indicators within defined geographical areas. One example is the Index of Multiple Deprivation, which uses data across seven domains to compare relative deprivation across small areas in England. The domains include:

- income levels
- employment levels
- education and skills within a population

i Box 9.4 The Marmot review

In 2010 in the UK, Sir Michael Marmot produced a seminal report, 'Fair Society Healthy Lives', which detailed evidence-based strategies for reducing health inequalities. The report acknowledged the social gradient in health, whereby an individual's social circumstances have an impact on health, and recognized that social inequalities lead to health inequalities. Marmot concluded that, to reduce health inequalities, action is required across the wider social determinants of health and that this should exhibit proportionate universalism (i.e., actions should be universal to all but proportionate in scale and intensity to the level of disadvantage). The report stated that actions to tackle inequalities should focus on six overlapping policy objectives:

1. giving all children the best start in life
2. enabling people of all ages to maximize their potential and acquire control over their lives
3. creating fair employment and work opportunities
4. ensuring that everyone has a healthy standard of living
5. creating and developing healthy, sustainable communities
6. strengthening preventative health measures.

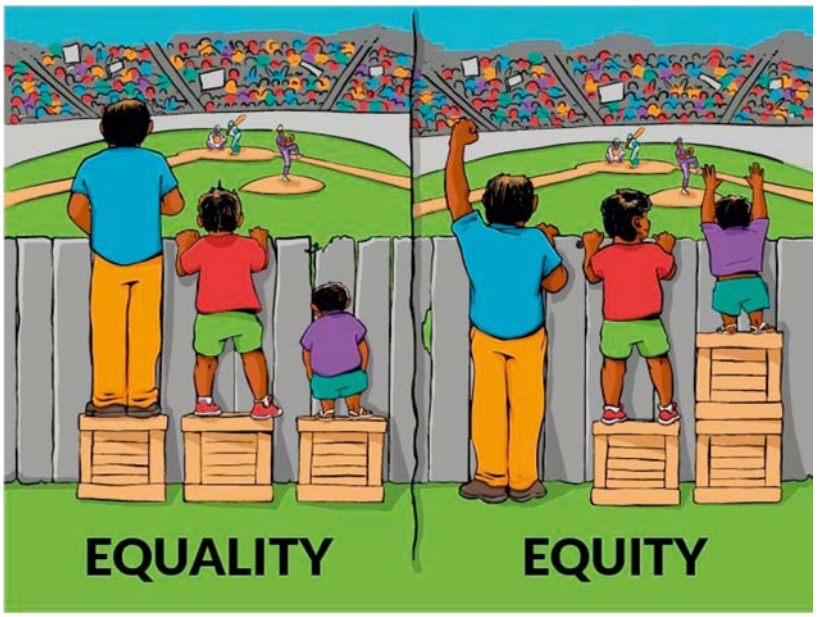


Fig. 9.4 Inequity and inequality. *(From Interaction Institute for Social Change. Artist: Angus Maguire.* Illustrating Equality vs Equity, 2016; *http://interactioninstitute.org/illustrating-equality-vs-equity/.)*

- levels of chronic ill-health and disability
- crime
- barriers to housing and local services
- the living environment.

A number of poor health outcomes have been found to correlate with deprivation. For example, children living in deprived areas have a higher prevalence of dental caries than those in less deprived areas. Water fluoridation is a public health intervention that improves oral health across the population but has been shown to have the most impact in those who are more deprived, thereby reducing health inequality. Other adverse outcomes associated with deprivation extend throughout the life-course and include low birth weight, higher infant mortality, increased chronic disease, and premature death. Data suggest that inequalities in health due to deprivation are widening. For example, over the past decade the prevalence of obesity in children aged 4 to 5 years is increasing disproportionately in those children living in the most deprived areas of England.

Politics and advocacy

The WHO defines health advocacy as 'a combination of individual and social actions designed to gain political commitment, policy support, social acceptance and systems support for a particular health goal or programme'. Advocacy is a crucial component of public health which is about making change. Targets of campaigns might be to change the law, to improve access to healthcare services, to improve standards in the food industry, to reduce climate change, or to change attitudes in society towards healthy lifestyle choices. Overcoming the powerful opposition of large corporations and/or governments is often a requirement of effective public health advocacy.

All policy decisions—not just those within the health sector but also those in other areas such as transport or housing—should be rigorously and transparently assessed to ensure that they strive to reduce health inequalities. ***Health impact assessments*** are used to understand what the health implications may be, as well as their distribution within society, to maximize positive health consequences (both indirect and direct) of projects or strategies and minimize any adverse ones.

Health advocacy has often had a larger impact on health than the introduction of new services. Examples of successful health advocacy include bans on smoking in confined public places and the reduction of salt and sugar content in processed food.

Smoking bans

Many countries have introduced bans on smoking in the workplace and in public spaces; a few (such as New Zealand and Finland) have announced a target to be 'tobacco-free' by a designated point in the future. International data from the USA and Europe have revealed a variety of improved health outcomes since the introduction of smoking bans, including reductions in preterm births, childhood hospital admissions for asthma, and emergency admissions for myocardial infarction. Despite unequivocal evidence of public health benefits, opponents of smoking bans have suggested they are examples of governments acting as a 'nanny state' or have expressed concern over the potential for a shift in smoking behaviour into private homes (which does not seem to have occurred). However, in general, smoking bans received public support, including among smokers, and much of the opposition came from large tobacco companies.

Attention is now turning to vaping. Vaping is much safer than smoking and has been supported as a tool to aid smoking cessation. However, vaping is not risk free. Concerns are growing that people, particularly if young, may start vaping despite not having previously been a smoker. There are also environmental considerations with the proliferation of non-reusable disposable vapes. Governments are introducing new vaping restrictions—including total sale bans (e.g., India) and prescription requirements (e.g., Australia). Other forms of restriction, such as on the flavour or re-usability, are also being proposed globally.

Improved nutritional content of food

Regularly exceeding the maximum amount of salt recommended for daily consumption is common across the world and is strongly associated with hypertension, cardiovascular disease, and premature mortality. Diets rich in simple sugars drive the development of obesity, insulin resistance, and diabetes, which are, in turn, risk factors for cardiovascular disease and many malignancies (see Chapter 37). Health advocacy efforts have been based on public education campaigns, clear labelling of the salt and sugar content of foods, and salt and sugar reduction targets for the food industry. Many Western diets are based on processed food, and an approach based on purely individual behaviour change is widely regarded as insufficient to tackle the problem. To make stronger gains the food industry needs to reduce salt and sugar further within prepared food in supermarkets and food outlets.

More widely, there is emerging evidence that the gut microbiome has an important impact on health and may be modified by the types of food we eat. Diets rich in plant and fermented products and low in ultra-processed foods are associated with better health outcomes. Achieving widespread change in a population's food intake requires government commitment, with imposed targets and incentives for the food industry.

Further reading

Bradford Hill A. The environment and disease: association or causation? *Proc R Soc Med* 1965; 58:295–300.
Doll R, Bradford Hill A. Smoking and carcinoma of the lung: preliminary report. *Br Med J* 1950; 2:739–748.
Tudor Hart J. The inverse care law. *Lancet* 1971; 297:405–412.
Marmot M, Atkinson T, Bell J et al. *Fair Society, Healthy Lives: The Marmot Review*. 2010; http://www.instituteofhealthequity.org/resources-reports/fair-society-healthy-lives-the-marmot-review/fair-society-healthy-lives-full-report-pdf.pdf.
Office for National Statistics. *Health State Life Expectancies by National Deprivation Deciles, England: 2018 to 2020*; https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/healthinequalities/bulletins/healthstatelifeexpectanciesbyindexofmultipledeprivationimd/2018to2020.
Spector TD, Gardner CD. Challenges and opportunities for better nutrition science-an essay by Tim Spector and Christopher Gardner. *BMJ*. 2020; 369:m2470.
Stevens A, Raftery J (eds). *Health Care Needs Assessment, 2nd series.* Oxford: Radcliffe; 1997.

PILLARS OF PUBLIC HEALTH

Although models of public health provision vary in different countries, services are often delivered at the level of local government or regional health authorities. National bodies are often responsible for collation of health statistics, support for national vaccination programmes, and advice to government on other aspects of health policy. In the USA this includes the Centres for Disease Control and Prevention (CDC) and in England, agencies include the Office of

Health Improvement and Disparities (OHID) and the United Kingdom Health Security Agency (UKHSA).

The work of public health services can be described according to the three pillars of public health:

- health improvement
- health protection
- healthcare public health.

Health improvement

While it might seem self-evident that preventing a disease is better than dealing with the consequences, there is also strong economic evidence to support the case for prevention. For example, a number of studies have demonstrated that smoking cessation services both improve population health and reduce healthcare costs overall. When considering approaches to prevention, they can be classified as primary, secondary, or tertiary (**Box 9.5**).

Wider determinants of health

Wider determinants of health (WDHs) are the broader social, economic, political, and environmental circumstances that influence health outcomes throughout life. Dahlgren and Whitehead (1991), subsequently modified by Barton and Grant (2006), developed an influential model of the main determinants of health, in which, at the core, are constitutional factors such as sex, age, and genetics; overlapping layers represent individual lifestyle factors, followed by the wider determinants (**Fig. 9.1**). The core attributes are relatively fixed but, as the layers of influence extend outwards, the determinants are amenable to change:

- The first layer represents individual ***lifestyle behaviours***, for example, diet, physical activity, smoking, and alcohol consumption.
- The second layer represents the ***communities in which individuals live***—social networks that influence health and the health behaviours we exhibit.
- Next, ***living and working conditions*** have an impact on health, for example, the quality of an individual's employment, housing, and education.

Box 9.5 Levels of prevention of disease

Primary

- Involves reduction in exposure to risk factors, focusing either on the whole population or on high-risk subgroups
- Examples include healthy lifestyle interventions for the prevention of cardiovascular disease, or workplace-based interventions to reduce asbestos exposure
- Focusing on the whole population may result in the ***prevention paradox***—such strategies often have a relatively small impact at an individual level, despite large benefits at a population level or for those at highest risk

Secondary

- Involves detecting disease at an early stage and undertaking prompt intervention to reduce the disease's impact
- Examples include screening programmes or prescription of statins for hypercholesterolaemia to reduce the likelihood of myocardial infarction or stroke

Tertiary

- Aims to reduce or mitigate long-term complications of a disease and maximize functioning
- Examples are HbA_{1c} monitoring in diabetes mellitus to minimize long-term micro- and macrovascular complications or neuro-rehabilitation after head injury

- Finally, the prevailing ***socioeconomic, cultural, and environmental conditions*** affect the health of the whole population, and these are influenced by local, regional, national, and international factors.

Air pollution

Air pollution is an example of a WDH, representing a huge environmental health risk globally. Poor air quality has a range of short- and long-term health impacts, including but not limited to cardiovascular and respiratory disease, emergency hospital admissions, and death. One of the first demonstrations of these effects was during the 'Great Smog of London' in 1952, which contributed towards many deaths. This led to the Clean Air Act of 1956, which limited the burning of solid fuel for heating in urban areas.

A range of pollutants are known to be harmful to health, including particulate matter and nitrogen dioxide. Although tackling air pollution requires individual efforts, the problem cannot be adequately addressed without local, national, and international policy change. Examples of interventions for tackling air pollution include:

- setting objectives and limits for different pollutants
- encouraging use of active transport instead of cars
- encouraging manufacturers to produce cleaner vehicles, and industry to adopt greener technologies
- introducing ultra-low emissions zones in urban areas and 'no-idling zones' around schools.

Poor air quality affects the whole population but also represents a significant health inequality in that more vulnerable members of society have higher exposure (e.g., those living in more deprived areas) or are more susceptible to harm (e.g., young children or the elderly).

Behaviour change

Behaviour change provides a crucial opportunity to improve health significantly. However, unhealthy behaviours are often complex and deeply ingrained. For example, the COM-B model recognizes that the interrelated components of *c*apability, *o*pportunity, and *m*otivation must all be addressed before *b*ehaviour is changed (**Fig. 9.5**). At a population level, instigating change requires holistic and sustained interventions across a range of policy types. Interventions may be classified as individual, structural, fiscal, or legal (**Fig. 9.6**), although there may be significant overlap between these; for instance, structural and fiscal approaches are often enshrined in legislature.

Individual interventions

Clinical consultations can be a prime setting to encourage individual-level behavioural change. The 'Making Every Contact Count' initiative seeks to encourage clinicians, and a broader range of professionals, to consider how they can build health promotion into their routine patient or client interactions. For example, professionals working in education or in the emergency services may identify issues of poor mental or physical health, isolation, or violence. They may have an initial conversation about the issues they have identified, before signposting to a range of other local services that may provide further support.

While brief advice is unlikely to instigate long-term change if delivered in isolation, it may raise the idea of change higher up the patient's agenda, particularly if delivered during a 'teachable moment' when an individual becomes especially aware of their health, such as during an unexpected hospital admission. There are several models to describe individual behaviour change, such as

Prochaska and DiClemente's 1983 Stages of Change model (Fig. 9.7). Individual-level behaviour change therapies have been found to be highly cost-effective: for example, brief advice given to adult patients in primary care about increasing their physical activity has been shown to cost £1730 per quality-adjusted life-year (QALY) gained.

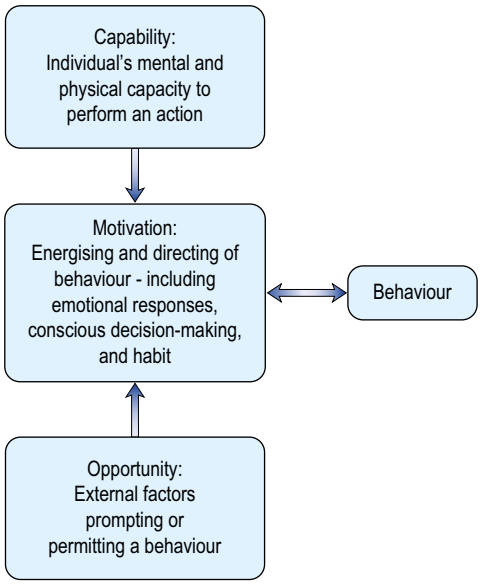


Fig. 9.5 The COM-B system. *(From Michie S, van Stralen M, West R. The behaviour change wheel: A new method for characterising and designing behaviour change interventions.* Implementation Science *2011; 6:42.)*

Fig. 9.6 Examples of initiatives to reduce smoking.

Structural interventions

Structural approaches to behaviour change may be achieved through 'nudging' individuals to default to healthier choices. The nudge theory of behavioural economics rose to prominence in 2008 with the publication of 'Nudge: Improving Decisions about Health, Wealth and Happiness' by Richard Thaler and Cass Sunstein, leading to Thaler being awarded the 2017 Nobel Prize for Economics. Nudge theory proposes that small tweaks to frame a choice more favourably or render an option easier are often sufficient to increase its uptake, and that decisions are not based purely on underlying rationale. A Cochrane review in 2015 concluded that individuals consistently eat more food when offered large portions or use larger items of tableware.

Fiscal interventions

Fiscal measures may also encourage behaviour change, through either the taxation of potentially harmful choices to render them less attractive or the subsidization of healthy ones to make them more favourable. The WHO encourages countries to tax sugary drinks, and many have implemented such legislation. Measures to improve healthy eating by subsidizing school dinners or offering free vitamin supplements to children and pregnant females may enhance health outcomes and also reduce inequality (Box 9.6).

Legal interventions

In addition to the introduction of taxes, legislative approaches to encourage behaviour change may include the regulation or prohibition of unhealthy activities. It was estimated in 2008 in the UK that over the previous 25 years since the wearing of seatbelts became compulsory, at least 60 000 lives had been saved and 600 000 severe injuries had been prevented. Legal approaches to behaviour change are often used once other methods have been tried without adequate success, since they may be perceived as paternalistic, attracting political controversy and significant lobbying from industry.

Further reading

Avineri E, Goodwin P. *Individual Behaviour Change: Evidence in Transport and Public Health*. London: Department for Transport; 2010; http://eprints.uwe.ac.uk/11211.

Falbe J, Thompson H, Becker C et al. Impact of the Berkeley excise tax on sugar-sweetened beverage consumption. *Am J Public Health* 2016; 106:1865–1871.

National Institute for Health and Care Excellence. *NICE Public Health Guideline 49: Behaviour Change: Individual Approaches*. NICE; 2014; https://www.nice.org.uk/guidance/ph49.

Owen L, Pennington B, Fischer A et al. The cost-effectiveness of public health interventions examined by NICE from 2011 to 2016. *J Public Health* 2017; 40:557–566.

Rose G. Strategy of prevention: lessons from cardiovascular disease. *Br Med J* 1981; 282:1847–1851.

Health protection

Health protection involves anticipating and controlling infectious, environmental, chemical, and radiological hazards to public health.

Fig. 9.7 Stages of change model. *(From Prochaska J, DiClemente C. Stages and processes of self-change of smoking: toward an integrative model of change.* J Consult Clin Psychol *1983; 51:390–395.)*

Box 9.6 Fiscal measures to improve diet and nutrition

Taxing harmful behaviours

- Poor oral health, including the presence of dental caries, is the most prevalent condition worldwide (Global Burden of Disease 2017). Reducing sugar consumption is predicted to improve oral health, as well as help tackle obesity.
- An excise tax on sugar-sweetened beverages ('sugar tax') is an example of a fiscal policy that aims to encourage manufacturers to reduce the sugar content of their products, as well as dissuading individuals from consuming products with high sugar levels.
- Such a tax has been introduced in a number of jurisdictions over recent years and initial results are promising. In the UK the government introduced a Soft Drinks Industry Levy in 2018 and it is estimated that half of manufacturers reformulated their drinks to avoid it. An evaluation of a similar tax in Berkeley, California, found that it was associated with a 21% decrease in sugar-sweetened beverages in low-income neighbourhoods, indicating that the policy has the potential not only to improve health but also to reduce health inequalities.

Subsidizing healthy behaviours

- It has been estimated that 1 billion people globally have vitamin D deficiency, and there has been a rise in cases of rickets in some countries.
- This can be prevented by eating a diet rich in vitamin D and calcium or by taking vitamin D supplements.
- The Healthy Start vitamins scheme was introduced in the UK in 2006 and provides free vitamin supplements (including vitamin D) for pregnant females and young children in low-income families.

There are three main areas of activity: control, monitoring (surveillance), and prevention.

Control

Control of infectious diseases

Rapid response to outbreaks of disease requires timely notification of new cases to the relevant public health authorities to enable controls to be established. The International Health Regulations (IHR) 2005 signify an agreement between 196 countries to build their capacities to detect, assess, and report public health events and to notify the WHO of any event that may constitute a 'public health emergency of international concern' (PHEIC). The WHO declared the novel coronavirus SARS-CoV-2 (COVID-19) as a PHEIC on 30 January 2020 and subsequently as a pandemic on 11 March 2020.

In England the most recent national legislation lists 34 diseases and 62 causative organisms as notifiable, and other infections not included on this list should also be notified if they could present significant harm to human health (such as emerging or new infections). Cases are notified based on clinical suspicion to the local health protection team, and notification should not be delayed while awaiting confirmation.

Health protection teams will then perform a risk assessment for each notification they receive and will introduce control measures specific to the disease and the context to mitigate identified risks, working in conjunction with environmental health officers from the local authority. Control measures may include:

- provision of infection prevention and control ('IPC') advice
- isolation of infected individuals
- post-exposure prophylaxis.

Identification of infected individuals through widespread testing, contact tracing, isolation of exposed individuals, closure of settings such as schools and workplaces, and reductions in movement with national and international travel restrictions were all implemented during COVID-19 to reduce infection rates.

Box 9.7 Middle East respiratory syndrome coronavirus (MERS-CoV)

- In 2015, MERS-CoV was reported to be one of the top emerging pathogens likely to cause severe outbreaks.
- It is a viral respiratory disease first reported in Saudi Arabia in 2012 and thought to be associated with bat coronaviruses.
- MERS-CoV has also been found in camels, and the WHO advises that individuals with co-morbidities should avoid contact with camels, refrain from drinking raw camel milk, and not eat undercooked camel meat.
- Patients may develop severe pneumonia and renal failure, and there is high mortality in 30% to 40% of cases.
- Thus far, all cases have been associated with countries in or near the Arabian Peninsula, and spread has been reported via close contact with infected cases or in healthcare settings.

Control of environmental hazards

Public health services are also notified of local environmental, chemical, or radiological hazards that threaten the health of the public. These threats are wide-ranging and include air pollution, drinking water contamination, and exposure to radiation. Public health professionals work with the relevant experts, emergency services, local government, and industry to identify and mitigate these hazards.

Monitoring (surveillance)

'Emerging infectious diseases' (EIDs) is a term that encompasses newly emerging pathogens, pathogens whose incidence is increasing and/or that are appearing in new populations, and pathogens that are re-emerging in different forms (e.g., multidrug-resistant [MDR] and extensively drug resistant [XDR] tuberculosis). Many of these EIDs are zoonoses (of animal origin), including, for example, human immunodeficiency virus (HIV) and Middle East respiratory syndrome coronavirus (MERS-CoV, **Box 9.7**). Other recent examples of EIDs include Zika virus associated with microcephaly; Ebola virus disease associated with outbreaks in West Africa and the Democratic Republic of Congo; and SARS-CoV-2.

The reasons why new infections appear and old ones return are multifactorial and include:

- ***population movement*** to new geographical areas, for example, displacement following war or natural disasters
- ***a rise in global trade and travel***
- ***ecological changes***, such as those driven by climate change or severe weather events (which can, for example, have an impact on vectors implicated in the transmission of pathogens); or changes in agricultural practices (which can, for example, alter the likelihood of coming into contact with sources of infection)
- ***technologies in healthcare***, for example, xenotransplantation or increased antibiotic use (associated with the development of antimicrobial resistance).

Well-developed surveillance mechanisms are key in the early detection of new and emerging pathogens that may represent a significant future threat to health. This monitoring facilitates outbreak investigations (**Box 9.8**), as well as informing broader infection prevention policy (**Box 9.9** and **Fig. 9.8**).

Box 9.8 *E. coli* outbreak investigation using whole-genome sequencing (WGS)

WGS is increasingly used in infectious disease surveillance. The identification of highly similar bacterial genomes can link individuals to the same outbreak, when there may not otherwise have been obvious similarities between cases. For example, in 2017, Public Health England surveillance identified an outbreak of 12 cases of *E. coli* O157 occurring within a 6-week window and carrying the same WGS. The cases lived across England, had no other known common features, and would not have been investigated collectively, were it not for WGS results. Using a detailed questionnaire, epidemiologists identified burgers from a particular retailer as a potential link. In collaboration with the supermarket and the Food Standards Agency, supermarket loyalty card records belonging to the affected individuals were analysed to identify the specific burger brand. The product was recalled and the supermarket contacted those known to have bought the burgers, advising them to return for a refund.

Box 9.9 Pertussis vaccination in pregnancy

Pertussis infection (whooping cough) is a significant infectious cause of morbidity and mortality in infants, particularly among those too young to have been protected by vaccination. After UK surveillance data identified that, during a 2012 pertussis outbreak the highest incidence of disease was among infants under 3 months of age, the Department of Health introduced an antenatal vaccination programme. Intrauterine transfer of maternal antibodies confers passive immunity, protecting infants until they are old enough for active immunity with routine vaccination. The antenatal programme has been shown to be effective, with a reduction in infant pertussis incidence (**Fig. 9.8**).

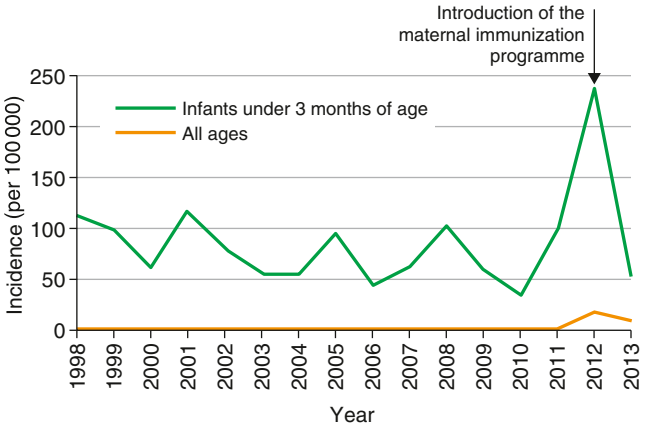


Fig. 9.8 Incidence of laboratory-confirmed cases of pertussis, England (1998–2013). *(By kind permission of Public Health England. The Green Book, Ch. 24, p. 4, Fig. 24.2. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/514363/Pertussis_Green_Book_Chapter_24_Ap2016.pdf.)*

Prevention

Prevention within health protection is concerned with reducing the risk of transmission of infectious diseases. Two key examples of this are hand hygiene and vaccination.

Hand hygiene is a non-pharmaceutical intervention (NPI) that is one of the simplest yet most effective public health (prevention and control) measures. Basic hand-washing in communities minimizes the spread of many infections (e.g., reducing the risk of *Escherichia coli* infection in petting zoos or disrupting the transmission of norovirus in schools) but is also the main way of combatting hospital-acquired infections and preventing the spread of antimicrobial resistance.

During the COVID-19 pandemic the importance of hand-washing was promoted alongside other NPIs to reduce the spread of infection. Individual-level NPIs introduced internationally to varying degrees included physical distancing of 1 to 2 metres between individuals and the use of face masks, including in community (non-healthcare) settings. Environmental NPIs also play a role in prevention and include cleaning and adequate ventilation of indoor spaces.

Box 9.10 UK vaccination universal schedule 2022

8 weeks	Diphtheria, tetanus, pertussis, polio, *Haemophilus influenzae b*, hepatitis B
	Meningococcal group B
	Rotavirus
12 weeks	Diphtheria, tetanus, pertussis, polio, *Haemophilus influenzae b*, hepatitis B
	Pneumococcal
	Rotavirus
16 weeks	Diphtheria, tetanus, pertussis, polio, *Haemophilus influenzae b*, hepatitis B
	Meningococcal group B
1 year	*Haemophilus influenzae b* and Meningococcal type C
	Pneumococcal
	Measles, mumps, and rubella
	Meningococcal group B
3 years 4 months	Diphtheria, tetanus, pertussis, polio
	Measles, mumps, and rubella
12–13 years	Human papillomavirus
14 years	Tetanus, diphtheria, and polio
	Meningococcal groups A, C, W, Y
65 years	Pneumococcal
70 years	Shingles
Seasonal[a]	Influenza: children aged 2–10 years and adults aged 50 and over
	COVID-19: adults aged 50 and over

[a]Please note eligibility for seasonal programmes has changed annually in the UK through the COVID-19 pandemic and these are the universal (non-risk) groups eligible in Winter 2022-23.

Vaccination

Vaccination has had a huge impact on the burden of infectious disease, second only to clean water provision. Vaccination programmes confer benefit to the individual and to the population as a whole (especially its most vulnerable members) through herd immunity, whereby a high proportion of population immunity reduces person-to-person transmission. The basic reproduction number (R_0) is the number of secondary infections produced by a typical infection in a totally susceptible population. For diseases with a high R_0 (e.g., measles), higher coverage is required to attain herd protection.

Vaccination programmes are tailored to the national prevalence of particular diseases (e.g., UK schedule, **Box 9.10**) and develop

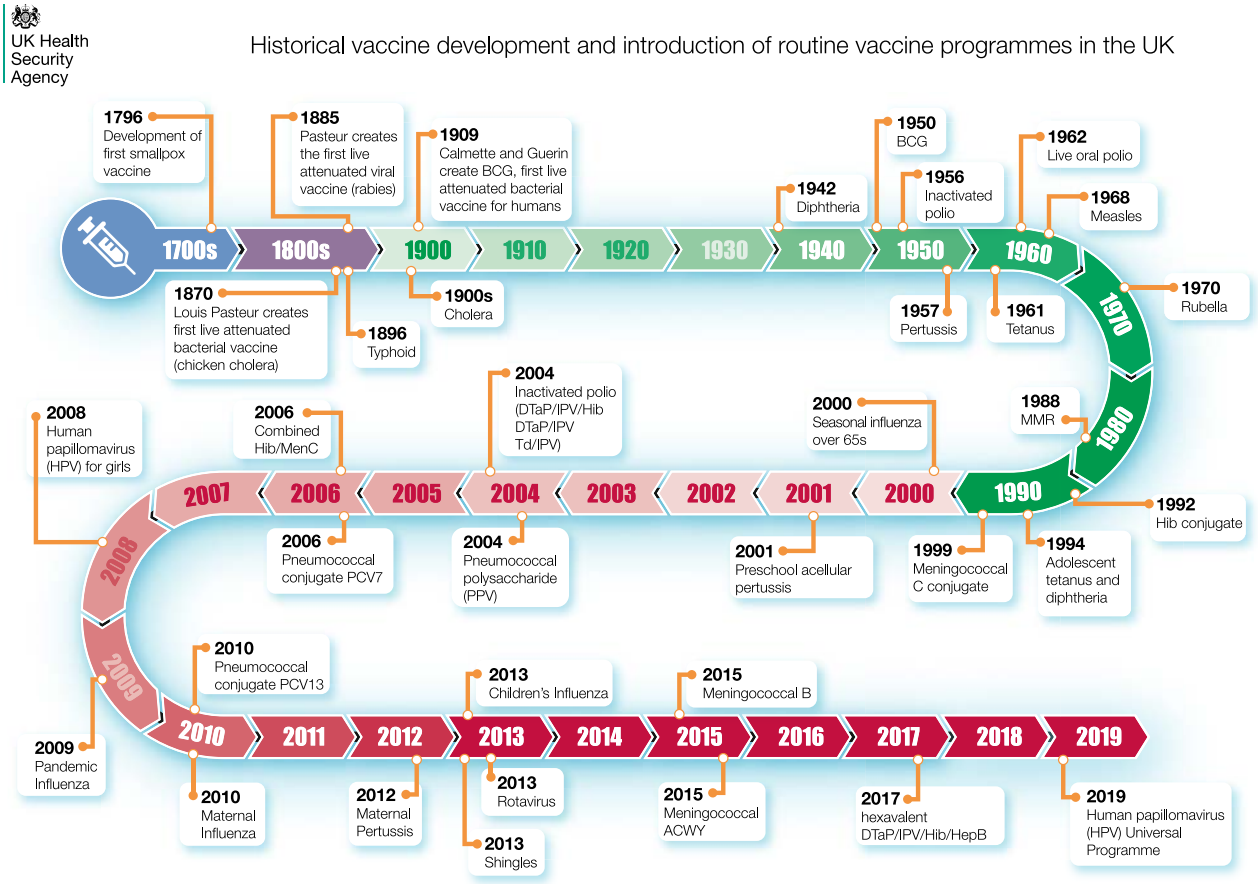


Fig. 9.9 **Development of the UK vaccination schedule.**

Box 9.11 Control, eradication, and elimination

- ***Control*** is defined as when the amount or impact of a disease within a population has reduced to a locally acceptable level.
- ***Eradication*** occurs when the disease incidence is zero worldwide and no environmental reservoir exists. This has been achieved only for smallpox.
- ***Elimination*** is when the disease incidence within a defined area is zero.

(Adapted from Dowdle WR, The principles of disease elimination and eradication. Bull World Health Organ *1998; 76[Suppl 2]:22–25.)*

Box 9.12 Notification of suspected measles cases

Recent outbreaks of measles in Europe and the USA have been seen primarily among individuals who have not received the measles, mumps, and rubella (MMR) vaccination. Post-exposure prophylaxis (PEP), if delivered promptly, can prevent contacts of known sufferers from developing measles or may attenuate their symptoms if they go on to develop the disease.

PEP can be delivered in two ways:

- ***Using a dose of MMR vaccine to induce active immunity.*** However, MMR is a live vaccine and should not be administered to immunocompromised individuals or pregnant females and is inappropriate for some infants.
- ***Using immunoglobulin to provide passive immunity.*** This is recommended for groups who cannot receive the MMR vaccine for the reasons listed above and who are often at greatest risk of morbidity and mortality from measles infection.

over time (Fig. 9.9). Selective vaccination programmes aim to protect those at highest risk of certain diseases (e.g., travel vaccines or hepatitis B vaccination for healthcare workers), while mass vaccination programmes aim to contain (control), eradicate, or eliminate disease (Box 9.11). Some vaccines provide protection *after* exposure in some cases (e.g., the measles, mumps, and rubella [MMR] vaccine, which can be used to help limit the spread of a measles outbreak by immunizing previously unvaccinated contacts of a known case, Box 9.12).

The success of a vaccination programme is driven as much, if not more, by its uptake among eligible populations as by its clinical efficacy in reducing infection and preventing transmission for an individual. Many factors may influence the acceptance of a vaccination offer (Fig. 9.10), and those that reduce uptake may coalesce in some communities, driving health inequalities. These social and behavioural barriers may spill over between vaccination programmes, with the uptake of routine childhood vaccinations falling globally during the COVID-19 pandemic, and can cluster with other behavioural risk factors for ill-health such as

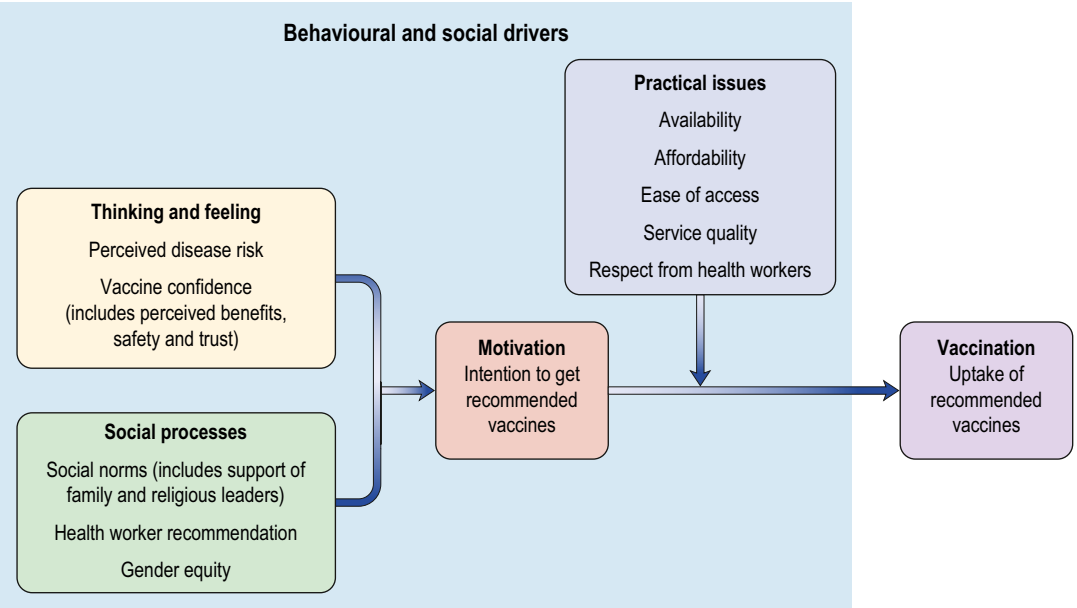


Fig. 9.10 **The WHO behavioural and social drivers (BeSD) of vaccination framework.**

smoking. Behaviour change models, such as COM-B (**Fig. 9.5**), can be used to analyse the reasons contributing to vaccine hesitancy within a community, and behavioural insights techniques can be used to design effective solutions. For example, pop-up clinic provision in community venues can address issues of capability (with knowledge transmitted through trusted community leaders), opportunity (ease of local access), and motivation (observing peers being vaccinated).

Unprecedented international coordination between academia, national governments, regulatory bodies, and the pharmaceutical industry resulted in rapid development and approval of a number of COVID-19 vaccinations. The first doses of COVID-19 vaccination were administered within a year of the global pandemic being declared. However, alongside vaccine hesitancy, another barrier to successful global vaccination uptake is inequitable distribution of vaccine supply.

Further reading

Andre FE, Booy R, Bock HL et al. Vaccination greatly reduces disease, disability, death and inequity worldwide. *Bull World Health Organ* 2008; 86:81–160.
Dowdle WR. The principles of disease elimination and eradication. *Bull World Health Organ* 1998; 76(Suppl 2):22–25.
Grant K, Byrne L. *Disease Detectives: Using Supermarket Loyalty Cards to Trace an E. Coli Outbreak*. 2018; https://publichealthmatters.blog.gov.uk/2018/09/26/disease-detectives-using-supermarket-loyalty-cards-to-trace-an-e-coli-outbreak/.
Local Government Association. *Applying behavioural insights to improve COVID vaccination uptake: a guide for councils*. 2021; https://www.local.gov.uk/publications/applying-behavioural-insights-improve-covid-vaccination-uptake-guide-councils#2-apply-behavioural-insights.
Morens DM, Fauci AS. Emerging infectious diseases: threats to human health and global stability. *PLoS Pathogens* 2013; 9:e1003467.
Morse SS. Factors in the emergence of infectious diseases. *Emerg Infect Dis* 1995; 191:7–15.
Parikh SR, Andrews NJ, Beebeejaun K et al. Effectiveness and impact of a reduced infant schedule of 4CMenB vaccine against group B meningococcal disease in England: a national observational cohort study. *Lancet* 2016; 388:2775–2782.
Public Health England. *Emerging Infections: How and Why They Arise*. 2019; https://www.gov.uk/government/publications/emerging-infections-characteristics-epidemiology-and-global-distribution/emerging-infections-how-and-why-they-arise.

Healthcare public health

Healthcare public health is the final of the three domains of public health, defined as being concerned with maximizing population benefits from healthcare interventions and reducing health inequalities. It involves the prioritization of limited resources, the prevention of disease, and the improvement of healthcare outcomes via 'design, access, utilisation and evaluation of effective and efficient health and social care interventions, settings and pathways of care' (Faculty of Public Health 2017).

Screening

Screening aims to reduce the harm caused by a disease and its complications by detecting the disease or its risk factors at an early enough stage for it to have a beneficial impact on outcome. Screening tests do not diagnose disease; they determine who should be investigated further for diagnosis and treatment. Screening can be:

- opportunistic, e.g., assessment of smoking status during a primary care visit
- part of a coordinated programme, with quality assurance and programme monitoring.

Although screening programmes are commissioned and delivered to improve health outcomes, screening has the potential to do harm, and so the benefits and risks need to be carefully considered, and screening should be implemented only when certain criteria are met (**Boxes 9.13 and 9.14**). In England, there are currently 11 population screening programmes (**Box 9.15**).

Screening for prostate cancer

An example of uncertainty around implementation of screening programmes is prostate cancer. This is the most common cancer in males in the UK, with over 50 000 cases annually. Although the rising incidence has links to increased case ascertainment through testing, as well as to increases in life expectancy, the mortality rate is highest in males over 90 years of age. It remains the second most common

cause of cancer death in males in the UK and is often aggressive in younger males, making it an attractive target for screening.

Prostate-specific antigen (PSA) is a glycoprotein that can be detected in blood and is associated with a range of conditions affecting the prostate, including but not limited to prostate cancer. This renders PSA testing of relatively poor sensitivity and specificity for a screening programme, with subsequent prostate biopsy for a definitive diagnosis not without harms. A 2018 systematic review demonstrated that PSA-based screening for prostate cancer would result in:

- a possible small reduction in disease-specific mortality over 10 years equating to 1 fewer death from prostate cancer per 1000 males screened
- no reduction in overall mortality, with 1 male hospitalized for sepsis, 3 experiencing urinary incontinence, and 25 experiencing erectile dysfunction from the same 1000 screened males.

The United Kingdom National Screening Committee does not currently recommend screening for prostate cancer, in contrast to EU recommendations (although few countries have an official population-based screening programme). In most high-income countries an opportunistic approach to PSA testing is taken on the basis of patient request, symptoms, or a high-risk family history—but this is viewed by many to be unsatisfactory and inequitable.

However, the balance between harms and benefits of screening for prostate cancer may shift with multiparametric magnetic resonance imaging (MRI) scanning, particularly if used in a targeted manner based on age, ethnicity, and a family history of prostate, breast, or ovarian cancer. Whether the UK has the capacity to deliver such a screening programme in an accessible and equitable way is currently debated and is in need of further evaluation.

Box 9.13 Wilson and Jungner criteria for screening tests

Wilson and Jungner (1968) developed the classical screening criteria to determine whether or not a condition/disease should be subject to screening:

- The condition should represent an important health problem.
- There should be an accepted treatment.
- Diagnostic and treatment facilities should be available.
- There should be a recognizable latent or early symptomatic stage.
- There should be an appropriate test or examination.
- The test should be acceptable to the population.
- The natural history of the condition should be well understood.
- There should be agreement on who to treat.
- The cost of case-finding should be economically balanced in relation to possible expenditure on medical care as a whole.
- Case-finding should be a continuous process.

Wilson JMG, Jungner G. Principles and practice of screening for disease. Geneva: World Health Organization; 1968.

Box 9.14 Screening test performance

Negative screening tests do not necessarily confirm the absence of disease. Screening test performance can be evaluated by assessing the disease status against the test result.

- ***False positives*** occur when individuals with no disease test positive.
- ***False negatives*** occur when individuals with disease test negative.
- Sensitivity and specificity describe how well the screening test performs against a gold standard (criterion) test:
  - ***Sensitivity*** is the ability of the test to detect all cases of disease.
  - ***Specificity*** is the ability of the test to detect all disease-free cases.
- Both positive and negative predictive values are dependent on the population disease prevalence:
  - ***Positive predictive value (PPV)*** is the probability of having the disease, given a positive test result.
  - ***Negative predictive value (NPV)*** is the probability of not having the disease, given a negative test result.

	Disease present	Disease absent	
Test positive	True positive (A)	False positive (B)	A+B
Test negative	False negative (C)	True negative (D)	C+D
	A+C	B+D	TOTAL

Sensitivity = A/(A + C) × 100
Specificity = D/(D + B) × 100
Prevalence = A + C/TOTAL × 100
PPV = A/(A + B) × 100
NPV = D/(D + C) × 100

Box 9.15 Screening programmes

Antenatal and newborn

- Sickle cell and thalassaemia
- Fetal anomaly
- Infectious diseases in pregnancy
- Newborn and infant physical examination
- Newborn blood spot
- Newborn hearing

Young person and adult

- Diabetic eye
- Abdominal aortic aneurysm
- Breast cancer
- Cervical cancer
- Bowel cancer

Antimicrobial resistance

This occurs when microorganisms develop resistance to commonly used antimicrobial drugs and are not suppressed by pharmacological agents that would normally inhibit their growth or kill them. Examples include methicillin-resistant *Staphylococcus aureus* (MRSA), XDR tuberculosis, and Carbapenemase-producing Enterobacteriaceae (CPE) (see Chapter 12). These multiply drug-resistant organisms are an emerging cause of hospital-acquired infections, which result in outbreaks, are associated with high morbidity and mortality, and have the potential to develop into even more drug-resistant strains (**Box 9.16**).

The cornerstones of prevention of the spread of antimicrobial resistance include:

- ***Early case recognition*** using robust surveillance mechanisms, for example, screening of all new admissions to hospital for asymptomatic carriage of such organisms.
- ***Antimicrobial stewardship:***
  1. Antibiotic prescribing practice is optimized so that antibiotics are prescribed only when clinically indicated
  2. Local prescribing guidelines are adhered to
  3. Use of broad-spectrum antibiotics is minimized
  4. Dose, route, and length of administration are monitored
  5. Antibiotics are stopped or switched promptly, as appropriate (see Chapter 12).

Box 9.16 Antimicrobial resistance

The advent of antimicrobial resistance is one of the biggest current threats to public health, as it has the potential to render individuals susceptible to a wide range of currently treatable microorganisms. As a result, 'low-risk' procedures (e.g., joint replacements) may become hazardous due to the infection risk, and treatment options for cancers or organ transplantation may become untenable if immunosuppression poses too great a threat. The O'Neill review commissioned by the UK government in 2014 estimated that, unless preventative action is taken, deaths from antimicrobial resistance could reach 10 million per year worldwide by 2050.

- ***Public education campaigns*** to try to reduce antibiotic demand
- ***Good hand hygiene*** on the part of healthcare workers
- ***Reduction of agricultural antibiotic use***
- ***Promotion of vaccination***
- ***Research*** to develop new classes of antimicrobial.

Health economics

Demand for healthcare often exceeds the finite resources available to deliver it. These finite resources include not only money but also the availability of space and appropriately trained staff. As not all demands for healthcare can be met, the concept of ***opportunity cost*** arises: if resource is used for one intervention, less will be available to do something else. The benefits gained from the chosen option must be greater than the benefits foregone by not choosing the alternate. Opportunity cost may be minimized by resourcing interventions with maximal cost-effectiveness.

Measuring outcomes

A prerequisite for maximal cost-effectiveness is maximal efficiency. Efficiency describes using the minimum resource to produce the same output or, alternatively, maximizing the output gained from using the same resource. When comparing interventions that produce the same output, such as an antihypertensive that, on average, reduces blood pressure by the same extent, the intervention that is cheaper will be most efficient (this is known as cost minimization). If the interventions produce different types of output, then a common measure of health is needed to permit comparison: if both treatments are being used for the same disease, then a disease-specific measure may be used.

Comparing treatments between completely different diseases requires a generic measure of the amount of health gained (termed '***utility***'). The two most commonly used measures of utility are the ***quality-adjusted life year (QALY)*** and the ***disability-adjusted life year (DALY)***. A QALY is 1 year of life in perfect health, whereas a DALY is the inverse: 1 year of healthy life that has been lost due to premature death or disability. The aim is to maximize the number of QALYs gained or minimize the number of DALYs lost.

Measuring cost-effectiveness

When the interventions being compared involve both different outputs and different costs, then a measure of utility in comparison to cost is produced. These cost–utility analyses produce an ***incremental cost-effectiveness ratio (ICER)***, which can then be compared by policy-makers to make funding decisions across the healthcare system (**Box 9.17**). The threshold at which interventions are deemed to be cost-effective depends on the resources available in a particular setting. For example, in England, funding recommendations are made by the National Institute for Health and Care Excellence (NICE), and a threshold of between £20 000 and £30 000 per QALY gained is currently used to determine whether an intervention should be recommended by NICE.

The ethics of funding decisions

Health economics does not only concern cost-effectiveness. The use of funds, and particularly of public funds, should be fair, as well as represent good value. The determination of 'fairness' is an ethical one, and contextual to the society in which the health system is operating. It is commonly said that the benefit of using QALYs is that they are uniform across conditions ('a QALY is a QALY is a QALY'), but whether this should be applied in all circumstances is debatable. For example:

- Is extending life by 1 month when a patient is only expected otherwise to live for 1 month the same as extending life by 1 month when a patient can be expected otherwise to live for 20 years?
- Should the amount a funder is willing to pay per QALY be the same for all diseases, when it is recognized that treatments for rare diseases are more expensive (to allow pharmaceutical companies to cover their research and development costs among a smaller sales volume)?

In the UK, specific rules have been introduced with respect to end-of-life treatments and to rare diseases in an attempt to answer these questions.

Box 9.17 Cost–utility calculations

Treatment A

- In comparison to current treatment, treatment A would extend patients' lives by a further 3 years, each at half the 'perfect' quality of life
- Net QALYs gained = 3 × 0.5 = 1.5
- Treatment A costs £45 000 per patient (one-off sum) more than the current treatment
- Cost–utility = 45 000/1.5 = £30 000 per QALY

Treatment B

- In comparison to current treatment, treatment B would not prolong patients' lives, but it would improve their quality of life from 30% to 50% (with a life expectancy from treatment of 10 years)
- Net QALYs gained = 10 × 0.2 = 2
- Treatment B costs £1600 per patient per year (ongoing)
- Cost–utility = (1600 × 10)/2 = £8000 per QALY

Therefore, treatment B is more cost-effective than treatment A.

Service delivery

Research has shown that active consideration of the needs of patient groups when planning how and where interventions are to be delivered has the potential to improve patients' health at a population level. Tailoring service design to patient needs can be observed in the use of outreach services for areas with high numbers of hard-to-reach individuals to increase accessibility. For example, homelessness is a risk factor for tuberculosis, but individuals who are homeless may be less likely to present to healthcare services with symptoms and so may remain undiagnosed; once diagnosed, they may be less likely to complete courses of treatment. Novel outreach services run to engage homeless people in the wider community have been described that use mobile X-ray diagnostic units to provide tuberculosis screening in the community.

Centralization

Where the volumes of patients with a particular condition are low the quality of care that patients receive may be improved by service centralization, which ensures they are seen by teams dealing with sufficient numbers of cases to make their care routine, not unusual. Services may be delivered using a 'hub-and-spoke' model, whereby problems of lower complexity are dealt with in local hospitals and problems of higher complexity or acuity are referred to a tertiary centre. For example, in London, city-wide stroke services were reorganized in 2010 with the introduction of a smaller number of 'hyperacute' stroke units where emergency care, including thrombolysis and mechanical

thrombectomy, is delivered; these are supported by a number of other local acute stroke units, to which patients are subsequently 'stepped down' for rehabilitation (see Chapter 29). A 2013 evaluation concluded that this service reorganization was associated with significantly improved patient survival and reduced healthcare costs.

Centralization has benefits including:

- bringing together specialized healthcare professionals and equipment
- enabling greater senior clinical presence out of hours
- helping to ensure that clinicians remain up to date with current evidence
- reducing unwarranted variation between units.

However, any centralization is a trade-off against increased travel time for patients (which might reduce accessibility and make services less equitable) and may be unpopular if there is a perception that local hospitals are being undermined.

Further reading

Ilic D, Djulbegovic M, Hung Jung J et al. Prostate cancer screening with prostate-specific antigen (PSA) test: a systematic review and meta-analysis. *Br Med J* 2018;362:k3519.

Hunter R, Davie C, Rudd A et al. Impact on clinical and cost outcomes of a centralized approach to acute stroke care in London: a comparative effectiveness before and after model. *PLOS One* 2013;8:e70420.

O'Neill J. *Tackling Drug-Resistant Infections Globally: Final Report and Recommendations*. 2016; https://amr-review.org/sites/default/files/160518_Final%20paper_with%20cover.pdf.

Public Health England. *Prostate Cancer Risk Management Programme (PCRMP): Benefits and Risks of PSA Testing*. 2016; https://www.gov.uk/government/publications/prostate-cancer-risk-management-programme-psa-test-benefits-and-risks/prostate-cancer-risk-management-programme-pcrmp-benefits-and-risks-of-psa-testing.

World Health Organization. *Global Action Plan on Antimicrobial Resistance*. 2015; http://apps.who.int/iris/bitstream/handle/10665/193736/9789241509763_eng.pdf;jsessionid=FB427A3145FC610B10462FA952207167?sequence=1.

Questions for this chapter can be found on Elsevier eBooks+.

10 Global and planetary health

Babulal Sethia, Dominik Zenner, and Parveen J. Kumar

Background 167
Overview 167
The global scale of the problem 167
Global health challenges 168
The global burden of disease 168
Infectious diseases and pandemics 168
Maternal and child health 168
Mental health 169
Surgical challenges 169
Social determinants of health 170
Poverty and food security 170
Education 171
Water and sanitation 171
Environmental degradation and pollution 171
Climate change 171
Human migration 172
Conflict and catastrophe 172
Global and planetary health governance 172
Organizations 172
Sustainable Development Goals 173
Economics and politics in global and planetary health 173
Human rights and the value of engagement in global health 173
The healthcare workforce and international recruitment ethics 173
Medical electives 173

CORE SKILLS AND KNOWLEDGE

Global and planetary health describes a high-level approach within medicine that recognizes that health is determined by challenges that transcend national boundaries, and looks at healthcare needs across the world as well as within individual nations. Such needs are complex, may be triggered by natural disasters or armed conflict, and must be addressed using a multiprofessional approach.

Efforts to improve global health are championed by non-governmental organizations such as the World Health Organization (WHO), the United Nations Children's Fund (UNICEF) and the World Food Programme, alongside individual charities and local government health systems, and interventions are often directed by international targets such as the United Nations' Sustainable Development Goals.

Opportunities for developing an understanding of global health include:

- Engaging actively in debates about the global economic and political determinants of health
- Considering the evidence for community-level interventions in improving health outcomes
- Travelling to see medicine being practised in a radically different context (e.g., as part of a medical school elective), and reflecting on the successes and ongoing challenges of local healthcare provision.

BACKGROUND

'Global health' (GH) has been described as 'an area of study, research and practice that places a priority on improving health and achieving equity in health for all people worldwide'. In 2015 this term was broadened to include 'planetary health', thereby encompassing the importance of the natural environment in supporting human health and wellbeing.

Significant improvements have been seen in many indices of GH (such as under-5 mortality), but momentum has been lost in recent years due to the effects of the COVID-19 pandemic, economic crises, and war.

In 2015, the United Nations General Assembly adopted 17 Sustainable Development Goals (SDGs) to replace the Millennium Development Goals. These offer a framework for ongoing efforts to improve global and planetary health (see **Fig. 10.1**).

Overview

Complex individual and community health needs are present in high- as well as low- and middle-income countries (LMICs). These may be particularly acute in times of crisis, for example, during or after armed conflict. Effective delivery of GH requires interprofessional collaboration between healthcare workers, politicians, economists, and scientists in pursuit of both individual wellbeing and population-based prevention and care.

Examples of successful interventions include campaigns for the provision of vaccines by the Global Alliance for Vaccines and Immunization (GAVI) and initiatives to reduce the economic exploitation of child labour (International Labour Organization [ILO]). The African Programme for Onchocerciasis Control (APOC) has transformed the lives of millions of people through the administration of a single, annual dose of the drug ***ivermectin***.

The global scale of the problem

Average global life expectancy has increased by over 6 years between 2000 and 2019 (66.8–73.4 years) but healthy life expectancy has not increased by the same level (58.3–63.7 years). In some countries, particularly in LMICs, life expectancy has actually decreased following years of economic austerity and the COVID-19 pandemic.

Some urgent global challenges:

- Around 5 million children under the age of 5 die each year; most deaths are preventable.
- Preterm birth complications account for >1 million deaths per year.

- Each day, 800 females die from preventable causes related to pregnancy and childbirth.
- Heart disease and stroke are the leading causes of death.
- Infectious diseases, including respiratory infection and tuberculosis (TB), as well as diarrhoeal disease, are significant contributors to mortality.
- An estimated 38.4 million people (2021) live with human immunodeficiency virus/acquired immunodeficiency syndrome (HIV/AIDS, 65% of the estimated 650 000 deaths from HIV/AIDS occurred in the World Health Organization [WHO] African region).
- Mental health disorders, for example, anxiety and depression, are main causes of ill health and disability worldwide.
- Tobacco use kills more than 7 million people each year and 80% of users live in a LMIC.
- The prevalence of obesity has nearly tripled since 1975. More than 1 billion adults, including 39 million children below 5 years of age and 340 million adolescents, were obese in 2022.
- The number of people living with diabetes is increasing. Approximately 1.5 million deaths were due to diabetes mellitus in 2019.
- Nearly 1.35 million people die from road traffic accidents every year.

Further reading

HM Government. *Global Health Framework: Working Together Towards a Healthier World.* 2023.
Frenk J, Moon S. Governance Challenges in Global Health. *N Engl J Med* 2013; 368:936–942.
Koplan JP, Bond TC, Merson MH et al. Towards a common definition of global health. *Lancet* 2009;373:1993–1995.
Whitmee S, Haines A, Beyrer C et al. Safeguarding human health in the Anthropocene epoch: report of The Rockefeller Foundation–Lancet Commission on planetary health. *Lancet* 2015;386:1973–2028.

GLOBAL HEALTH CHALLENGES

The global burden of disease

The Global Burden of Disease (GBD) study is a collaborative effort of over 5000 researchers in over 200 countries, coordinated by the Institute for Health Metrics and Evaluation (IHME). It provides regular estimates on mortality and disability-adjusted life years (DALY). DALY summarize years of life lost and years lived with disability as a result of different diseases. These are further described in Chapter 9.

This metric allows comparison of the burden of disease across society, including mortality, morbidity, disability, injuries, and risk factors. While DALYs can be affected by demographic changes, especially by changing population age profiles, they are very useful when comparing health in different populations.

While infectious diseases remain a major global disease burden, the latest iteration of the GBD study (2019) demonstrated a continuation of the global transition of health burden to non-communicable disease.

There is a global mismatch in the use of funds, with most research and development going to countries with low disease burden. Classically described as the '10 to 90 gap', with 90% of resource going to countries with only 10% of the disease burden (Fig. 10.2), it has been re-evaluated according to disease areas, but the main message remains valid.

Infectious diseases and pandemics

With more than 10 million cases and around 1.5 million deaths globally, TB causes the largest single-agent burden of disease and is the 13th most common cause of death overall. Deaths from diarrhoeal diseases, particularly among children, are even more common (1.53 million annually).

As demonstrated by numerous recent epidemics, including the recent COVID-19 pandemic, the world remains vulnerable to pathogens with pandemic potential. The new WHO Hub for Pandemic and Epidemic Intelligence is an example of new infrastructure that is being developed to mitigate future pandemic risk.

Maternal and child health

Maternal health

Although maternal mortality dropped by about 44% between 1990 and 2015, approximately 800 females die from preventable causes related to pregnancy and childbirth every day, 95% of these in LMICs.

Of major concern, 287 000 females died in childbirth in 2020, 95% in LMIC and most of these deaths were avoidable (WHO). The risk of maternal mortality is highest in adolescent girls under 15 years old. The major medical causes of death are haemorrhage, hypertensive disease of pregnancy, and infections. Hypertensive disease can be complicated by pre-eclampsia (see Chapter 48). Infections can be caused, for example, by bacteria, HIV, malaria, and syphilis. Female genital mutilation (FGM), illegal in most countries, may also lead to

Fig. 10.1 **The 17 Sustainable Development Goals.**

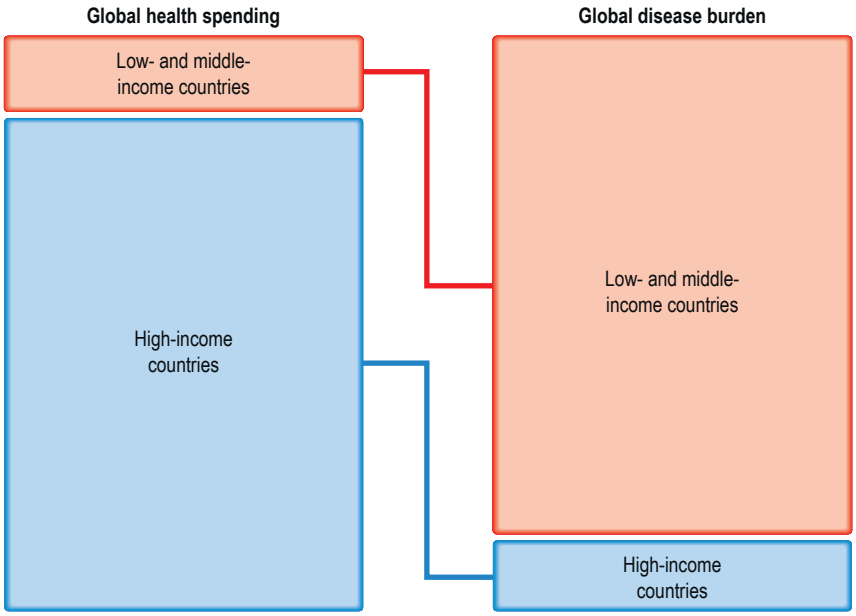


Fig. 10.2 Mismatch between health spending and resources. (*From Gottet P, Schieber G.* Health Financing Revisited: A Practitioner's Guide. *Washington: World Bank; 2006.*)

obstructed labour, with potentially fatal consequences. As most maternal and neonatal deaths occur around the time of delivery, the need for basic and emergency care before and during labour and delivery is paramount. The presence of a skilled birth attendant has been shown to reduce both maternal and neonatal mortality in LMICs. Sadly, this service is unavailable to approximately half of the pregnancies that are unplanned or unwanted, especially in adolescents.

Child health

Globally, the under-5 years mortality rate, especially in sub-Saharan Africa and Central and South-east Asia, has decreased (from 93 to 38 per 1000 live births between 1990 and 2021, UNICEF) mainly due to the prevention of pneumonia, diarrhoea, and malaria. Most of these deaths could have been prevented by access to simple and affordable interventions, including safe drinking water and vaccinations. About 45% are linked to malnutrition.

Children die from common, otherwise non-fatal, childhood ailments. Mortality figures in children under 5 have also dropped from 12.8 million in 1990 to 5 million in 2021 (WHO). Priority actions to mitigate this, advocated by the WHO, include the availability and accessibility of high-quality care in the ante-, peri-, and postnatal periods.

Vaccination

The World Health Assembly produced a framework to prevent millions of deaths by more equitable access to vaccines and developed a 'Decade of Vaccines Global Vaccine Action Plan 2011–2020'. GAVI (Global Alliance for Vaccines and Immunisation) plays a critical role in this area by financing and facilitating the delivery of vaccine platforms. By 2019, 86% of children had received the DPT vaccine (WHO) but immunization coverage decreased to 81% following the COVID-19 pandemic.

Child labour

The global number of children engaged in child labour declined by 38% between 2012 and 2017 (246 million to 152 million). Forty-eight percent of these children (73 million) were engaged in hazardous work, mainly in agriculture. More recently, the number of children in child labour has risen to 160 million (2021). Of these, approximately 9 million were forced into child labour by the COVID-19 pandemic (ILO).

Mental health

Mental health conditions or substance misuse disorders are estimated to cause the loss of 418 million DALYs globally (16% of global DALYs) and are on the increase. They are a leading cause of long-term disability (WHO).

Depression and anxiety disorders remain major clinical challenges. Approximately 5% of the adult population experience depression worldwide. Suicide is a very common cause of death among those aged 15 to 29 years (700 000 deaths annually). Over three-quarters of these deaths occur in LMICs.

The most important risk factor for suicide is a previous attempt, and unsuccessful suicide attempts are 15 to 20 times that of actual suicides. This offers the opportunity to implement effective mental health interventions. The COVID-19 pandemic has added to the external stressors, including conflict, forced migration, and poverty, which have resulted in increased anxiety, depression, and suicide. Children and adolescents have been disproportionately affected.

The global incidence of dementia is rising, and 53 million people live with this condition (2023). About two-thirds of patients live in LMICs. This creates a substantial social and financial burden. It is therefore an important cause of disability and death as well as a WHO-recognised public health priority.

Surgical challenges

Accidents

Approximately 4.4 million people die each year as a result of accidents and violence; 1.3 million result from road traffic accidents, almost all of these (93%) occurring in LMICs (WHO). In children and young adults (aged 5–29 years), road traffic accidents, homicide,

Box 10.1 International guidelines for essential trauma care ('rights of the injured')

- Obstructed airways cleared and maintained
- Impaired breathing supported until the injured person is self-ventilating
- Pneumothorax and haemothorax promptly relieved
- Bleeding stopped promptly
- Shock recognized and treated with intravenous fluid replacement
- Traumatic brain injury treated with timely decompression of space-occupying lesions
- Intestinal and other abdominal injuries promptly addressed
- Disabling extremity injuries corrected
- Unstable spinal cord injuries managed appropriately with early immobilization
- Appropriate rehabilitative services available
- Medications for the above and for pain control readily available

(Adapted from Mock C, Joshipura M, Goosen J et al. Overview of the Essential Trauma Care Project. World J Surg *2006; 30:919–929.)*

and suicide are among the most common causes of death. The WHO estimates that, by 2030, trauma from road traffic accidents will be the third most common cause worldwide of both mortality and disability (measured in DALYs).

Trauma

The Guidelines for Essential Trauma Care (WHO 2004) established a core list of 11 essential trauma services (**Box 10.1**). Implementation of these recommendations has been hampered by deficiencies in planning and infrastructure. These need to be addressed by national governments. Recent (2018) evaluation of these guidelines suggests that they remain appropriate recommendations for trauma system development.

Surgery

Around 5 billion people lack access to safe and affordable surgical and anaesthesic care when needed, especially in LMICs. Many conditions need swift surgical intervention. Less than 80% of the population in many countries can access 'bellwether procedures' (Caesarean section, laparotomy, treatment of open fracture) within 2 hours. The Lancet Commission on Global Surgery estimated that 16.5 million deaths from surgically treatable conditions occurred in 2010.

Oral disease

Oral pathology affects 45% of the world's population (3.5 billion people). From 1990 to 2023, case numbers have been increasing and now exceed, by over 1 billion, the combined numbers of the five major NCDs (Global Oral Health Status Report 2023). The remit of oral health has now been incorporated into the aspirations for universal health coverage by 2030.

Further reading

GBD 2019 Diseases and Injuries Collaborators. Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet* 2020; 396:1204–1222.
Collins PY, Patel V, Joestl SS et al. Grand challenges in global mental health. *Nature* 2011; 475:27–30.
Lucero-Prisno III DE, Shomuyiwa DO, Kouwenhoven MBN, et al. Top 10 public health challenges to track in 2023: Shifting focus beyond a global pandemic. *Public Health Challenges* 2023; 2:e86.
Meara JG, Leather AJM, Hagander L et al. The Lancet Commission on Global Surgery. Global Surgery 2030; evidence and solutions for achieving health, welfare, and economic development. *Lancet* 2015;386:569–624.

Box 10.2 The nine main targets of the Sustainable Development Goal for Health (Goal 3)

1. By 2030, reduce global maternal mortality ratio to less than 70 per 100 000 live births.
2. End preventable deaths of newborns and children under 5 years of age.
3. End the epidemics of AIDS, TB, malaria, and neglected tropical diseases.
4. Reduce premature mortality from non-communicable diseases by one-third while promoting mental health and wellbeing.
5. Strengthen the prevention and treatment of substance abuse.
6. Ensure universal access to sexual and reproductive healthcare.
7. Substantially reduce the number of deaths and illnesses from hazardous chemicals and air, water, soil pollution and contamination.
8. By 2020, halve the number of deaths and injuries from road traffic accidents.
9. Achieve universal health coverage including access to essential medicines and vaccines.

SOCIAL DETERMINANTS OF HEALTH

The drivers of health inequities reside in the social, economic, and political environments (see Chapter 9). The WHO defines the social determinants of health as 'the conditions in which people are born, grow, work, live and age'.

The health and illness of populations is significantly determined by individual, hereditary, and lifestyle risk factors. These are shaped by the distribution of money, power, and resources at global, national, and local levels. The social gradient of health follows the socioeconomic pattern from the top to the bottom: the lower the individual is within their socioeconomic position, the worse their health. This is seen in HICs, as well as in LMICs. The socioeconomic status of a person is their social position in society, as determined by their education, income, and occupation. There is now clear evidence to justify national policies that aim to reduce health inequity and the health divide across all countries. The reduction in health inequities should become one of the main criteria used to assess the effectiveness of health systems and governments.

Poverty and food security

Maslow's hierarchy of human needs offers an approach to human development in which growth in higher personal domains (such as self-actualisation) is dependent on more basic needs being met, starting with the physiological: clean air, water, food, protection from cold, and shelter. Poverty and food security therefore have a direct and indirect bearing on human health. Extreme poverty (defined as <US$ 1.90 per day) decreased from 33% in 1990 to 10% of the world population in 2015 (including 385 million children). Efforts continue to reduce this incidence to less than 3% by 2030. The SDGs (SDGs 1, 2, 3, 12, 13, and 15; see **Fig. 10.1**) relate to this ambition. The nine key priorities relating to SDG 3 ('Good health and wellbeing') have a broad scope (**Box 10.2**).

Poverty, defined as living on less than US$ 5.50 per day, was seen in nearly half of the world population in 2018. This is increasing in sub-Saharan Africa and in conflict-affected states. Chronic illness and disease as well as associated healthcare expenditure can contribute significantly to poverty and extreme poverty. In 2019, over one billion people suffered catastrophic health spending (where health spending exceeded 10% of the total household budget, SDG 3.8.2).

Improving child nutrition remains a global imperative, as nearly half of the deaths in children under 5 years of age are attributable

to malnutrition. In this age group, according to data from UNICEF (2022), stunting affects 148 million children. This problem can be mitigated by appropriate interventions during pregnancy and increased access to food before the age of 2 years.

Poverty and food security illustrate the interdependence between global and planetary health. Globally, one-third of all food produced, equivalent to 1.3 billion tonnes, is wasted each year (WHO). Crop deterioration due to deficiencies in transport and storage is estimated to account for 25% of food wastage in LMICs. Food production is compromised when agricultural land is directed towards alternative priorities, such as industrial development. Conversely, unsustainable land exploitation for crop production can aggravate food security in the long term.

Education

Education has a major impact on the health of a nation. The more years that are spent in schooling, the better the health outcomes. 'Health literacy' describes an individual's ability to read and understand basic information to allow them to make decisions about their healthcare. It requires basic literacy and numeracy skills and can have a major impact on nutrition and the control of disease, for example, by encouraging the simple process of hand washing. Education helps to promote healthier lifestyles, both by improving nutrition and development and by reducing the risks associated with infectious diseases. As a result, unemployment falls while family and community wellbeing is improved.

Females have a crucial role in the welfare of their families and the development of a nation's good health by looking after the household and caring for the children, elderly, and sick.

Nearly one-third of females in LMIC begin childbearing in adolescence (≤19 years old). Forty percent of girls with a first birth at 14 years of age or younger will have a third birth before the age of 20. (UNFPA). These pregnancies have a high rate of deaths from complications of pregnancy and childbirth (see Chapter 48). Data on maternal and child mortality provide a strong argument for females to control all aspects of their own reproductive health. The cost to a country's economy of adolescent pregnancy, as a share of gross domestic product (GDP), can be as high as 30%. The education and welfare of females should be major issues in any society.

Water and sanitation

Poor water and sanitation are major causes of early mortality, particularly in children. Access to water and sanitation is a UN-recognised human right (SDG 6.1). Only 74% of the global population have access to 'safely managed drinking water services' and only 54% use 'safely managed sanitation services' (UNESCO, 2023). Almost 500 million persons collect their water from unprotected sources, including wells and ponds, rivers or streams (WHO). Where water is not directly available, it is usually collected by females, often from distant sources. This is a hazardous activity, as the females may be subject to physical attack.

A secondary effect of improved access to water is that the time saved in collecting water can be spent on income generation, food production, education, and activities that lead to social and health benefits.

Environmental degradation and pollution

In the context of conflict, the United Nations International Strategy for Disaster Reduction (UNISDR) defines environmental degradation as 'the reduction of the capacity of the environment to meet social and ecological objectives and needs'.

Environmental degradation leads to the deterioration of the quality of both biological and physical resources, such as air, water, and soil. This, together with the increase in human population, has resulted in the destruction of ecosystems and habitats, and approximately 68% falls in vertebrate wildlife populations over the last 5 decades. Fish stocks have been reduced by overfishing. Massive deforestation, for agricultural purposes, has led to the loss of carbon absorptive capacity, and the effects of climate change have contributed to forest fires and air pollution. Melting oceanic glaciers have caused a rise in sea level and widespread floods. All of these have affected vital ecosystems, for example, pollinators, pests, and soil organisms, which are all essential for food production and agriculture.

Fresh water scarcity, particularly in LMICs, is a serious problem. One in three of the world population face water shortages. This is aggravated by pollution through the use of domestic and industrial sewage, fertilizers, pesticides, and chemicals.

International agreements, for example, the yearly Conference of the Parties (COP) meetings, develop approaches to environmental protection and resource management including those on deforestation, climate temperature, and preservation of cleaner air and water. Despite advances in environmental technologies (e.g., carbon capture, wind farms), political progress has been slow in agreeing and meeting targets.

Climate change

Climate change poses major threats to human health either directly (heatwaves, floods, fires, droughts) or indirectly (agricultural losses, mass migration); for example, a 1°C rise in mean temperature in India would result in the loss of 7 million tons of wheat (Fig. 10.3). The annual United Nations Climate Change Conference (COP) reviews progress in addressing climate change. The 2015 Paris accord (COP 21), ratified by 195 parties, agreed to target a reduction in global temperature rise to below 2°C by 2100. The 2022 report of the Intergovernmental Panel on Climate Change (IPCC) emphasized the urgent need to further reduce global warming and limiting climate change to a 2°C rise in global temperatures (Fig. 10.4).

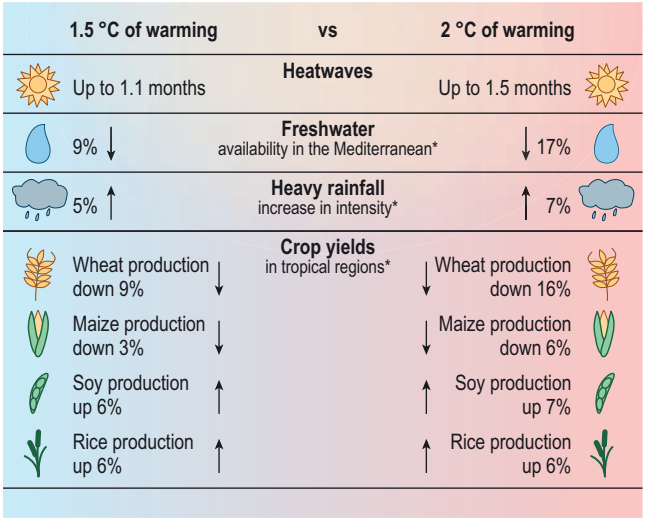


Fig. 10.3 The effects of global warming. *(All data from Schleussner CF, Tabea K, Lissner EM et al. Differential climate impacts for policy-relevant limits to global warming: the case of 1.5°C and 2°C.* Earth Syst Dynam *2016; 7:327–351.)*

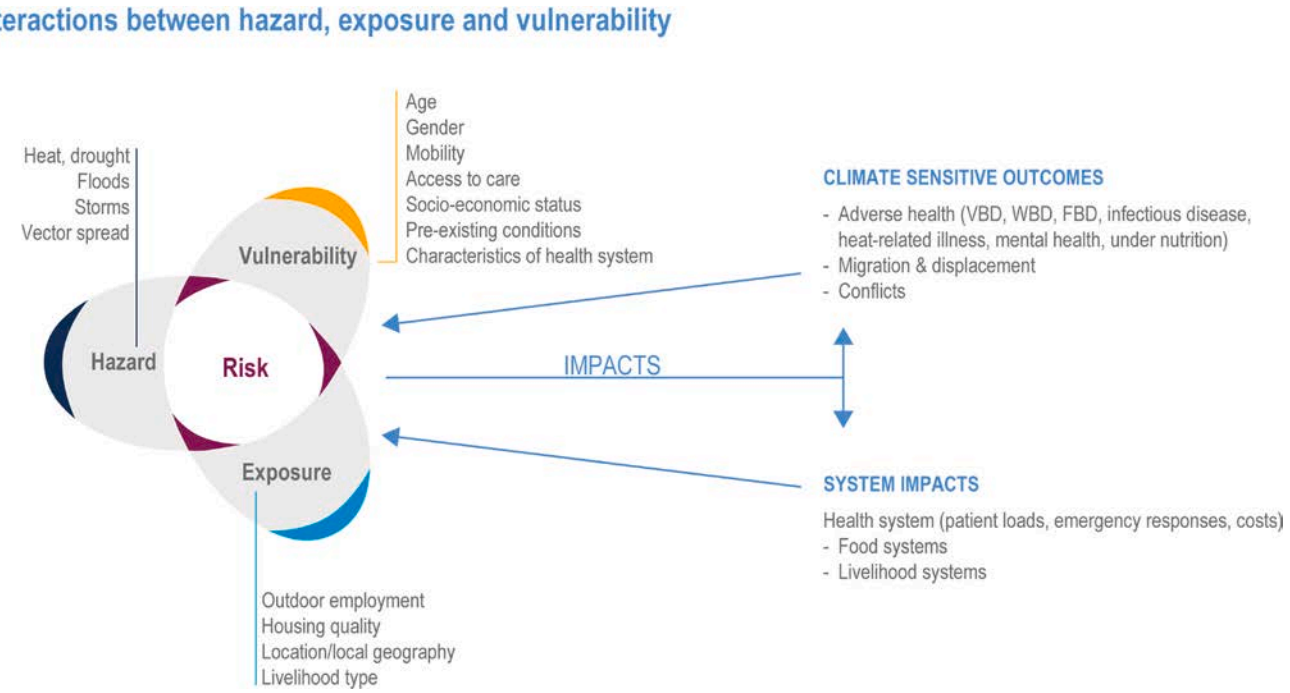


Fig. 10.4 Interactions between hazard, exposure, and vulnerability that generate impacts on health systems and outcomes, with selected examples. *FBD*, Food-borne disease; *VBD*, vector-borne disease; *WBD*, water-borne disease.

Human migration

Migration, a well-recognised social determinant of health, has been a key global phenomenon throughout human history. It continues to serve as important survival mechanism, in response to adverse environmental, economic, or social factors. Migration is likely to increase with climate change and the impact of local challenges (e.g., food shortages).

An estimated 281 million persons lived outside their country of origin in 2020. With global population growth, the number of migrants has increased significantly over the last decades (IOM). Individuals experiencing forced migration (refugees and asylum seekers) have significantly increased to 108 million (UNHCR, 2022). The health needs of migrants are dependent on their route of migration, legal status and access to services in the country of arrival, and wider socioeconomic, cultural, and linguistic factors. In many countries, migrants' health needs are not adequately met.

CONFLICT AND CATASTROPHE

Recent years have seen a number of natural disasters. Examples include earthquakes in Nepal (2015) and Indonesia (2018), hurricanes in the Caribbean (2019), flooding in Pakistan (2022), and forest fires worldwide. Conflicts persist in Ukraine, the Middle East, Ethiopia, and numerous other parts of the world. Effective global action following these unexpected events is frequently compromised by deficiencies in the disaster response. These deficiencies can be mitigated with appropriate training and long-term structural planning.

Conflicts and catastrophes severely disrupt healthcare provision, especially for females and children. The use of rape as a weapon of war magnifies this tragedy. Disruption to education and the physical and psychological sequelae of conflict lead to long-term societal healthcare burdens, especially in mental health.

Further reading

Donkin A, Goldblatt P, Allen J et al. Global action on the social determinants of health. *BMJ Global Health* 2018; 3:e000603.
Pörtner H-O, Roberts DC, Tignor MMB et al., eds. *Climate Change 2022: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC).* 2022. Available online at www.ipcc.ch.
Chaaban J, Cunningham W. *Measuring The Economic Gain of Investing in Girls: The Girls Effect Dividend. Policy Research Working Paper 5753:* World Bank 2011; available at: http://documents.worldbank.org/curated/en/730721468326167343/Measuring-the-economic-gain-of-investing-in-girls-the-girl-effect-dividend.
Maslow AH. A theory of human motivation. *Psychol Rev* 1943; 50:370–396.

GLOBAL AND PLANETARY HEALTH GOVERNANCE

Organizations

International support for GH initiatives may be directed at disease-specific projects (vertical care models) or aligned to national health system development. A number of organizations (**Box 10.3**) address the key areas for health development and infrastructural support, including the response to emergency situations (war and natural disasters).

These areas encompass service delivery, patient care, education and training, research, equipment, medicines, and human resources. Investment in healthcare systems promotes public health benefits by the provision of vaccination, hygiene, and sanitation, as well as major infrastructural projects.

The WHO is a specialized agency of the United Nations with a remit for international public health. Its constitution enshrines 'the highest attainable standard of health' as a fundamental right

Box 10.3 Some organizations involved in global aid

International organizations
- World Health Organization (WHO)
- Global Fund to fight AIDS, TB and Malaria
- Joint United Nations Programme on HIV/AIDS (UNAIDS)
- World Bank
- Department for International Development (DFID; UK)
- President's Emergency Plan for AIDS Relief (PEPFAR; USA)
- Global Alliance for Vaccines and Immunization (GAVI)

Scientific and educational organizations
- Mainly disease specific, e.g., microbiology, diabetes, infectious diseases
- Consortium of Universities for Global Health (CUGH)
- Global Health through Education, Training and Service (GHETS)

Advocacy and policy organizations
- Centre for Strategic and International Studies (CSIS)
- Global Health Policy Centre
- Global Alliance for Chronic Diseases (GACD)
- Global Health Council
- Global Health Technologies Coalition (GHTC)
- The Earth Institute (Columbia University)
- Kaiser Family Foundation (KFF)
- US Global Health Policy
- Tropical Health and Education Trust (THET)

Foundations
- Bill and Melinda Gates Foundation
- Wellcome Trust
- UN Foundation (UNF)
- Accordia Global Health Foundation

Other non-governmental organizations
- Save the Children
- United Nations International Children's Emergency Fund (UNICEF)
- Red Cross
- Oxfam
- International Rescue Committee (IRC)
- Médecins Sans Frontières (MSF)

of every human being. It has occasionally adopted legally binding health regulations, using its member state representatives through the World Health Assembly. An example is the Framework Convention on Tobacco Control (2003), ratified by 177 countries. This created binding norms for the demand and supply of tobacco products.

Sustainable Development Goals

The 17 SDGs (**Fig. 10.1**) emphasize the fact that achievement of the good health and wellbeing goal (SDG 3) is inextricably linked to other major factors including politics, economics, and agriculture. The right to health contains four elements: *availability* (of programmes of public health), *accessibility* (of health facilities and services in a non-discriminatory fashion), *acceptability* (ethical and cultural requirements), and *good-quality care*.

Economics and politics in global and planetary health

Funding for initiatives in the GH development of LMIC (also known as Development Assistance for Health, DAH) has, historically, originated from multiple sources. DAH has supported interventions in specific diseases such as malaria (see Chapter 42), TB (Chapter 42), and HIV/AIDS (Chapter 44). In contrast, an ongoing funding gap may have contributed to the slow response of the WHO and others to the 2014 outbreak of Ebola virus in Africa. Transnational collaboration is especially necessary in the face of pandemic threats.

Improvements in healthcare commonly result in economic growth. For example, a 10% reduction of malaria in endemic areas is associated with a 0.3% increase in GDP. Treatment of HIV-positive patients with anti-retroviral drugs results in net economic benefits through increased productivity and a reduction in the costs of medical care. The cumulative loss of economic productivity between 2015 and 2030, in the absence of a significant scaling up of global surgical services, is estimated at US$ 12.3 trillion. Any failure to invest in health and health systems is a threat to future global prosperity, particularly in poorer countries.

Political decisions regarding investment in healthcare and health systems must also focus on infrastructure, including food and agriculture, the environment and human rights issues, especially the rights of females. The pursuit of 'pro-poor' policies that place the poor at the centre of development policy promotes global prosperity.

Human rights and the value of engagement in global health

The Universal Declaration of Human Rights (1948), while not legally binding, serves as a 'common standard for all peoples and all nations'. It has given rise to two new legally binding covenants: the International Covenant on Civil and Political Rights and the International Covenant on Economic, Social and Cultural Rights. Additionally, the rights of children and females and the Right to Health are enshrined in global human rights legislation. This international body of law with its declarations and conventions is regularly consulted to protect individuals and populations.

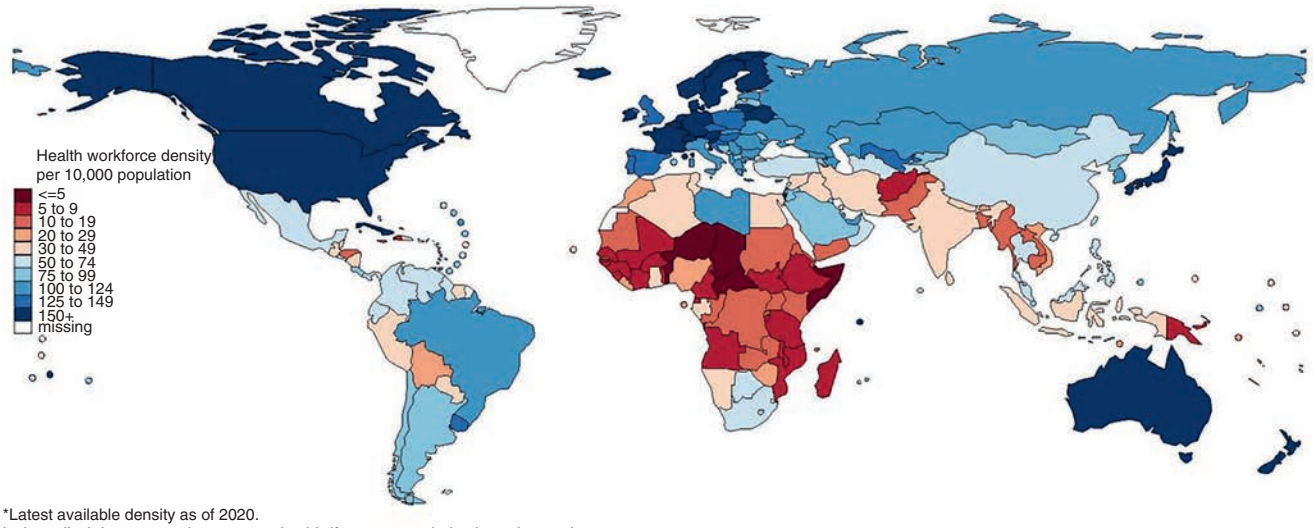
The healthcare workforce and international recruitment ethics

There is a worldwide shortage and highly unequal global distribution of healthcare workers from all disciplines (**Fig. 10.5**). Current estimates suggest that an additional 10 million healthcare workers will be required by 2030 to meet global demand. This shortage has been accentuated, to the disadvantage of LMIC, when healthcare workers migrate to wealthier countries for economic reasons. The WHO Global Code of Practice (2010) provides a framework for best practice and monitoring of global healthcare worker recruitment.

Medical electives

Engagement in GH challenges helps to promote patterns of behaviour that benefit healthcare workers and their recipients. These behaviours include altruism, team-working, and appreciation of cultural diversity.

Many medical students and doctors undertake periods of time visiting and working in unfamiliar environments, whether in their home country or abroad. This is mutually beneficial to all participants, provided that a culture of shared learning is embraced. All such visits, commonly termed 'medical electives', should have clear objectives and measurable educational outcomes. They should



*Latest available density as of 2020.
incl. medical doctors, nursing personnel, midwifery personnel, dentists, pharmacists

Fig. 10.5 Worldwide shortage of healthcare workers. There are 57 countries with a critical shortage of 2.4 million health service providers (doctors, nurses, and midwives). Africa has 25% of the world's healthcare burden and 1.3% of the providers. (*From Bonoi M, Kunjumen T, Nair TS et al. Fig. 2. The global workforce stock and distribution in 2020 and 2030: a threat to equity and 'universal' health coverage?* BMJ Global Health *2022;7:e009316. Doi:10.1136/bmjgh-2022-009316.)* Map lines delineate study areas and do not necessarily depict accepted national boundaries.

not compromise the working arrangements for host organizations. Pre-departure preparation should include consideration of culture, ethical challenges, and security issues. All visitors must work within their approved competencies, be appropriately supervised, and comply with national guidance on good medical practice. Maximum benefit from medical electives is achieved when there is a mutual commitment to long-term partnership.

Further reading

Resch S, Korenromp E, Stover J et al. Economic returns to investment in AIDS treatment in low and middle income countries. *PLoS One* 2011; 6:e25310.
Bustreo F, Okwo-Bele JM, Kamara L. World Health Organization perspectives on the contribution of the Global Alliance for Vaccines and Immunization on reducing child mortality. *Arch Dis Child* 2015; 100:S34–S37.
International Labour Organization (ILO). *Ending Child Labour by 2025: A Review of Policies and Programmes.* Geneva: ILO 2017.

Significant websites

https://www.healthdata.org/research-analysis/gbd *Institute for Health Medtrics and Evaluation Global Burden of Disease study.*
https://pandemichub.who.intv *World Health Organization hub for pandemic and epidemic intelligence.*
https://www.bps.org.uk/psychologist/maslows-hierarchy-needs-sixth-level *British Psychological Society, Maslow's Hierarchy of Needs.*

Bibliography

Crisp N. *Turning the World Upside Down Again: Global Health in a Time of Pandemics, Climate Change and Political Turmoil*. 2022.
Global Health through Education, Training and Service (GHETS). *Health Information for All.* https://www.hifa.org/support/supporting-organisations/global-health-through-education-training-and-service-ghets (accessed Aug 21, 2023).
Grimm M, Sipangule K, Thiele R, Wiebelt M. *Changing Views on Growth: What Became of Pro-Poor Growth? PEGNet Policy Brief, No. 1/2015.* 2015. https://www.econstor.eu/bitstream/10419/122093/1/pegnet-policy-brief-2015-01.pdf.
Kumar P, Sethia B. editors. *Essentials of Global Health—1st Edition.* 2018.
Marmot M, Allen J, Bell R et al. World Health Organization. European review of social determinants of health and health divide. *Lancet* 2012; 38:1011–1029.
Marmot M, Allan J, Boyce T, Goldblatt P, Morrison J. *Health Equity in England: The Marmot Review 10 Years On—The Health Foundation.* https://www.health.org.uk/publications/reports/the-marmot-review-10-years-on (accessed Aug 21, 2023).
Romanello M, Napoli CD, Drummond P et al. The 2022 report of the Lancet Countdown on health and climate change: health at the mercy of fossil fuels. *Lancet* 2022; 400: 1619–1654.
Tropical Health and Education Trust (THET). Working in partnership to support health workers across the world. *THET*. https://www.thet.org/ (accessed Aug 21, 2023).
United Nations. *Universal Declaration of Human Rights.* United Nations. https://www.un.org/en/about-us/universal-declaration-of-human-rights (accessed Aug 21, 2023).

Questions for this chapter can be found on Elsevier eBooks+.

11 Acute medicine

Nicholas Robert Murch, Barbara Lachana Onen, John Booth, John Patrick Donaghy, and Laura Kay Harrington

Background 175
Patient flow through AIM 176
Patient-centred care and admission avoidance 177
Multidisciplinary care model and approaches to effective patient flow 178
In-reach to ED: enhancing acute patient care 178
The role of the AIM clinician 178
Clinical skills 179
The initial patient assessment: the 'medical clerking' 179
Risk assessments for common acute medical conditions 180
Communicating decisions and referrals 180
Discharge planning 182
Liaison with acute care MDT 182
Procedures 182
Core clinical presentations in AIM 192
Cardiac arrest 192
Anaphylaxis 198
Chest pain 198
Palpitations 201
Breathlessness 204
Haemoptysis 218
Nausea and vomiting 220
Diarrhoea 220
Haematemesis and melaena 223
Abdominal swelling 226
Abnormal blood results 227
Acute kidney injury 228
Hyperkalaemia 231
The unconscious patient 232
Dizziness and blackouts 236
Headache 239
Seizures 249
Delirium 250
Functional neurological disorder 253
Polysymptomatology 254
Swollen joint 254
The unwell diabetic patient 257
Deep vein thrombosis and pulmonary embolism 264
The unwell cancer patient 272
Rash 274

CORE SKILLS AND KNOWLEDGE

Acute internal medicine (AIM) is a relatively new hospital specialty but has grown rapidly to become the fulcrum of acute medical care within the modern hospital. The evolution of AIM has ushered in numerous paradigm shifts in models of acute care, with increasing emphasis on ambulatory and same-day emergency care, patient-centred solutions such as 'hospital at home', joined-up working with allied health professionals and across the primary and secondary care interface, and effective patient flow both within the hospital and to/from the community.

Acute physicians are the hospital's 'specialist generalists'. Key skills in AIM include:

- Understanding the hospital pathways and infrastructure needed to help navigate each patient smoothly and effectively through their healthcare journey
- Acquiring a broad range of clinical procedure skills required in the investigation and treatment of medical emergencies
- Mastering the focused 'problem-based', symptom-driven patient assessments required to assign a working diagnosis, decide the appropriate location of ongoing care, and determine initial management
- Managing a broad range of medical emergencies which span the entire spectrum of subspecialty medicine.

Opportunities for learning include shadowing the acute medicine 'on-call' team and taking the opportunity to 'clerk' patients with a broad range of presenting symptoms, following ward rounds and joining multidisciplinary team meetings on the acute medical unit and, where possible, undertaking 'simulation training' to allow exposure to a range of clinical scenarios and procedures in a safe, controlled and supported learning environment.

BACKGROUND

Acute internal medicine (AIM) is a relatively new specialty that has grown out of general internal medicine (GIM), with a recognition that the acute care of medically unwell patients, and the interface with primary care, is a specialty in its own right that needs dedicated resources and a bespoke training programme. Wards such as the acute medical unit (AMU) were established, with specialist AIM in-reach into the emergency department (ED), and outreach to medical, specialty, and surgical wards.

AIM has been evolving since its inception three decades ago, and is rapidly growing in the UK, as well as in those nations that model their medical system on the National Health Service. The 'hospitalist' movement has emerged in parallel in the USA with a similar focus on acute inpatient medical care. One major difference is that hospitalists often look after patients for their whole inpatient

stay, whereas in AIM, the tendency is to diagnose and safely discharge in a timely fashion, or, when appropriate, refer to longer stay specialties, including GIM.

The AIM model revolves around intensive senior clinician-led care, often including twice daily review, improving efficiency and reducing length of stay for medical admissions. Additional roles have been absorbed into the specialty including ambulatory care, same day emergency care (SDEC), acute frailty hubs, virtual wards, and 'hospital at home' services, with AIM physicians often working closely with other specialties such as geriatric medicine. Increasingly, enhanced care units (ECUs), with higher ratios of nursing care, monitored beds, and the ability to provide advanced therapies including respiratory or basic inotropic support, may be co-located on the AMU. As such, AIM finds itself at a key intersection between primary care, EDs, critical care, and specialty services including surgery (**Fig. 11.1**).

AIM encompasses a large portion of the medical curriculum, and the AMU has become a valuable area for teaching and learning for both students and professionals from across the interprofessional team. The acute medical service is where the initial contact with patients often occurs, affording the opportunity to formulate *de novo* differential diagnoses and management plans, hone diagnostic skills, and improve clinical acumen, and as such it is a popular placement for a range of healthcare students.

PATIENT FLOW THROUGH AIM

The AMU can be viewed as a 'header tank' that fills from primary care or ED, and then empties out through discharge pathways or transfers to other specialized wards including intensive therapy units (ITUs, see Chapter 13). Smaller 'header tanks' can be interconnected with the AMU, including SDEC (which is often co-located), virtual wards, or 'hospital at home'. Given the variety of patient presentations to AIM, medical specialty team in-reach or co-location with cardiology (Chapter 17), geriatric medicine (especially with frailty services, Chapter 49), stroke medicine (Chapter 29), and

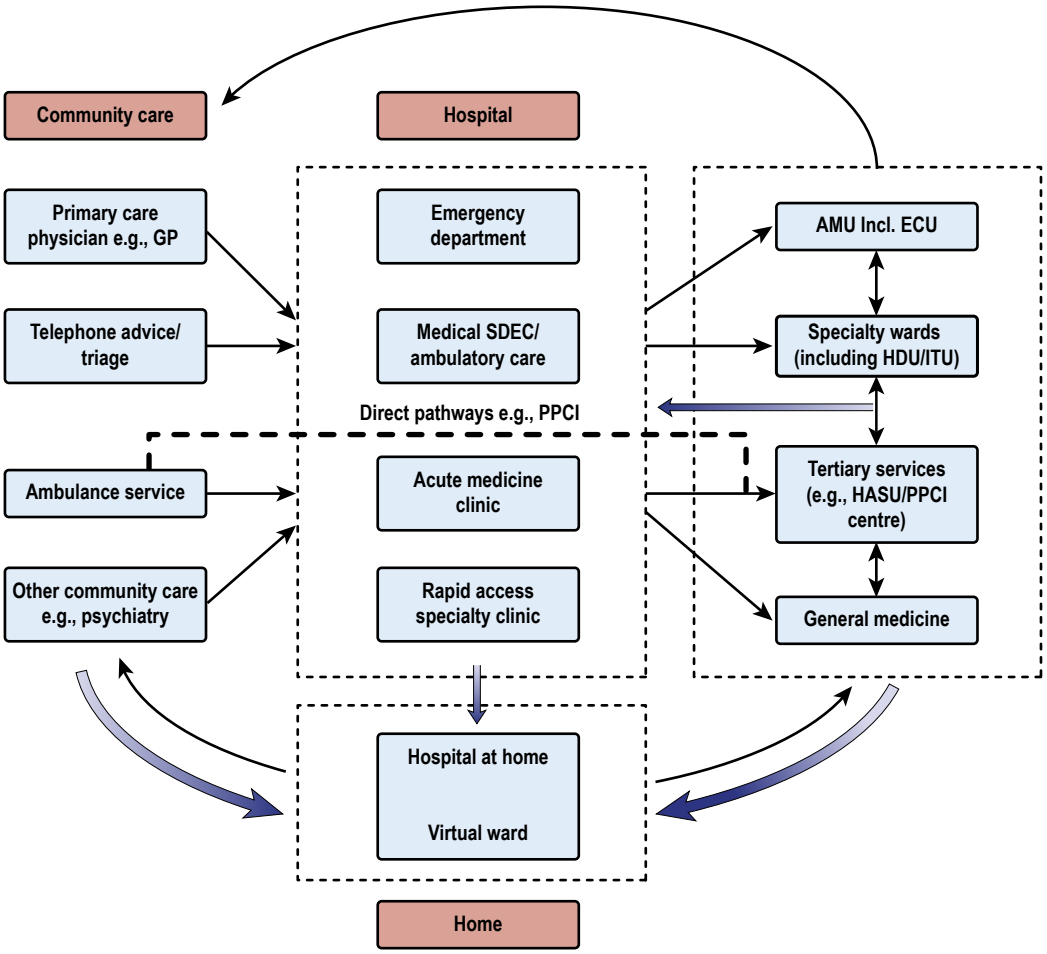


Fig. 11.1 Model of acute medical care. Patients enter acute care systems via a number of community routes, including GP consultation or direct contact with the ambulance service. The venue of acute medical assessment may be determined by triage (e.g., by discussion with an acute physician) or defined emergency care pathways (e.g., direct ambulance transfer to a Primary PCI centre for a patient with STEMI). Subsequent patient flow may be to a hospital ward (e.g., AMU or specialty ward), domiciliary treatment (e.g., 'hospital at home'), or back to community care, depending on the predicted care needs, individual circumstances of the patient and local resources. The most efficient contemporary models leverage close collaboration between primary and secondary care clinicians and allied health professionals and utilize bidirectional flow between the different silos of care. *HASU,* Hyperacute stroke unit; *PPCI,* primary percutaneous intervention.

respiratory medicine (Chapter 20) is common. By its nature, there needs to be sufficient capacity throughout the system to ensure that patient flow does not 'stagnate' and the AMU can continue to fill appropriately. Prioritization of investigations including blood taking and dedicated imaging, is important to ensure that flow is maintained. Ready availability of advice and specialty management pathways from other medical specialties (e.g., liaison psychiatry (Chapter 30)) is also critical to a successful AMU.

AMUs provide care for patients who are acutely unwell, therefore, their beds can have differing levels of staffing and observation or monitoring. Acute medical teams often help decide where patients are best placed depending on their needs and the resources available (**Box 11.1**). Infection control measures and patient and staff safety are always important in these decisions. Access to higher-level care is also important—AMUs work closely with teams in high-dependency units, ITU, and coronary care units and often collaborate in running ECUs embedded within the AMU. Each AMU is planned in response to the hospital's unique layout, staffing, resources available, and local needs; therefore, the physical configuration and clinical operation may vary greatly.

In the UK, The Royal College of Physicians (RCP) advises centralizing the care of patients either in their first 48 hours in hospital, or requiring more enhanced or higher dependency therapy. This should ideally occur within a shared hospital footprint of acute medical services comprising the 'acute care hub':

- ED
- SDEC and ambulatory care
- AMU (indicative length of stay up to 48 h)
- Short-stay beds (indicative length of stay up to 72 h)
- ITU.

Increasingly, AIM provides outreach to specialty wards, surgical liaison, and in-reach to ED to enhance patient-centred, focused care and break down the boundaries between 'silos' of care created by more traditional models of hospital operation.

Patient-centred care and admission avoidance

In recent years, there has been a global push to increase sustainability and alleviate the strain on acute healthcare systems generated by changing population demographics and increased demand on emergency care. Ambulatory care, SDEC, 'zero day admission conversions', and virtual wards/'hospital at home' models have emerged as credible alternatives to traditional inpatient medical care.

Box 11.1 Different levels of adult critical care

Level 0: Those patients whose needs can be addressed on a standard acute hospital ward with normal staffing.

Level 1: Those patients that are deemed to have the potential to deteriorate, so areas may have enhanced staffing ratios e.g., AMU. (Sometimes this may be labelled 'Level 1 plus' if it has bedside monitoring, and input from appropriate specialists including critical care outreach e.g., an ECU).

Level 2: Those patients, often with single organ failure, requiring a higher ratio of care, often one nurse to two patients (e.g., ECU or HDU/ITU).

Level 3: Those patients often with more than one organ failing, often requiring one-to-one nursing (e.g., ITU).

Ambulatory care and SDEC

Ambulatory care, often manifested as SDEC, enables selected urgent care scenarios that do not require significant time in a hospital bed to be made semi-elective, with dedicated clinic slots. Patients are able to return home in between attending an ambulatory service for daily or as-needed monitoring, investigations, or treatments, examples of which may include administration of intravenous antibiotics or fluids or subcutaneous injections such as low-molecular-weight heparin. The healthcare team assesses patients and those who clinically deteriorate may be admitted to hospital, while those who improve are discharged to their regular healthcare provider for follow-up.

Ambulatory/SDEC models have been shown to be cost-effective in high-income countries, reducing the burden on medical wards without compromising patient outcomes, while simultaneously improving the patient experience.

Virtual wards and 'hospital at home'

These models extend the care continuum beyond the physical boundaries of the hospital, allowing patients to receive treatment and monitoring traditionally associated with a hospital admission, in their own homes.

Virtual wards

Virtual wards operate through a combination of telemedicine, remote monitoring technology, and digital communication platforms. Healthcare professionals, including doctors and nurses, remotely monitor a patient's incoming data looking for signs of deterioration or changes that may require intervention. Based on the data and the patient's specific care plan, doctors provide medical advice, prescribe medication, and offer guidance on self-care techniques. During the COVID-19 pandemic, virtual wards proved valuable, playing a role in reducing the burden on hospitals and healthcare facilities. This prevented unnecessary hospital admissions, allowing hospitals to focus their resources on severe cases and emergencies. This system of care also minimized the risk of COVID-19 transmission to healthcare workers and patients. Other conditions potentially lending themselves to virtual ward management after careful patient triage include other acute respiratory infections, heart failure, acute kidney injury (AKI), and atrial fibrillation.

Digital communication platforms, such as video consultations or messaging applications, facilitate direct communication between patients and healthcare providers. A home visit, or patient visit to an SDEC facility can also be arranged for physical review if required during this period.

'Hospital at home'

'Hospital at home' provides an alternative to traditional inpatient care, allowing patients to receive hospital-style medical treatment and monitoring in their own homes by leveraging a combination of telemedicine, remote monitoring technologies, and skilled healthcare professionals. Patients are carefully selected based on specific criteria that ensure their safety and suitability. Patients then receive regular visits from a multidisciplinary team (MDT) of healthcare workers, which may include physicians, nurses, therapists, and other specialists/community teams, depending on the individual patient's needs. During these visits, necessary medical interventions are carried out, for example, administering treatments such as intravenous antibiotics, fluids, nebulizers or oxygen, conducting examinations, or adjusting medications.

Receiving care at home can reduce the stress and disruption associated with hospitalization. This is particularly helpful for frail elderly patients as it allows them to recover in a familiar environment and reduces incidence of delirium, hospital-acquired infections, and deconditioning associated with traditional hospital stays. The flexibility and convenience of home care can contribute to improved patient satisfaction and overall wellbeing.

Multidisciplinary care model and approaches to effective patient flow

In the fast-paced and complex environment of acute healthcare, effective patient flow and MDT working are crucial elements in ensuring high-quality care and optimal patient outcomes.

The multidisciplinary care model

The multidisciplinary care model involves a collaborative approach to patient care, where healthcare professionals from various disciplines work together to provide comprehensive and coordinated treatment. This team-based approach ensures that patients receive holistic care that addresses their medical, psychological, and social needs. It encompasses the participation of doctors, nurses, pharmacists, therapists, social workers, and other allied health professionals. By pooling their expertise and perspectives, the MDT can develop tailored care plans, improve patient outcomes, and enhance the overall patient experience.

Approaches to effective patient flow

Care pathways and standardized protocols

Implementing care pathways and standardized protocols helps to streamline patient flow by providing a structured and consistent approach to care. These pathways outline the optimal sequence of assessments, investigations, and interventions for specific patient conditions, promoting efficient and timely care delivery. Standardized protocols ensure that healthcare professionals follow evidence-based guidelines, reducing unnecessary variations in practice and enhancing patient safety.

Early senior review and decision-making

Early senior review and decision-making expedites formulation of definitive treatment plans, facilitates appropriate referrals, and enables early decisions to admit or discharge patients and thus optimize patient flow. This approach reduces delays and prevents bottlenecks in the care pathway.

Daily MDT meetings

Regular MDT meetings are effective for coordination and communication among healthcare professionals. These meetings provide a platform for sharing information, discussing complex cases, and aligning treatment plans. By collectively reviewing patient progress, identifying barriers to discharge, setting expected dates for discharge, and coordinating resources, the team can address any issues promptly.

Discharge planning and early allied health professional involvement

Discharge planning should commence **concurrently with admission** to the hospital. Early involvement of allied health professionals, such as physiotherapists, occupational therapists, and social workers, can help with assessment of patients' functional status, identify any potential discharge barriers, and develop appropriate and timely discharge plans.

Bed management and capacity planning

Efficient bed management and capacity planning help optimize patient flow in the acute setting. This involves monitoring and predicting bed availability, coordinating transfers, and ensuring appropriate allocation of resources. Transfer of patients to the 'right bed' with the 'right specialty' ensures the best care of patients from specialist teams and nurses. Proactively managing bed capacity and collaborating with the MDT can reduce delays, improve patient flow, and minimize long hospital stays.

In-reach to ED: enhancing acute patient care

In-reach programs facilitate rapid assessment and early intervention for patients with acute medical conditions. Direct involvement of specialists such as physicians, intensivists, and surgeons in patient care within the ED reduces delays in diagnosis and treatment, improves patient outcomes, and brings the timely expertise of different specialties directly to the patient's bedside. In-reach initiatives also ensure a smooth transition of patients from ED to specialized wards within the hospital and result in better handover to specialist teams and improved coordination of care. In-reach also helps in the **early identification of high-risk patients** who require immediate intervention or higher-intensity care such as ITU. AIM clinicians are trained to recognize deteriorating patients promptly and provide appropriate management.

THE ROLE OF THE AIM CLINICIAN

AIM physicians enjoy the diversity and breadth of medicine incorporated within the specialty: varied medical pathophysiology; diagnostic challenges; impactful quality improvement work; high-quality teaching and training in an enriched learning environment; caring for critically ill patients; working under pressure and using advanced communication skills; all while constantly being invigorated by new clinical, ethical, and academic challenges.

AIM physicians are 'specialist generalists' who thrive on multidisciplinary working in a busy, multidisciplinary, pathology-rich teaching environment. There is often a breadth of subspecialist skills within a single acute medicine team that AIM physicians acquire during higher specialty training, for example, echocardiography, education, leadership and management, obstetric medicine, and acute oncology. The AMU is a central hub encompassing disparate care pathways for patients and this has resulted in AIM physicians becoming holistic, highly skilled, and adaptable in their approach to patient care especially with regard to critical decision-making such as inpatient admission vs outpatient management and treatment escalation planning. The AIM physician leads and coordinates patient care across the acute unselected medical take including in-reach and support to ED, the AMU, SDEC, ambulatory care unit, and 'hospital at home' services.

CLINICAL SKILLS

The initial patient assessment: the 'medical clerking'

The 'medical clerking' is a critical process in the care of hospitalized patients, forming a cornerstone piece of documentation for the patient's admission (**Box 11.2**). The clerking will be referenced at any point of transition in the patient's journey through the hospital, such as transferring to a different ward, and is frequently the first clinical document that is reviewed by the out-of-hours team when a patient suddenly deteriorates. It is therefore vital that the clerking is a concise description of the patient, their acute presentation, and their background health.

The clerking involves the systematic collection of patient information in order to formulate a working differential diagnosis and rational strategy for investigation, plan admission when needed, and initiate thoughtful discharge planning. A well-executed clerking sets the stage for provision of high-quality patient care, facilitates efficient decision-making, and promotes effective communication between teams.

Gathering patient information

A detailed discussion of the medical interview, history-taking strategies, and physical examination can be found in Chapter 1. The resulting 'medical clerking' document should include the following domains:

- State the patient's ***presenting symptom:*** this should be a brief headline only e.g., 'chest pain', 'shortness of breath'. Then record a detailed history of presenting symptom noting relevant associated features or important 'negatives' (i.e., absent symptoms).
- Record a detailed background of the patient including:
  - ***Past medical and surgical history***
  - ***Current medications*** including over-the-counter medications and herbal remedies
  - ***Allergies and intolerances***
  - ***Family history***
  - ***Social history***
    - Include information on the patient's occupation, smoking (including vaping), alcohol history and recreational drug use (N.B. Cannabis is often not viewed as a drug, so may need to ask specifically), and, when relevant, travel and sexual history, and pets.
  - ***Home situation***
    - Include information on the patient's living conditions and function e.g., the patient may have a package of care in place to support their activities of daily living; the patient may have specific nutrition requirements such as thickened fluids or a soft diet.
- ***Review of systems:*** particularly useful when presenting symptoms are non-specific.
- ***Collateral history*** from an eyewitness (e.g., a family member or carer) is invaluable in scenarios where a patient is unable to communicate (e.g., coma, acute confusion, or pre-existing cognitive impairment) or experienced a period of reduced awareness (e.g., a blackout).
- Record the findings from a comprehensive ***physical examination***, focusing on pertinent findings related to the patient's presenting problem and underlying conditions.
- Review available patient data including ***previous medical records and investigations***, such as laboratory results and imaging studies. Depending on each hospital's record system, there can be a wealth of information to be gathered electronically, and it is important to consider which pieces of information are important to highlight at this point in the patient's admission. For example, if a patient is presenting with shortness of breath, it may be useful to know if they have had an echocardiogram in the last six months.
- Consider available ***point-of-care testing*** (POCT) to add to your initial assessment of the patient's presentation e.g., arterial/venous blood gas, electrocardiogram (ECG), urinalysis, point-of-care ultrasound (POCUS), forced vital capacity (FVC).

Formulating an admission plan

Identify and record the patient's primary problem(s) and establish working diagnoses, or a differential diagnosis if further targeted investigation is required.

- Consider the information gathered from the patient's background to determine their **Clinical Frailty Scale** (CFS) score (see **Fig. 49.2**). This is an important measure when considering further steps in treatment planning, as discussed further below.
- Certain conditions have established associated **risk scores** which should be calculated on admission to plan care (see section below: 'Risk assessments for common acute medical conditions').
- Formulate an **initial investigation and management plan**, addressing urgent clinical concerns and initiating targeted investigations and interventions.

Box 11.2 The admission medical clerking

Information gathering

History taking
- Presenting problem
- History of presenting problem
- Past medical and surgical history
- Allergies and intolerances
- Social history (occupation, alcohol, smoking, substance use, travel, and/or sexual history)
- Home circumstances (level of function, care provision, dietary requirements)
- Review of systems
- Collateral history if relevant

Physical examination

Medical data review
- Hospital records and clinic letters
- Previous investigations

Point-of-care testing (e.g., arterial or venous blood gas analysis, ECG, etc.)

Synthesis

Establish working diagnosis, or differential diagnosis to guide investigations and initial management

Determine Clinical Frailty Score

Apply established risk scores (e.g., GRACE score with ACS)

Actions

Initiate urgent treatment as appropriate

Request relevant diagnostic tests (e.g., CT, echocardiogram)

When admitting to hospital:
- Set goals of admission
- Consider level of care (**Box 11.1**)
- Plan frequency and type of observations and bedside measurements (e.g., CBG)
- Consider TEP and discussing CPR status
- Carry out VTE risk assessment in line with local policy

Set estimated discharge date and clinical criteria for discharge early in admission

ACS, Acute coronary syndrome; *CBG,* capillary blood glucose; *TEP,* treatment escalation plan; *VTE,* venous thromboembolism.

- Immediate **involvement of other teams** may be required e.g., critical care outreach for the critically ill patient, interventional cardiology for high-risk acute coronary syndromes; in other cases, referrals may be planned for later in the patient's admission e.g., endocrinology input for the unwell diabetic patient.
- Establish the **level of monitoring** and care required based on the patient's clinical condition (incorporating data from Fig. 13.1), planned nature and frequency of observations, and perceived risk of deterioration (see Box 11.1).
  - Within a ward environment, clearly stipulate the frequency of observations and other relevant clinical measurements e.g., urine output, pain scores, neurological observations.
- Set specific **goals of care** and discuss them with the patient and their family, ensuring shared decision-making.
- Consider discussing and documenting a **Treatment Escalation Plan (TEP)**. This should state the appropriate level of treatment for a patient should they deteriorate during their admission, taking into account the patient's wishes (including valid advance directives, or proxy opinions in cases where power of attorney for health has been decided), co-morbidities, and functional status. This includes considering appropriateness of cardiopulmonary resuscitation (CPR) in the event of cardiorespiratory arrest. For patients in whom CPR is determined not to be appropriate, it is important that a 'Do Not Attempt Cardiopulmonary Resuscitation' (DNACPR) decision is documented clearly. Every patient with a DNACPR decision should have an accompanying TEP.
  - DNACPR/TEP discussions can be challenging and require clear and compassionate communication in a safe and supportive environment for staff and patients. The point of admission, i.e., the medical clerking, is an opportunity to have these discussions with patients and their loved ones, though it will not always be the most appropriate or optimal time to do so. It is reasonable to delay this discussion until later in the admission if it is safe to do so.
  - Occasionally, patients may present to hospital with advanced, incurable disease associated with distressing symptoms that are not being adequately controlled at home. Recognizing the dying patient allows early involvement of specialist palliative care teams (Chapter 50) for effective symptom control, access to 'fast track' community support, or discharge to hospice services.
- Evaluate the patient's **risk of venous thromboembolism (VTE)** based on local risk assessment models and, when appropriate, prescribe mechanical or pharmacologic prophylaxis.

Additional considerations

Certain patients will have additional communication needs, for example, patients with learning disability; patients with dementia; patients with hearing impairment. In these circumstances every effort should be made to optimize communication for the patient and, when appropriate, an advocate should be sought to complement any history provided by the patient.

When a patient presents with acute cognitive impairment, it is important to establish their baseline cognition via an appropriate collateral history and quantify the degree of impairment at the point of admission with a suitable scoring system, for example, abbreviated mental test score (AMTS); or 4AT score (see Chapter 49).

When a language barrier is present between staff and patients, an independent translation service should be sought to obtain the information needed for the medical clerking, for example, an in-person interpreter or a telephone translation service.

Discharge planning

A proportion of patients seen by the medical admissions team will be planned for discharge on the day of admission, although many will require at least 24 hours of treatment and/or monitoring in hospital. For patients requiring admission, it is important to state their criteria for discharge (this may require the assistance of a senior AIM clinician).

The patient's presentation presents an opportunity to plan necessary investigations, although some of these investigations may be appropriately performed as an outpatient (e.g., investigations directed towards an incidental finding of microcytic anaemia). These investigations should be flagged in the admission plan to ensure that necessary referrals/follow-up arrangements are not subsequently overlooked.

Risk assessments for common acute medical conditions

The ability to estimate the risk that a patient presenting with a particular symptom or condition will experience an adverse outcome (e.g., death or disability) is a powerful tool to assist the AIM clinician in making decisions to admit to a hospital bed, facilitate early discharge or harness ambulatory care or domiciliary care pathways. Some common risk assessment tools applied by AIM clinicians are described in Box 11.3.

Ambulatory score

The ambulatory score (AMB) is a validated scoring system used in the UK to triage suitability for ambulatory care in medical patients and incorporates patient age, ability to travel, need for intravenous treatment, mental status, physiology (National Early Warning Score), and need for recent hospitalization. Patients with an AMB score ≥5 are classed as suitable for consideration of ambulatory care with 92% sensitivity in predicting discharge within 12 hours of index assessment.

Glasgow Admission Prediction score

The Glasgow Admission Prediction (GAP) score is a validated scoring system used in the UK to triage patients admitted to hospital from ED and incorporates age, physiology, triage severity category, route of referral, mode of arrival in ED, and need for recent admission. Patients with a GAP score ≤17 were classified as suitable for consideration of ambulatory care (the statistically significant threshold predictive of discharge).

Neither the AMB nor GAP score was derived to predict clinical outcomes and have limitations in identifying patients who may remain at risk of clinical deterioration, and so should only ever be used as an **adjunct** in clinical decision-making.

Communicating decisions and referrals

Clear and concise communication is an important part of medical practice in promoting effective decision-making and safe patient care and it contributes to improved patient outcomes, reduced length of stay, and good continuity of care.

SBAR

SBAR (Situation, Background, Assessment, Recommendation) is an effective structure for communication among healthcare professionals. SBAR ensures all important information is clearly relayed, reduces the risk of misinterpretation, and thus enhances patient

Box 11.3 Risk assessment tools used for common conditions presenting to the acute medical service

Condition	Risk assessment tool	Explanation
Respiratory		
Pneumonia	CURB-65	The CURB-65 score is used to assess the severity of community-acquired pneumonia. It evaluates factors such as **c**onfusion, **u**rea level, **r**espiratory rate, **b**lood pressure, and age (>**65**) to support treatment decisions.
Gastrointestinal		
Upper gastrointestinal bleed	Glasgow-Blatchford Bleeding score	The Glasgow-Blatchford Bleeding score helps stratify the risk from upper gastrointestinal bleeding, prioritize endoscopic interventions, and identify candidates for ambulatory management.
Neurological		
Subarachnoid haemorrhage	Ottawa score	The Ottawa Subarachnoid Haemorrhage (SAH) Rule is used to assess the probability of SAH in patients with acute headache.
Alcohol withdrawal	CIWA score	The Clinical Institute Withdrawal Assessment for Alcohol Scale (CIWA) evaluates the severity of alcohol withdrawal symptoms and those who require pharmacological treatment.
	AUDIT-C score	Alcohol use disorders identification test for consumption (AUDIT-C) identifies 'at-risk' drinkers.
Cardiovascular		
Acute coronary syndrome	TIMI score	The Thrombolysis in Myocardial Infarction (TIMI) score assesses the risk of adverse cardiac events in patients with acute coronary syndrome.
	GRACE score	The Global Registry of Acute Coronary Events (GRACE) score predicts the risk of death or myocardial infarction in patients with acute coronary syndrome.
Atrial fibrillation	CHA_2DS_2-VASc score	This score is used to estimate the risk of stroke in patients with atrial fibrillation.
	HAS-BLED score	The HAS-BLED score assesses the risk of bleeding in patients with atrial fibrillation on anticoagulation therapy.
	ORBIT score	Outcomes Registry for Better Informed Treatment (ORBIT) of Atrial Fibrillation is used to assess the risk of bleeding from anticoagulation for patients with atrial fibrillation.
Pulmonary embolism	Wells score	The Wells score is used to assess the probability of pulmonary embolism. It considers clinical symptoms, signs, risk factors, and alternative diagnoses to guide further diagnostic testing.
	Geneva score	The Geneva Score is another tool for assessing the probability of pulmonary embolism. It evaluates clinical criteria, such as signs and symptoms, as well as risk factors.
	PESI score	The Pulmonary Embolism Severity Index (PESI) score estimates the severity and risk of adverse outcomes in patients with pulmonary embolism. It incorporates factors such as age, co-morbidities, and vital signs.
	PERC score	The Pulmonary Embolism Rule-out Criteria (PERC) score helps exclude the need for further testing where pulmonary embolism is considered but the patient is low-risk. It evaluates clinical criteria to determine if additional investigations are necessary.
DVT (deep vein thrombosis)	Wells score	The Wells score is also used to assess the probability of deep vein thrombosis. It takes into account clinical features, risk factors, and alternative diagnoses when determining the need for further investigation.

safety. It is designed such that the person initiating the communication should have the pertinent information to hand, as well as know what the expectation is of the interaction.

Handover processes

Effective handover of patients between shifts and teams is crucial for continuity of care and patient safety, as changeover in medical and nursing staff is a potential source of error. A well-executed handover ensures that important details are accurately communicated, preventing vital information from being missed or overlooked. Utilizing succinct summaries during handovers allows for efficient transfer of knowledge, including vital patient information, current treatment plans, and any pending tasks. Written handovers (ideally using dedicated electronic handover databases) should be used whenever possible to maximize clarity and accountability. Handovers should also extend to communication with practitioners outside of hospital, for example general practitioners (GPs) or ongoing care providers when a patient is transferred to a different hospital, ensuring a smooth transition and uninterrupted follow-on care.

Passports for complex patients

Patients with complex needs, such as those with learning disabilities, sickle cell disease, or frequent hospital re-attenders, should have easily accessible passports outlining their individualized plans of action upon admission. This passport, either integrated into the

hospital records system or carried by the patient, serves as a concise reference for healthcare professionals, ensuring prompt and appropriate care delivery. By having a predefined plan in place, potential risks and challenges can be addressed proactively, enhancing patient safety and satisfaction.

Discharge planning

Commencing early discharge planning

Discharge planning should commence alongside acute care at the point of admission. Establishing an estimated date of discharge (EDD) and clinical criteria for discharge (CCD) enables the entire healthcare team to work productively towards a targeted goal. Research has shown that EDD implementation effectively reduces the length of hospital stays, benefiting both patients and the healthcare system.

Recent NHS initiatives developed to assist patients in progressing effectively through their hospital journey include:

- the SAFER care bunder (**S**enior review, **A**ll patients have EDD and CCD, Early **F**low between AMU and hospital wards, **E**arly discharges ideally before midday, MDT **R**eview of all patients with hospital stay of >7 days) and
- 'Red and Green days' ('red days' being those where little value is added to clinical care—to be avoided—and 'green days' a day of value, helping to move the patient towards discharge).

Multidisciplinary discharge planning

Discharge planning for complex patients should involve all members of the MDT and is a key function of daily 'board rounds' and AMU MDT meetings. Early inclusion of patients and families in discussions regarding discharge planning empowers them to participate in the decision-making process, consider their discharge options, and mentally prepare for life outside the hospital.

Comprehensive discharge summaries

Written discharge plans, formulated by members of the MDT, play a role in ensuring a smooth transition from hospital to discharge destination. Summaries should be provided to patients in a legible format and in a language they can understand and should include the diagnosis, investigations, medication information (including any changes to their long-term prescription made while in hospital), follow-up instructions, and potential risks or warning signs to watch for. Verbal communication should accompany the written summary, ensuring patients are well-informed and empowered to manage their health post-discharge.

Liaison with acute care MDT

Acute frailty in reach

Identifying acutely unwell frail patients from the medical 'take', using screening tools such as the Clinical Frailty Scale (CFS), enables early Comprehensive Geriatric Assessment (CGA, Chapter 49) and ensures tailored multidisciplinary input from the earliest point in their admission. Acute frailty services (usually geriatricians and geriatric clinical nurse specialists) will often provide in-reach to the AMU or ED. CGA encompasses not only physical assessment but also includes functional, social, environmental, and psychological domains enabling the formulation of an individualized holistic care plan with the intention to promote optimal outcomes and prevent unnecessary hospital admissions or re-admissions.

Chest physiotherapy for respiratory conditions

In cases of acute respiratory distress, specialist chest physiotherapists can assist in the assessment and management of patients with conditions such as pneumonia, chronic obstructive pulmonary disease (COPD), bronchiectasis, interstitial lung disease, and asthma. Their involvement in airway clearance techniques, breathing exercises, and respiratory support can optimize lung function, reduce complications, and improve patient comfort and recovery.

Drugs and alcohol liaison service

Collaboration with drugs and alcohol liaison professionals (Chapter 30) when managing patients with substance abuse disorders provides expertise in assessing and managing withdrawal syndromes, coordinating appropriate detoxification protocols, and providing support for patients struggling with addiction. Liaising with this team helps address underlying substance abuse issues and provides the necessary resources for ongoing care and rehabilitation, potentially reducing repeated admissions or further morbidity.

Specialist teams for asthma and diabetes

Asthma and diabetes clinical nurse specialists often offer valuable insights into the management of these chronic conditions during acute exacerbations, by adjusting medication regimens, providing education on self-management, and facilitating appropriate follow-up care.

Critical care outreach

Critical care outreach teams (Chapter 13) can comprise critical care physicians, nurses or respiratory therapists, providing advice and support for patients either exhibiting signs of clinical deterioration, or proactively in those patients who have recently stepped down from higher levels of care. Their input can help guide decisions regarding escalation of care, early initiation of critical care interventions, and transfer to ITU.

PROCEDURES

AIM physicians are proficient in a range of procedural skills linked to the urgent investigation and treatment of acute medical presentations (**Box 11.4**). Competency in these procedures is gained through initial training and supervision in simulation settings using anatomical models or mannequins, before undertaking the procedure under direct supervision and guidance in real patients.

Box 11.4 Core procedural skills in AIM

Procedure
Central venous cannulation (internal jugular and femoral)
Intraosseous access to circulation for resuscitation
Intercostal drain for pneumothorax/effusion
Knee aspiration
Lumbar puncture
Abdominal paracentesis
Setting up non-invasive ventilation or CPAP
Arterial line insertion
Point-of-care ultrasound

For all procedures, always check local guidelines and familiarize yourself with equipment available at that site before beginning. If a Local Safety Standard for Invasive Procedures (LocSSIP) exists, this should be adhered to and the appropriate checklist completed. Many AMUs have a dedicated 'procedure room', designed to optimize space and lighting and facilitate clean technique. A 'non-sterile' assistant is often required to aid with gowning and preparation of equipment.

The following principles apply to the preparation for and practice of all acute medical procedures.

Review blood test results and medications

Prior to any intervention, review clotting tests (prothrombin time/INR and activated partial thromboplastin time) and full blood count (haemoglobin, platelets). Review the drug chart for antiplatelets or anticoagulants and consider whether these may need to be suspended prior to the procedure. Safe thresholds for intervention differ between procedures (e.g., requirements are more stringent for lumbar puncture (LP) than central venous catheter insertion). If using lidocaine (lignocaine) during any procedure, the maximum cumulative dose must not exceed 3 mg/kg, repeated not more often than every 4 hours (for reference, 1 mL of 1% lidocaine solution = 10 mg; 1 mL of 2% lidocaine solution = 20 mg).

Review imaging

Review any relevant imaging (e.g., CXR or CT prior to intercostal chest drain insertion, CT brain prior to LP) both to confirm the indication for the procedure and exclude anatomical contraindications.

Patient consent

Informed consent should include an explanation of the procedure and its purpose and potential complications, as well as the completion of a consent form signed by both patient and operator. In emergency situations, it may not be possible to take formal consent before proceeding, in which case the rationale behind a 'best interests' decision should be recorded in the notes.

Equipment

All the necessary equipment to complete the procedure from start to finish should be gathered prior to commencing. Many hospitals purchase bundled 'procedure packs' containing common equipment including a sterile drape, gauze swabs, syringes and needles, scalpal, suture, and dressings. Non-sterile equipment such as the procedure trolley and ultrasound machine should be thoroughly cleaned and disinfected prior to use; a sterile plastic 'ultrasound sleeve' is placed over the probe with the help of an assistant once the operator has donned their sterile gown and gloves. Procedure-specific 'kits' (e.g., central venous catheter kit) contain both the catheter/drain itself, and also the necessary introducer needle, guidewire, and dilators required for insertion; these should be opened carefully by the assistant, allowing the 'sterile' operator to remove the contents and place them on their sterile field (see below).

Aseptic technique

Aseptic technique should be practised to minimize risk of introducing infection. This should include use of a dedicated and frequently cleaned procedure room where possible, thorough hand washing, donning of surgical gown and sterile gloves, and creation of a 'sterile field' (usually on a procedure trolley) where instruments and equipment are laid out for use during the procedure. For all procedures, skin should be cleaned thoroughly with a disinfectant solution (usually chlorhexidine) and a sterile fenestrated drape used to isolate the relevant area before commencing.

Documentation

On completion of the procedure, a structured entry should be placed in the notes. Details should include the volume and strength of local anaesthetic used (to avoid any inadvertent overdose if further procedures are scheduled), any problems or immediate complications encountered during the procedure, a record of the precise medical device that has been inserted and any relevant after-care instructions.

Central venous cannulation, ***intraosseous access to the circulation***, ***intercostal chest drain insertion***, ***LP***, and ***abdominal paracentesis*** are described in detail in **Boxes 11.5–11.9**, respectively. The Seldinger technique (insertion of a catheter over a metal guidewire) is common to several of these procedures and is illustrated in **Fig. 11.2** for the example of central venous catheter insertion. Joint aspiration is described in **Box 32.7** and arterial cannulation in **Box 13.2**. Setting up non-invasive ventilation (NIV) or continuous positive airway pressure (CPAP) is described in the section 'Advanced respiratory support', below.

Point-of-care ultrasound

Point-of-care ultrasonography (POCUS) is an advanced diagnostic ultrasound skill that is increasingly performed and interpreted by AIM clinicians at the bedside. In the UK, the Focused Acute Medicine Ultrasound (FAMUS) accreditation involves three modules: thoracic, abdominal (liver, renal, and bladder), and vascular. It can be used as a rapid diagnostic tool, especially in emergency and acute medicine, and to support the safe undertaking of invasive procedures (see **Box 11.10**). Growing availability of ultrasound technology in EDs and AMUs, together with an expanding pool of trained front-line clinicians means POCUS is increasingly used to make rapid and confident diagnoses in emergency scenarios where traditional radiological investigations may delay diagnosis and treatment.

Advanced respiratory support

Advanced respiratory support techniques include ***high-flow nasal oxygen*** (HFNO), ***CPAP***, and ***bi-level positive airway pressure*** (BiPAP, also referred to as non-invasive ventilation or NIV). These techniques are often delivered in ECUs embedded within AMU, respiratory wards, or critical care units. CPAP and BiPAP have additional indications in a chronic setting in the community and patients receiving these treatments at home may be seen frequently on the acute medical take. Both CPAP and BiPAP reduce cardiac output and may potentially cause haemodynamic instability and so, close monitoring of these patients is required. It is also important to consider that wearing a CPAP or BiPAP mask may reduce the oral intake of a patient and insensible losses will be higher. Careful fluid balance and consideration of supplemental i.v. fluids are important, along with oral hygiene and care, nasogastric tube feeding, pressure dressings, and patient positioning.

Indications for continuous positive airway pressure, bi-level positive airway pressure, and high-flow nasal oxygen

Common indications for initiating CPAP, BiPAP, and HFNO are summarized in **Box 11.11**.

Box 11.5 Central venous cannulation (CVC)

(also **Figs. 11.2 and 11.3**)

Indications

- Administration of medications that require central venous access e.g., inotropes, high-concentration electrolytes
- Fluid balance monitoring with central venous pressure (CVP)
- Intravenous access (long term or difficult peripherally)

Pre-procedure

- Informed consent should include: risks of infection, bleeding (arterial puncture, haematoma, haemothorax), pain, damage to surrounding structures (including pneumothorax), thrombosis, and failure to insert the catheter successfully
- For internal jugular vein placement, position patient with head down-tilt in the Trendelenburg position if they can tolerate it, with head turned away from side of planned insertion: this ensures maximum venous filling and reduces the risk of complications including carotid puncture or air embolism but may exacerbate respiratory distress and is potentially dangerous in those with raised intracranial pressure
- Scan area with US to define anatomy and ensure patency of the intended vein for CVC insertion prior to embarking on procedure

Procedure

- Under ultrasound guidance infiltrate lidocaine cutaneously, subcutaneously, and around the internal jugular vein (4 mls of 2% solution is generally sufficient)
- While lidocaine takes effect, flush all lumens of the line with saline and then clamp all lumens except the Seldinger port
- Under ultrasound guidance, take introducer needle attached to syringe and insert into the internal jugular vein
- When blood is freely aspirated, remove syringe and immediately insert guidewire—this should pass easily
- **Keeping hold of the inserted wire**, remove the needle. Ensure the wire stays in the vein as you do this
- Use a scalpel to make a small incision in the skin. This should be made cutting away from the wire so as not to risk damaging it
- Pass the dilator over the wire and gently but firmly dilate a tract through to the internal jugular. At this stage, there may be some bleeding so ensure to have some swabs ready
- Remove the dilator and pass the central line over the Seldinger wire. Do not advance the line until you have hold of the end of the wire
- Once the central line is in place, remove the wire
- Aspirate and flush all lumens and re-clamp and apply lumen caps
- Suture the line with 4 points of fixation (anchoring each hole on the catheter's hub to the skin)
- Attach plastic cap if available to secure in place
- Dress with a clear dressing so the insertion point can be clearly seen without removing the dressing, labelling with the date of insertion

Post-procedure

- Attach central line to pressure bag to allow CVP monitoring if indicated
- Run a blood gas to ensure a venous sample
- CXR to confirm correct placement and to exclude pneumothorax

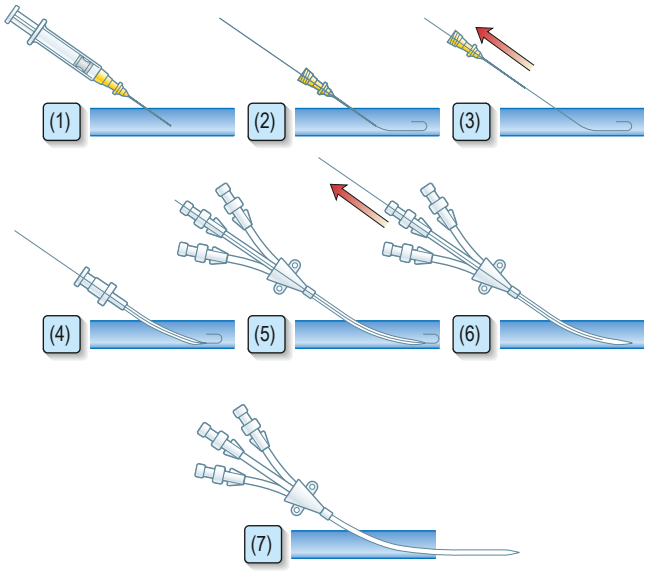


Fig. 11.2 Seldinger technique—insertion of a catheter over guidewire. This catheter insertion technique is used in a broad range of procedures including central venous cannulation, intercostal drain insertion, and paracentesis. In this example, the Seldinger technique is being used to insert a central venous catheter, although the same sequence of steps is common to all 'wire-guided' procedures. **(1)** Puncture vessel. **(2)** Advance guidewire. **(3)** Remove needle. **(4)** Dilate vessel. **(5)** Advance catheter over guidewire. **(6)** Remove guidewire. **(7)** Catheter *in situ*.

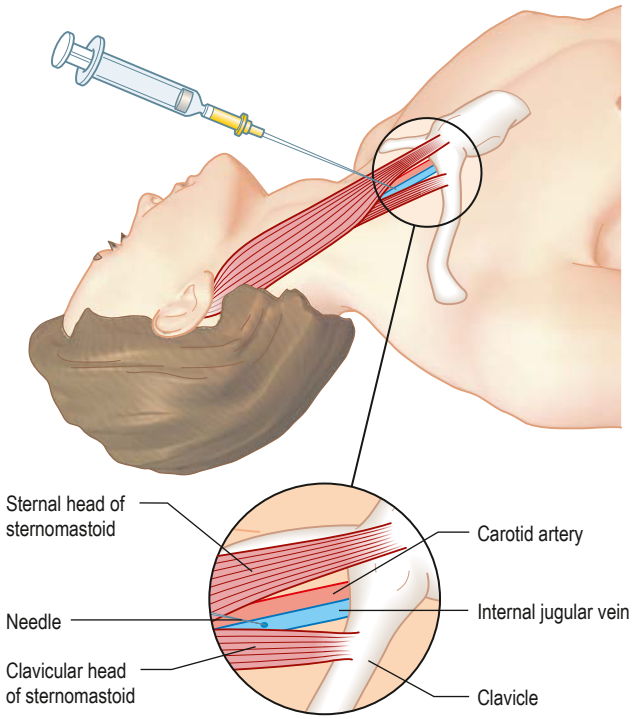


Fig. 11.3 Cannulation of the right internal jugular vein.

High-flow nasal oxygen

Also referred to as high-flow nasal therapy (HFNT) or Optiflow, HFNO is a technique of providing warmed and humidified oxygen at a higher flow rate in patients with respiratory failure when attempting to avoid intubation or to help after extubation. Oxygen administration by HFNO is distinct from standard low-flow nasal cannulae. When patients are administered low-flow nasal oxygen, the oxygen flow rates are typically between 2 and 10 L/minute (L/M). When a patient is spontaneously breathing, the inspiratory flow rate (IFR) is 20–40 L/M. Once the IFR surpasses the flow of oxygen coming from the nasal cannulae, room air is entrained which reduces the fraction of inspired oxygen (FiO_2). The effective delivered oxygen concentration is usually 25%–30%, if a patient is

Box 11.6 Intraosseous (IO) access to circulation for resuscitation

Intraosseous (IO) access refers to the use of the medullary space of a long bone to administer medications and may be obtained by using a drill insertion device (usually an EZ-IO driver) for insertion of a specialized needle and catheter, usually in an emergency e.g., cardiac arrest.

Indications

- In time-critical and/or life-threatening situations such as shock, resuscitation, or trauma settings, peripheral cannulation has been unsuccessful
- To act as an interim measure to administer life-saving medication e.g., by paramedics or pre-hospital teams until a more definitive intravenous access can be secured on arrival to hospital

Site of insertion

- The two most common sites for IO insertion in an adult are the proximal tibia and the proximal humerus
- Proximal tibia: 2 finger breadths below the patella and 1–2 cm medial to the tibial tuberosity (**Fig. 11.4**)
- Proximal humerus: insertion at the surgical neck of the humerus—place the patient's arm with the elbow adducted with the hand over the umbilicus. From the midshaft of the humerus, palpate upwards to locate the greater tubercle or 'surgical neck' of the humerus—this is the insertion site
- Other IO insertion sites include the distal tibia, femur, iliac crest, and sternum but are only usually used in isolated circumstances e.g., advanced trauma

Pre-procedure

- Obtain and document informed consent if the patient is able; otherwise document why the procedure is indicated or in the patient's best interests
- Choose insertion site (usually proximal tibia or humerus as above)
- Select the correct IO needle size and catheter: there are three different sizes, all 15G diameter but varying in length depending on the insertion location and size of the patient (**Fig. 11.5**). When inserting into the proximal humerus, the 45 mm (yellow) should always be used. In the proximal tibia, the 25 mm (blue) is most commonly used

Procedure

- Connect selected needle and catheter set to drill (turning clockwise to connect)
- Locate, palpate, and stabilize insertion site
- Remove needle cap and insert perpendicular ensuring one black line (5 mm) remains visible on the needle externally to the skin once the needle tip rests against the bone
- Once bone has been reached, gently drill and apply gentle but firm pressure to penetrate the bone cortex until a 'give' is felt on entry to the medullary space
- Disconnect the driver from the needle tip and remove the stylet by turning anti-clockwise
- Using the IO stabilizer dressing if available, secure the insertion
- Aspirate blood samples for analysis e.g., a venous blood gas at this time if required (which must be analyzed in the lab as opposed to point-of-care machines as IO samples frequently block conventional point-of-care systems)
- Connect primed extension set and flush with saline

Post-procedure

- If the patient is conscious, infusion of lidocaine (0.5 mg/kg, maximum dose 40 mg) can be given for pain relief
- When administering further IO infusions, these typically require more pressure than i.v. and therefore 50 mL syringes may be used to manually infuse therapies such as fluids. Ensure any bolus medications such as adrenaline are followed with a 10 mL saline flush; pressure delivery systems may be useful

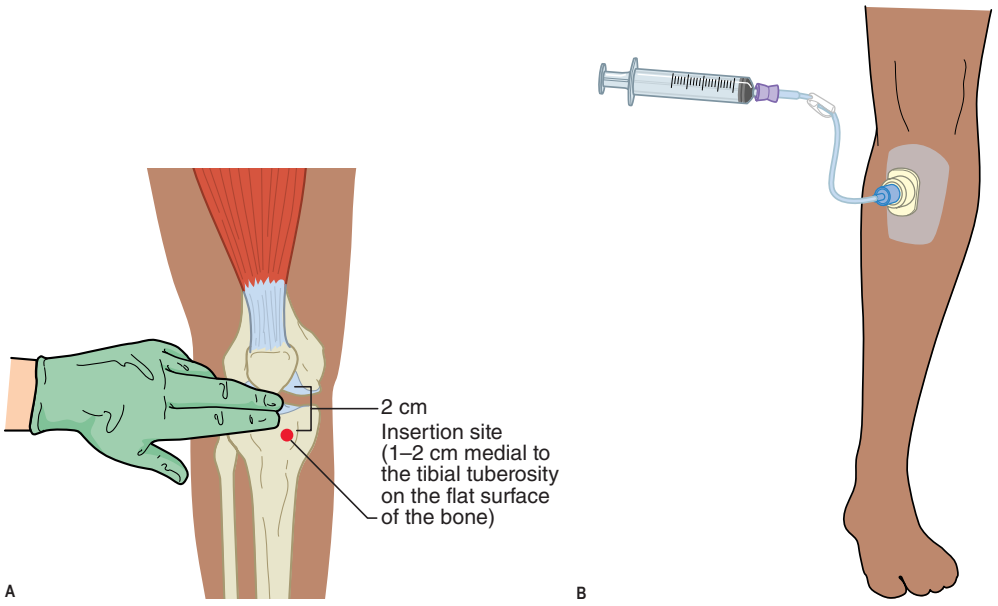


Fig. 11.4 Proximal tibia insertion site for IO access. **(A)** Landmarks for insertion. **(B)** IO needle once *in situ*.

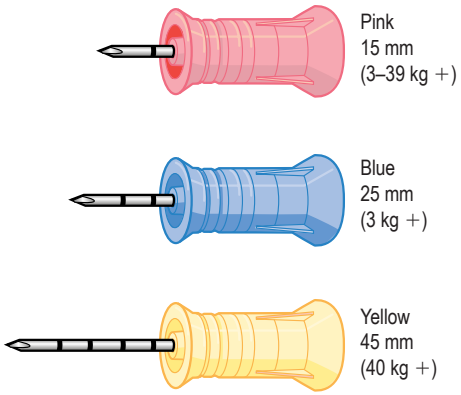


Fig. 11.5 Intraosseous needles. Colour coding corresponds to the EZ-IO system which is commonly used in the UK. Blue needle is most commonly used in adults, pink in children, and yellow in patients with excess soft tissue.

Box 11.7 Intercostal chest drain insertion (Seldinger technique)

Indications

- Diagnosis and/or therapeutic relief of a pleural effusion
- Drainage of empyema (source control of infection)
- Drainage of pneumothorax
- Other indications include in trauma settings such as a haemopneumothorax or haemathorax (but this is more often treated with a surgical chest drain as opposed to a Seldinger chest drain)

All pleural procedures should involve direct/indirect thoracic ultrasound at the time of procedure. Exceptions to this include tension pneumothorax or emergency drainage of a large pleural effusion causing significant breathlessness (peri-arrest).

Pre-procedure

- Informed consent should include: risks of infection, difficulty breathing or low oxygen levels after insertion, air leak into chest wall, bleeding, pain, damage to surrounding structures, low blood pressure, and failure of successful insertion/drainage
- Check blood tests: patients should have an INR of <1.5, and a platelet count of $>50 \times 10^9$/L.
- Review imaging prior to the procedure (either a PA CXR or CT thorax)
- Position patient: typically, pleural aspiration is performed in one of three different positions:
  - lateral decubitus
  - semi-recumbent
  - leaning forwards
- Assessment of surface anatomy: ideally, pleural aspiration should take place within the 'safe triangle' (**Fig. 11.6**). If insertion in the safe triangle is not possible, a different site can be identified, but ensure that this is 6 cm lateral to the spine to avoid injury to an intercostal artery
- Confirm site of insertion using ultrasound and mark skin
- Assemble drainage bottle. Pour 500 mL sterile water up to the underwater seal line (often marked 'start level') and leave with cap off, ready for tubing to be attached later

Procedure

- At the pre-marked US-guided site, infiltrate lidocaine (4 mls of 2% is generally sufficient). Site of entry should be 'above the rib below' in order to avoid damage to the neurovascular bundle. Use a 25G orange needle to anaesthetize the skin, and then a 21G green needle to anaesthetize the underlying soft tissue and intercostal muscle. Advance the needle perpendicular to the skin, until pleural fluid may be aspirated—or air if pneumothorax. Once the pleura has been reached and anaesthetized, withdraw the needle slowly, infiltrating the 'tract' as you draw back
- Attach the introducer needle to a 10 mL syringe and insert along the anaesthetized tract, aspirating as you advance. You may feel slight resistance as the needle passes through the pleura. Once aspirating pleural fluid or air, advance the needle a further 0.5 cm, before stopping
- Unscrew the syringe, leaving the needle in place. Ensure to keep a hand on the needle, close to the skin surface when removing the syringe to avoid accidentally advancing the needle further. Place your gloved thumb over the open end of the introducer needle to block the outflow of pleural fluid or air
- Insert guidewire through the needle—this should pass easily.
- Keeping hold of the guidewire, remove the introducer needle, leaving just the guidewire in place
- Make a small incision using a scalpel at the incision point where the wire enters the chest.
- Pass the dilator over the guidewire using a twisting motion to help separate the intercostal muscle fibres. Ensure that the dilator is only inserted as far as necessary—no further than needed originally to aspirate pleural fluid: a safety guard can be adjusted on the dilator to reduce risk of excessive dilation and lung injury. If using an 18Fr drain this will need to be done three times, once with each of the successively larger dilators
- Keeping hold of the guidewire, remove the dilator and thread your drain over the guidewire and through the chest wall into the pleural space. Continue inserting the drain until you reach the 12 cm mark
- Remove the guidewire and attach the three-way tap in OFF position
- Suture the drain in place, ensuring it is 'anchored to the skin', and apply dressing
- Connect the chest drain tubing to both the drain and the drainage bottle (**Fig. 11.7**), and open the three-way tap. You should see pleural fluid if draining effusion, or bubbles if draining pneumothorax. If you do not see bubbles for a pneumothorax, ask the patient to cough

Post-procedure

- A CXR post-procedure is recommended if unexpected complications are encountered such as pain, air-aspiration (not including that from a pneumothorax), or if pleural fluid is not draining easily
- Ensure the nursing staff are informed and are familiar with chest drain care; a chest drain chart should be started following chest drain insertion
- If samples are required, conventional tests would include: pleural fluid MC&S, fluid pH, glucose, protein, LDH, AFB smear and mycobacterial culture, and cytology (ideally 60 mL)

receiving 2–4 L/M of nasal oxygen. In contrast, HFNO uses oxygen flows of 50–100 L/M. With this technique, the high flows delivered via the specially designed nasal cannulae exceed the patient's IFR. Therefore, there is lower entrainment of room air which allows the delivery of a higher FiO_2.

HFNO should be considered in hypoxaemic acute respiratory failure of any cause, but particularly in type I respiratory failure. This includes pneumonia, COVID-19 pneumonitis, post-operatively and post-extubation, following chest trauma and for patients with impaired ventilation due to pain. HFNO may also be used to reduce hypoxia during induction of anaesthesia, during procedures under sedation, and can be a helpful adjunct to patients receiving CPAP or BiPAP, to allow breaks from face mask-based therapies. HFNO may be an option for patients not considered suitable for invasive ventilation, or for those requiring palliative care.

Continuous positive airway pressure

CPAP is delivered via a tight-fitting mask over the nose, the nose and mouth, or via a whole head hood system. It may also be delivered via a tracheostomy tube. The mask or hood is attached to a ventilator and delivers a set positive pressure throughout the respiratory cycle, although the physiological effect is equivalent to applying positive end-expiratory pressure (PEEP). The aim of CPAP is to splint open airways, reduce alveolar collapse, enable alveolar recruitment, and to increase the functional residual capacity to help reduce the work of breathing. CPAP is often used in pulmonary oedema in addition to or after optimized medical management has failed.

Bi-level positive airway pressure

BiPAP is also delivered via a tight-fitting mask but the ventilator delivers two different airway pressures: inspiratory and expiratory. The expiratory positive airway pressure (EPAP) is equivalent to PEEP (as delivered by CPAP). The inspiratory positive airway pressure (IPAP) is a higher pressure which aims to augment the patient's respiratory effort. BiPAP is used to treat **type 2 respiratory failure** and most commonly exacerbations of COPD (see Chapter 20). Close monitoring and often serial arterial blood gases are required to identify improvements or deteriorations and titrate treatment. In the event of escalating pressure requirements, respiratory and/or ITU team involvement should be sought promptly.

Prior to commencing CPAP or BiPAP, it is important to consider contraindications (see **Box 11.12**) and ensure that there is a clear

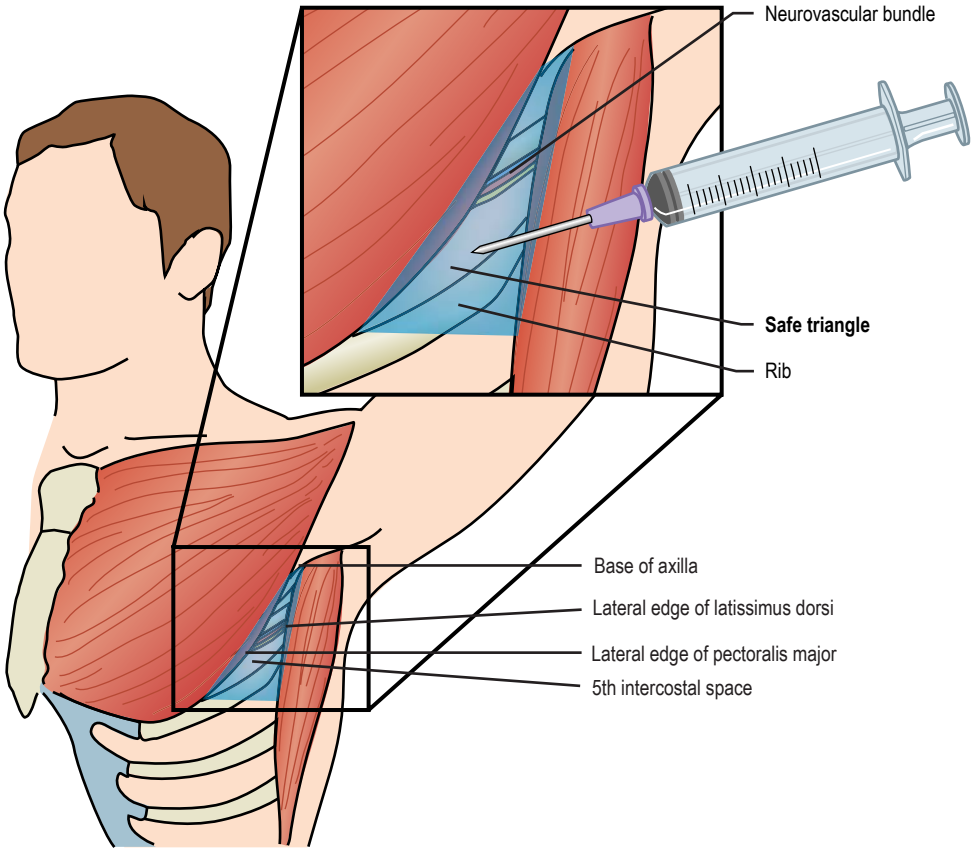


Fig. 11.6 **Intercostal drain insertion.**

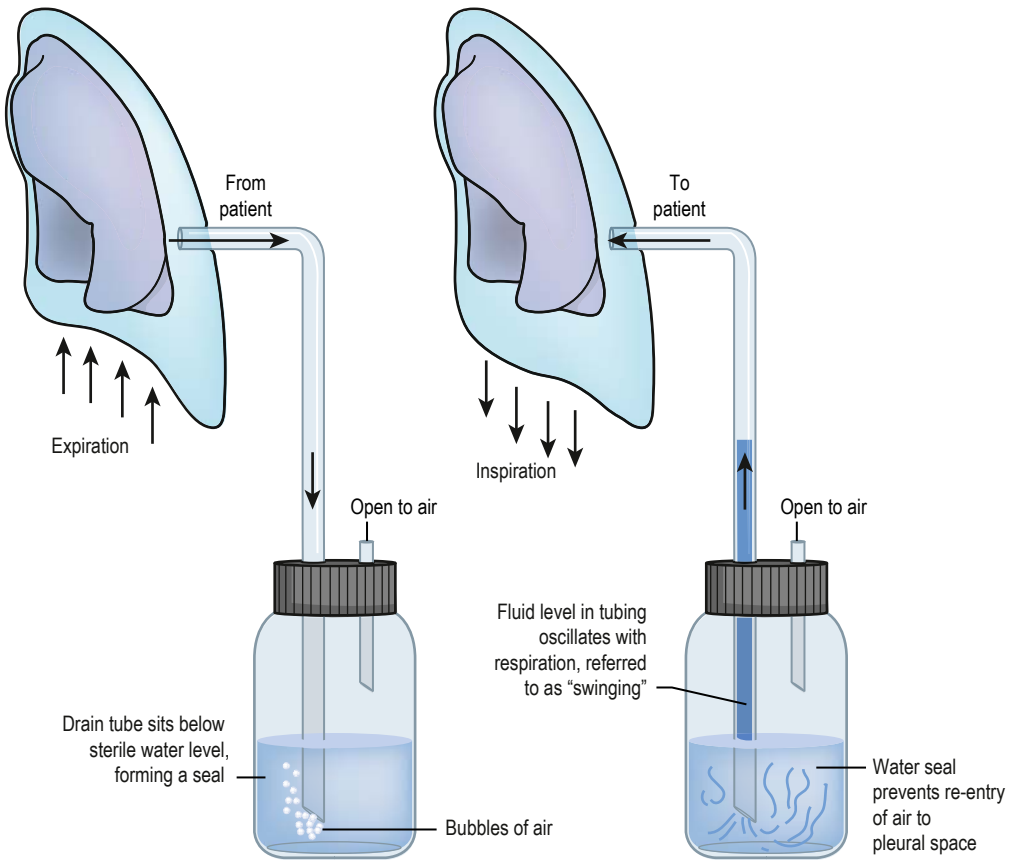


Fig. 11.7 **Intercostal drain connected to drainage system with underwater seal.**

Box 11.8 Lumbar puncture (LP)

Indications
- Diagnosis of neurological disease (e.g., meningoencephalitis)
- Measurement of CSF pressure (e.g., idiopathic intracranial hypertension)
- Removal of CSF therapeutically (e.g., idiopathic intracranial hypertension)
- Intrathecal injection of drugs (e.g., chemotherapy)

Pre-procedure
- Informed consent should include: risk of headache/worsening of headache, localized back pain, bleeding or bruising, transient neurological symptoms such as shooting pain down the legs at the time of the procedure, and failure of the procedure
- Check blood tests: patients should have an INR of <1.4, and a platelet count of >50. A minimum interval of 12 h after prophylactic doses of low-molecular-weight heparin (in patients with normal renal function) is advised
- Imaging, including a CT head (if available) should be reviewed prior to the procedure. A CT scan is mandatory in unconscious patients and those with papilloedema or other symptoms or signs of raised ICP. Removal of CSF in a patient with a mass lesion in the brain or cord may result in caudal herniation of the cerebellar tonsils (coning) and death, and LP is contraindicated in such settings. There are exceptions where an LP may be undertaken in the context of raised ICP (e.g., for treatment of idiopathic intracranial hypertension)
- Patient position: Typically LP is performed in one of two positions:
  - lateral recumbent
  - sitting and leaning forwards

  CSF opening pressure can only be accurately measured in the lateral recumbent position
- Assessment of surface anatomy: the spinal cord ends at L1, and the needle insertion point must be below this to avoid spinal cord injury. The ideal needle insertion sites for LP are between L3/L4, or L4/L5. Use surface anatomy to help identify the correct point in which to perform the procedure (Fig. 11.8)
  - Palpate the posterior superior iliac crests bilaterally. Imagine a line across the lower back connecting these two points. Follow this medially until you reach the spine, and this is typically at the level of L3/L4. Palpate this vertebral space, and the one below (L4/L5) to identify the widest space to perform the procedure

  LP is contraindicated if there is skin infection over needle insertion site.

Procedure (Fig. 11.9)
- Correct positioning for both clinician and patient is vital. Most clinicians complete this procedure sitting down, so ensure your chair is at the right height and is comfortable
- Palpate and mark the insertion site (ultrasound may be useful here)
- Infiltrate lidocaine (up to 10 mls of 1% is usually sufficient) over the pre-marked area, with needle tip pointing towards the umbilicus. Use a 25G orange needle to anaesthetize the skin and then a 21G green needle and continue to anaesthetize the underlying subcutaneous tissue.
- Take your spinal needle (typically black 22G), and insert this, bevel up, through the marked and anaesthetized area in the direction of the umbilicus. Ensure to advance slowly, and that the stylet is in place at all times
- When you pass through the spinous ligament you will feel a change in resistance. At this point, remove the stylet gently and observe for flow of CSF. If you do not visualize this, rotate the needle slightly and recheck
- Note the appearance of the CSF (clear, cloudy, yellow in xanthochromia or red) and then attach a three-way tap and manometer, and turn the tap so that it is open between the needle and the manometer. Allow the fluid to fill the manometer and record the opening pressure (if required). Normal CSF pressure is 60–150 mm of CSF
- To take diagnostic samples, an assistant should hold the bottles underneath the three-way tap. For biochemistry, microbiology, or xanthochromia 0.25 mL or approximately 6 drops of CSF is usually sufficient. If sending cytology, a larger volume (up to 10 mL) increases diagnostic yield
- Once you have collected the samples, remove the three-way tap and **replace the stylet.** Withdraw the needle and stylet together and cover the insertion site with a dressing

Post-procedure
- To minimize the risk of post-lumbar puncture headache:
  - encourage the patient to remain hydrated and caffeinated drinks can be considered
  - lying as flat as able depending on clinical setting can also help avoid this
- Samples sent will depend on indication for procedure but may include: protein, glucose, Gram stain, culture and sensitivity, xanthochromia, cytology, oligoclonal band analysis, virology, or special stains (e.g., AFB). Remember to send a simultaneous paired blood glucose sample

treatment escalation or de-escalation plan in place with a timetable of regular reviews. Initiation of respiratory support can be frightening for patients and it is vital that the indications, process, and timeline are discussed beforehand to help alleviate this.

Setting up CPAP/BiPAP

What to consider before starting

- Consent: explain the procedure, its practicalities, and potential complications to the patient
- Treatment may be uncomfortable and poor tolerance is one important cause of failure. Reassure the patient that treatment will be increased gradually and emphasize the potential benefits
- TEP/DNACPR status should be reviewed prior to commencing CPAP/BiPAP, but note that having a DNACPR does not in itself preclude the use of advanced respiratory support or invasive ventilation
- Consider contraindications to treatment (Box 11.12).

Setting up treatment

Prepare
- Select the correctly sized mask and Velcro stretch straps for the patient
- It is essential to seek the assistance of a nurse experienced in CPAP/BiPAP
- The pressure from CPAP/BiPAP often leads to pneumophagia resulting in bloating and nausea. Give an antiemetic/peppermint capsule and consider NG tube insertion to reduce gastric pressure.

Putting on the mask
- An incorrect mask fitting is one of the most common errors in initiating CPAP/BiPAP
- Select the pressure settings with the mask not on the patient
- Position the mask near the patient's face so they can feel the air blowing into their face without the mask touching them
- Hold part of the mask on their face, giving them time to adjust
- When ready, hold the mask to their face allowing them to acclimatize gradually to the pressure
- Secure the mask using the Velcro stretch straps to ensure a proper fit.

Initial settings

CPAP
- Initial pressure settings are usually 5, 7.5, or 10 cmH_2O
- Some machines allow a slow increase in the pressure over time (e.g., 10 min) to allow patients to adjust, otherwise, manually increase the pressure to the desired setting

BiPAP
- The aim is to commence BiPAP at settings of 12 cmH_2O/4 cmH_2O (IPAP of 12, EPAP of 4) and escalate the IPAP in

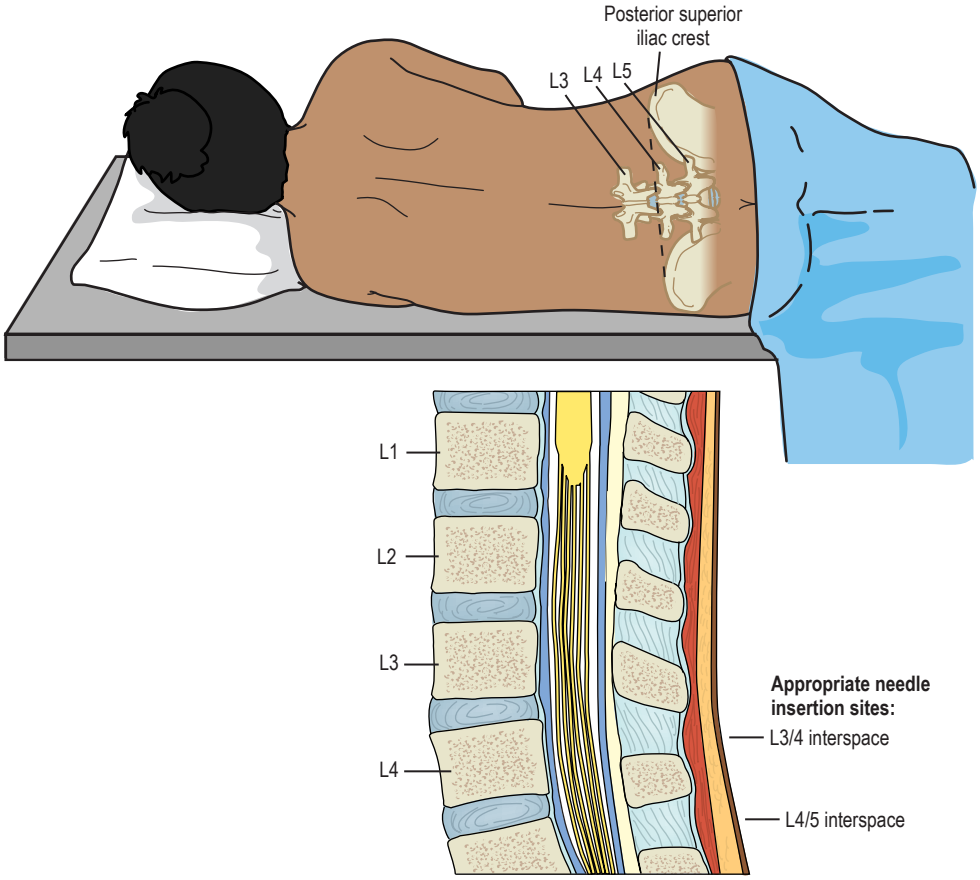


Fig. 11.8 **Lumbar puncture: patient position and anatomy.**

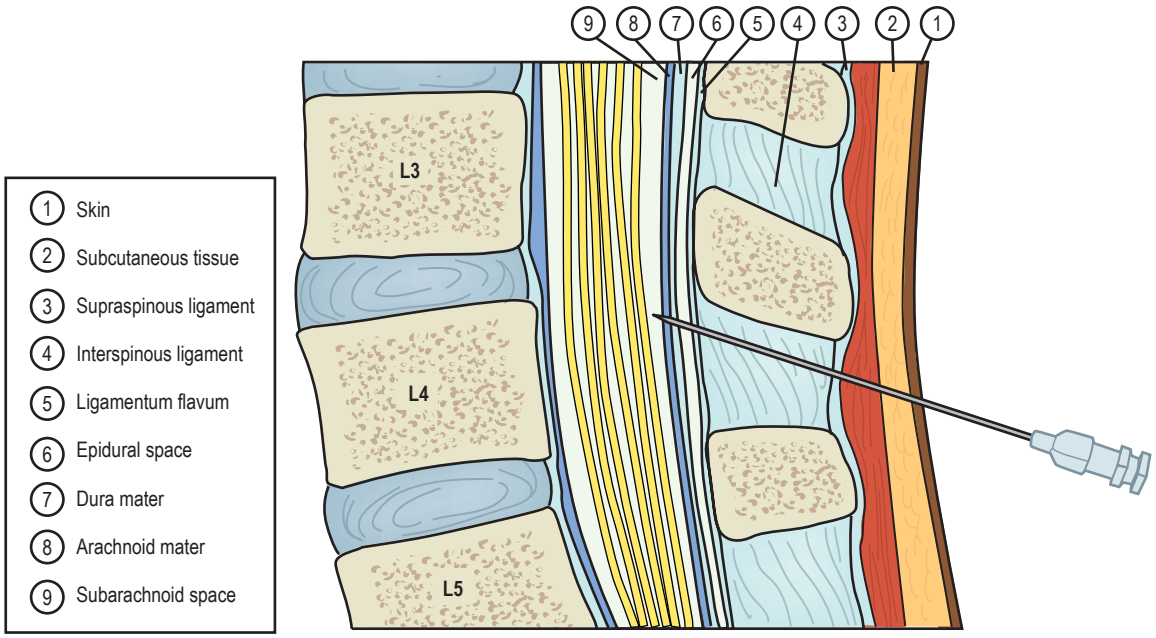


Fig. 11.9 **Anatomical layers relevant to passage of the lumbar puncture needle.**

Box 11.9 Abdominal paracentesis

Indications

- Diagnostic evaluation of ascites (including exclusion of spontaneous bacterial peritonitis)
- Therapeutic relief of gross ascites (liver cirrhosis, malignancy)

Pre-procedure

- Informed consent should include: pain, bleeding or bruising, low blood pressure (especially in cases where it is predicted larger volumes of ascites may be removed), infection (either of the drain site itself or introduction of infection to the peritoneal cavity/peritonitis), damage to surrounding structures, blocked tubing or failure of drainage
- Coagulopathy and thrombocytopenia are often present in liver disease and though it is a caution, it is not a contraindication to paracentesis or drainage. Always discuss with the local gastroenterology/haematology team if uncertain
- Depending on indication, consider ordering human albumin solution (HAS) 20% prior to drain insertion
- Best practice is to use ultrasound at the bedside to identify and mark a suitable pocket of fluid for drainage. In cases where ascites is clinically obvious, the procedure may be performed blind, although an ultrasound scan is mandatory prior to drainage of malignant ascites, where there is doubt regarding the cause of abdominal distension or suspicion that fluid may be loculated
- Patient position: the procedure is usually performed with the patient supine
- Surface anatomy: the patient should be asked to empty their bladder. A drainage site is usually selected in either the left or right iliac fossa, to avoid contact with an enlarged liver or spleen, and at least 10 cm from the midline to ensure avoidance of the inferior epigastric artery

Procedure (Fig. 11.10)

- Infiltrate lidocaine (4 mls of 2% is generally sufficient) over marked, US-guided site. Use an orange 25G needle to anaesthetize the skin and a green 21G needle to anaesthetize the underlying subcutaneous tissue down to the peritoneal membrane (you will often aspirate ascites at this point). Once the peritoneum has been anaesthetized, withdraw the needle slowly, infiltrating the 'tract' as you draw back

If performing a diagnostic ascitic tap:

- Take a 20 mL syringe with a green 21G needle and aspirate as you advance. Stop when you aspirate ascites and fill up your syringe
- Withdraw the needle and place an adhesive island dressing or similar to cover the area

If performing a therapeutic ascitic drain:

- Assemble your drain kit: a paracentesis catheter with three-way tap (which usually requires preparation by insertion of the long introducer needle into the catheter, slowly so as to not pierce the tubing, and removal of the external plastic sheath), a 20 mL syringe, a scalpel, along with preparation of the drainage tubing by connecting to a drainage bag with stand
- Ensure the three-way tap is turned off and connect a 20 mL syringe to the end of the introducer needle (which is now within the paracentesis catheter)
- Use a scalpel to make a small cut in the skin where previously marked and anaesthetized
- Insert the introducer needle and drain perpendicular to the skin through the premade cut and slowly advance, aspirating the syringe periodically, until entering the peritoneal cavity
- Once able to aspirate ascitic fluid, advance a further centimetre and then keep the long introducer needle still while sliding the paracentesis catheter forward, off and over the needle until flush with the skin. Then withdraw the needle leaving the drain *in situ*
- Secure with 2–3 cannula-type adhesive dressings (the steri-strips can be useful)
- Attach a 2 L drainage bag
- Open three-way tap and ascitic fluid should start draining. You can use the 3-way tap to draw off your required samples

Post-procedure

- For cirrhotic ascites, ensure 20% HAS is prescribed and instructions for monitoring and infusion clearly handed over to the nursing staff (usually 100 mL of 20% HAS for each 2 L ascites drained)
- Ensure removal after no more than 6 h (drains are sometimes left *in situ* for longer when draining malignant ascites, although the risk of infection rises proportionally to the duration of drainage)

2–5 cm increments at a rate of approximately 5 cmH_2O every 10 minutes with a usual pressure target of 20 cmH_2O or until a therapeutic response is achieved or the limit of patient tolerability has been reached

- Setting changes should be guided by serial ABGs and the respiratory team should be involved to help guide this.

Monitoring of patients on CPAP/BiPAP

- Try to stay with the patient for at least five minutes to ensure they are comfortable, using this time to start optimizing settings
- ABGs should be taken as a minimum at 1, 4, and 12 hours after the initiation treatment; consider placement of an arterial line to facilitate this without repeated arterial puncture if the patient is being managed in an ECU/HDU setting
- A decision as to whether to escalate to invasive mechanical ventilation should normally be taken within 4 h of initiating CPAP/BiPAP
- With both CPAP and BiPAP the FiO_2 can be titrated to achieve a target P_aO_2 or saturation

Escalation or de-escalation of respiratory support

Escalation plans when commencing CPAP/BiPAP generally fall into one of four categories:

- If treatment fails, escalate to intubation
- If treatment fails, consider palliation if it is not appropriate to proceed to intubation
- Continue treatment if there is a response to therapy and seek specialist respiratory input with regards to duration of therapy and weaning
- It is inappropriate to commence treatment and palliate from the outset.

How to wean CPAP/BiPAP

- Initially weaning should be during the day with extended periods off the ventilator for meals, physiotherapy, nebulized therapy, etc.
- After successfully weaning during the day, most patients will require an additional night with treatment continued
- Depending on the indication, some may wean quicker than others e.g., CPAP to treat pulmonary oedema may often be weaned quickly if the response to medical therapy is good
- Usual practice involves a 2–5 day weaning period
  - Day 2: Continue treatment for 16 h (including 6–8 h overnight)
  - Day 3: Continue treatment for 12 h (including 6–8 h overnight)
  - Day 4–5: Treatment may be discontinued unless continuation is clinically indicated.

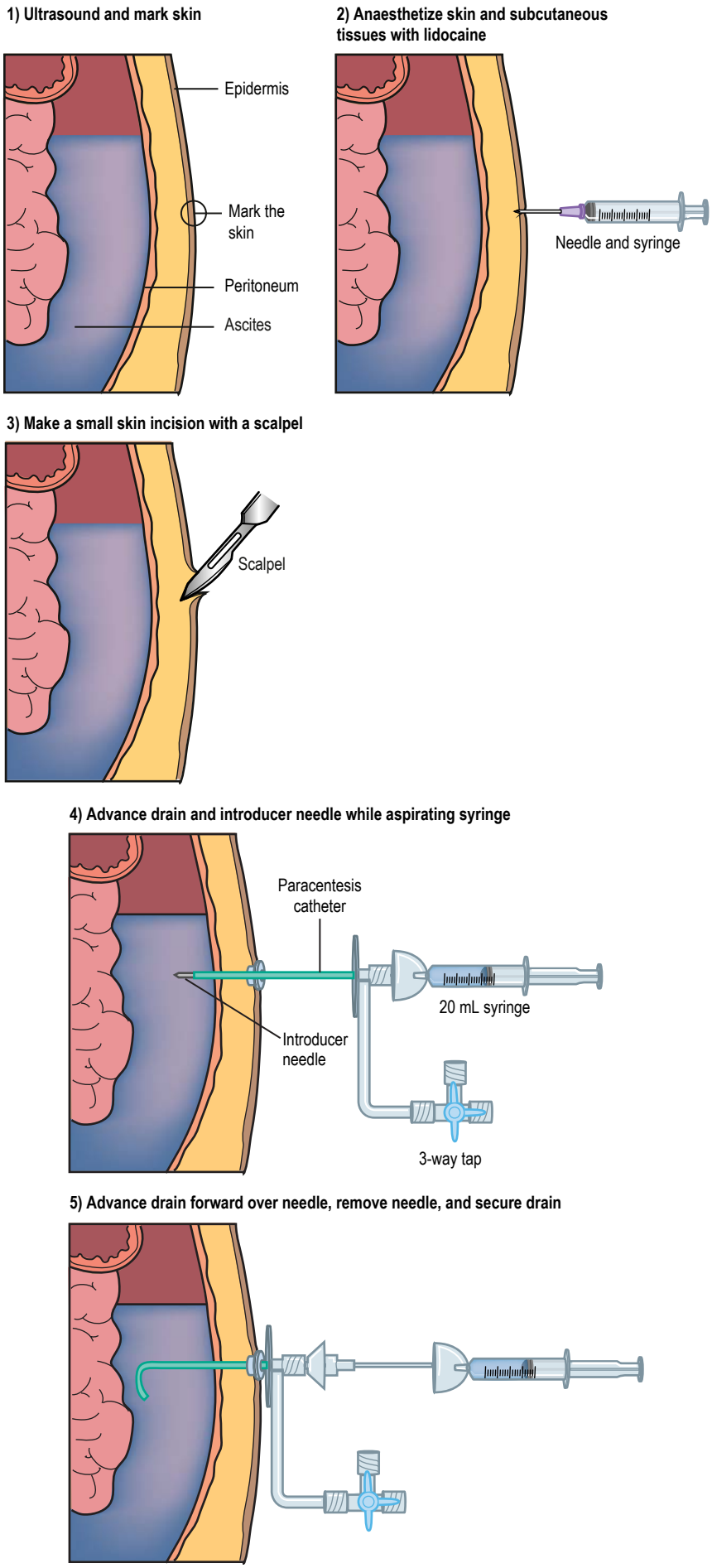


Fig. 11.10 Abdominal paracentesis.

CORE CLINICAL PRESENTATIONS IN AIM

Below are 25 common clinical presentations seen in the context of AIM. Other common emergency presentations are dealt with in specialty-specific chapters elsewhere in the book, but four other chapters are focused specifically on acute presentations requiring a tailored approach:

- sepsis (Chapter 12)
- acute poisoning (Chapter 15)
- environmental emergencies (Chapter 16)
- stroke (Chapter 29).

Cardiac arrest

Every year in the UK, there are approximately 100 000 unexpected deaths occurring within 24 hours of the development of cardiac symptoms. About half of these deaths are almost instantaneous. Most deaths are due to ventricular fibrillation or rapid ventricular tachycardia; a small proportion are caused by severe bradyarrhythmias. Coronary artery disease accounts for approximately 80% of sudden cardiac deaths in Western society. Transient ischaemia is suspected of being the major trigger factor; however, only a small proportion of survivors have clinical evidence of acute myocardial infarction.

When cardiac arrest occurs, basic life support must be started immediately. The longer the period of respiratory and circulatory arrest, the lower is the chance of restoring healthy life. The chain of survival (Fig. 11.11) includes:

- early recognition of cardiac arrest
- early activation of emergency services
- early CPR
- early defibrillation and early advanced life support
- high-quality post-resuscitation care.

Basic life support

The first step in basic life support (BLS) is to ensure the safety of the victim and rescuer. The next is to ascertain that the victim is unresponsive by shaking them and shouting into one ear. If no response is obtained, help should be sought immediately prior to commencement of BLS. If the victim has absent or abnormal breathing, then cardiac arrest is confirmed and BLS should be started (Box 11.13).

Box 11.10 Point-of-care ultrasound (POCUS)

Body system	Core diagnoses	Core skill
Thoracic	Pulmonary oedema Pneumonia Pleural effusion Pneumothorax	Site mark for drainage of pleural effusions (as per British Thoracic Society guidance)
Abdominal/renal	Hydronephrosis Bladder distension Abdominal free fluid	Site mark for abdominal paracentesis/ascitic tap
Lower limb	DVT (rule in)	Three-point compression test
Peripheral vascular access		Ultrasound-guided peripheral vascular access

Box 11.12 Contraindications to ward-based non-invasive ventilation

Absolute

- Severe facial deformity
- Facial burns/trauma
- Asthma
- Vomiting
- Fixed airway obstruction
- Undrained pneumothorax
- Bowel obstruction

Relative

- Glasgow Coma Scale score <8[a] or if patient is unable to protect their own airway
- Hypotension (systolic blood pressure <90 mmHg)
- Copious secretions
- Confusion or agitation[a] (unless caused by hypoxia or hypercapnia which would be reversed by NIV)
- Severe cognitive impairment
- Recent upper GI surgery

With regard to treatment of type 2 respiratory failure (i.e., BiPAP/NIV):

- pH < 7.15[a]
- pH < 7.25 and additional adverse features[a]
- NIV is not the treatment of choice for patients with consolidation on CXR but is sometimes used if escalation to intubation and ventilation is deemed inappropriate

[a]Consider NIV in a critical care setting.

Box 11.11 Indications for CPAP, BiPAP, and HFNO

	CPAP	BiPAP	HFNO
Type of respiratory failure	Type 1 (hypoxemic) respiratory failure	Type 2 (hypercapnic) respiratory failure	Type 1 (hypoxemic) respiratory failure
Arterial blood gas values	P_aO_2 < 8 kPa	pH < 7.35, P_aCO_2 >6.5 kPa	P_aO_2 < 8 kPa
Aetiology	Pulmonary oedema Viral pneumonitis Obstructive sleep apnoea (OSA) Postoperative respiratory failure Weaning from invasive ventilation	Chronic obstructive pulmonary disease (COPD) Bronchiectasis Pulmonary oedema Neuromuscular disease-causing weakness (e.g., motor neurone disease (MND)) Obesity hypoventilation syndrome (OHS) OSA Chest wall deformity (e.g., scoliosis) Weaning from invasive ventilation	Pneumonia Viral pneumonitis Asthma Bronchiectasis Acute respiratory distress syndrome (ARDS) Postoperative respiratory failure Weaning from invasive ventilation

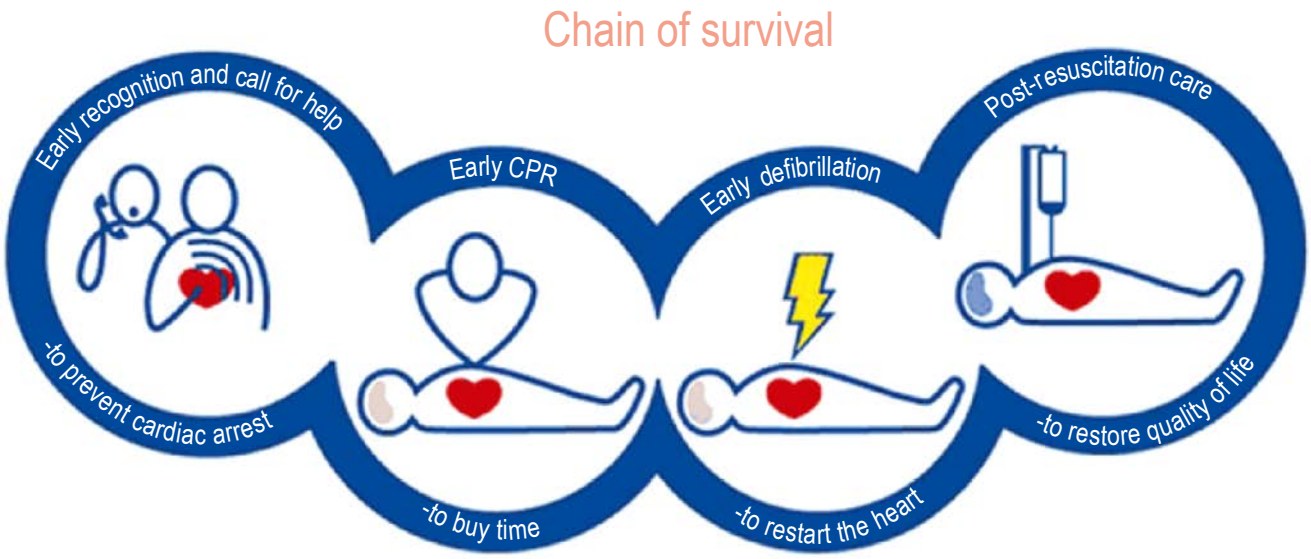
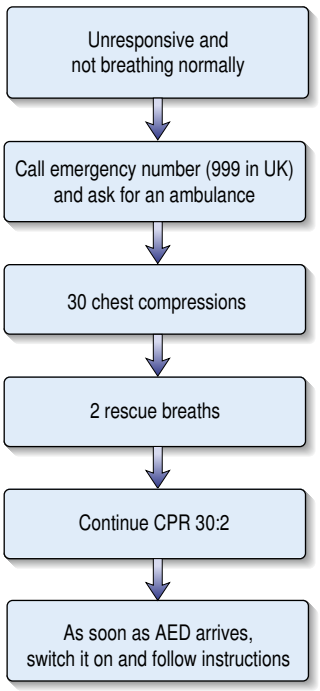


Fig. 11.11 **Cardiopulmonary resuscitation chain of survival.**

Box 11.13 Adult basic life support (out of hospital)



Airway

Debris (e.g., blood and mucus) in the mouth and pharynx should be removed, as should loose or ill-fitting dentures. The airway should be opened gently by flexing the patient's neck and extending the head ('sniffing the morning air' position). This manoeuvre is not recommended if a cervical spine injury is suspected. Any obstruction deep in the oral cavity or upper respiratory tract may need to be removed using abdominal and/or chest thrusts (see Box 20.12).

Circulation

Most adult cardiac arrest is due to a primary cardiac disorder, such as acute coronary syndrome or arrhythmia, that results in circulatory collapse. Pulse detection can be difficult, and if the victim is unresponsive, with absent or abnormal breathing, external chest compression should be started immediately. The heel of one hand is placed over the centre of the patient's chest and the heel of the second hand is placed over the first with the fingers interlocked. The arms are kept straight and the sternum is rhythmically depressed by 5–6 cm at a rate of approximately 100–120 per minute, allowing for complete recoil between compressions.

Chest compressions do not massage the heart; rather the thorax acts as a pump and the heart provides a system of one-way valves to ensure forward circulation. Respiratory and circulatory support is continued by providing two effective breaths for every 30 cardiac compressions (30:2 for one or two persons in an adult patient, 15:2 in paediatric patients). It is easier for the lay public to give compressions without interruption ('hands only CPR'), with mouth-to-mouth resuscitation considered less effective, and with obvious potential infection control consequences. CPR may be considered an aerosol-generating procedure, so adequate personal protective equipment should be considered in certain conditions including COVID-19 infection.

CPR maintains adequate cerebral and coronary perfusion pressures. The initiation of early CPR by the lay public at the site of the arrest saves lives. There is evidence to suggest that if the person performing compressions tires, the quality of resuscitation deteriorates. Mechanical CPR devices are available but not yet universally used; they may be of most use in more prolonged resuscitation attempts, for example after thrombolysis in pulmonary embolism (PE). Cardiac enzyme levels will likely rise after CPR.

Breathing

After 30 compressions, the rescuer opens the victim's airway by tilting the head backwards ***(head tilt)*** and pulling the chin forwards (***chin lift*** or ***jaw thrust***). The rescuer then pinches the victim's nostrils firmly, takes a deep breath, and seals their lips around the mouth of the victim. Two effective breaths are given, each over 1 second. In paediatric patients, respiratory arrests are more common

and patients should be given rescue breaths and a minute of CPR before a sole rescuer leaves the victim to seek help. CPR should not be interrupted to reassess the victim unless they start to show signs of life and begin to breathe normally.

Advanced cardiac life support

By the time effective life support has been established, more help should have arrived and advanced cardiac life support (ACLS) can begin. This consists of ECG monitoring, advanced airway management (endotracheal intubation or supraglottic airway tube) and establishment of an intravenous infusion in a large peripheral or central vein (an intraosseous needle [see **Box 11.6**] may be used if intravenous access is not possible). As soon as feasible, the cardiac rhythm should be established, as this determines which pathway of the European Resuscitation Council and the Resuscitation Council UK ACLS algorithm is followed (**Fig. 11.12**). This can be determined with an automated external defibrillator (AED), or the paddles or limb leads of a standard defibrillator.

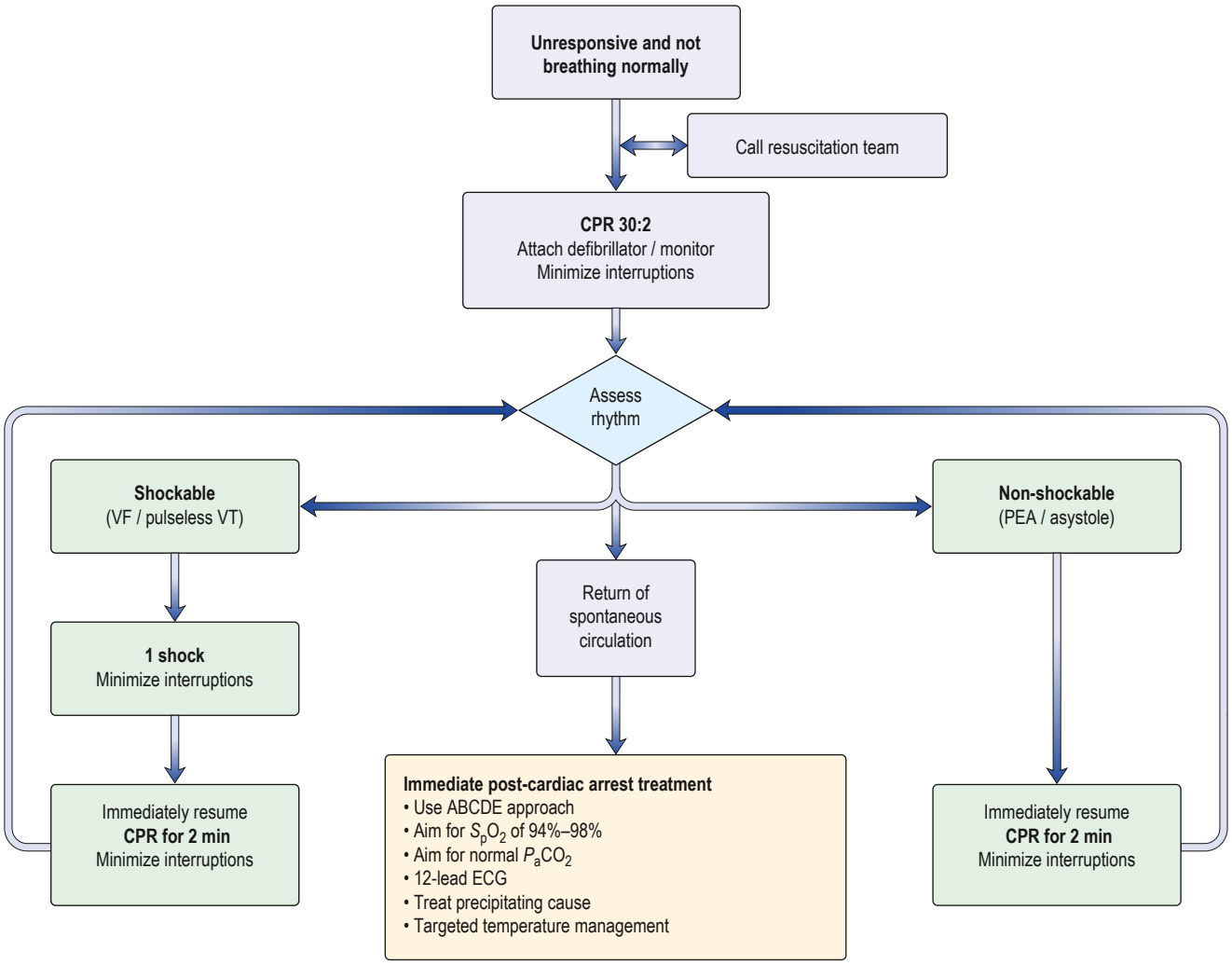


Fig. 11.12 Adult advanced life support algorithm. *ABCDE,* Airway, Breathing, Circulation, Disability, Exposure; *CPR,* cardiopulmonary resuscitation; *PEA,* pulseless electrical activity; *VF,* ventricular fibrillation; *VT,* ventricular tachycardia. *(Reproduced with permission from the Resuscitation Council; http://www.resus.org.uk.)*

- If the ECG shows a shockable rhythm—ventricular fibrillation or pulseless ventricular tachycardia—then an unsynchronized shock of 150–200 J biphasic (360 J monophasic) is delivered without delay via paddles or self-adhesive pads, followed immediately by 2 min of CPR.
- For a non-shockable rhythm—asystole or pulseless electrical activity—2 minutes of CPR is delivered with 1 mg of intravenous adrenaline (epinephrine).

For both sides of the algorithm, it is vital to maintain CPR, ensure oxygenation, and exclude or treat reversible causes—the ***'four Hs and four Ts':***

- ***Hypoxia*** should be minimized by ventilating the patient with oxygen and a bag-valve mask or advanced airway (endotracheal intubation or supraglottic airway), ensuring that there is bilateral air entry and chest expansion. With an advanced airway, CPR should continue, with a ventilation rate of 10 per minute without interrupting cardiac massage.
- ***Hypovolaemia*** is a frequent cause of pulseless electrical activity due to haemorrhage. Intravenous volume should be replaced.
- ***Hyper-*** or ***hypokalaemia*** may cause ECG abnormalities and should be detected by biochemical testing. Intravenous calcium chloride may be helpful in hyperkalaemia or hypocalcaemia. Acidosis should be managed with effective ventilation.
- ***Hypothermia*** should be excluded with a low-reading thermometer and treated with external or internal warming.
- ***Thrombosis of the heart*** and massive PE may cause pulseless electrical activity and patients should be considered for thrombolysis, thrombectomy, or primary PCI as appropriate, and if available.
- ***Tension pneumothorax*** may occur during central venous cannulation, or following chest trauma including barotrauma. Diagnosis is clinical (deviated trachea, hyper-resonant chest, absent breath sounds, ultrasound) and needle thoracocentesis or thoracostomy may be required. The 'safe triangle' (**Box 11.7**), using a specialist longer 10G non-kinking thoracocentesis catheter, is the preferred site of insertion and has a lower failure rate than the traditional approach of inserting a 14G or 16G cannula in the 2nd intercostal space, midclavicular line.
- ***Tamponade*** should be excluded with echocardiography; if present, it should be treated with pericardiocentesis ('Pericardiocentesis' section in Chapter 17).
- ***Toxins*** may have been ingested by accident or deliberate self-harm, and standard care, as well as specific antidotes should be used in appropriate patients.

POCUS is increasingly used by competent members of the resuscitation team as an adjunct to diagnosis of cardiac tamponade and tension pneumothorax, but should be deployed skilfully and not at the expense of effective chest compressions.

Defibrillation

This technique is used to convert ventricular fibrillation or pulseless ventricular tachycardia to sinus rhythm. When the defibrillator is discharged, a high-voltage field envelops the heart, depolarizing the myocardium and allowing an organized heart rhythm to emerge. Electrical energy is discharged through two paddles with gel pads or adhesive pads placed on the chest wall.

The paddles are placed in one of two positions:

- One paddle is placed to the right of the upper sternum and the other over the cardiac apex.
- One paddle is placed under the tip of the left scapula and the other over the anterior wall of the left chest.

All personnel should stand clear of the patient. The person performing defibrillation has responsibility for ensuring the safety of the patient and of the other people present. Conventional defibrillators employ a damped monophasic waveform. Biphasic defibrillators, which require less energy, are increasingly common. AEDs, which recognize ventricular fibrillation automatically, deliver a shock if indicated. These are available in some public places (in the UK, their location is signalled by a specific sign, **Fig. 11.13**). It is the responsibility of all healthcare practitioners to be familiar with the range of defibrillators they may be called on to use in their workplace.

Post resuscitation—targeted temperature management

Early studies suggested that 24 hours of therapeutic hypothermia (32°C–34°C for 12–24 h) might potentially improve outcomes in unconscious adult patients with spontaneous circulation after an out-of-hospital cardiac arrest due to ventricular fibrillation. A recent large international trial (TTM2) randomizing 1900 patients to either hypothermia (33°C) or normothermia (avoidance of temperatures >37.8°C) in the 28-hour period following out-of-hospital cardiac arrest did not demonstrate a difference in either all-cause mortality or functional outcomes, casting doubt over the routine employment of therapeutic hypothermia after cardiac arrest. The current European Resuscitation Council guidelines, however, do continue to recommend a target temperature between 32°C and 36°C in patients who remain unresponsive (GCS < 8) for the 24-hour period following ROSC.

Neurological recovery appears to be more favourable in patients with purposeful movements and electroencephalogram activity within 3 days of a cardiac arrest.

Therapeutic procedures used for peri-arrest rhythms

Direct current cardioversion

Tachyarrhythmias that do not respond to medical treatment or are associated with haemodynamic compromise (heart failure, infarction, shock, syncope—'HISS') may be converted to sinus rhythm by the use of a transthoracic electric shock. A short-acting general anaesthetic is used. Muscle relaxants are not usually given.

Fig. 11.13 **UK sign indicating the location of automated external defibrillators.**

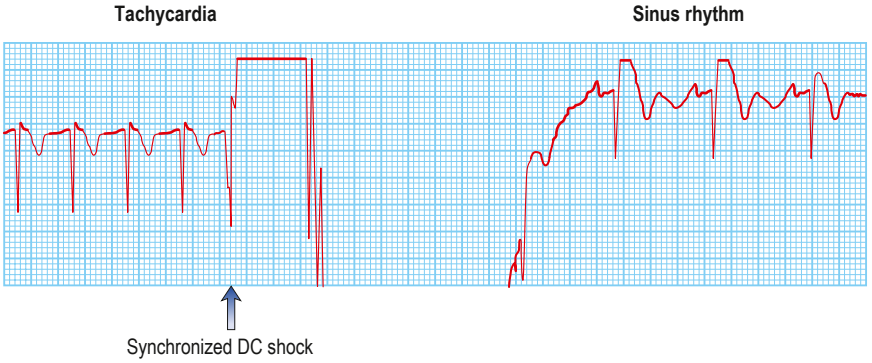


Fig. 11.14 Direct current (DC) cardioversion of a supraventricular tachycardia to sinus rhythm.
The direct current shock is delivered synchronously with the QRS complex.

When the arrhythmia has definite QRS complexes, delivery of the shock should be timed to coincide with the downstroke of the QRS complex (synchronization, Fig. 11.14). The machine being used to perform the direct current cardioversion (DCC) will do this automatically if the appropriate button is pressed. There is a crucial difference between defibrillation and cardioversion: a non-synchronized shock is used to defibrillate. Accidental defibrillation of a patient who does not require it may itself precipitate ventricular fibrillation.

Typical indications for DCC include:

- atrial fibrillation (often refractory so consider higher voltage)
- atrial flutter
- sustained ventricular tachycardia
- junctional tachyarrhythmias.

If atrial fibrillation or flutter has been present for more than 48 hours, it is necessary to anticoagulate the patient adequately for 3 weeks before elective cardioversion to reduce the risk of embolization. The duration of anticoagulation after successful cardioversion for atrial fibrillation is a complex issue and depends on a number of factors: it should be given for at least 4 weeks after the procedure and may well be given for much longer.

Digoxin toxicity may lead to ventricular arrhythmias or asystole following cardioversion. Therapeutic digitalization does not increase the risks of cardioversion, but it is conventional to omit digoxin several days prior to elective cardioversion in order to be sure that toxicity is not present.

Cardiac enzyme levels may rise after a cardioversion.

Temporary cardiac pacing

Therapeutic cardiac pacing is employed in any patient with sustained symptomatic or haemodynamically compromising bradycardia. Bradycardias may be due to either a slow intrinsic heart rate (e.g., sinus node dysfunction) or atrioventricular (AV) block. Reversible causes of bradycardia should be addressed in parallel (e.g., electrolyte abnormalities, overdose of beta-blockers or digoxin) and the use of a positive chronotropic medication such as isoprenaline may be required as an interim measure.

Transvenous pacing is the preferred method in patients with symptomatic bradycardias. In summary, a thin (French gauge 5 or 6), bipolar pacing electrode wire is inserted via an internal jugular vein, a femoral vein, or a subclavian vein and is positioned at the right ventricular apex using cardiac fluoroscopy. The energy needed for successful pacing (the pacing threshold) is assessed by reducing the energy until the pacemaker fails to stimulate the tissue (loss of capture). The output energy is then set at three times the threshold value to prevent inadvertent loss of capture. If the threshold increases above 5 V, the pacemaker wire should be resited. A temporary pacemaker unit (Fig. 11.15) is almost always set to work 'on demand'—to fire only when a spontaneous beat has not occurred. The rate of temporary pacing is usually 60–80 per minute.

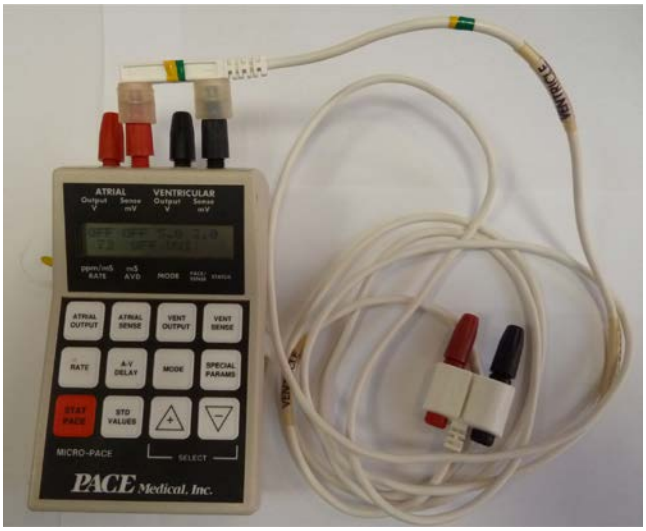


Fig. 11.15 Pacemaker. A temporary pacemaker unit.

Transcutaneous pacing may be life-saving when a cardiac arrest is precipitated by bradycardia or in instances of severe symptomatic bradycardia where transvenous pacing is not readily available. In this method the myocardium is depolarized by current flow between two large adhesive electrodes positioned anteriorly and posteriorly on the chest wall. Most modern monitor/defibrillator devices have transcutaneous pacing capability. Transcutaneous pacing is uncomfortable for the conscious patient (so safe sedation and appropriate analgesia should be sought) but can usually be tolerated until a temporary transvenous pacemaker is inserted.

Insertion of the various types of permanent pacemaker is discussed in Chapter 17.

Ventricular tachyarrhythmias

Ventricular tachyarrhythmias can be discussed under the following headings:

- life-threatening ventricular tachyarrhythmias
  - sustained ventricular tachycardia
  - ventricular fibrillation
  - torsades de pointes
- normal heart ventricular tachycardia

- non-sustained ventricular tachycardia
- ventricular premature beats (ectopics).

Some of these conditions are caused by cardiac channelopathies, congenital disorders resulting from mutations that affect the function of cardiac ion channels and hence the electrical activity of the heart. They include Brugada syndrome, congenital long QT syndrome, short QT syndrome, catecholaminergic polymorphic ventricular tachycardia (CPVT) and idiopathic ventricular fibrillation, Torsades de pointes, normal heart VT, non-sustained VT, and ventricular premature beats (discussed further in Chapter 17).

Sustained ventricular tachycardia

Sustained ventricular tachycardia (>30 sec) often results in ***presyncope*** (a feeling that one might collapse), ***syncope***, ***hypotension***, and ***cardiac arrest***, although it may be remarkably well tolerated in some patients. Examination reveals a pulse rate typically between 120 and 220 b.p.m. Usually, there are clinical signs of atrioventricular dissociation (i.e., intermittent cannon 'a' waves in the neck and variable intensity of the first heart sound).

The ECG shows a rapid ventricular rhythm with broad (often ≥0.14 sec), abnormal QRS complexes (**Fig. 11.16**). AV dissociation may result in visible P waves, which appear to march through the tachycardia, capture beats (an intermittent narrow QRS complex owing to normal ventricular activation via the AV node and conducting system) and fusion beats (intermediate between ventricular tachycardia beat and capture beat).

Supraventricular tachycardia (SVT) with bundle branch block may resemble ventricular tachycardia on the ECG. However, if a broad-complex tachycardia is due to SVT with either right or left bundle branch block, the QRS morphology should resemble a typical right bundle branch block or left bundle branch block pattern (see **Figs. 17.39A and 17.40**). Other ECG criteria to differentiate ventricular tachycardia from SVT with aberrancy are indicated in **Box 11.14**. Some 80% of all broad-complex tachycardias are due to ventricular tachycardia and the proportion is even higher in patients with structural heart disease. Therefore, in all cases of doubt, ventricular tachycardia should be diagnosed.

Management

Treatment may be urgent, depending on the haemodynamic situation. If the patient is haemodynamically compromised (e.g., hypotensive or pulmonary oedema), emergency DC cardioversion may be required. On the other hand, if the blood pressure and cardiac output are well maintained, intravenous therapy with amiodarone is often started, although upon specialist advice, other therapies such as intravenous beta-blockers (esmolol) or class I drugs (e.g., lidocaine) can be considered. DC cardioversion is necessary if medical therapy is unsuccessful, although it is important to consider correcting any potential reversible contributory causes including electrolyte abnormalities or toxins, to reduce the risk of relapse.

Ventricular fibrillation (VF)

This involves very rapid and irregular ventricular activation with no mechanical effect. The patient is pulseless and becomes rapidly unconscious; respiration ceases (cardiac arrest). The ECG shows shapeless, rapid oscillations and there is no hint of organized complexes (**Fig. 11.17**). It is usually provoked by a ventricular ectopic beat. Ventricular fibrillation rarely reverses spontaneously. The only effective treatment is electrical defibrillation. Basic and advanced cardiac life support is needed. Again, correcting any reversible causes may help with management.

Box 11.14 ECG distinction between supraventricular tachycardia with bundle branch block and ventricular tachycardia

Ventricular tachycardia is more likely than supraventricular tachycardia with bundle branch block where there is:

- a history of ischaemic heart disease
- a very broad QRS (>0.14 sec)
- atrioventricular dissociation
- a bifid, upright QRS with a taller first peak in V_1
- a deep S wave in V_6
- a concordant (same polarity) QRS direction in all chest leads (V_1–V_6)

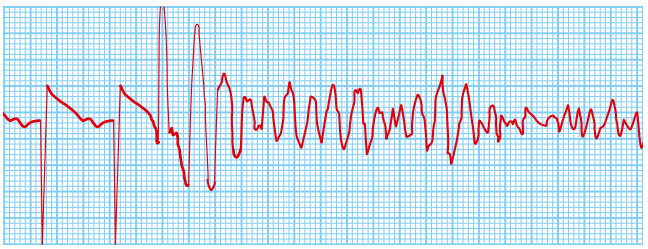


Fig. 11.17 Ventricular fibrillation. Two beats of sinus rhythm are followed by a ventricular ectopic beat that initiates ventricular fibrillation. The ST segment during sinus rhythm is elevated, owing to acute myocardial infarction in this case.

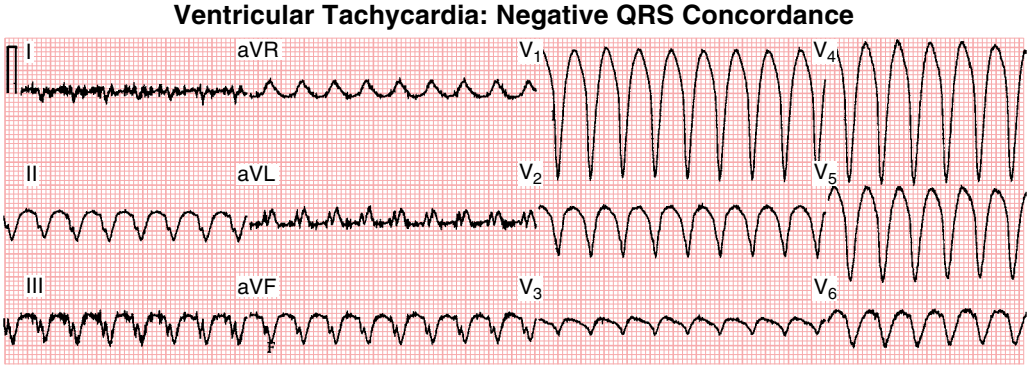


Fig. 11.16 Ventricular tachycardia.. Note the concordance (same polarity) of QRS complexes in all chest leads. *(Adapted from Goldberger's Clinical Electrocardiography: A Simplified Approach, Tenth Edition.)*

If the attack of VF occurs during the first day or two of an acute myocardial infarction, it is probable that prophylactic therapy will be unnecessary. If the ventricular fibrillation was not related to an acute infarction, the long-term risk of recurrent cardiac arrest and sudden death is high.

Survivors of these ventricular tachyarrhythmias are, in the absence of an identifiable reversible cause (e.g., acute myocardial infarction, severe metabolic disturbance), at high risk of sudden death. Implantable cardioverter–defibrillators are first-line therapy in the management of these patients (see Chapter 17).

Further reading

Resuscitation Council UK Professional resources. Accessed at: https://www.resus.org.uk/.

Anaphylaxis

Anaphylaxis is a serious allergic reaction that is rapid in onset and may cause death. It arises as an acute, generalized IgE-mediated immune reaction involving specific antigen, mast cells, and basophils. The reaction requires priming by the allergen, followed by re-exposure. To provoke anaphylaxis, the allergen must be systemically absorbed, after either ingestion or parenteral injection. A range of allergens that provoke anaphylaxis has been identified (**Box 11.15**).

Anaphylaxis is rare, but the symptom/sign constellation ranges from widespread urticaria to cardiovascular collapse, laryngeal oedema, airway obstruction, and respiratory arrest leading to death:

- Fatal reactions to penicillin occur once every 7.5 million injections.
- Between 1 in 250 and 1 in 125 individuals have severe reactions to bee and wasp stings, and a death takes place every 6.5 million stings, with 60–80 deaths per year in North America and 5–10 in the UK.

Central to the pathogenesis of anaphylaxis is the activation of mast cells and basophils, with systemic release of some mediators and generation of others. The initial symptoms may appear innocuous: tingling, warmth, and itchiness. The ensuing effects on the vasculature lead to vasodilation and oedema. The consequence may be no more than a generalized flush, with urticaria and angio-oedema. More serious sequelae are hypotension, bronchospasm, laryngeal oedema, and cardiac arrhythmia or infarction. Death may occur within minutes.

Serum platelet-activating factor (PAF) levels correlate directly with the severity of anaphylaxis, whereas PAF acetylhydrolase (the enzyme that inactivates PAF) correlates inversely and is significantly lower in peanut-sensitive patients with fatal anaphylactic reactions.

Box 11.15 Sources of allergens known to provoke anaphylaxis

Foods
- Nuts: peanuts (protein—*Arachis hypogaea* Ara h 2), Brazil, cashew
- Shellfish: shrimp (allergen Met e 1), lobster
- Dairy products
- Egg
- More rarely: citrus fruits, mango, strawberry, tomato

Venoms
- Wasps, bees, yellow-jackets, hornets

Medications
- Antisera (tetanus, diphtheria), intravenous iron, latex, some antibiotics

Management

Early recognition and treatment are essential (**Box 11.16**). Longer-term avoidance and desensitization strategies for patients who experience anaphylaxis are discussed in Chapter 6.

Further reading

BMJ Learning, 'Hospital presentations: Anaphylaxis'. Accessed at: https://new-learning.bmj.com/course/10065912.

Resuscitation Council UK: *Emergency Treatment of Anaphylactic Reactions, Guidelines for Healthcare Providers*. Accessed at: https://www.resus.org.uk/library/additional-guidance/guidance-anaphylaxis/emergency-treatment.

Chest pain

Chest pain is a common presenting symptom with a large number of potential causes (**Box 11.17**). Causes can be broken down anatomically into cardiac, vascular (aortic or pulmonary) GI, musculoskeletal, neuropathic, or respiratory. Patients with pathology below the diaphragm may also experience referred pains within the chest.

Clinical assessment

Initial emergency assessment should follow the 'ABCDE' approach to ensure that serious or life-threatening perturbations of physiology are identified promptly. In unwell patients, clinical assessment, point-of-care testing, and emergency resuscitative management measures may need to commence in parallel, with definitive treatment initiated once the diagnosis is clear.

History

- Determining the **character of pain** is vital: ischaemic pain is typically 'oppressive' in character and described as a 'heaviness', 'tightness', or 'pressure' behind the sternum and may radiate to the left shoulder, arm, or jaw. Pleuritic pain is sharp and worse on breathing and indicates pleural inflammation or irritation (commonly

Box 11.16 Treatment of acute anaphylaxis

Clinical features
- Bronchospasm
- Facial and laryngeal oedema
- Hypotension
- Nausea, vomiting, and diarrhoea

Management
- ABCDE (airway, breathing, circulation, disability, exposure)
- Position the patient lying flat with the feet raised
- Ensure the airway is free
- Give high-flow oxygen
- Monitor blood pressure
- Establish venous access
- Administer 0.5 mg *intramuscular* adrenaline (epinephrine) 0.5 mL of 1:1000 adrenaline, i.e., 1 mg/mL, and repeat after 5 min if shock persists. Note that antihistamines and steroids are not currently routinely indicated as per Resus Council UK
- If refractory, consider a peripheral low-dose i.v. adrenaline infusion 0.5–1.0 mL/kg/h with specialist guidance, and titrate to effect. Further i.m. adrenaline every 5 minutes may be needed until an i.v. infusion is established
  - If hypotension persists, give 1–2 L of intravenous fluid
  - If hypoxia is severe, assisted ventilation may be required, however, be aware that the airway may be oedematous and may become compromised
  - Take blood for tryptase levels within 1–2 h ideally (aids diagnosis)

Box 11.17 Differential diagnosis of acute chest pain

Cause	History clues	Examination clues	Investigations	Other comments
ACS (unstable angina, NSTEMI/STEMI)	Oppressive pain radiating from chest to arm, neck, or jaw. Heavy/pressure. May be atypical especially in older patients/those with diabetes	Often normal but may be sweaty (diaphoresis)	ECG: Dynamic changes may accompany pain; ST elevation/depression, LBBB, Q waves if delayed presentation. Normal ECG does not exclude ACS CXR: Often normal, may have cardiomegaly, signs of LVF Troponin may be raised if NSTEMI/STEMI (but elevations may take up to 12 hours after pain to become apparent)	Risk stratification e.g., GRACE score (Box 17.33). If NSTEMI or STEMI refer urgently to cardiology
Aortic dissection	Pain often sudden onset and radiating to the back or abdomen. May cause TLOC. May have history of aortic or connective tissue disease	Asymmetrical or absent pulses. Discrepancy in BP between limbs. Murmur of AR (EDM) may be evident if dissection involves aortic valve. Hypotension suggests tamponade or uncontained rupture and requires immediate surgery	ECG: May be normal, often shows LVH, may have similar changes to ACS if coronary blood supply is compromised CXR: Often normal, may have cardiomegaly or widened mediastinum, signs of LVF. Pericardial effusion (globular heart) if aortic root involved Troponin often elevated D-dimer often raised	Maintain an index of suspicion for aortic dissection as treating incorrectly for ACS may be fatal. May be apparent on echo, CT angiography diagnostic. Referral to surgical team imperative
Gastrointestinal (gastro-oesophageal reflux, diffuse oesophageal spasm)	Nocturnal pain, especially when lying on back/right-hand side. Acid brash in throat. Delayed response to GTN. Relief with belching. Dysphagia or dyspepsia. Relationship to meals	May have epigastric tenderness	ECG and CXR usually normal Blood tests: Usually normal, but severe anaemia could elevate troponin and cause a type II MI	Consider raising head end of bed/sleeping on left-hand side. Avoiding spicy food, alcohol, or certain medications such as NSAIDs/steroids may help Consider treatments such as alginates/H_2-receptor antagonists (e.g., famotidine) or PPIs (e.g., lansoprazole)
Musculoskeletal (e.g., costochondritis)	Pain worse on movement. History of straining or recent viral illness	Reproducible with movement of chest especially against opposition Joint tenderness may be present	ECG, CXR, and blood tests usually normal	Vitamin D replacement, if low levels noted, may be helpful for some patients Simple analgesia as treatment
Neuropathic (e.g., herpes zoster infection)	Pain over one dermatome/radiates. May feel like electric shock. Previous or current shingles	Rash or scar of shingles	ECG, CXR, and blood tests usually normal	Consider medications for neuropathic pain including amitriptyline/gabapentin/pregabalin
Pericarditis	Retrosternal sharp stabbing pain that radiates to arm or shoulder, may be relieved by leaning forwards. Cardiac surgery/MI in past 2 months may suggest Dressler syndrome	Pericardial friction rub may be heard left lower sternal border on inspiration (if patient leaning forwards)	ECG: Diffuse saddle-shaped ST elevation with accompanying PR depression CXR: May be normal, if effusion then globular cardiac silhouette Elevated troponin suggests myopericarditis and warrants further investigation	Consider uraemia, chronic inflammatory diseases e.g., SLE or familial fevers Consider aspirin/NSAIDs/colchicine. If effusion may need to exclude tamponade
Pulmonary embolism	Usually pleuritic pain if small peripheral clot, pain from a central clot may be non-specific quality	Tachypnoea, tachycardia, possibly new AF. If large clot then raised JVP and TR (right ventricular strain)	ECG: Tachycardia most common, followed by AF. Right axis deviation, RBBB may suggest large burden of clot. S1Q3T3 suggestive of PE but rarely seen CXR: Oligaemic lung fields (Westermark sign), elephant trunk pulmonary arteries may indicate pulmonary hypertension. Small pleural effusions or wedge-shaped lung infarction may be noted D-dimer positive, Troponin and BNP being raised suggestive of more severe disease	CTPA usually used, or V/Q SPECT scan considered if pregnant/contrast allergy/CKD Echo may be helpful if suspect right heart strain and consideration of thrombolysis or thrombectomy

Continued

Box 11.17 Acute causes chest pain: Differential diagnoses—cont'd

Cause	History clues	Examination clues	Investigations	Other comments
Respiratory pathology e.g., pneumonia/pleural disease	Pleuritic pain often accompanied by other symptoms of respiratory disease (e.g., cough, productive of sputum, fevers)	Pleural friction rub indicates pleural inflammation of any cause. Additional lung signs may point to cause (e.g., bronchial breathing and coarse crackles in pneumonia)	ECG usually normal CXR: Lobar collapse/consolidation, pleural effusion/reaction/ pneumothorax Raised inflammatory markers in pneumonia, also note D-dimer and troponin may be elevated in sepsis	Treatment is that of underlying disease. Colchicine may be of use in inflammatory pleural disease

ACS, Acute coronary syndrome; *AR,* aortic regurgitation; *EDM,* early diastolic murmur; *GRACE,* Global Registry of Acute Coronary Events; *LBBB,* left bundle branch block; *LVF,* left ventricular failure; *LVH,* left ventricular hypertrophy; *RBBB,* right bundle branch block; *TLOC,* transient loss of consciousness; *TR,* tricuspid regurgitation.

pneumonia, PE, or pneumothorax). A 'tearing' sensation or pain of sudden onset radiating to the back should always raise suspicion of aortic dissection. The pain of pericarditis is often sharp, may be worsened by breathing or lying flat, and is classically eased by leaning forwards. Pain from gastro-oesophageal causes is usually easily recognized ('heartburn') although diffuse oesophageal spasm may mimic cardiac ischaemic pain. Atypical pain descriptions are more common in elderly patients and those with diabetes.

- **Onset:** pain of very sudden onset often implies either a vascular (cardiac ischaemia, aortic dissection, PE) or mechanical (pneumothorax, rib fracture) cause.
- Consider **aggravating and relieving factors:** does the patient describe a prodrome of chest pain provoked by exertion and relieved by rest? (angina—although myocardial infarction (MI) is often the first manifestation of coronary artery disease). Is there a relationship to meals? (gastro-oesophageal reflux disease). Improvement in pain after sublingual GTN is often held to reflect ischaemic pain, but carries a poor predictive value for this diagnosis.
- Seek **associated symptoms:** cough with production of sputum or fever may signify pneumonia. Haemoptysis may reflect pulmonary infarction (PE) or infection.
- Consider the **past medical history:** has the patient undergone previous coronary revascularization procedures (e.g., angioplasty or stenting)? Have they had recent investigations for coronary ischaemia (e.g., angiography)? Do they have a known risk factor for PE (e.g., recent major surgery, immobility) or aortic dissection (e.g., Marfan's syndrome or history of bicuspid aortic valve)?
- History should include an assessment of **clinical risk factors for ischaemic heart disease** (including smoking history, hypertension, diabetes, obesity, and hypercholesterolaemia).

Physical examination

Examination should be systematic and include a full assessment of respiratory and cardiovascular systems (including peripheral vascular examination) along with focused examination of the abdomen and relevant neurology. Tachycardia and tachypnoea are non-specific but often signify the presence of significant cardiac or respiratory pathology. Patients with ischaemic chest pain may be diaphoretic (sweating, pale) or reveal stigmata of risk factors for coronary disease (e.g., tar staining of fingertips, xanthelasmata) but examination is often normal.

- A difference in systolic blood pressure >20 mmHg between the arms or pulse deficits should raise suspicion of aortic dissection.
- General examination should include a search for signs of atheromatous disease in other territories (absent foot pulses, vascular bruits) and DVT (asymmetrical calf swelling or tenderness). Signs of conditions which may unmask or aggravate angina (e.g., anaemia, thyrotoxicosis, aortic stenosis) should also be noted.
- Atrial fibrillation (characterized by an irregularly irregular pulse) may either complicate or aggravate (in the context of a rapid ventricular response rate) cardiac ischaemia, while AV block (bradycardia) may develop in patients with an inferior MI.
- A dermatomal vesicular rash implies shingles, although dermatomal pain may precede the onset of rash.
- Point tenderness at the margin of the sternum is typical of costochondritis, although many patients with chest pain will describe local tenderness on palpation and this should hence be interpreted with caution.
- Elevated JVP may indicate right ventricular strain (submassive PE) or compression (cardiac tamponade, tension pneumothorax) or congestive cardiac failure (e.g., complicating a myocardial infarction). These patients are typically very unwell and may be hypoxic or haemodynamically unstable.
- Patients with pleuritic chest pain may have an audible pleural rub (indicating pleural inflammation) while a pericardial rub may be audible in patients with pericarditis.
- Chest examination may yield signs consistent with a specific pulmonary diagnosis (e.g., bronchial breathing and coarse crackles in a patient with pneumonia, reduced breath sounds, and hyperresonant percussion note in a patient with pneumothorax). Care should be taken to examine all lung areas, including the apices (supraclavicular) and axillae.
- Abdominal examination will identify epigastric tenderness (sometimes interpreted by a patient as 'chest' pain) and possible sources of subdiaphragmatic referred pain (e.g., cholecystitis, pancreatitis).

Investigations

- **Oxygen saturations** will identify a hypoxic patient at the bedside.
- **Core laboratory tests** may reveal a low haemoglobin (anaemia), elevated white blood cell count and CRP (infection), or metabolic abnormalities. Renal function testing is required for GRACE score calculation (a widely used ACS risk model, **Box 17.33**). A pregnancy test should be conducted in premenopausal female patients.
- **ECG** is a high-yield investigation in patients presenting with chest pain. Careful examination may disclose signs of ischaemia (ST elevation or depression, new LBBB or T-wave inversion), pericarditis (diffuse saddle-shaped ST elevation accompanied by PR segment depression), or PE (right axis deviation, RBBB or S1Q3T3 pattern). ACS may be complicated by both tachy- or bradyarrythmias. Serial ECG monitoring (initially every 15 min) is indicated in patients with persistent ischaemic pain.
- **Cardiac troponins** rise rapidly (hours) in myocardial infarction and remain elevated for several days. Levels correlate with cardiac myocyte damage and positive predictive value (PPV) for

myocardial infarction increases with higher levels (levels >5× the upper reference limit have a PPV >90% for acute MI). Serial levels will discriminate acute cardiac injury (rising troponin level) from chronic cardiomyocyte damage (static levels). New high-sensitivity troponin T assays have enabled development of rapid 'rule-out' algorithms in low-risk individuals although protocols still vary widely between hospitals. Importantly, troponin rises are not specific for cardiac ischaemia and may be seen when cardiac myocyte damage occurs for other reasons (e.g., myocarditis or PE).

- **D-dimer testing** may be indicated (see section 'Deep vein thrombosis and pulmonary embolism').
- **Chest X-ray** (ideally postero-anterior) may reveal consolidation (pneumonia), mediastinal widening (aortic dissection), and signs of congestive cardiac failure or pneumothorax. Globular heart enlargement suggests presence of a pericardial effusion. Bones and soft tissues should be inspected carefully for fractures or presence of surgical emphysema.
- **POCUS or bedside echo** could be used to look for pneumothoraces, pleural effusions, consolidation, dissections, biliary dilatation or right ventricular overload, and pericardial disease. In experienced hands, it may also support assessment of intravascular filling or in diagnosis of a DVT in a patient with possible PE.
- **Further testing** may include departmental echocardiography (to assess for left ventricular function, the presence of regional wall motion abnormalities or pericardial effusion), CT scanning (e.g., diagnosis of aortic dissection or pulmonary embolus), or emergency coronary angiography (STEMI or very high-risk NSTE-ACS).

Management

Management of the patient with chest pain varies with the underlying cause.

Patients with ischaemic chest pain require continuous cardiac monitoring and often need intravenous analgesia and antiemetics for symptom control. Patients predicted to require emergency coronary angiography (STEMI or very high-risk NSTE-ACS) are often transported directly by paramedic crews to cardiac hubs, but otherwise require immediate discussion with local primary PCI services. Aspirin (300 mg) is indicated in all patients with ACS and often administered by paramedics.

Management of acute coronary syndromes is discussed in Chapter 17 (see 'Acute coronary syndrome' section).

Palpitations

Palpitations are a common presenting problem across all age groups and are a frequent reason for referral to acute medical services. Palpitations can be defined as the awareness of a hard, irregular, or rapid heartbeat. A key factor in patients deciding to seek medical advice is that this subjective awareness of cardiac activity is experienced as unpleasant or distressing. Palpitations may be benign or physiological (e.g., in stressful situations or pregnancy), or pathophysiological (e.g., arrhythmias). The majority of patients experiencing palpitations have a very good prognosis and so the challenge of the acute physician is to identify the minority who have an underlying arrhythmia or other significant disorder.

Differential diagnosis

Causes of palpitations can be classified as either arrhythmic or non-arrhythmic (**Box 11.18**). The latter includes any cause of sinus tachycardia. Arrhythmic causes of palpitations each, in turn, have their own list of potential causes (e.g., for atrial fibrillation, causes may include ischaemia, sepsis, thyrotoxicosis, alcohol, electrolyte disturbances, etc.).

? Box 11.18 Causes of palpitations

Arrhythmic	Non-arrhythmic (usually sinus tachycardia)
Sinus arrhythmia (rarely, usually asymptomatic)	Cardiopulmonary • Heart failure • Myocarditis • Myocardial infarction • Pulmonary embolus • Hypoxia (any cause)
Ectopic beats (atrial or ventricular)	Anaemia
Vasovagal syncope (more commonly experienced as dizziness)	Endocrine • Thyrotoxicosis • Phaeochromocytoma • Carcinoid syndrome
Tachyarrhythmias • Atrial fibrillation or flutter • Atrial tachycardia • Supraventricular tachycardias (including pre-excitation syndromes such as Wolff-Parkinson-White) • Ventricular tachycardia • Torsade de pointes • Ventricular fibrillation (the patient will not be conscious at presentation)	Drugs • Caffeine • Recreational stimulants (e.g., cocaine, amfetamine) • β_2 agonists (e.g., salbutamol), certain decongestants (e.g., pseudoephedrine) • Nicotine • Alcohol or therapeutic drug withdrawal)
Bradyarrhythmias (less commonly felt as palpitations than tachyarrhythmias) • Sick sinus syndrome • Heart (AV) block	Physiological • Stress/anxiety • Fever • Pregnancy

Clinical assessment

As with all medical emergencies, an initial focused assessment following the ABCDE approach should proceed in parallel with information gathering. In the case of ongoing tachy- and bradyarrhythmias accompanied by haemodynamic compromise ('peri-arrest rhythms'), immediate interventions are required (see 'Therapeutic procedures used for peri-arrest rhythms').

When taking a history, care should be taken to ascertain precisely what the patient means by 'palpitations'; in some cases, careful probing may establish that the patient is in fact experiencing another symptom for example, breathlessness. **Box 11.19** outlines helpful questions to ask patients who present with palpitations.

Investigations

- **Laboratory investigations:** Full blood count (exclude anaemia), urea and electrolytes, CRP (infection), thyroid function tests (TFTs).
- **12 lead ECG:** A high-yield investigation. Where the palpitations are ongoing, this will demonstrate the associated cardiac rhythm. In patients where the palpitations have resolved, look for electrical abnormalities which may point towards the diagnosis (e.g., the short PR interval and 'delta wave' of Wolff-Parkinson-White syndrome; signs of conduction disease such as prolonged PR interval or bundle branch block).
- **24- or 48-hour Holter monitor** or 7-day event recorder depending on frequency. Required for patients with intermittent

Box 11.19 Questions for patients presenting with palpitations

1. What were you doing at the time of onset (or in the hours before)? Including an enquiry about potential precipitants including caffeine, prescription and recreational drug use.
2. Are the palpitations constant or do they come and go?
Many patients are asymptomatic by the time of presentation. Ongoing symptoms of 'palpitations' in the absence of tachycardia may indicate inaccurate labelling. A sensation of transient 'jumping' or 'jolting' may indicate ectopic beats.
3. Did the palpitations start suddenly or gradually? Did they stop suddenly or gradually?
Paroxysmal arrhythmias are typically of sudden onset. This is particularly true of SVTs, where both onset and 'offset' may be abrupt. Palpitations due to sinus tachycardia often build up and resolve more gradually.
4. Can you describe the palpitations or 'tap out' the rhythm? Was it regular or irregular?
This is particularly helpful in discriminating paroxysmal atrial fibrillation from regular tachycardias.
5. Any 'red flag' symptoms e.g., syncope or presyncope, chest pain, breathlessness. Or other associated symptoms (e.g., tremor, flushing, diarrhoea, wheezing)?
The presence of red flag symptoms points towards a significant underlying disorder (usually an arrhythmia). Other symptoms may point towards a systemic disorder (e.g., thyrotoxicosis, carcinoid syndrome, or phaeochromocytoma).
6. What is happening in your life at present? (stress, pregnancy, other illness)
7. Is there any personal or family history of cardiac disease, or family history of sudden cardiac death?

symptoms where a reasonable index of suspicion exists for a paroxysmal arrhythmia.
- Consider **echocardiogram** (if personal or family history of cardiac disease, presence of red flags in history, audible murmur, or other signs of structural heart disease)
- Other targeted investigations may be required both to investigate the cause of an arrhythmia if one is identified, or if clinical assessment has suggested another specific cause of palpitations (e.g., phaeochromocytoma).

Management

Management varies with underlying cause.

Patients with 'red flag' symptoms (syncope, chest pain, breathlessness) where the cause is not immediately identified or rectifiable, will typically need admission for a period of observation with continuous recorded monitoring of cardiac rhythm (telemetry).

'Low-risk' patients in whom palpitations have resolved, do not have red flag symptoms and with a normal 12-lead ECG can usually be investigated further via outpatient ambulatory pathways.

Management of peri-arrest rhythms and ventricular tachyarrhythmias is discussed earlier in the chapter (see 'Therapeutic procedures used for peri-arrest rhythms').

Atrial fibrillation

Atrial fibrillation (AF) is commonly seen on the acute medical take either in isolation, or more commonly secondary to other acute illnesses. The context in which AF presents will determine its management, and treating the underlying driving cause is often as important as measures directed towards the arrythmia itself.

For a description of the causes, pathogenesis, classification, and clinical features, please refer to, 'Atrial fibrillation' section in Chapter 17. The European Society of Cardiology suggest an ABC mnemonic for management of atrial fibrillation (**A**void stroke, **B**etter symptom control, manage **C**ardiovascular risk factors, and detect **C**oncomitant diseases).

Acute management

When confronted with a patient with acute-onset atrial fibrillation (see **Fig. 17.42A**), considerations should include identifying and treating any precipitant or underlying cause, controlling symptoms (either by slowing ventricular rate or cardioverting to sinus rhythm, as appropriate), and minimizing risk of stroke due to embolization of left atrial thrombus.

Treating the underlying cause

When atrial fibrillation is due to an acute precipitating event, such as alcohol toxicity, chest infection, or hyperthyroidism, the provoking cause should be treated which will often result in spontaneous cardioversion. In the absence of life-threatening features, rate control strategies are thus usually preferred, although in patients with critical illness and/or severe left ventricular systolic dysfunction, i.v. amiodarone (which has both rate and rhythm-controlling properties) may be used.

Controlling symptoms

AF accompanied by a rapid ventricular rate causes unpleasant symptoms which may interfere with patients' activities and if sustained may lead to tachycardia-induced cardiomyopathy. Likewise, older patients with other associated co-morbidities may tolerate tachycardias poorly if sustained.

Strategies for the acute management of atrial fibrillation are:
- ***Ventricular rate control***, achieved by drugs that block the AV node.
- ***Cardioversion***, achieved electrically by synchronized DC shock (see 'Therapeutic procedures used for peri-arrest rhythms'), or medically by intravenous infusion of an antiarrhythmic drug such as flecainide or amiodarone.

The choice of whether to deploy rate or rhythm control for a patient presenting acutely depends on:
- how well the arrhythmia is tolerated (is cardioversion urgent?)
- the time of onset of the arrythmia (more or less than 48 h—in practice this may be uncertain, in which case proceed as if onset was more than 48 h prior)
- whether spontaneous cardioversion is likely (previous history? reversible cause?)—if likely, a rate-control strategy should be favoured.

Patients with life-threatening haemodynamic instability

When caused directly by new-onset atrial fibrillation, life-threatening features including shock, syncope, myocardial ischaemia, or severe heart failure are indications for emergency synchronized DC cardioversion. While action should still be taken to minimize risk of embolic stroke as best possible (see later), the balance of benefits and risks favours immediate restoration of sinus rhythm.

Patients without life-threatening haemodynamic instability

Ideally, expert help should be sought (e.g., from a cardiologist) to enable the highest quality decision-making; numerous factors influence treatment choice in atrial fibrillation (including patient age and co-morbidities, presence or absence of left ventricular dysfunction or other structural heart disease, concern for pre-excited atrial

fibrillation conducted via an accessory pathway and important medication contraindications such as asthma precluding beta-blocker use, see 'Atrial fibrillation (rate control) section' in Chapter 17.

In patients with a clear onset of new AF symptoms **with onset less than 48 hours**, the risk of left atrial thrombus formation is deemed to be low and either rate or rhythm control strategies may be used.

In patients with symptom onset **more than 48 hours previous (including symptoms suggesting paroxysmal AF), or where the time of onset is unclear**, rate control should be employed due to the unacceptable risk of left atrial thrombus, embolism of which may be precipitated by reversion to sinus rhythm.

Rate control may be achieved with beta-blockers (e.g., metoprolol or bisoprolol), non-dihydropyridine calcium channel antagonists (e.g., verapamil), or digoxin. The choice of intravenous or oral route should be made based on the perceived urgency of rate control, tolerance of symptoms, and factors which may impact bioavailability (e.g., unavailability of the oral route after surgery). Beta-blockers are contraindicated in asthma and avoided in AHF/pulmonary oedema. Calcium channel antagonists should not be used in patients with heart failure and a reduced ejection fraction. Renal function should be checked prior to digoxin use, and dose reduction may be required.

The optimal heart rate target in patients with AF is unclear, although clinical trials indicate that a lenient target (<110 bpm) is not inferior to a strict target (<80 bpm) and the former is thus a reasonable goal for initial management.

Patients managed initially with rate control may subsequently be considered for elective DC cardioversion after a period of at least 3 weeks' therapeutic anticoagulation, which is successful in approximately 80% of patients. If cardioversion is urgent and the patient is not on any anticoagulation, some expert centres will perform transoesophageal echocardiography to exclude the presence of atrial thrombus to facilitate an earlier procedure.

Rhythm control (cardioversion to sinus rhythm) in a patient without emergency indications for DC cardioversion, is usually attempted by pharmacologic means. Options include flecainide (contraindicated in patients with ischaemic heart disease or significant structural heart disease including left ventricular hypertrophy), propafenone (to be avoided in patients with hypotension, recent acute coronary syndromes, severe heart failure or aortic stenosis, or prolonged QT interval), or amiodarone (usually the preferred agent in heart failure). Amiodarone is usually administered i.v. but may cause phlebitis, and should be given via a central venous catheter or large peripheral vein. Baseline thyroid function tests should be checked. Flecainide is effective by both oral and i.v. routes, and is a common 'pill in the pocket' solution for patients with repeated episodes of AF in the absence of structural heart disease.

Longer-term, cardiac ablation (see Chapter 17) is increasingly used as a rhythm control strategy in AF.

Aside from duration of symptoms, factors which may influence the decision to pursue rhythm or rate control may include younger age, first episode, absence of co-morbidities (favouring rhythm control), or the presence of cardiac or non-cardiac disease likely to favour further episodes (favouring rate control).

Preventing stroke

AF is associated with a fivefold increased in stroke risk, attributable to the formation and embolism of left atrial thrombus and risk of clot formation rises with time spent in AF.

Current UK NICE guidelines recommend that in the absence of contraindications, patients presenting with acute AF are offered treatment with low-molecular-weight heparin, pending further risk stratification and risk/benefit assessment. CHA_2DS_2VASc scores (see **Box 17.13**) and HAS-BLED score (see **Box 17.14**) should subsequently be used to compute individualized thrombosis and bleeding risks in all patients.

Onset less than 48 hours

Where the patient fulfills criteria for anticoagulation based on CHA_2DS_2VASc score and benefit is considered to outweigh risks, the following patients should be offered ongoing anticoagulation:

- those in whom sinus rhythm is not restored within the same 48-hour period
- those with a high risk for recurrence of AF (e.g., structural heart disease, previous recurrences).

DOACs are now generally the preferred agents for anticoagulation in non-valvular AF; vitamin K antagonists are reserved for those with contraindications or valvular AF (defined as prosthetic heart valves or moderate-to-severe mitral stenosis).

Onset more than 48 hours or uncertain

All such patients who fulfill criteria for anticoagulation based on CHA_2DS_2VASc score and benefit is considered to outweigh risks, should be offered ongoing anticoagulation. For patients in whom future elective DC cardioversion is planned, full anticoagulation with a DOAC or warfarin (with confirmation that the patient has attained a therapeutic INR of 2–3) is required for at least 3 weeks prior to the procedure and 4 weeks after the procedure, regardless of CHA_2DS_2VASc score.

Atrial fibrillation with pre-excitation

AF with pre-excitation results from antegrade atrioventricular conduction via an accessory pathway and may be the presenting manifestation of Wolf-Parkinson-White (WPW) syndrome. Pre-excited AF is always an emergency and may precede ventricular fibrillation and sudden death. Consider this diagnosis in young patients with AF, with very rapid ventricular rates (often >200 bpm) and broad QRS complexes, the morphology of which often changes from beat to beat, reflecting the aberrant pathway of conduction (see **Fig. 17.41D**). Baseline ECGs (to identify typical WPW morphology) are rarely available, and thus a high index of suspicion is required. Importantly, **all AV nodal-blocking drugs (including beta-blockers, verapamil, and digoxin) should be avoided, as they potentially enhance accessory pathway conduction**. Amiodarone may also enhance pathway conduction. The European Society of Cardiology recommends ibutilide (a class III antiarrhythmic), although this drug is not available in the UK and flecainide is most commonly used under expert guidance. Emergency DC cardioversion facilities should be readily available.

Atrial flutter

Management of a symptomatic acute paroxysm of atrial flutter (see **Fig. 17.42B**) is by electrical ***cardioversion***. Patients who have been in atrial flutter for more than 1–2 days should be treated in a similar manner to those with atrial fibrillation and require ***anticoagulation*** for 3 weeks prior to cardioversion, using AV nodal modulating drugs (beta-blockers, calcium channel antagonists) to control rate as best possible in the interim. Acute pharmacological cardioversion can be achieved using class Ic (flecainide, propafenone) or certain class III antiarrhythmic agents (dofetilide, ibutilide; these have better efficacy than in atrial fibrillation but are not available in many countries).

In the long term, catheter ablation (Chapter 17) is highly effective in interrupting flutter pathways and preventing episodes.

Paroxysmal supraventricular tachycardia

Although technically speaking, all arrythmias arising from the conduction system above the ventricles are 'supraventricular' in origin (including atrial arrythmias, see **Box 17.12**), the term 'paroxysmal supraventricular tachycardia' (or simply 'SVT') is usually used in reference to tachycardias arising from the atrioventricular junction: either AV nodal re-entrant tachycardias (AVNRT, where the re-entry circuit is located within the AV node or AV re-entrant tachycardias (AVRT, where a macroscopic 'accessory pathway' for re-entry is present, see 'Atrioventricular junctional tachycardias' section in Chapter 17). These most commonly affect young people, often with age of first presentation <30 years and usually no or minimal underlying structural heart disease.

Clinical features

The leading symptom of most SVTs is rapid regular palpitations, usually with abrupt onset and sudden termination, which can occur spontaneously or be precipitated by simple movements. A common feature is termination by Valsalva manoeuvres. In younger individuals with no structural heart disease, the rapid heart rate can be the main pathological finding. Other symptoms may include anxiety, dizziness, dyspnoea, neck pulsation, central chest pain, and weakness. Polyuria may occur because of the release of atrial natriuretic peptide in response to increased atrial pressures during the tachycardia. Prominent jugular venous pulsations due to atrial contractions against closed AV valves may be observed during AVNRT.

Syncope has been reported in 10%–15% of patients, usually just after initiation of the arrhythmia or in association with a prolonged pause following its termination. It is more common if the person is standing. However, in older patients with concomitant heart disease, such as aortic stenosis, hypertrophic cardiomyopathy, and cerebrovascular disease, significant hypotension and syncope may result from moderately fast ventricular rates.

Management

In an emergency, distinguishing between AVNRT and AVRT may be difficult but is usually not critical, as both tachycardias respond to the same treatment. Patients presenting with SVTs and haemodynamic instability (e.g., hypotension, pulmonary oedema) require emergency cardioversion. If the patient is haemodynamically stable, vagal manoeuvres, including right carotid massage (see **Box 17.8**), the Valsalva manoeuvre, and facial immersion in cold water, can be successfully employed.

If physical manoeuvres have not been successful, intravenous adenosine should be tried. Give 6 mg by i.v. push initially, followed if needed by 12 mg 1–2 minutes later. If required, a second dose of 12 mg can be given after a further 1–2 minutes. Adenosine is a very short-acting (half-life <10 sec), naturally occurring purine nucleoside that causes complete heart block for a fraction of a second following intravenous administration. It is highly effective at terminating AVNRT and AVRT or unmasking underlying atrial activity, but rarely affects ventricular tachycardia. The side-effects of adenosine are very brief but may include chest pain, sense of impending doom, bronchospasm, flushing, or heaviness of the limbs. Adenosine should be used with caution in patients with a history of asthma. An alternative treatment is verapamil 5–10 mg i.v. over 5–10 min or beta-blockers (esmolol, propranolol, metoprolol). Verapamil must not be given after beta-blockers *or* if the tachycardia presents with broad (>0.12 sec) QRS complexes.

The longer-term management of junctional tachycardias is discussed under 'Long-term management' subsection of 'Atrioventricular junctional tachycardias' section in Chapter 17.

Further reading

European Society of Cardiology 2020 Guidelines for Management of Atrial Fibrillation. Accessed at: https://www.escardio.org/Guidelines/Clinical-Practice-Guidelines/Atrial-Fibrillation-Management.

Breathlessness

Breathlessness (dyspnoea) is one of the most common presenting symptoms encountered on the acute medical take. It is usually experienced as difficult or laboured breathing and can result from a range of underlying medication conditions (**Box 11.20**). Respiratory and cardiovascular disorders predominate with other potential causes including severe anaemia, neuromuscular disease, obesity, deconditioning, and rarely functional breathing disorders. Care should be taken to differentiate true dyspnoea from other symptoms which a patient may report as 'breathlessness' (e.g., palpitations).

Clinical assessment

Evaluation should begin with a **focused history**, although emergency measures (e.g., application of supplemental oxygen if the patient is hypoxic) may need to commence in parallel.

- **Onset and duration of symptoms:** sudden onset dyspnoea may implicate a 'mechanical' (e.g., pneumothorax) or vascular cause (e.g., acute embolism [PE], or acute pulmonary oedema resulting from myocardial infarction or arrhythmia)
- **Aggravating or relieving factors:** dyspnoea due to left ventricular failure usually worsens when lying flat (orthopnoea); dyspnoea due to acute asthma may be eased by inhaled or nebulized bronchodilators
- Identify associated symptoms:
  - **Pleuritic chest pain** (worse on deep inspiration) suggests pneumonia, pneumothorax, or PE
  - **Wheeze** may reflect bronchospasm, or accompany pulmonary oedema ('cardiac asthma')
  - Expectoration of **purulent sputum** suggests bronchitis or pneumonia

Box 11.20 Common emergency causes of breathlessness

Respiratory

- Acute asthma exacerbation
- COPD exacerbation
- Pulmonary infection (bronchitis/pneumonia)
- Viral pneumonitis (COVID-19, influenza)
- Pneumothorax
- Pleural effusion
- Upper airway obstruction (e.g., aspiration, anaphylaxis)

Cardiovascular

- Heart failure
- Pulmonary embolism
- Cardiac tamponade

Other

- Anaemia
- Obesity
- Deconditioning
- Metabolic acidosis ('air hunger' e.g., in DKA)
- Panic attack
- Neuromuscular disorders (e.g., myasthenia gravis)

- **Haemoptysis** may imply infection, infarction (i.e., PE), pulmonary oedema, malignancy (e.g., a bronchial tumour causing lung collapse), or rarely pulmonary haemorrhage

- Enquire about relevant medical history, including pre-existing respiratory conditions, cardiac diseases, or other co-morbidities that may contribute to breathlessness. Does the patient have home oxygen, nocturnal CPAP, or BiPAP (NIV)?
- Ask about previous episodes and their severity, including any times when the patient may have required admission to ITU and any interventions required (e.g., intubation). Much information may be gathered from a thorough review of old notes, although it is important to keep an open mind (e.g., in a patient with frequent hospitalizations due to COPD exacerbations who presents with sudden dyspnoea, consider PE or pneumothorax)
- Additionally, ask about smoking history, recent travel, occupational exposures, pets, and social history

The **physical examination** should begin with a detailed respiratory examination, looking for digital clubbing, central or peripheral cyanosis, signs of deep venous thrombosis (DVT), presence of tracheal deviation, quality of breath sounds, the presence of wheezing or crackles, and signs of respiratory distress. Evaluate the cardiovascular system, checking for pallor, oedema, signs of heart failure, and murmurs. Depending on the suspected cause of breathlessness, examine other systems as appropriate.

- Presence of a polyphonic expiratory wheeze suggests bronchospasm (acute asthma or some exacerbations of COPD), or pulmonary oedema ('cardiac asthma')
- A gallop rhythm (third heart sound), elevated jugular venous pressure (JVP), and fine bi-basal inspiratory crackles suggest acute left ventricular failure
- Bronchial breathing, coarse crackles, and fever suggest pneumonia
- Hypoxia and tachycardia despite a clear chest should raise suspicion of PE
- Reduced breath sounds on one side accompanied by a 'stony dull' percussion note suggests a pleural effusion; reduced breath sounds accompanied by a hyper-resonant percussion note suggest pneumothorax
- Muffled heart sounds with hypotension and neck vein engorgement suggest pericardial effusion with tamponade
- Drowsiness, bounding pulse, and flapping tremor (asterixis) suggest the presence of type II respiratory failure
- Neurological weakness may signify a neuromuscular disorder
- Pallor may suggest anaemia

Investigations

- **Oxygen saturations** will identify a hypoxic patient at the bedside
- **Peak expiratory flow rate** (PEFR) is essential in asthma exacerbation to help quantify severity
- **ABG** will quantify hypoxia and is essential to detect type II respiratory failure (hypercapnia). ABG also allows detection of metabolic acidosis (e.g., DKA) and provides other point-of-care laboratory data (e.g., haemoglobin)
- **COVID-19/influenza lateral flow or antigen test**
- **ECG:** This may reveal arrhythmia or ischaemia (as a cause of acute pulmonary oedema). Low voltage complexes should raise suspicion of pericardial effusion. A right ventricular strain pattern, right bundle branch block, or the 'S1, Q3, T3' pattern may be present in PE, although all are insensitive for the diagnosis, and sinus tachycardia may be the only finding
- **CXR:** This is a high-yield investigation in the breathless patient and will provide confirmation of a range of diagnoses including pneumonia, pulmonary oedema, lobar collapse, and pleural effusion (see Chapter 8). Think of PE in the hypoxic patient with a normal CXR. Pneumothorax may be subtle: turning the CXR film 90 degrees may help in detecting the boundary between lung edge and non-aerated lung
- **Core laboratory tests** may reveal a low haemoglobin (anaemia), elevated white blood cell count and C-reactive protein (infection), or metabolic abnormalities (e.g., renal failure as a cause of pulmonary oedema)
- **POCUS:** This can be useful in identifying pneumothorax: look for lack of lung sliding, absence of comet-tails, and the 'lung point' sign. POCUS can also be used to assess for pleural effusion, pulmonary oedema (B lines), or evidence of right ventricular strain in submassive PE
- **Further diagnostic tests** may include computed tomography pulmonary angiogram (CTPA) or ventilation–perfusion (VQ) scan if PE is suspected or echocardiography, to assess left ventricular function and valvular integrity in acute pulmonary oedema and detect pericardial effusion

Management of the breathless patient varies with the underlying cause, although the ABCDE (airway, breathing, circulation, disability, exposure) approach to initial emergency care is common to all scenarios and should be applied in parallel to the clinical evaluation and diagnostic work-up.

Acute severe asthma

The term acute severe asthma is used to mean an exacerbation of ('bronchial') asthma that has not been controlled by the use of standard medication (inhaled bronchodilators ± corticosteroids). Patients frequently have a prior diagnosis of chronic asthma, although acute severe asthma may, on occasion, be the first presentation of disease. A discussion of the epidemiology and pathophysiology of bronchial asthma can be found in Chapter 20.

Clinical assessment

Assess for evidence of respiratory distress, air entry, wheeze, and for tracheal deviation (superadded tension pneumothorax).

Patients with **acute severe asthma** typically have:

- an inability to complete a sentence in one breath
- a respiratory rate of ≥25 breaths/min
- a tachycardia of ≥110 b.p.m. (pulsus paradoxus is not useful, as it is present in only 45% of cases)
- a PEFR of 33%–50% of predicted normal or best.

Features of **life-threatening attacks** are:

- a silent chest, cyanosis, or feeble respiratory effort
- exhaustion or altered level of consciousness
- bradycardia, hypotension, or arrhythmia
- a PEFR of <33% of predicted normal or best (often <150 L/min in adults) or S_pO_2 of <92%.

Investigations

- All patients should have their initial observations measured and supplementary oxygen offered to any patient who is hypoxic.
- Bedside PEFR is important in defining severity and monitoring treatment response.
- Send respiratory viral swab and COVID-19 and influenza rapid tests, consider atypical pneumonia screen if appropriate depending on the history.
- CXR should be performed in all those requiring admission to exclude pneumothorax and pneumonia.

Arterial blood gases should only be measured in asthmatic patients requiring admission to hospital who look to be tiring, with particular attention paid to the P_aCO_2. It is important to note that a 'normal' P_aCO_2 (4.6–6 kPa, 35–45 mmHg) is a worrying feature in an asthmatic patient suggesting the patient is tiring and is a marker of a life-threatening attack. ***Pulse oximetry*** is useful in monitoring oxygen saturation during the admission and can reduce the need for repeated arterial puncture.

Features suggesting very severe life-threatening attacks are:

- a high P_aCO_2 of >6 kPa
- severe hypoxaemia: P_aO_2 < 8 kPa despite treatment with oxygen
- a low and/or falling arterial pH.

Management (see Box 11.21)

Management consists of nebulized short-acting bronchodilators; nebulized antimuscarinics (e.g., ipratropium bromide) are also helpful. Intravenous hydrocortisone should be given together with prednisolone (40–50 mg daily) orally. When treating an acute severe exacerbation of asthma, an initial dose of oral prednisolone should be given alongside intravenous hydrocortisone. Failure to do so is a common cause of 'late deterioration' due to the shorter half-life of hydrocortisone (6–8 h). If symptoms are not controlled, consider a single dose of intravenous magnesium sulphate (1.2 g–2 g infusion over 20 min).

Intravenous β_2-adrenoceptor agonists or aminophylline may be considered. Ventilation is required for patients who deteriorate despite this initial regimen.

If the patient is improving, discharge can be considered if PEFR is >75% of best or predicted and PEFR diurnal variability <25%. Inhaler technique should be assessed prior to discharge and the patient should have inhaled corticosteroid therapy started if new diagnosis (or inhaler reviewed if repeated admissions in a short period of time). In the UK, this is often guided by specialist respiratory nurses who will also provide the patient with their own peak flow metre and asthma action plan.

Box 11.21 Treatment of acute severe asthma

At home

1. The patient is assessed. Tachycardia, a high respiratory rate and inability to speak in sentences indicate a severe attack.
2. If the peak expiratory flow rate (PEFR) is <150 L/min (in adults), an ambulance should be called.
3. Nebulized salbutamol 5 mg or terbutaline 10 mg is administered.
4. Hydrocortisone 200 mg i.v. is given.
5. Oxygen 40%–60% is given if available.
6. Prednisolone 40–50 mg is given orally.

In hospital

1. The patient is reassessed.
2. Oxygen 40%–60% is given.
3. The PEFR and O_2 saturation are measured.
4. Nebulized salbutamol 5 mg or terbutaline 10 mg is repeated and administered 4-hourly.
5. Nebulized ipratropium bromide 0.5 mg is added to nebulized salbutamol/terbutaline.
6. Hydrocortisone 200 mg i.v. is given.
7. Prednisolone is continued at 40–60 mg orally daily for at least 5 days.
8. Arterial blood gases are measured; if the P_aCO_2 is >8 kPa, ventilation may be required.
9. A chest X-ray is performed to exclude pneumothorax.
10. If there is no improvement, i.v. magnesium sulphate is given at 1.2–2 g over 20 min.
11. If there is still no improvement, urgent transfer to the intensive therapy unit is arranged.

The patient should be discharged with a minimum of 5 days of oral prednisolone 40 mg once a day. Consider prescribing a proton pump inhibitor as gastric protection particularly in older patients. The patient should see their GP within 2 working days and should have follow-up with a respiratory physician within 4 weeks.

Further reading

https://www.brit-thoracic.org.uk/quality-improvement/guidelines/asthma/.

Exacerbation of chronic obstructive pulmonary disease (COPD)

Acute exacerbations of COPD may be precipitated by a viral or bacterial infection. Patients may have symptoms of cough, wheeze, and dyspnoea. Type I or type II respiratory failure (the latter often acute-on-chronic) may occur as a consequence of a COPD exacerbation.

Many patients will have a prior diagnosis of COPD (confirmed by spirometry) or have previously been hospitalized for exacerbations. Underlying COPD should be suspected in a patient with a significant smoking history (>20 pack/years) and a background history of *progressive* dyspnoea with reduction in exercise tolerance.

A discussion of the epidemiology and pathophysiology of COPD can be found in Chapter 20.

Clinical assessment

Clinical signs of COPD exacerbation vary widely and in particular the presence of wheeze, which ranges from severe to absent. Patients may use accessory muscles of respiration and signs of thoracic hyperinflation may be present (e.g., increased cricoid-sternal distance). The traditional paradigms of disease phenotype (the 'pink puffer'—cachexic with marked tachypnoea and pursed-lip breathing—and the 'blue bloater'—cyanosed and oedematous, resulting from cor pulmonale and right heart failure) do correlate loosely with propensity to type I and type II respiratory failure, respectively, but are insufficiently sensitive to guide clinical decisions.

Investigations

- PEFR is of limited use in assessing severity of COPD exacerbation
- **ABG** is essential in order to detect type II respiratory failure
- **Viral respiratory swab, COVID-19 and influenza rapid testing**, and **sputum** cultures should be sent
- **ECG** will detect concomitant atrial fibrillation (common in this group)
- **CXR** will exclude pneumonia and pneumothorax and may yield incidental diagnosis (e.g., bronchial malignancy). An appropriate index of suspicion should be maintained for PE as risk factors (e.g., immobility) commonly coexist in this population

Management (see also Fig. 11.19)

- ***Oxygen therapy.*** COPD is by far the most common cause of respiratory failure. In managing respiratory failure, the main goal is to improve the P_aO_2 by continuous oxygen therapy. A fixed-percentage mask (Venturi mask, Fig. 11.18) is used to deliver controlled concentrations of oxygen. Initially, 24% oxygen is given, and the concentration of inspired oxygen can be gradually increased, provided the P_aCO_2 does not rise unacceptably. In chronic type II respiratory failure, recognized by elevated P_aCO_2 and plasma bicarbonate (indicating renal compensation), application of *excessive* supplemental oxygen will nearly always cause a further rise in P_aCO_2. Historically this was ascribed to dependency on 'hypoxic respiratory drive' to

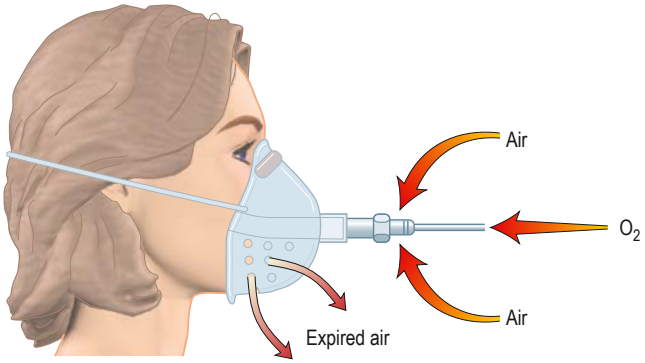


Fig. 11.18 'Fixed-performance' device for administration of oxygen to spontaneously breathing patients (Venturi mask). Oxygen is delivered through the injector of the Venturi mask at a given flow rate. A proportionate amount of air is entrained and the inspired oxygen can be predicted accurately. Masks are available that deliver 24%, 28%, and 35% oxygen.

maintain adequate ventilation, although displacement of carbon dioxide from haemoglobin by oxygen and a loss of pulmonary hypoxic vasoconstriction (resulting in increased pulmonary 'dead space') are more plausible mechanisms. Patients at risk of hypercapnia (particularly those with chronic type II respiratory failure, on long-term oxygen therapy, or with previous need for NIV or invasive ventilation) should be managed with controlled oxygen therapy to maintain the saturations within a target range of 88%–92%. It is important to monitor arterial blood gases closely if there is any risk of decompensated type II respiratory failure. If there is evidence of respiratory acidosis (pH < 7.35 with an elevated P_aCO_2), despite medical management, the patient should be considered for NIV unless there are any contraindications (**Fig. 11.19**).

- ***Corticosteroids, antibiotics, and bronchodilators.*** These should be administered in the acute phase of an exacerbation but decisions on long-term use should wait until the patient has recovered.
- ***Removal of retained secretions.*** Patients should be encouraged to cough up secretions. Chest physiotherapy is helpful in achieving adequate chest clearance.

Type II respiratory failure in COPD

- ***Respiratory support:*** BiPAP (see 'Advanced respiratory support') should be offered if a patient has a persistent respiratory acidosis with a pH of <7.35. Randomized controlled trials have demonstrated that BiPAP reduces the need for intubation and lowers mortality. Indications and contraindications are shown in **Boxes 11.11 and 11.12**. Assisted ventilation with an endotracheal tube is occasionally necessary for patients with COPD who have severe respiratory failure but only when there is a clear precipitating factor and the overall prognosis is reasonable. Assessing the likelihood of reversibility in an acute setting can present a difficult ethical problem.

Acute heart failure

Acute heart failure (AHF) occurs with the rapid onset of symptoms and signs of heart failure secondary to abnormal cardiac function, causing elevated cardiac filling pressures. This leads to severe dyspnoea, and fluid accumulates in the interstitium and alveolar spaces of the lung (pulmonary oedema). AHF is the leading cause of hospital admission in people above the age of 65 years; it has a poor prognosis, with a 60-day mortality rate of nearly 10% and a rate of death or rehospitalization of 35% within 60 days. In patients with ***acute pulmonary oedema***, the in-hospital mortality rate is 12% and by 12 months, this rises to 30%. Poor prognostic indicators include a high (>16 mmHg) pulmonary capillary wedge pressure (see Chapter 13), low serum sodium concentration, increased left ventricular end-diastolic dimension on echo, and low oxygen consumption.

Often heart failure will be categorized into heart failure with reduced ejection fraction (left ventricular ejection fraction (LVEF) less than 40%) or preserved (LVEF > 40%) ejection fraction.

The aetiology of AHF is similar to that of chronic heart failure (CHF) (see **Box 17.17**):

- ***Ischaemic heart disease*** patients present with an acute coronary syndrome or develop a complication of myocardial infarction, such as papillary muscle rupture or ventricular septal defect requiring surgical intervention.
- ***Valvular heart disease*** patients also present with AHF due to valvular regurgitation in endocarditis or prosthetic valve thrombosis. A thoracic aortic dissection may produce severe aortic regurgitation.
- ***Hypertension*** patients present with episodes of 'flash' pulmonary oedema despite preserved left ventricular systolic function.
- ***Acute and chronic kidney disease*** both involve fluid overload and reduced renal excretion, which will produce pulmonary oedema.
- ***Atrial fibrillation*** is frequently associated with AHF and may require emergency cardioversion.

Several clinical syndromes of AHF can be defined (**Box 11.22**). In a clinical environment, both the Killip score (used to assess severity of AF post-MI, see Chapter 17, 'Complications of myocardial infarction') and Forrester classification (based on right heart catheterization findings) provide therapeutic and prognostic information.

Pathophysiology

The pathophysiology of AHF is similar to that of CHF with activation of the renin–angiotensin–aldosterone axis and sympathetic nervous system. In addition, prolonged ischaemia (e.g., in acute coronary syndromes) results in myocardial stunning (see Chapter 17—'Hibernating myocardium and myocardial stunning'), which exacerbates myocardial dysfunction but may respond to inotropic support. If myocardial ischaemia persists, the myocardium may exhibit hibernation (persistently impaired function due to reduced coronary blood flow), which may recover with successful revascularization.

Clinical assessment

History

- Enquire about the patient's medical history, including the presence of hypertension, diabetes, renal disease or coronary artery disease, previous cardiac surgeries, or percutaneous coronary interventions and a prior history of CHF.
- For patients with CHF, check adherence with medications (including diuretics) and enquire regarding recent treatment changes. Do they know their recent body weight and, if so, has it increased (indicating hypervolaemia)?
- An antecedent history of chest pain or palpitations may point towards the underlying cause of AHF.
- Note any recent illnesses or infections, as they can precipitate acute pulmonary oedema in a patient with CHF.

Physical examination

- Measure vital signs, with particular attention to heart rate and rhythm, blood pressure, respiratory rate and oxygen saturation, to assess respiratory and haemodynamic status. Blood pressure

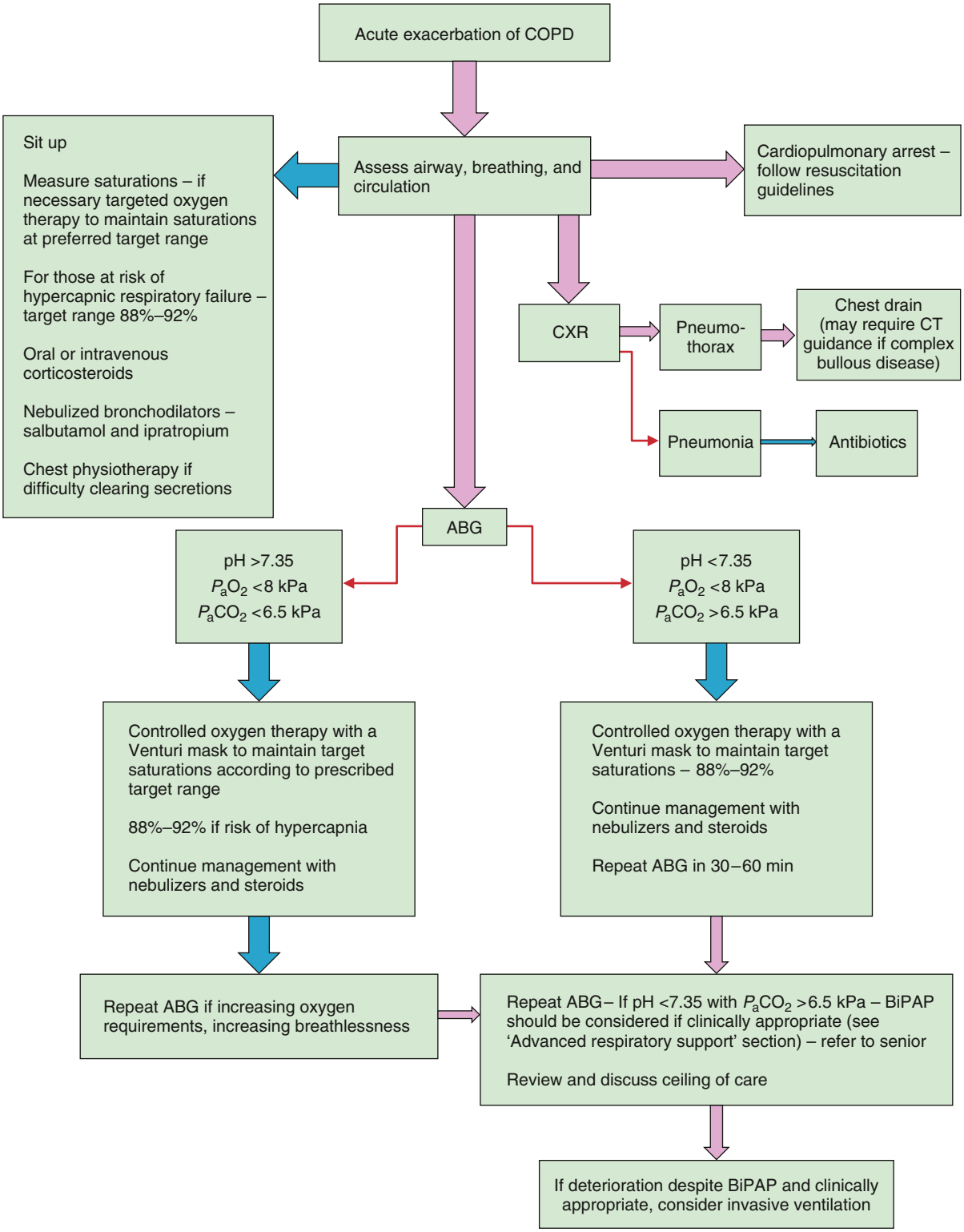


Fig. 11.19 Algorithm for the treatment of respiratory failure in COPD. *BiPAP,* Bilevel positive airway pressure.

is often markedly elevated in patients with acute pulmonary oedema and does not always indicate a hypertensive aetiology; hypotension, conversely, is a worrying sign.

- Evaluate JVP and look for signs of hypervolaemia, including peripheral and sacral oedema and ascites (absent in very acute presentations without background CHF).
- Auscultate for abnormal heart sounds, such as murmurs or the gallop rhythm of a 3rd of 4th heart sound.
- Bilateral fine inspiratory crackles, initially in the dependent zones (lung bases), but extending throughout all lung zones in severe cases, are the most characteristic sign of pulmonary oedema. Pleural effusions (bilateral, but often more prominent

Box 11.22 Clinical syndromes of acute heart failure (AHF)

Type	Clinical features
Acute decompensated heart failure	Mild features of heart failure e.g., dyspnoea
Hypertensive AHF	High blood pressure, preserved left ventricular function, pulmonary oedema on CXR
Acute pulmonary oedema	Tachypnoea, orthopnoea, pulmonary crackles, oxygen saturation <90% on air, pulmonary oedema on CXR
Cardiogenic shock	Systolic blood pressure <90 mmHg, mean arterial pressure drop >30 mmHg, urine output <0.5 mL/kg per hour, heart rate >60 b.p.m.
High-output heart failure	Warm peripheries, pulmonary congestion, blood pressure may be low e.g., septic shock
Right heart failure	Low cardiac output, elevated jugular venous pressure, hepatomegaly, hypotension

(Modified from the European Society of Cardiology.)

on the right) may also be present in acute decompensated heart failure.

Investigations

Initial investigations should include:

- ***Laboratory investigations:*** serum creatinine and electrolytes, full blood count, blood glucose, cardiac enzymes and troponin, C-reactive protein (CRP), and consideration of D-dimer.
- ***Plasma BNP*** or ***NT-proBNP*** (BNP > 100 pg/mL or NT-proBNP >300 pg/mL is suggestive of heart failure).
- ***A 12-lead ECG*** will identify acute coronary syndromes, left ventricular hypertrophy, atrial fibrillation, some valvular heart disease, and left bundle branch block.
- ***A CXR*** may demonstrate cardiomegaly, pulmonary oedema, pleural effusions, or non-cardiac disease.
- ***Transthoracic echocardiogram (TTE)*** should be performed without delay to assess ventricular structure and function and possibly identify the cause.

Management

The goals of treatment in a patient with AHF include:

- immediate relief of symptoms and stabilization of haemodynamics (short-term benefits)
- reduction in length of hospital stay and rates of re-admissions to hospital
- reduction in mortality from heart failure.

Patients with AHF who are unwell should be managed in a high-dependency area (e.g., coronary care unit) with regular measurement of temperature, heart rate, blood pressure, and cardiac monitoring. All require prophylactic anticoagulation with low-molecular-weight heparin (LMWH), unless contraindicated.

Individuals with haemodynamic compromise may need arterial lines for invasive blood pressure monitoring and arterial gas sampling, central venous cannulation (intravenous medication, inotropic support, monitoring of central venous pressure) and pulmonary artery catheterization (calculation of cardiac output/index, peripheral vasoconstriction, and pulmonary wedge pressure), and consideration of NIV.

Initial therapy (Fig. 11.20 and Box 11.23) includes oxygen and diuretics (e.g., i.v. furosemide 40 mg). Given intravenously, furosemide

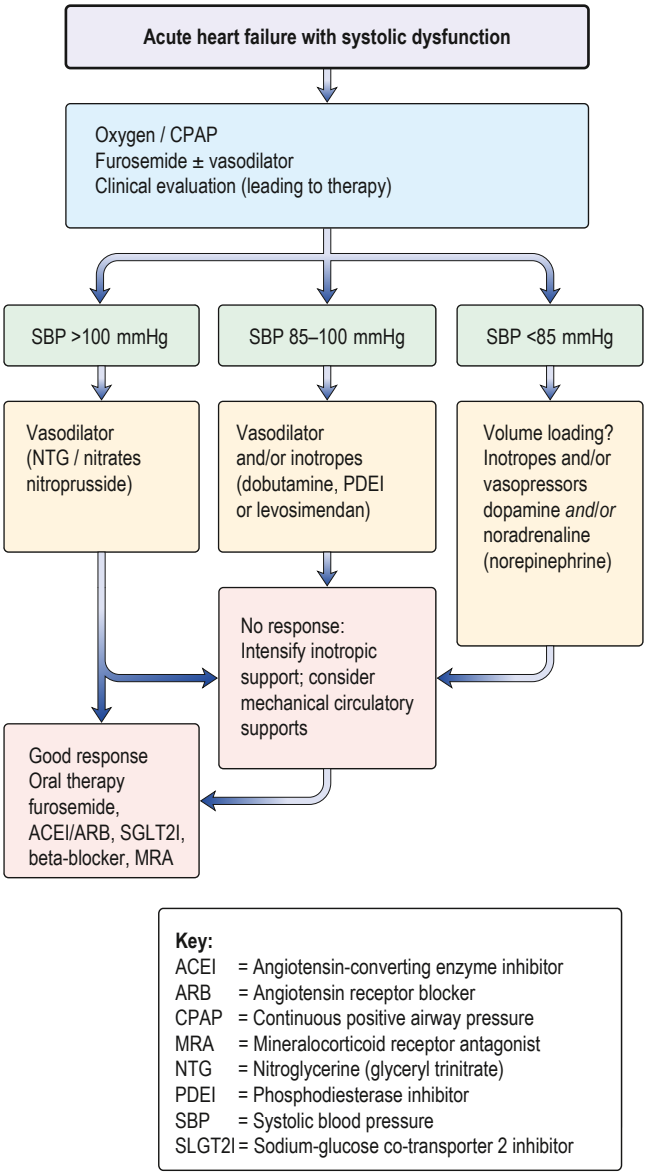


Fig. 11.20 Algorithm for the management of acute heart failure with systolic dysfunction. *(Modified from Nieminen MS, Böhm M, Cowie MR et al. Executive summary of the guidelines on the diagnosis and treatment of acute heart failure: the Task Force on Acute Heart Failure of the European Society of Cardiology.* Eur Heart J *2005; 26:384–416, Fig. 6, with permission from Oxford University Press and the European Society of Cardiology.)*

has venodilating properties resulting in clinical benefit before diuresis becomes apparent. If intravenous nitrates (e.g., glyceryl trinitrate infusion 10–200 μg/min) are required (e.g., concomitant myocardial ischaemia, severe hypertension), careful monitoring of the blood pressure is mandatory. Inotropic support with dobutamine, phosphodiesterase inhibitors (milrinone), or levosimendan can be added in patients who do not respond to initial therapy. Specialist options for treating severe AHF which do not respond to initial management include extracorporeal membrane oxygenation (ECMO), miniature cardiac pumps, and (less used now) intra-aortic balloon pumps; these are discussed further in 'Temporary mechanical cardiac support' in Chapter 17.

If there is limited improvement with bolus furosemide, furosemide infusion can be considered.

Box 11.23 Pharmacological therapy in acute heart failure

Drug	Dose	Indications/mechanism of action
Myocardial oxygenation		
Oxygen	35%–50% inspired oxygen concentration	Ensure airway is patent and maintain arterial saturation at 95%–98%
Non-invasive positive pressure e.g., CPAP		Use if failing to maintain arterial saturation Increases pulmonary recruitment and functional residual capacity—reduces work of breathing
Intubation/mechanical ventilation		Use if patient is failing to maintain arterial saturation and is fatigued (reduced respiratory rate, increased arterial CO_2, confusion)
Opiate		
Morphine	2.5–5.0 mg i.v. (with antiemetic metoclopramide 10 mg i.v.)	Use in agitated patient Relieves dyspnoea, venous and arterial dilation
Antithrombin		
Low-molecular-weight heparin	E.g., Enoxaparin 1 mg/kg s.c. ×2 daily (ACS) or 40 mg s.c. daily (prophylaxis)	Use in patients with AHF, ACS, or atrial fibrillation, or for DVT prophylaxis Caution if creatinine clearance <30 mL/min
Vasodilators		
Glyceryl trinitrate	10–200 μg/min i.v. infusion	Reduces pulmonary congestion; at low doses causes venodilation, reducing preload; at high doses causes arterial vasodilation, reducing afterload Ensure BP > 85–90 mmHg
Sodium nitroprusside	0.3–5 μg/kg per min i.v. infusion (initially 0.5–1.5 μg/kg per min, adjust by 0.5 μg/kg per min every 5 minutes, stop if response unsatisfactory with maximum dose in 10 mins)	Use in severe AHF where there is predominantly high afterload e.g., hypertensive AHF Needs arterial BP monitoring as risk of profound hypotension
Diuretic		
Furosemide	Bolus 40–100 mg i.v. or infusion 5–40 mg/h	Low doses produce vasodilation, reduce right atrial pressure and PCWP, and promote diuresis Need to monitor sodium, potassium, and creatinine
Inotropes		
Dopamine	Low dose <2 μg/kg per min	Low dose acts on peripheral dopamine receptors to produce vasodilation (renal, splanchnic, coronary, cerebrovascular) and may improve diuresis although evidence of significant improvements in outcomes is lacking
	Medium dose >2 μg/kg per min	Medium dose acts on β-receptors to increase myocardial contractility and cardiac output
	High dose >5 μg/kg per min	High dose acts on α-receptors, causing vasoconstriction and increasing total peripheral resistance (increases afterload, PAP)
Dobutamine	2–20 μg/kg per min (patients on beta-blockers may need high dose)	Stimulates β_1 and β_2-receptors, producing vasodilation. Increases heart rate and cardiac output, and also diuresis as haemodynamics are improved
Milrinone	Bolus 25–75 μg/kg over 10–20 min then 0.375–0.75 μg/kg per min	Inhibits phosphodiesterase and maintains cAMP Increases cardiac output and stroke volume, reduces PAP/PCWP/total peripheral resistance/BP
Levosimendan	Bolus 6–12 μg/kg over 10 min then 0.05–2 μg/kg per min	Positive inotropic drug with vasodilator effects by increasing sensitivity of contractile proteins to calcium and opening potassium channels Increases cardiac output and reduces PCWP
Vasopressors		
Noradrenaline (norepinephrine)	0.2–1.0 μg/kg per min	Stimulates α-receptors Increases total peripheral resistance and BP
Adrenaline (epinephrine)	Bolus 1 mg at resuscitation then 0.05–0.5 μg/kg per min	Stimulates α-, β_1-, and β_2-receptors Increases cardiac output, heart rate, total peripheral resistance, and BP
Cardiac glycoside		
Digoxin	0.5 mg i.v. repeated after 2–6 h	Inhibits myocardial sodium/potassium ATPase, leading to increased calcium and sodium exchange Increases cardiac output and slows AV conduction Use in AF; avoid in ACS

ACS, Acute coronary syndrome; *AF,* atrial fibrillation; *AHF,* acute heart failure; *ATPase,* adenosine triphosphatase; *AV,* atrioventricular; *BP,* blood pressure; *cAMP,* cyclic adenosine monophosphate; *CPAP,* continuous positive airway pressure; *DVT,* deep venous thrombosis; *i.v.,* intravenous; *PAP,* positive airway pressure; *PCWP,* pulmonary capillary wedge pressure; *s.c.,* subcutaneous.

Patients with profound hypotension may require inotropes and vasopressors to improve haemodynamic status and alleviate symptoms but these have not been shown to improve mortality. Non-invasive continuous positive airway pressure/positive pressure ventilation (CPAP/NIPPV; see 'Advanced respiratory support') has been shown to provide earlier improvement in dyspnoea and respiratory distress than standard oxygen via mask; mortality is, however, unaffected.

Patients who present with mild to moderate decompensation of chronic heart failure can be considered for outpatient management of their condition if there is access to intravenous furosemide, regular monitoring via hospital at home or ambulatory care services, and input from heart failure teams (specialist nurses and cardiologists) to aid with titrating/optimizing medication. Point-of-care echo and lung/abdominal ultrasound can be useful adjuncts to assess volume status alongside clinical assessment of the patient. Daily weight monitoring, along with renal function and clinical fluid status assessments, is important.

Discharge

Hospitalization with AHF carries a high risk for future re-admission to hospital, the risk of which can be reduced by instigating a follow-on community care plan. Nurse-led community heart failure teams are fundamental to this process and referral should be made prior to discharge. Once euvolaemia is achieved, it is helpful to document a 'dry weight' to enable patients to monitor weight, identify early fluid reaccumulation and adjust diuretic doses under medical guidance.

Evidence-based treatments for CHF are discussed further from 'Management of heart failure' in Chapter 17, and comprise:

- lifestyle measures
- multidisciplinary interventions including cardiac rehabilitation
- cardiac resynchronization therapy ('Cardiac resynchronization therapy' section in Chapter 17)
- four main drug treatment classes:
  - ACE inhibitor, ARB, or combined angiotensin receptor-neprilysin inhibitors
  - beta-blocker
  - mineralocorticoid receptor antagonist (MRA)
  - SGLT2 inhibitor.

These pharmacological agents should be initiated in hospital where appropriate and up-titrated proactively in the weeks following admission. The STRONG-HF trial demonstrated that up-titration of ACEi/ARB, beta-blocker, and MRA to optimal doses within 2 weeks of discharge resulted in reduced all-cause mortality and 180-day hospitalization rates compared to usual management, with benefits seen irrespective of baseline left ventricular ejection fraction. This study predated routine use of SGLT2 inhibitors in heart failure although it is widely assumed that these should also be initiated early. Combined use of ACEi/ARB and MRA carries a risk of hyperkalaemia and AKI and requires careful oversight and monitoring by heart failure teams.

Further reading

2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure. Accessed at: https://www.escardio.org/Guidelines/Clinical-Practice-Guidelines/Acute-and-Chronic-Heart-Failure.

2023 Focused Update of the 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure. Accessed at: https://www.escardio.org/Guidelines/Clinical-Practice-Guidelines/Focused-Update-on-Heart-Failure-Guidelines.

Mebazaa A, Davison B, Chioncel O et al. Safety, tolerability and efficacy of up-titration of guideline-directed medical therapies for acute heart failure (STRONG-HF): a multinational, open-label, randomised, trial. *Lancet.* 2022; 400:1938–1952.

Thomson-Anderson A, Fawcett N, Naseer S et al. A tertiary care ambulatory heart failure pathway managing one-third of all admissions including older patients with similar quality to inpatient management. *Clin Med (Lond).* 2022; 22(Suppl 4):59–60.

? Box 11.24 Common causes of respiratory viral infections

- Influenza virus
- Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2; causative agent of COVID-19)
- Respiratory syncytial virus (RSV)
- Parainfluenza virus
- Adenovirus
- Rhinovirus
- Metapneumovirus
- Other non (SARS-CoV-2) coronaviruses e.g., Middle East respiratory syndrome coronavirus (MERS/MERS-CoV)

Acute viral illness

Acute viral illnesses are a common cause of breathlessness presenting to the acute medical take, either via induction of pneumonitis or through exacerbation of a pre-existing respiratory condition (e.g., asthma). Common pathogens are listed in **Box 11.24**.

Clinical assessment

History

- Enquire about fever, cough, shortness of breath, coryza, myalgia, and joint ache
- Ask about recent exposure to individuals with similar symptoms
- Enquire about loss or disturbance of taste or smell (frequently reported with COVID-19)
- Ask about travel and vaccination history (COVID-19/flu vaccine)
- Take full systems history as viral respiratory illnesses can also present with systemic manifestations (a description of the clinical syndromes associated with individual pathogens can be found in Chapter 42)

Physical examination

- Check vital signs, including temperature, heart rate, and oxygen saturation
- Evaluate respiratory rate and effort and identify signs of respiratory distress
- Chest examination findings may include signs of consolidation (crackles, bronchial breathing) or polyphonic wheeze if small airway inflammation predominates. Examination findings may be minimal even in the presence of severe hypoxia
- Assess any other systems that might have been affected as described from the history

Investigations

- **Point-of-care tests:**
  - Rapid PCR or antigen tests for COVID-19 and influenza
  - Respiratory viral swabs for other respiratory viruses
- **Blood tests:**
  - Full blood count (FBC) and basic metabolic panel to check electrolyte levels. Lymphopaenia is common with respiratory viral pathogens
  - Procalcitonin and CRP
- **CXR** to evaluate for evidence of pneumonitis, pneumonia, or other lung pathology
- **Thoracic CT scan** in selected cases for detailed lung assessment. The possibility of PE should be considered in patients with COVID-19 infection which may warrant CT-PA.

Positive PCR or antigen test results indicate the presence of a specific viral pathogen. Sometimes the cause of an acute respiratory viral illness is not identified.

Pairing CRP with procalcitonin (PCT, regarded as a marker of bacterial infection) may help guide the rational use of antibiotic therapy in the setting of acute viral illness.

Management

Principles of managing an acute viral respiratory infection are outlined in **Box 11.25**. A more detailed discussion of management of acute COVID-19 pneumonitis can be found in Chapter 20.

Virtual wards and telehealth

Virtual wards are an emerging and valuable tool in managing respiratory viral illnesses, including COVID-19. Virtual wards facilitate remote monitoring of patients with mild to moderate symptoms who do not require immediate hospitalization. Pulse oximeters and thermometers are provided to patients for home monitoring and regular 'check-ins' are conducted to assess evolution of symptoms.

Prevention

Vaccination

It is important to highlight the value of annual influenza and COVID-19 vaccinations for at-risk individuals.

Patient education

Patients should be advised of respiratory hygiene, respiratory isolation, and staying home when ill. They should be advised about symptoms of deterioration and when to seek medical help.

Community mitigation strategies

For some respiratory viruses, mask mandates, social distancing, and travel restrictions have been implemented to reduce spread of disease with use of public health campaigns for awareness and prevention. Surveillance and research on respiratory viruses have informed vaccination strategies and public health responses.

Further reading

https://www.nice.org.uk/guidance/ng191/resources/covid19-rapid-guideline-managing-covid19-pdf-51035553326.
https://www.who.int/teams/health-care-readiness/covid-19.

Pneumonia

Pneumonia is defined as inflammation of the substance of the lungs. It is usually caused by bacteria but can also be caused by viruses and fungi. Clinically, it usually presents as an acute illness with cough, purulent sputum, breathlessness, and fever, together with physical signs or radiological changes compatible with consolidation of the lung (**Fig. 11.21**). However, it can present with more subtle symptoms, particularly in the elderly.

Pneumonia is usually ***classified*** by the setting in which the patient has contracted the infection, for example:

- ***community-acquired pneumonia*** in a person with no underlying immunosuppression or malignancy
- ***hospital-acquired pneumonia*** (sometimes called 'healthcare-associated pneumonia', reflecting the role of other institutions such as nursing homes)
- ***aspiration pneumonia***, associated with the aspiration of food material or stomach contents into the lungs, and caused by impaired swallowing
- ***pneumonia in immunocompromised patient***, acquired through either a genetic defect, immunosuppressive medication, or acquired immunodeficiency, as in human immunodeficiency virus (HIV) infection
- ***ventilator-acquired pneumonia***, acquired through mechanical ventilation in a critical care unit.

Community-acquired pneumonia

Community-acquired pneumonia (CAP) occurs across all ages but is more common at the extremes of age. *Streptococcus pneumoniae* is the most common cause overall; however, in 30%–50% of cases, no organism is identifiable, while in about 20% more than one organism is present. Infection can be localized, when the whole of one or more lobes is affected ('lobar pneumonia'), or diffuse, when the lobules of the lung are mainly affected, often due to infection centred on the bronchi and bronchioles ('bronchopneumonia'). Factors that increase the risk of developing CAP are shown in **Box 11.26**.

Clinical features

The clinical presentation varies according to the immune state of the patient and the infecting agent. Features include:

Box 11.25 Management of an acute viral respiratory infection

Isolation and infection control measures
- Rapid identification of the viral pathogen
- Strict infection control protocols (PPE, hand hygiene)
- Visitor restrictions if in hospital and contact tracing when necessary

Supportive care
- Hydration, rest, and symptom management
- Antipyretic medications (paracetamol, ibuprofen if not contraindicated)
- Hospitalization for severe cases or specific viruses

Antiviral medications
- Considered for severe cases or vulnerable populations e.g., immunocompromised
- Choice of antiviral depends on the specific viral pathogen

Antibiotics and bacterial coinfections
- On occasion, when superadded bacterial infection is suspected, antibiotics may be considered as guided by clinical judgement and diagnostic tests

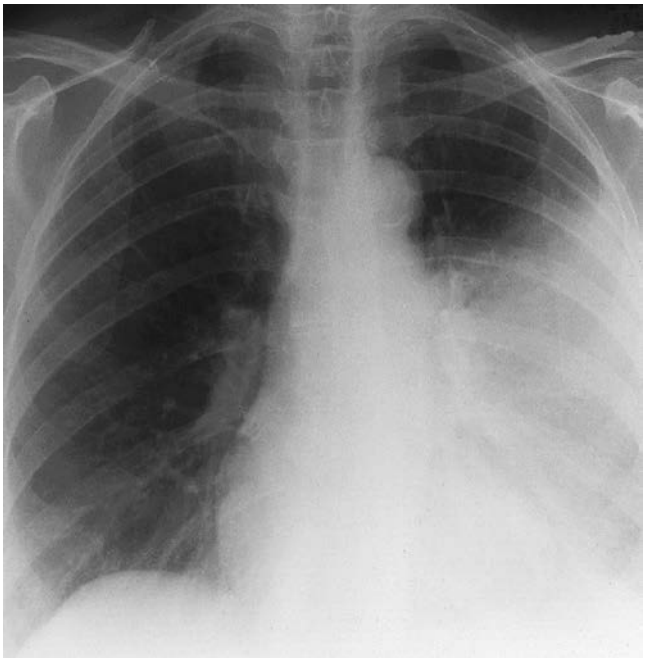


Fig. 11.21 CXR showing lobar pneumonia.

- ***a dry or productive cough***, sometimes with haemoptysis
- breathlessness
- ***fevers***, which, if swinging, may indicate empyema
- ***chest pain*** may be experienced, commonly pleuritic in nature and due to inflammation of the pleura; a pleural rub may be heard early on in the illness
- extrapulmonary features (**Box 11.27**).

In the elderly, CAP can present with confusion or non-specific symptoms such as recurrent falls. CAP should always be considered in the differential diagnosis of sick elderly patients, given their frequently atypical presentation.

Initial assessment

The type and extent of investigations depend on the severity of the illness, which also guides where the patient should be managed and predicts their outcome. Diagnostic microbiological tests are not needed in mild infection, which should be treated at home with standard oral antibiotics (amoxicillin, or clarithromycin for those with a history of penicillin allergy). Where patients have mild disease, CXR is not routinely recommended unless they fail to improve after 48–72 hours.

Severity is commonly assessed by the ***CURB-65*** or the ***CRB-65 score***; the CRB-65 score is used in the community where the serum urea level is not usually available (**Box 11.28**). These give a guide to the likely risk of fatal outcome but antibiotic choice must always be tempered by clinical assessment and judgement, taking into account other factors associated with increased rates of mortality (see 'other markers' in **Box 11.28**). These scoring systems are also helpful as clinical adjuncts to decisions regarding intravenous versus oral antibiotics, discharge, admission, and ambulatory management.

Box 11.26 Risk factors for community-acquired pneumonia

- ***Age:*** <16 or >65 years
- ***Co-morbidities:*** HIV infection, diabetes mellitus, chronic kidney disease, malnutrition, recent viral respiratory infection
- ***Other respiratory conditions:*** cystic fibrosis, bronchiectasis, chronic obstructive pulmonary disease, obstructing lesion (endoluminal cancer, inhaled foreign body)
- ***Lifestyle:*** cigarette smoking, excess alcohol, intravenous drug use
- ***Iatrogenic:*** immunosuppressant therapy (including prolonged corticosteroids)

Box 11.27 Extrapulmonary features of community-acquired pneumonia

- ***Myalgia, arthralgia, and malaise*** are common, particularly in infections caused by *Legionella* and *Mycoplasma*
- ***Myocarditis and pericarditis*** are cardiac manifestations of infection, most commonly in *Mycoplasma* pneumonia
- ***Headache*** is common in *Legionella* pneumonia. Meningoencephalitis and other neurological abnormalities also occur but are much less common
- ***Abdominal pain, diarrhoea, and vomiting*** are common. Hepatitis can be a feature of *Legionella* pneumonia
- ***Labial herpes simplex*** reactivation is relatively common in pneumococcal pneumonia
- ***Other skin rashes***, such as erythema multiforme and erythema nodosum, are found in *Mycoplasma* pneumonia. Stevens–Johnson syndrome (see 'Stevens-Johnson Syndrome and toxic epidermal necrolysis' later in this chapter) is a rare and potentially life-threatening complication of pneumonia

Investigations

All patients admitted to hospital with suspected CAP should have a CXR, blood tests, and microbiological tests.

Chest X-ray (CXR)

Radiological abnormalities can lag behind clinical signs. A normal CXR on presentation should be repeated after 2–3 days if CAP is suspected clinically. The CXR must be repeated 6 weeks later to rule out an underlying bronchial malignancy causing pneumonia due to bronchial obstruction.

Blood tests

Full blood count, serum creatinine and electrolytes, biochemistry, and CRP are helpful.

- ***Strep. pneumoniae.*** White cell count is usually >15 × 5^9/L (90% leucocytosis neutrophils). Inflammatory markers are significantly elevated: erythrocyte sedimentation rate (ESR) >100 mm/h; CRP >100 mg/L.
- ***Mycoplasma.*** White cell count is usually normal. In the presence of anaemia, haemolysis should be ruled out (direct Coombs' test and measurement of cold agglutinins, see Chapter 38).
- ***Legionella.*** There is often lymphopenia without marked leucocytosis, hyponatraemia, hypoalbuminaemia, and high serum levels of liver aminotransferases.

Other tests

- ***Sputum culture and blood cultures*** are required for all patients who have moderate to severe CAP, ideally before antibiotics are administered. In *S. pneumoniae* infection, positive blood cultures indicate more severe disease with greater mortality.
- ***Arterial blood gas analysis*** is necessary if oxygen saturation is <94%. Raised serum lactate suggests the presence of sepsis (Chapter 12).
- ***Urinary antigen testing*** for suspected pneumococcal or legionella pneumonia.
- ***An HIV test*** should be offered to all patients with pneumonia since it is a common indicator condition of previously undiagnosed HIV infection.

Box 11.28 CURB-65 score and other markers of severe community-acquired pneumonia

CURB-65:
C: ***Confusion*** present (abbreviated mental test score <8/10)
U: (plasma) ***Urea*** level >7 mmol/L
R: ***Respiratory rate*** >30 breaths/min
B: systolic ***Blood pressure*** <90 mmHg; diastolic <60 mmHg
65: ***age >65***

1 point for each of the above:

- ***Score 0–1:*** Treat as outpatient
- ***Score 2:*** Admit to hospital
- ***Score 3+:*** Often require care in intensive treatment unit

Mortality rates increase with increasing scores.

Other markers

- ***CXR:*** more than one lobe involved
- P_aO_2: <8 kPa
- ***Low albumin:*** <35 g/L
- ***White cell count:*** <4 × 10^9/L or >20 × 10^9/L
- ***Blood culture:*** positive
- Other ***co-morbidities***
- ***Absence of fever*** in the elderly

Management

Initial management and assessment should follow the guidelines for management of sepsis (see Chapter 12), particularly when a patient appears to have a moderate to severe pneumonia. In general:

- ***Oxygen.*** Supplemental oxygen should be administered to maintain saturations between 94% and 98% (provided the patient is not at risk of carbon dioxide retention, due to loss of hypoxic drive in COPD). In patients with known COPD, oxygen saturations should be maintained between 88% and 92%.
- ***Intravenous fluids.*** These are required in hypotensive patients showing any evidence of volume depletion and hypotension.
- ***Antibiotics.*** The first dose of antibiotic should be administered *within 1 hour of identifying any high-risk criteria* and treatment should not be delayed while investigations are awaited. Parenteral antibiotics should be switched to oral once the temperature has settled for a period of 24 h, provided there is no contraindication to oral therapy. If patients fail to respond to initial treatment, microbiological advice should be sought and alternative diagnoses considered. The antibiotic regimen should be adjusted specifically once culture and sensitivity results are available (Fig. 11.22).
- ***Thromboprophylaxis.*** If the patient is admitted for >12 h, subcutaneous LMWH should be prescribed and thromboembolism deterrent (TED) stockings should be fitted, unless contraindications exist.
- ***Physiotherapy.*** Chest physiotherapy is not needed unless sputum retention is an issue.
- ***Nutritional supplementation.*** Need is assessed by a dietician, particularly in severe disease.

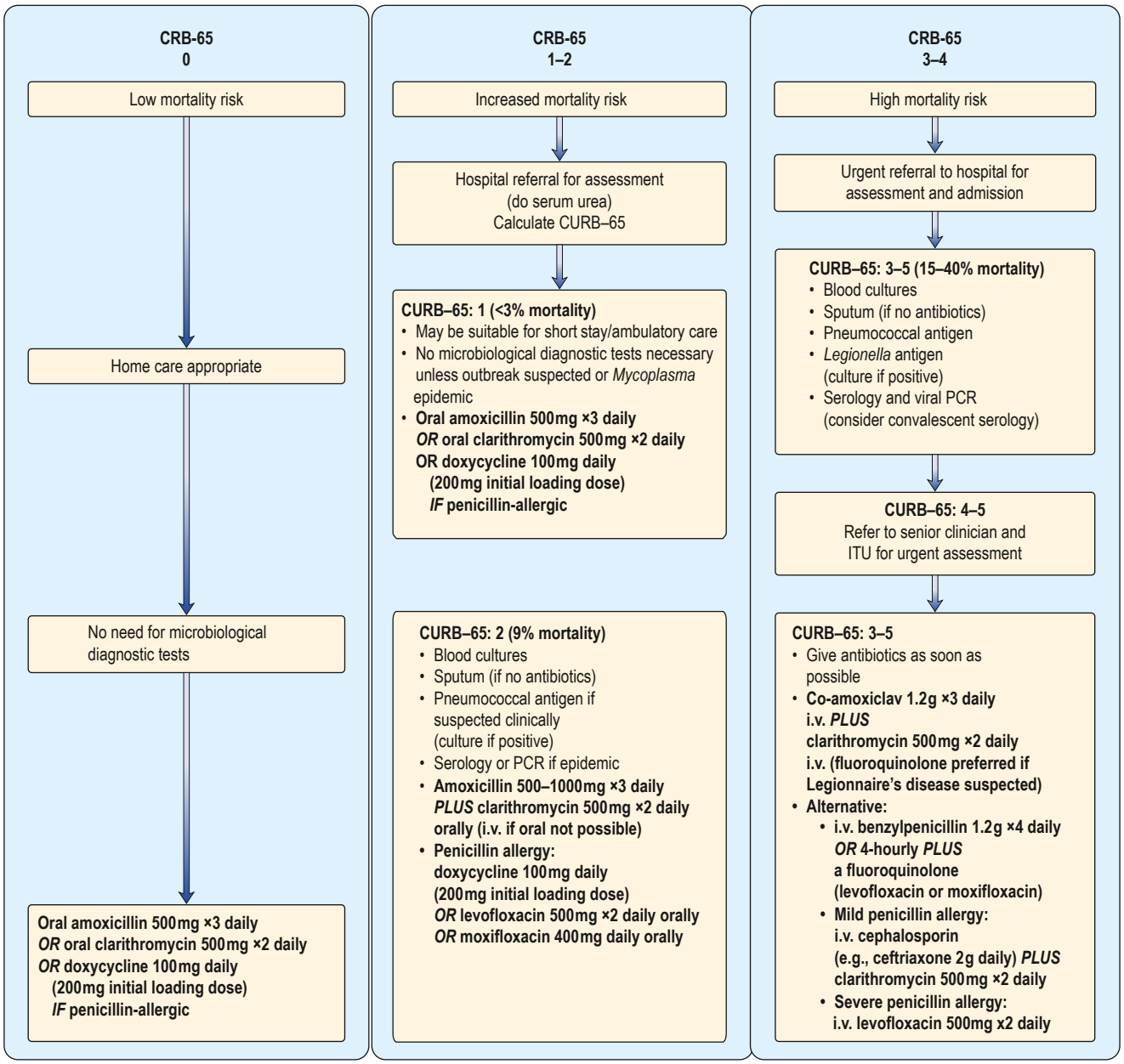


Fig. 11.22 Algorithm for the assessment and treatment of community-acquired pneumonia.
CRB, Confusion, respiratory rate, blood pressure (CURB includes urea); *ITU,* intensive treatment unit; *PCR,* polymerase chain reaction.

Box 11.29 Causes of slow-resolving pneumonia

Incorrect or incomplete antimicrobial treatment
- Underlying antibiotic resistance
- Inadequate dose/duration
- Non-adherence
- Malabsorption

Complication of community-acquired pneumonia
- Parapneumonic pleural effusion (exudative)
- Empyema
- Lung abscess

Underlying neoplastic lesion or other lung disease
- Obstructing lesion
- Bronchoalveolar cell carcinoma
- Bronchiectasis
- Tuberculosis

Alternative diagnosis
- Pulmonary thromboembolic disease
- Cryptogenic organizing pneumonia
- Eosinophilic pneumonia
- Pulmonary haemorrhage

Box 11.30 Complications of pneumonia

General
- Respiratory failure
- Sepsis—multisystem failure

Local
- Pleural effusion
- Empyema
- Lung abscess
- Organizing pneumonia
- Immunothrombosis *in situ*

- ***Analgesia.*** Simple analgesia, such as paracetamol or an NSAID (unless contraindicated), helps treat pleuritic pain, thereby reducing the risk of further complications due to restricted breathing because of pain (e.g., sputum retention, atelectasis, or secondary infection).

Causes of a slow-resolving pneumonia are outlined in **Box 11.29**.

Complications of pneumonia (Box 11.30)

Parapneumonic effusion and empyema

- Pleural effusions complicate around one-third to one-half of CAP. The majority of these are simple exudative effusions but empyema (purulent fluid in the pleural space) may develop. Early indications of empyema are ongoing fever and rising or persistently elevated inflammatory markers despite appropriate antibiotic therapy. Pleural aspiration should be performed under ultrasound guidance and sent for Gram stain, culture, fluid protein, glucose, and LDH (with comparison to serum levels). Light's criteria (see **Box 20.32**) can be applied to assess whether an effusion is transudative or exudative. Pathogens are often detectable and sensitivity analysis can guide antimicrobial therapy.
- An exudative effusion with pleural fluid pH of <7.2 is strongly suggestive of empyema. If an empyema develops, the fluid should be urgently drained to prevent further complications such as development of a thick pleural rind and prolonged hospital admission. Empyema increases mortality risk. The duration of antibiotic administration will usually need to be extended. Thoracic surgical intervention may be necessary.

Box 11.31 Causes of lung abscess and causative organisms

Causes
- Aspiration pneumonia
- Tuberculosis
- *Staphylococcus aureus* or *Klebsiella pneumoniae*
- Septic emboli usually containing staphylococci
- Inadequately treated community-acquired pneumonia
- Spread from an amoebic liver abscess
- Bronchial obstruction by an endoluminal cancer
- Foreign body inhalation

Common causative organisms
- *Klebsiella pneumoniae*
- *Staphylococcus aureus*
- Gram-negative enteric bacilli
- *Mycobacterium tuberculosis*
- *Streptococcus milleri*
- Anaerobic bacteria (post aspiration)
- *Haemophilus influenzae*

Lung abscess

This term is used to describe localized suppuration within the lung associated with cavity formation visible on the chest radiograph or CT scan, often with a fluid level (which indicates an air–liquid interface). There are several causes of lung abscess (**Box 11.31**). Clinical features include persisting or worsening pneumonia associated with the production of large quantities of sputum, which is often foul-smelling owing to anaerobic organisms. There is usually a swinging fever. Malaise and weight loss are frequent. There may be few signs on physical examination, although clubbing occurs in chronic suppuration. Patients have a normocytic anaemia and raised inflammatory markers (ESR/CRP). CT scanning is diagnostic, and bronchoscopy can be undertaken to obtain samples or remove foreign bodies. Treatment should be guided by available clinical judgement and culture results and is often prolonged (4–6 weeks). Surgical drainage is sometimes necessary.

Prevention

Cigarette smoking is an independent risk factor for CAP; if the patient still smokes, cessation advice and support should be given.

Vaccination against influenza and pneumococcus is recommended for at-risk groups. All patients over the age of 65 who have not previously been vaccinated and are admitted with CAP should have the pneumococcal vaccine before discharge from hospital.

Pneumonia in other settings

Hospital-acquired pneumonia

Hospital-acquired pneumonia (HAP) is defined as new onset of cough with purulent sputum, along with a compatible CXR demonstrating consolidation, in patients who are beyond 2 days of their initial admission to hospital or who have been in a healthcare setting within the last 3 months (including nursing or residential homes, as well as acute care facilities such as hospitals). HAP is the second most common form of hospital-acquired infection after urinary tract infections and carries a significant mortality risk, particularly in the elderly or those with comorbidities such as stroke, respiratory disease, or diabetes. In HAP, the causative organisms differ from those causing CAP (**Box 11.32**). Viral or fungal pathogens only affect immunocompromised hosts.

Ventilator-associated pneumonia

This occurs in the context of mechanical ventilation in a critical care setting. Often multidrug-resistant Gram-negative organisms

Box 11.32 Organisms implicated in hospital-acquired pneumonia

- Gram-negative bacteria (*Pseudomonas* spp., *Escherichia* spp., *Klebsiella* spp.)
- Anaerobic bacteria (*Enterobacter* spp.)
- *Staphylococcus aureus* (including methicillin-resistant *S. aureus*)
- *Acinetobacter* spp.

(such as *Acinetobacter baumannii*) are responsible, requiring careful selection of an appropriate antibiotic in association with a clinical microbiologist.

Aspiration pneumonia

Acute aspiration of gastric contents into the lungs can produce an extremely severe and sometimes fatal illness owing to the intense destructiveness of gastric acid. This can complicate anaesthesia, particularly during pregnancy (Mendelson's syndrome). Because of the bronchial anatomy, the most common sites for aspirated material to end up are the right middle lobe and the apical or posterior segments of the right lower lobe. Persistent pneumonia is often due to anaerobes and may progress to lung abscess or even bronchiectasis if protracted. It is vital to identify any underlying problem, as aspiration will recur without appropriate corrective measures.

Treatment should be directed specifically against positive cultures if available. If not, then co-amoxiclav is used for mild to moderate disease, which covers Gram-negative and anaerobic bacteria. Treatment needs to be escalated when there is a lack of response or cases are severe.

Pneumonia in immunocompromised patients

Patients who are immunosuppressed (either iatrogenically or due to a defect in host defences) are at risk not only from all the usual organisms that can cause pneumonia but also from opportunistic pathogens that would not be expected to cause disease. These opportunistic pathogens can be commonly occurring microorganisms (that are ubiquitous in the environment) or bacteria, viruses, and fungi that are less common (see **Box 44.13**). The symptom pattern may resemble CAP or be more non-specific. A high degree of clinical suspicion is therefore necessary when assessing an ill patient who is immunocompromised.

Pneumocystis jirovecii pneumonia

Pneumocystis pneumonia is one of the most common opportunistic infections encountered in clinical practice. It affects patients on immunosuppressant therapy, such as long-term corticosteroids, monoclonal antibodies, or methotrexate for autoimmune disease; those on anti-rejection medication after a solid organ or haematopoietic stem cell transplantation; and those infected with HIV. Individuals with CD4 counts of less than 200/mm^3 are at particular risk. *Pneumocystis jirovecii* is found in the air, and pneumonia arises from re-infection rather than reactivation of persisting organisms acquired in childhood.

Clinically, the pneumonia is associated with a high fever, breathlessness, and dry cough. A characteristic feature on examination is rapid desaturation on exercise or exertion. The typical radiographic appearance is one of diffuse bilateral alveolar and interstitial shadowing beginning in the perihilar regions and spreading out in a butterfly pattern. Other CXR appearances include localized infiltration, nodules, cavitation, or a pneumothorax but in approximately 10% of cases is normal. Empirical treatment is justified in very sick high-risk patients; wherever possible, however, the diagnosis should be confirmed by indirect immunofluorescence on induced sputum or bronchoalveolar lavage fluid. First-line treatment of *Pneumocystis* pneumonia is with high-dose co-trimoxazole (see Chapter 44), with adjunctive corticosteroids in patients with HIV infection.

Further reading

BMJ Best Practice. *Community-acquired pneumonia (non COVID-19)*. Accessed at: https://bestpractice.bmj.com/topics/en-gb/3000108.
British Thoracic Society. *2015-Annotated BTS Guideline for the Management of CAP in Adults (2009): Summary of Recommendations*; BTS 2015; Accessed at: https://brit.thoracic.org.uk/guidelines-and-quality-standards/community-acquired-pneumonia-in-adults-guidance/annotated-bts-guideline-for-the-management-of-cap-in-adults-2014/.
File TM, Ramirez JA. Community acquired pneumonia, *N Engl J Med.* 2023; 389:632–641.
Mansell A, Niederman LA. Aspiration pneumonia. *N Engl J Med.* 2019; 380:651–663.
National Institute for Health and Care Excellence. *Clinical Guideline 191: Pneumonia in Adults: Diagnosis and Management*. NICE 2014. Accessed at: https://www.nice.org.uk/guidance/cg191.

Pulmonary embolism

Refer to the section 'Deep venous thrombosis and pulmonary embolism'.

Pneumothorax

'Pneumothorax' means air in the pleural space. Primary spontaneous pneumothoraces occur predominantly in young people. Traditionally, patients are tall, thin, and male but shape, size, and sex often do not follow this rule. Primary pneumothoraces are usually caused by rupture of a pleural bleb, typically apical, and are thought to be due to congenital defects in the connective tissue of the alveolar walls. Both lungs are affected with equal frequency.

Secondary pneumothoraces occur in conjunction with pre-existing lung disease such as COPD, infection, and cystic fibrosis. Iatrogenic pneumothoraces are caused by instrumentation to the thorax, such as central venous line insertion, percutaneous or transbronchial lung biopsy, or trauma.

Clinical features

Patients commonly present with sudden onset of pleuritic chest pain and breathlessness. On careful questioning, they may be discovered to have had a milder version of these symptoms in the past but not sought medical attention. Clinical signs are of reduced or absent breath sounds over the affected lung and a hyper-resonant percussion note.

A high index of suspicion should be maintained for tension pneumothorax, an uncommon but immediately life-threatening condition, most commonly seen after trauma or in the context of mechanical ventilation, where air enters the pleural space and is unable to escape, meaning intrapleural pressure progressively increases. With tension pneumothorax, the trachea will deviate progressively away from the affected side, and embarrassment of right ventricular venous return results in jugular venous engorgement, hypotension, and ultimately cardiac arrest. Immediate treatment is required (hence the old adage that a clinician should never see an X-ray showing tension pneumothorax) although POCUS can provide real-time confirmation of the diagnosis. Emergency decompression is preferably using a specialist longer 10G non-kinking thoracocentesis catheter, inserted in the 'triangle of safety' and has a lower failure rate than the traditional approach of inserting a 14G or 16G cannula in the 2nd intercostal space, mid-clavicular line. Subsequent open thoracostomy or chest tube insertion is required. Chest compressions will help to treat pneumothorax, so if the patient is in cardiopulmonary arrest, then this is prioritized.

Investigations

Plain CXR is the baseline investigation and the size of pneumothorax should be recorded (Fig. 8.5). There are several ways to measure: in North America the apex to cupola distance tends to be used, while UK guidelines recommend the pleura to edge-of-lung distance. If a patient's first pneumothorax resolves and there is no recurrence, cross-sectional imaging with a CT thorax is unnecessary. However, if the CXR is abnormal on resolution, a CT scan should be requested. With the common use of cannabis and other recreational drugs, bullae and associated emphysema are being more commonly seen in young people.

POCUS can also be used to confirm and localize a pneumothorax.

Management

Pneumothoraces can be managed in a number of ways (**Fig. 11.23**), depending on the cause and the severity of symptoms on presentation. Treatment includes simple observation, pleural aspiration, intercostal drain, and surgical management, both in the more acute situation and electively.

- ***Observation:*** For asymptomatic patients with a primary spontaneous pneumothorax, or those with minimal symptoms (no significant pain, breathlessness, or physiological compromise), conservative management can be considered regardless of the size of the pneumothorax. Likewise, for patients with mild

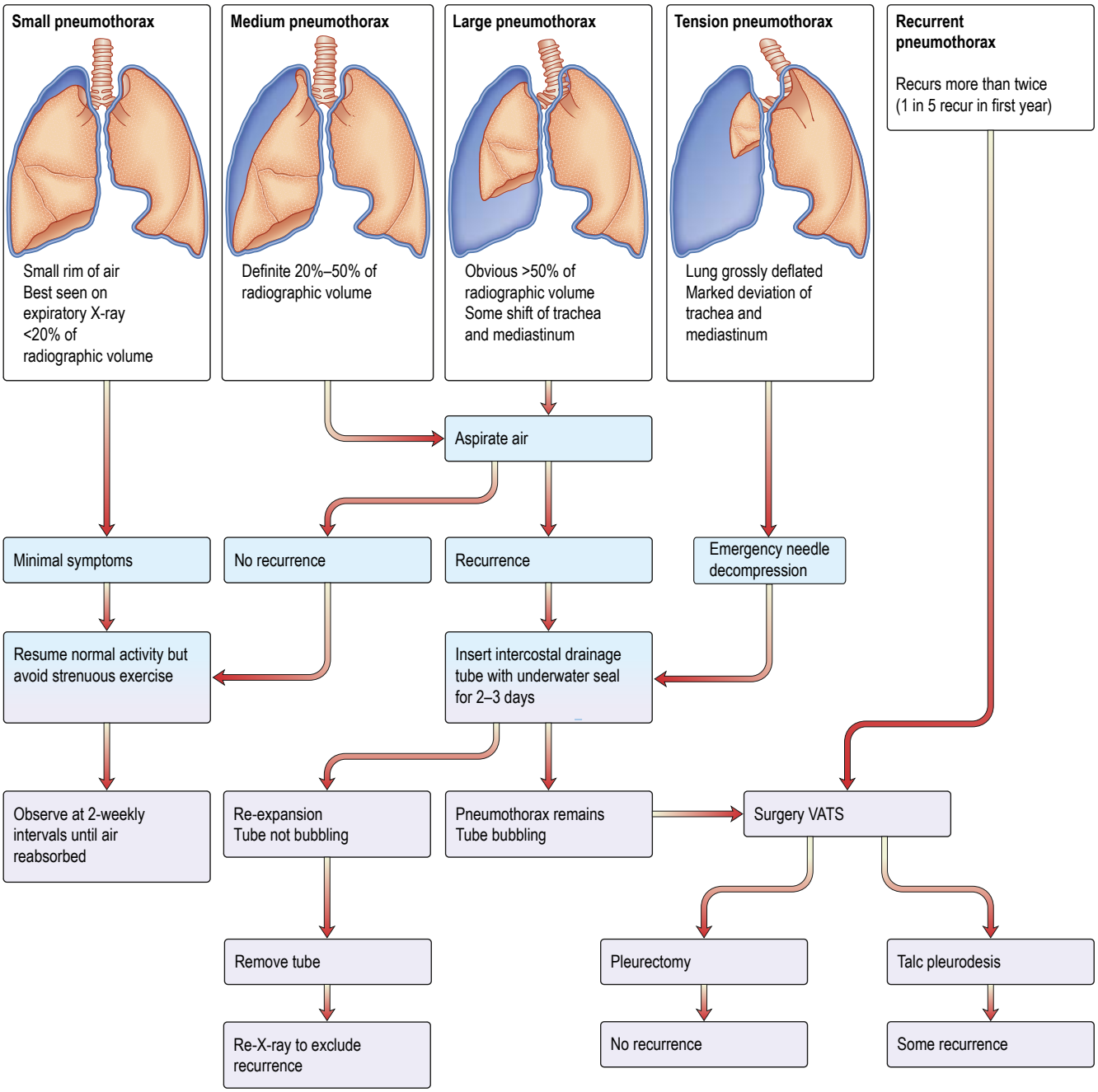


Fig. 11.23 Pneumothorax: an algorithm for management. Recent British Thoracic Society Guidelines (2023) recommend considering initial conservative management in asymptomatic/minimally symptomatic patients with primary spontaneous pneumothorax of any size (not tension pneumothorax). *VATS,* Video-assisted thoracoscopic surgery.

symptoms and a pneumothorax of <2 cm simple observation is enough. Patients can be discharged from the ED with instructions to return if their symptoms become worse. They should be advised not to fly until the pneumothorax has completely resolved and never to scuba dive unless they have had surgical pleurodesis (Box 11.33). They have a 30% chance of a recurrent pneumothorax. Patients often ask about sport and exercise, and while there are no real restrictions apart from diving, they should generally rest for a few days before exercising in moderation and should avoid contact sports for a week or two. Outpatient follow-up should be arranged within 10 days.

- ***Aspiration:*** This can be considered in symptomatic patients with a primary spontaneous pneumothorax of >2 cm or >1 cm in a secondary pneumothorax. It is undertaken using a 16 French gauge needle attached to a three-way tap and a 50 mL syringe (Box 11.34). Generally, aspiration should not be attempted in most cases of traumatic pneumothorax. It should be followed by a repeat CXR, discharge advice (as described), and outpatient follow-up.
- ***Pleural intercostal drain:*** If aspiration fails to inflate the lung satisfactorily, the patient is very breathless on admission, there is evidence of a tension pneumothorax or the cause is related to trauma, an intercostal drain should be inserted, normally in the mid-axillary line 4th intercostal space (see **Box 11.7**). The tube should be connected to an underwater drain and bottle, which are kept below the level of the patient's chest. It should be checked daily for evidence of infection at the insertion site and for kinks and leaks along its course.
- ***Surgical pleurodesis and bleb resection:*** An open surgical pleurodesis via thoracotomy or VATS procedure should be considered when the lung fails to re-inflate. Elective surgery may be appropriate after a first episode of pneumothorax in those with tension pneumothorax or who work in high-risk professions where the risk of recurrence needs to be maximally reduced (e.g., airline pilot) and after a second episode in others.
- ***Smoking:*** Patients should be strongly advised to stop smoking both tobacco and recreational drugs. Cannabis, in particular, can cause marked emphysematous changes in younger people, which leads to subcutaneous blebs and bullae.
- ***Psychological effects:*** Medical staff underestimate the concern that patients often feel, knowing that they have a medical condition that can potentially recur at any time. Some become very anxious and often request surgical pleurodesis after their second or third recurrence.

Box 11.33 Advice to patients after pneumothorax

- No flying for 1 week after complete resolution
- No diving
- Smoking cessation
- 30%–50% chance of recurrence
- Future management: consider surgery

Box 11.34 Simple aspiration of pneumothorax

1. Explain the nature of the procedure and obtain consent.
2. Infiltrate 2% lidocaine (maximum dose 3 mg/kg, not to exceed 200 mg) down to the pleura in the second intercostal space in the mid-clavicular line.
3. Push a 3–4 cm 16-gauge cannula through the pleura.
4. Connect the cannula to a three-way tap and 50 mL syringe.
5. Aspirate up to 2.5 L of air. Stop if resistance to suction is felt or the patient coughs excessively.
6. Repeat the chest X-ray (in expiration) in the X-ray department.

Ambulatory pleural drainage devices have been shown to be a feasible alternative to conventional inpatient management in patients with primary spontaneous pneumothorax in the absence of significant chronic lung disease, albeit with an increased risk of complications such as enlarging pneumothorax or device dislodgement, and may gather popularity in coming years.

Further reading

British Thoracic Society Guideline for pleural disease. *Thorax* 2023; 78:1143–1156.
Courtney Broaddus V. Clearing the air—a conservative option for spontaneous pneumothorax. *N Engl J Med.* 2020; 382:469–470.
Halifax R, McKeown E, Sivakumar P et al. Ambulatory management of primary spontaneous pneumothorax: an open-label, randomised controlled trial. *Lancet*. 2020, 396:39–49.

Haemoptysis

Haemoptysis is defined as the expectoration of blood originating from the lower respiratory tract below the glottis. It has the potential to lead to life-threatening complications and requires prompt and careful evaluation of its severity and the patient's overall status (ABCDE approach) as well as of the underlying cause.

Massive haemoptysis

Massive haemoptysis—the expectoration of a significant volume of blood—presents a substantial risk and should be addressed urgently. Definitions of massive haemoptysis vary but include expectoration of blood equal to or exceeding either 500 mL within 24 hours or a rate of 100 mL per hour, or any haemoptysis impairing gas exchange or resulting in haemodynamic instability. It is important to note that haemoptysis often involves aerated blood, which can be mistaken for a lesser severity.

Common causes of haemoptysis

Haemoptysis can be broadly categorized into five main groups based on the underlying aetiology: infective, neoplastic, vascular, autoimmune, and drug-related (Box 11.35).

It is important to discriminate true haemoptysis from 'pseudo-haemoptysis'. Examples of pseudo-haemoptysis include misinterpretation of haematemesis as haemoptysis, or coughing of blood originating from sources other than the lower respiratory tract, such as the nasopharynx, which would require different treatment approaches.

Box 11.35 Common causes of haemoptysis

Category	Common causes of haemoptysis
Infective	Pneumonia Bronchiectasis Tuberculosis Lung abscess Mycetoma
Neoplastic	Lung cancer (primary or secondary) Bronchial adenoma
Vascular	Pulmonary embolism Arteriovenous malformations Pulmonary hypertension Mitral stenosis (severe) Pulmonary oedema
Autoimmune	Granulomatosis with polyangiitis Anti-glomerular basement membrane disease (anti-GBM disease, Goodpasture's syndrome)
Drug-related	Anticoagulants and antiplatelet agents

Clinical assessment

History

- **Characterize the haemoptysis:** Establish the quantity and characteristics of the blood expectorated, such as whether it is streaked, mixed with coloured sputum or pure blood. This aids in determining both the severity and source of the haemoptysis (e.g., the haemoptysis of pulmonary oedema is often described as 'pink and frothy').
- Ask questions aimed at **ruling out pseudo-haemoptysis** (e.g., was the haemoptysis preceded by epistaxis or vomiting?)
- Ask about **bleeding at other sites**, which may point to a more generalized bleeding tendency
- Ask about **associated symptoms:**
  - Cough with coloured sputum and fever suggest infection
  - Pleuritic chest pain might suggest pulmonary infarction (PE)
  - A prodrome of constitutional symptoms including anorexia, weight loss, or chronic cough might suggest malignancy or tuberculosis
- Daily cough productive of sputum or a history of childhood respiratory disease may suggest bronchiectasis
- Ask about skin rashes, ocular or nasal symptoms, and joint aches which might suggest vasculitis
- Take a **smoking history** as a risk factor for lung cancer/chronic lung disease
- Enquire about relevant **co-morbidities** (e.g., heart failure or valvular disease, coagulation disorders)
- Seek a **family history** of bleeding disorders, cancer, or hereditary conditions that may contribute to haemoptysis (e.g., hereditary haemorrhagic telangiectasia, HHT)
- **Review medications** (particularly antiplatelets or anticoagulants)
- Enquire about **occupational or environmental exposures**, recent **travel**, and **risk factors for VTE**.

Physical examination

Physical examination should include an assessment for possible haemodynamic or respiratory compromise and should focus primarily on cardiovascular and respiratory systems.

- Generalized cachexia suggests malignancy or chronic infection
- Fingernail clubbing suggests bronchogenic carcinoma, chronic suppurative lung disease (bronchiectasis or lung abscess), or pulmonary TB
- Purpura may suggest underlying vasculitis or a generalized bleeding disorder
- Examine the skin for telangiectasia suggesting HHT
- Lymphadenopathy in the neck or axillae may suggest malignancy or infection
- Look for signs of DVT in the legs (which may imply PE)
- Listen for the low-pitched 'rumbling' mid-diastolic murmur of mitral stenosis (which may lead to bronchial venous bleeding, so-called 'pulmonary apoplexy')
- Examine the chest:
  - A localized wheeze may indicate an obstructing endobronchial lesion
  - A pleural rub may be present in patients with pneumonia or pulmonary infarction
  - Signs of lung collapse (reduced breath sounds, dull percussion noted, deviation of trachea towards affected side) suggest an endobronchial lesion
  - The presence of coarse crackles or bronchial breathing may indicate infection, but can also be heard over zones of pulmonary haemorrhage

Investigations

- **Oxygen saturations** will identify a hypoxic patient at the bedside.
- **Core lab tests** should include a full blood count to check haemoglobin and exclude thrombocytopaenia. Coagulation studies may identify a generalized bleeding tendency. Elevated white blood cell count and CRP suggest infection. Significant AKI raises the possibility of a pulmonary-renal vasculitis.
- **Additional blood tests** may include D-dimer (in the diagnostic pathway for PE), BNP or NT-proBNP (congestive cardiac failure), or autoantibody testing (e.g., ANCA, anti-GBM antibody).
- **Chest X-ray** is a high-yield investigation in the work-up of haemoptysis and may show a solitary lung lesion or lobar collapse (suggestive of malignancy), an apical cavity (suggestive of TB), or signs of congestive cardiac failure. Consolidation may be due to infection or pulmonary haemorrhage; pulmonary vasculitis often results in bilateral airspace shadowing ('diffuse alveolar haemorrhage'). Pulmonary cavities may result from malignancy, infection (including TB), or vasculitis (usually granulomatosis with polyangiitis).
- Other tests may include urinalysis (haemoproteinuria may suggest renal vasculitis), sputum microscopy, culture and cytology (including AFB smear for mycobacteria), and respiratory viral panel.
- **Further tests** required may include CT chest (diagnosis of bronchiectasis, characterization of lung diseases including cancer and TB, identification of mycetoma), CT pulmonary angiogram (diagnosis of PE), echocardiography (diagnosis of left ventricular dysfunction, pulmonary hypertension or valvular disease), or bronchoscopy.

Management of massive haemoptysis

Boxes 11.36 and 11.37 summarize the steps that should be taken in management of a patient with massive haemoptysis. Timely intervention, including airway protection, resuscitation, and consultation with respiratory or interventional radiology specialists, is important.

Box 11.36 Management of massive haemoptysis

Management of massive haemoptysis	Description
Correct patient positioning	Position the patient according to the affected lung, aligning them to lie on the same side as the lesion or abnormality observed on imaging.
Establish a patent airway	Ensure a clear airway through suctioning to facilitate effective ventilation and oxygenation.
Maintain gas exchange and cardiovascular function	Implement supportive measures to sustain adequate gas exchange and cardiovascular stability. Red cell transfusion may be required.
ITU referral	In cases of massive haemoptysis, referral to ITU is often necessary for specialized management and monitoring. Intubation, to secure the airway and allow for effective suctioning, and mechanical ventilation may also be required.
Control the bleeding	Definitive treatments to control bleeding include arterial embolization and bronchoscopic therapy (**Box 11.37**). The anti-fibrinolytic drug tranexamic acid may help temporize while planning definitive treatment.

Box 11.37 Diagnostics and therapeutic procedures employed in massive haemoptysis

Test	Description
CT angiogram	This is helpful in confirming active bleeding and identifying a bleeding point and targets for treatment, but is diagnostic only
Direct angiography ± arterial embolization	Direct angiography performed by an interventional radiologist will confirm an active site of bleeding and facilitate arterial embolization as an option by which to control bleeding.
Flexible bronchoscopy	Usually the first choice for diagnostic bronchoscopy, particularly if the patient is stable.
Rigid bronchoscopy under general anaesthesia	Provides larger suction channels and facilitates jet ventilation (a form of specialized mechanical ventilation), preferred in cases of catastrophic bleeding.
Lasertherapy	Applied bronchoscopically, can be used to manage bleeding tumours in some cases.

Nausea and vomiting

Nausea and vomiting are common symptoms encountered in clinical practice, resulting from a wide range of underlying aetiologies. Management of patients presenting with nausea and vomiting is focused on identifying and treating the underlying cause.

The origins of nausea and vomiting are diverse (Box 11.38), encompassing metabolic factors, infectious agents, surgical or mechanical triggers, and adverse reactions to both foods and medications. Additionally, these symptoms can manifest in the absence of organic disease.

Clinical assessment

History should include details on the onset, duration, and frequency of symptoms and any associated features; for example, abdominal pain, weight loss. Enquire about potential triggers or exacerbating factors; for example, changes in diet, pregnancy. Consider features which may indicate a need for urgent intervention; for example, severe abdominal pain, haematemesis. Obtain a review of systems which may provide clues to the underlying cause; for example, GI symptoms such as diarrhoea and constipation, and neurological symptoms such as headache and dizziness.

Consider the patient's medical history, focusing on conditions known to cause nausea and vomiting; for example, peptic ulcer disease. Enquire about new medications, alcohol and drug use, and travel history.

Perform a thorough physical examination, specifically assess hydration status and examine the abdomen for tenderness, distension, organomegaly, masses, or abnormal bowel sounds.

Investigations

When the underlying cause remains elusive, further investigations such as imaging (e.g., abdominal ultrasound or CT) or gastroscopy may become necessary. Appropriate specialty referral should be considered.

Management

Management hinges on identification and treatment of the underlying cause. Supportive measures, applicable to all patients, may involve oral or intravenous rehydration and antiemetic therapy (see Box 22.7).

Box 11.38 Common causes of nausea and vomiting

Infective	Gastroenteritis (viral, bacterial)
Metabolic	Diabetic ketoacidosis Hypercalcaemia Uraemia
Abdominal	Mechanical bowel obstruction Paralytic ileus Biliary obstruction Pancreatitis Gastric outlet obstruction/pyloric stenosis Gastro-oesophageal reflux Peptic ulcer disease Gastroparesis
Neurological	Raised intracranial pressure (e.g., brain tumour) Migraine Brainstem or cerebellar syndromes
Vestibular	Vertigo (any cause)
Drugs	Opiates Chemotherapy Antibiotics
Psychiatric	Bulimia nervosa
Other	Cyclical vomiting syndrome Cannabis smoking Alcohol (toxicity, or 'hangover') Myocardial infarction Hyperemesis gravidarum

Diarrhoea

Diarrhoea is formally defined as more than 300 g per day of dry faeces passed in three motions or more, however, its meaning has changed somewhat in everyday language such that it is often used to describe loose stool, urgency, or watery stool of any amount.

Clinical assessment

Patients presenting to acute medical services would usually have diarrhoea that falls into the acute (less than four weeks duration) category. Travel, social, and past medical history can be key to establishing a cause, while many medications can also cause loose stool, also making a detailed drug history important. Loose stool can also be a non-specific sign of general illness including sepsis. Box 11.39 summarizes the causes of acute diarrhoea and clues to resolving the differential diagnosis.

Investigations

- **Arterial or venous blood gas:** lactic acidosis may reflect volume depletion in severe diarrhoeal illnesses, or bowel ischaemia.
- **Core blood tests:** FBC, urea and electrolytes, creatinine, calcium, phosphate, LFTs, and CRP. Severe or prolonged diarrhoea may result in hypomagnesaemia. Additional tests depending on circumstance may include thyroid function, haematinics, and coeliac serology.
- **Stool tests:** stool cultures and *Clostridium difficile* glutamate dehydrogenase (GDH) and toxin testing are important first-line investigations. Low faecal calprotectin (<50 µg/g) excludes inflammatory bowel disease (IBD), but elevated levels are not specific and may occur with infectious colitis. Stool microscopy for ova, cysts, and parasites is indicated in the returning traveller and immunocompromised patients; the diagnostic

Box 11.39 Acute diarrhoea on the medical take: causes and diagnostic clues

Cause	Bloody stool?	History clues	Examination clues	Investigation clues
Infective colitis	Can be bloody	Short history (<14 days, even duration <24 h if viral), recent travel, social history of illness, visiting farms (*E. coli* 0157). Receptive anal sex may suggest Shigella or other STIs. Immunocompromised may suggest CMV, cryptosporidium, or fungal infections.		Eosinopenia may suggest typhoid species incl. *Salmonella*
Clostridium difficile infection	No	Recent antibiotics Previous *C. diff* carriage PPI use		WCC may be elevated (>25 × 10^9/L) May have toxic megacolon (transverse colon >6 cm width) *C. diff*. toxin or PCR positive
Drug induced	No	Causative drugs including recent antibiotics, immunotherapies, laxatives, overflow from anticholinergics or opiates, prokinetic medications	Faecal impaction	
Crohn's disease	Often bloody	Personal or family history, bimodal age peaks 15–40 and 50–60 Ulcers potentially from mouth to perianal region, mucus, and pus in stool Extraintestinal features may include episcleritis, seronegative arthritis	Lesions potentially from mouth (ulcers) to perianal region (skin tags, fistulae, scarring, sinuses) Skins lesions e.g., pyoderma gangrenosum or erythema nodosum Scars from previous surgery	Positive faecal calprotectin >50 μg/g
Ulcerative colitis	Often bloody	Personal or family history, younger age peak 15–30. History of primary sclerosing cholangitis Extraintestinal features may include uveitis, iritis, seronegative arthritis	Skins lesions e.g., erythema nodosum	May have toxic megacolon Positive faecal calprotectin >50 μg/g
Diverticular disease	Often bloody	Often older age group with history of constipation or previous diverticular disease		
Microscopic colitis	No	Female predominance, usually older age groups		By definition, no macroscopic disease such as ulcers on colonoscopy, requires biopsies to diagnose
Bacterial overgrowth	No	Colonization with bacteria after previous radiation, small bowel surgery, blind gut loop		Often B_{12} deficiency alongside that of fat-soluble vitamins (ADEK)
Bile salt malabsorption/ pancreatic insufficiency	No	Steatorrhoea Previous surgery e.g., cholecystectomy/ terminal ileum resection Alcohol misuse/pancreatitis	Scars from previous surgery	Faecal elastase low in pancreatic insufficiency
Coeliac disease	No	History of anaemia	Dermatitis herpetiformis Perioral stomatitis from iron deficiency	Anti-TTG/gliadin antibodies (check not IgA deficient)
Colon cancer	Can be bloody	Altered bowel habit, urgency. Family history of cancer	Mass on DRE	
Ischaemic colitis	Often bloody	Frequently abdominal pain, may be sudden onset. Cardiovascular risk factors or history of atrial fibrillation	Localized peritonism common	Neutrophil leucocytosis, hyperlactataemia unresponsive to fluid therapy
Irritable bowel syndrome	No	Largely diagnosis of exclusion		

yield for amoebiasis is optimal when examining a very fresh sample.

- **Abdominal X-ray:** detection of toxic megacolon in patients with severe *Clostridium difficile* or flare of IBD.
- Other investigations may be required if there is diagnostic uncertainty or symptoms fail to settle. These may include flexible sigmoidoscopy, abdomino-pelvic CT, small bowel MRI, or capsule endoscopy. Specialist input from

gastroenterology or infectious diseases teams may be required.

Management

Where suspicion exists for an infective cause, or the cause is unclear, patients with acute diarrhoea should be nursed in a side-room with appropriate barrier precautions and signage.

Acute gastrointestinal infections

These are common and often self-limiting in nature. Passage of bloody stool may indicate perianal disease (often bright red and coating the stool), or presence of an invasive organism causing enteritis (blood mixed with stool). Examples of invasive organisms include shigella (which may be acquired as a sexually transmitted infection), salmonella, campylobacter (both often called 'food poisoning'), or more rarely *Entamoeba histolytica* (if travel related; see Chapter 42). Empirical antibiotics, for example, azithromycin may be warranted although are usually reserved for those with severe or prolonged symptoms, or risk factors for severe illness such as immunocompromise. The management of acute diarrhoeal illness is usually supportive: treat dehydration with i.v. or oral fluids (oral rehydration solutions may be needed), check electrolytes, and replace as required.

Clostridium difficile (CD) colitis

This results from infection with a toxin-secreting strain of *Clostridium difficile* (CD) and typically follows broad-spectrum antibiotic treatment, depleting the normal colonic bacterial flora and allowing rapid multiplication. Previous carriage of CD (usually detected by glutamate dehydrogenase testing) or a very high white cell count may be clues to the diagnosis.

The initial treatment of CD is usually with oral metronidazole or oral vancomycin. There is a risk of toxic megacolon, which can be diagnosed on plain imaging or CT, and is a surgical emergency. Occasionally other medications (e.g., fidaxomicin), or even stool transplantation, taken as either frozen capsule or via nasogastric tube, are required to treat severe or chronic disease.

Drug-induced diarrhoea

Many medications can provoke a change in bowel habit with common offenders including metformin, opiates (constipation with overflow), laxatives, antiemetics (including metoclopramide), and weight loss medications (such as orlistat or acarbose). Non-infective diarrhoea is also a common side-effect of antibiotic therapy or for those patients receiving enteral feeding.

Non-infective colitis may present after treatment with immune checkpoint inhibitors for malignancy. Commonly, onset is approximately 5 to 10 weeks after the second or third dose of treatment. Infective causes may still need to be excluded. Affected patients may exhibit other organ-specific toxicity including adrenalitis, hypophysitis, thyroiditis, etc. Treatment is with high-dose immunosuppression for example, prednisolone 1 mg/kg or i.v. methylprednisolone.

Inflammatory bowel disease flare

Chronic inflammatory bowel diseases (IBD) that cause acute relapsing disease ('flares') include Crohn's disease and ulcerative colitis (UC) and are most often seen in younger patients (less than 40 years). Crohn's disease is characterized by transmural patchy inflammatory changes that can affect any portion of the bowel from the mouth (ask about ulcers) to the anus (ask about perianal disease). Ulcerative colitis is, as the name suggests, confined to the colon and manifests with mucosal inflammation that is usually confluent. Extraintestinal manifestations are well recognized in both conditions, including erythema nodosum, arthritides, and eye involvement, among others.

Clinical assessment

Patients often present with episodes of diarrhoea (often with blood or mucus) as well as weight loss, abdominal pain, and bloating. It is sometimes difficult to differentiate an infectious aetiology from an inflammatory one, and often the two coexist. Usually, a purely infective cause will settle after a few days whereas inflammatory causes may persist and recur. While patients with flares of UC will most commonly present with bloody diarrhoea, patients with Crohn's disease exhibit a much broader range of presentations including strictures (signs of intestinal obstruction), terminal ileitis (sometimes mimicking appendicitis), perforations, enteric fistulae (to bladder, vagina or skin) and perianal disease (including abscesses) and so need close examination and observation, and often a low threshold for cross-sectional imaging (e.g., CT abdomen and pelvis). The clinical status of patients should be monitored regularly (fever, tachycardia, stool frequency) and specialized gastroenterological advice should be sought early.

Investigations

FBC, CRP, liver function tests, urea and electrolytes and magnesium should be performed. Serial abdominal X-rays are required if concern exists for toxic megacolon (a thin-walled colon with diameter of >6 cm suggests impending perforation and requires urgent surgical referral). Stool cultures should be sent to exclude enteric infection. The incidence of concomitant *C. difficile* infection in patients admitted with severe colitis is increasing. This is associated with a significant increase in morbidity and must be sought and treated promptly.

Early flexible sigmoidoscopy should be performed to confirm diagnosis and biopsies taken to evaluate for cytomegalovirus. Additional small bowel assessments (e.g., MRI) may be required in Crohn's disease. Full colonoscopy should **not** be undertaken during severe attacks due to unacceptably increased risk of perforation and instead examinations limited to an unprepared flexible sigmoidoscopy.

Management

Ulcerative colitis

Patients with acute severe colitis (see **Box 11.40**) should be admitted to hospital and treated initially with hydrocortisone 100 mg i.v. 6-hourly, and with prophylactic doses of s.c. LMWH to prevent VTE. Corticosteroids should not be delayed while waiting for stool culture or *C. difficile* toxin results, although if *C. difficile* is strongly suspected, oral vancomycin should be commenced in parallel.

Success or failure of medical treatment of a severe attack of UC must be judged by an experienced gastroenterologist and colorectal surgeon. If patients have not responded to intravenous steroids within 3 days, either salvage medical therapy or surgery is required. If patients respond to intravenous steroids, they should be switched to oral prednisolone after approximately 5 days, which they can be weaned off over 8–10 weeks. All patients who have been admitted for severe colitis should be considered for long-term maintenance therapy with a thiopurine (azathioprine/mercaptopurine), biological agents such as anti-TNF or newer immunomodulatory therapies (see Chapter 22).

Crohn's disease

Management should always be in conjunction with a gastroenterology with IBD expertise and will vary widely depending on the acute

Box 11.40 Definition and management of a severe attack of ulcerative colitis

Definition (Modified Truelove and Witts criteria)

- Stool frequency: >6 stools/day with blood +++
  AND systemic toxicity defined by one of:
- Fever: >37.8°C
- Tachycardia: >90 beats/min
- Anaemia: <105 g/L haemoglobin
- CRP: >30 mg/L

Management

- Admit to hospital
- Test bloods: FBC/U&E/magnesium/CRP/LFTs
- Abdominal X-ray (assess for toxic megacolon)
- Exclude enteric infection (stool culture, CDT assay)
- Confirm diagnosis with unprepared limited flexible sigmoidoscopy (include histology to evaluate for CMV)
- Assess fluid status
- Give prophylactic anticoagulation
- i.v. hydrocortisone 100 mg, 4 times daily
- Monitor daily:
  - Stool frequency
  - Abdominal X-ray
  - Bloods (FBC/CRP/albumin)

clinical manifestation (e.g., intestinal obstruction vs. diarrhoea vs. abscess vs. enterocutaneous fistula, etc.). Early nutritional assessment and support is vital. Hospitalized patients should receive prophylactic s.c. LMWH for prevention of VTE.

In all cases, close liaison with a specialist IBD multidisciplinary service is required for longer-term management or for those patients with severe acute complications requiring anything other than a short hospital stay.

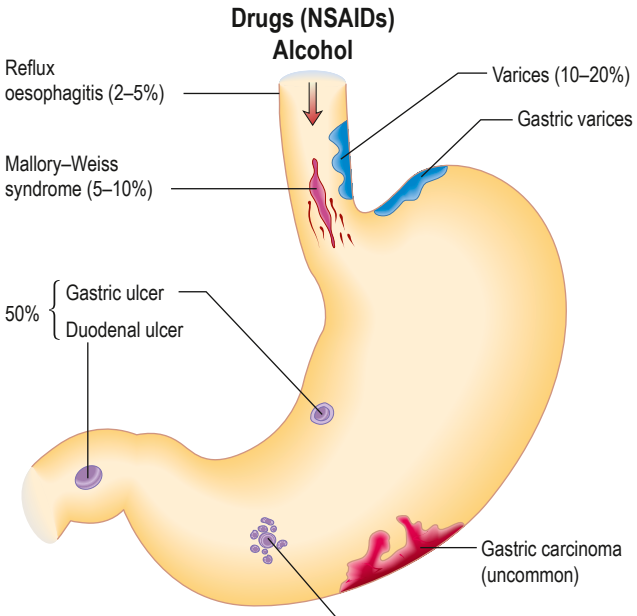
Haematemesis and melaena

Acute upper GI bleeding impacts approximately 1.33 individuals per 1000 in the population, equating to 85 000 cases annually. Notably, the 2015 NCEPOD report on GI haemorrhage management highlighted that 40% of patients receiving treatment for acute upper GI bleeding were inpatients initially admitted for a different medical condition.

The cardinal features are haematemesis (the vomiting of blood) and melaena (the passage of black tarry stools, the black colour being due to blood altered by passage through the gut). Melaena can occur with bleeding from any lesion proximal to the right colon. Rarely, melaena can also result from bleeding from the right colon.

Following a bleed from the upper GI tract, unaltered blood can appear *per rectum*, but the bleeding must be massive and is almost always accompanied by shock. The passage of dark blood and clots without shock is most often due to lower GI bleeding. Of note, stool may be frankly bloody or maroon with massive or brisk upper GI bleeding.

The **causes of upper GI bleeding** are summarized in Fig. 11.24. Peptic ulceration is the leading and most common cause of serious and life-threatening bleeding. Other important causes include oesophageal or gastric varices, arteriovenous malformations, tumours, and oesophageal (Mallory–Weiss) tears. Drugs including aspirin and NSAIDs can induce ulcers and erosions, with the risk of GI haemorrhage particularly marked in the elderly. Combining glucocorticoids and NSAIDs amplifies the incidence of GI haemorrhage. Anticoagulants and other antiplatelet agents do not cause acute GI haemorrhage *per se*, but bleeding from any cause is greater if the patient is anticoagulated.



Other uncommon causes
Gastric antral vascular ectasia (GAVE)
Hereditary telangiectasia (Osler–Weber–Rendu syndrome)
Pseudoxanthoma elasticum
Blood dyscrasias
Dieulafoy gastric vascular abnormality
Portal gastropathy
Aortic graft surgery with fistula

Fig. 11.24 Causes of upper gastrointestinal haemorrhage. The approximate frequency is also given. *NSAID,* Non-steroidal anti-inflammatory drug.

Clinical assessment

Evaluation should begin with focused history and examination, although emergency management measures (e.g., establishing intravenous access, resuscitation, and obtaining blood samples for crossmatching) may need to commence in parallel.

History

- Take a description of the **nature and quantity of blood** vomited or passed per rectum. Whenever possible a sample of vomitus or faeces should be inspected directly: the 'faeculant' vomiting of small bowel obstruction may be mistaken for the 'coffee ground' appearance of altered/coagulated blood; likewise the green-black/greyish stools of patients taking oral iron supplements may be mistaken for melaena
- **Onset and time course of symptoms:** A preceding history of vomiting/retching may indicate a Mallory-Weiss tear. A prodrome of epigastric discomfort relieved by eating or antacids may indicate a duodenal ulcer (the pain from which is often absent when haemorrhage occurs)
- **Medication use:** NSAIDs, aspirin, anticoagulants, antiplatelet agents
- **Alcohol history**
- **Previous GI bleeding** or endoscopic examinations
- **History of liver disease** (risk of varices) or **coagulopathy**

Physical examination

- **Vital signs:** Tachycardia and orthostatic blood pressure changes suggest moderate to severe blood loss; hypotension suggests life-threatening blood loss (hypotension may be late finding in healthy young adult)
- **Rectal examination** is performed to assess stool colour (melaena vs. haematochezia vs. brown)
- Significant abdominal tenderness accompanied by signs of peritoneal irritation (e.g., involuntary guarding) suggests perforation
- Look for signs of CLD (Chapter 23)

Investigations

- Laboratory tests should include full blood count, renal and liver function tests, clotting screen, blood group/antibody screening, and in all cases of severe bleeding or high-risk patients, cross-matching of red cell units.
- Venous blood gas sampling for rapid Hb measurement and lactate.
- In very brisk upper GI bleeds, the initial haemoglobin may be misleading and serial FBC samples are required.

Risk assessment

The routine use of validated risk scores for upper GI bleeding facilitates optimal triage decisions regarding timing of endoscopy.

The **Rockall score** (Boxes 11.41 and 11.42) predicts risk of rebleeding and death based on combined clinical and endoscopy findings. The 'pre-endoscopy Rockall' score (calculated based on age, circulatory status, and co-morbidities) is also validated to predict mortality in patients prior to undergoing gastroscopy.

The **Glasgow-Blatchford score** (Box 11.43) uses the level of plasma urea, haemoglobin, and clinical markers, but not endoscopic findings, to determine the need for intervention such as blood transfusion or endoscopy in GI bleeding.

Patients with very low pre-endoscopy Rockall scores (0–1) or Glasgow-Blatchford scores (0) may be suitable for ambulatory management if appropriate local pathways exist.

Scoring systems should never be relied upon exclusively for assessment of risk and urgency and should always be combined with common sense and clinical judgement.

The following factors affect the risk of rebleeding and death:

- age
- evidence of co-morbidity e.g., cardiac failure, ischaemic heart disease, CKD, and malignant disease
- presence of the classical clinical features of shock (pallor, cold peripheries, tachycardia, and low blood pressure)
- endoscopic diagnosis e.g., Mallory–Weiss tear, peptic ulceration, gastric antral vascular ectasia (GAVE; Fig. 11.25)
- endoscopic stigmata of recent bleeding e.g., adherent blood clot, spurting vessel
- clinical signs of CLD.

Bleeding associated with liver disease is often severe and recurrent if it is from varices. Liver failure can develop.

Management

Immediate management

This is shown in Box 11.44. In addition, stop NSAIDs, aspirin, clopidogrel, warfarin, and DOACs if patients are taking them (Box 11.45). Stopping antiplatelets may be hazardous (e.g., in patients with coronary artery stents) and can provoke thrombosis; discuss this urgently with a cardiologist.

Many hospitals have multidisciplinary specialist teams with agreed major haemorrhage protocols and the latter should be followed. Patients should be managed in monitored beds or ITU depending on the severity of bleeding. Oxygen should be given and the patient should be kept nil by mouth until an endoscopy has been performed.

Patients with large bleeds and clinical signs of shock require urgent resuscitation (see Fig. 13.22).

Blood volume

See Box 11.46. The major principle is to restore the blood volume rapidly to normal via two or more large-bore intravenous cannulae; plasma expanders or crystalloids are given until red cells become available. Transfusion of red cell concentrates is used with

Box 11.42 Rockall scores post-endoscopy

Rockall score	Predicted probabilities	
	Rebleeding risk (%)	Mortality (%)
0	5	0
1	3	0
2	5	0
3	11	3
4	14	5
5	24	11
6	33	17
7	44	27
8+	42	41

Box 11.41 Rockall risk assessment score in non-variceal upper gastrointestinal haemorrhage

Variable	Score			
	0	1	2	3
Age (years)	<60	60–79	>79	–
Circulation	BP >100 mmHg	BP > 100 mmHg	BP < 100 mmHg	–
	Pulse < 100 beats/min	Pulse >100 beats/min	Pulse > 100 beats/min	
Co-morbidity	None	–	Cardiac disease, any other major co-morbidity	Chronic kidney disease, liver failure, disseminated malignancy
Endoscopic diagnosis	Mallory–Weiss tear, no lesion	All other diagnoses	Malignancy of the upper gastrointestinal tract	–
Major SRH	None, or dark spots	–	Blood in the upper gastrointestinal tract, adherent clot, or spurting vessel	–

BP, Blood pressure (systolic); *SRH*, stigmata of recent haemorrhage.

Box 11.43 The Glasgow-Blatchford bleeding score

Criterion	Threshold	Score
Blood urea (mmol/L)	6.5–7.9	2
	8.0–9.9	3
	10.0–24.9	4
	>25	6
Haemoglobin in males (g/L)	120–129	1
	100–119	3
	<100	6
Haemoglobin in females (g/L)	100–119	1
	<100	6
Systolic blood pressure (mmHg)	100–109	1
	90–99	2
	<90	3
Pulse	≥100/min	1
Melaena at presentation	Yes	1
Syncope at presentation	Yes	2
Hepatic disease	Yes	2
Cardiac failure	Yes	2

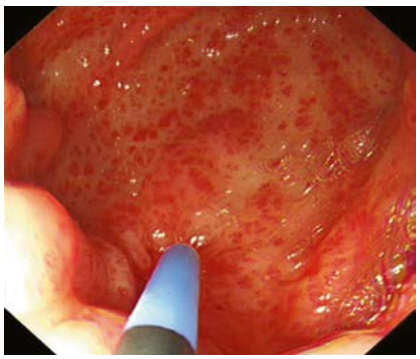


Fig. 11.25 Endoscopic view of gastric antral vascular ectasia (GAVE). An argon plasma coagulation catheter is shown at 6 o'clock.

a proposed transfusion threshold of 70 g/L. A higher threshold (90 g/L) may be appropriate in those with significant cardiopulmonary co-morbidities (e.g., coronary artery disease) or the elderly.

Transfusion must be monitored to avoid overload leading to heart failure, particularly in the elderly. The pulse rate and venous pressure are guides to adequacy of transfusion. A central venous pressure line is inserted for patients with organ failure who require blood transfusion, and in those most at risk of developing heart failure.

Haemoglobin levels are generally a poor indicator of the need to transfuse because anaemia does not develop immediately as haemodilution has not taken place. In most patients, the bleeding stops, albeit temporarily, so that further assessment can be made.

Endoscopy

Endoscopy will usually diagnose, stratify risk, and enable therapy to be performed if needed. Endoscopy should be carried out as soon as possible after the patient has been resuscitated. Patients with Rockall scores of 0 or 1 pre-endoscopy or Glasgow-Blatchford score of 0 may be candidates for immediate discharge (see below) and outpatient endoscopy the following day, depending on local policy.

Box 11.44 Management of acute gastrointestinal bleeding

- Take history and perform examination. Note co-morbidity
- Monitor pulse and BP half-hourly
- Take blood for haemoglobin, urea, electrolytes, liver biochemistry, coagulation screen, group, and crossmatching (2 units initially)
- Establish intravenous access—two large-bore i.v. cannulae (16 gauge or larger)
- Give oxygen therapy
- Give blood transfusion/crystalloid/colloid as necessary to correct haemodynamic instability. Indications for blood transfusion include:
  a. Shock (pallor, cold nose, systolic BP below 100 mmHg, pulse >100 beats/min)
  b. Haemoglobin <70 g/L in low-risk patients, <90 g/L in high-risk patients (elderly or cardiovascular co-morbidities)
- Consider need to activate major haemorrhage/massive transfusion protocols
- Keep 'nil by mouth'
- Perform urgent endoscopy in shocked patients/liver disease
- Continue to monitor pulse and BP
- Re-endoscope for continued bleeding/hypovolaemia
- Arrange surgery if bleeding persists

Box 11.45 Management of major upper GI bleed in patients with coagulopathy

- Initial resuscitation
- Blood transfusion with care
- Platelets if platelets <50 *and* actively bleeding/haemodynamically unstable and/or platelet dysfunction (e.g., chronic aspirin therapy)
- Fresh frozen plasma if fibrinogen <1 g/L *or* PT/INR or APTT > 1.5 times normal
- Prothrombin complex concentrate if taking warfarin and actively bleeding
- The half-lives of DOACs are short and in most cases GI bleeds will settle with resuscitation, blood products, and cessation of the drug; with massive or life-threatening bleeds, consider use of reversal agents (idracizumab for dabigatran, andexanet alfa for rivaroxaban/apixaban) in discussion with haematology. If specific reversal agents are not available, prothrombin complex concentrate can be considered.

Box 11.46 Blood transfusion in upper GI bleeding

- Avoid over-transfusion in variceal bleeds
- Hypovolaemic patients with normal Hb may need blood
- Haemodynamic instability despite crystalloid resuscitation
- Hb <90 g/L in high-risk patients (e.g., elderly, coronary artery disease)
- Hb <70 g/L in low-risk patients

Endoscopic treatment of acute GI bleeding is discussed further in Chapter 22, see 'Endoscopic management of acute gastrointestinal bleeding'.

Balloon tamponade may be performed as a temporizing measure for patients with uncontrollable haemorrhage likely due to varices using any of several devices (e.g., Sengstaken–Blakemore tube). Tracheal intubation is necessary if such a device is to be placed; ensure proper device placement prior to inflation to avoid oesophageal rupture.

Drug therapy

Use of acid suppression treatments (proton pump inhibitors) **before endoscopy** in patients with non-variceal upper GI bleeding is currently controversial. UK NICE guidance recommends *against* this in light of limited evidence of impact on mortality, rebleeding risk,

or need for surgery. The practice remains common, however, and may be associated with a reduction in the need for endoscopic interventions.

The antifibrinolytic agent, tranexamic acid, is no longer widely used in upper GI bleeding in light of a recent large randomized controlled trial (HALT-IT trial) showing no benefit on mortality but increased risk of VTE.

If suspecting *variceal* bleeding, immediate management should include:

- Full septic screen and prophylactic antibiotics
- Terlipressin 2 mg i.v., four times daily
- Erythromycin 3 mg/kg i.v. (20 min) (or metoclopramide 10 mg i.v.) in patients with severe bleeding; 30–90 min prior to endoscopy

After endoscopy, an intravenous PPI (e.g., omeprazole 80 mg followed by infusion 8 mg/h for 72 h) should be given to all patients with actively bleeding ulcers or ulcers with a visible vessel, as it reduces rebleeding rates and the need for surgery.

Uncontrolled or repeat bleeding

Endoscopy should be repeated to assess the bleeding site and to treat, if possible. Embolization by an interventional radiologist may be necessary if the bleeding persists. If this is not available locally or is unsuccessful, surgery is used to control the haemorrhage primarily.

Discharge policy

The patient's age, diagnosis on endoscopy, co-morbidity, the presence or absence of shock, and the availability of support in the community should be taken into consideration. In general, all patients who are haemodynamically stable and have no stigmata of recent haemorrhage on endoscopy (Rockall score post endoscopy ≤1) can be discharged from hospital within 24 hours. All shocked patients and patients with significant co-morbidity need longer inpatient observation.

Further reading

British Society of Gastroenterology acute upper GI bleed care bundle. Accessed at: bsg.org.uk/clinical-resource/bsge-acute-upper-gi-bleed-care-bundle.

The HALT-IT Trial Collaborators. Effects of a high-dose 24-h infusion of tranexamic acid on death and thromboembolic events in patients with acute gastrointestinal bleeding (HALT-IT): an international randomised, double-blind, placebo-controlled trial. *Lancet.* 2020; 395:1927–1936.

Laine L. Timing of endoscopy in patients hospitalized with upper gastrointestinal bleeding. *N Engl J Med*. 2020; 382:1361–1363.

Abdominal swelling

A swelling in the abdomen can be *focal*; for example, hepatomegaly, splenomegaly, hernia; or *generalized*; for example, ascites or small bowel obstruction (Box 11.47). The common causes of generalized abdominal swelling can usually be easily separated through history/examination and readily available laboratory and imaging tests:

- ***Ascites:*** Dull percussion note over fluid (dullness shifts on rolling), presence of a 'fluid thrill', confirmed with POCUS or departmental USS.
- ***Bowel obstruction:*** Symptoms include colicky abdominal pain, vomiting, and (absolute) constipation but vary depending on the level of obstruction; tympanic percussion note over gaseous distension; exaggerated high-pitched ('tinkling') bowel sounds in mechanical bowel obstruction; typical bowel gas patterns of large or small bowel obstruction on abdominal X-ray (see 'Plain radiography/X-rays' section, Chapter 8)
- ***Constipation:*** Faecal impaction should be recognized by digital rectal examination. Patients may experience 'overflow diarrhoea' which may obscure diagnosis. Risk factors include immobility, prolonged hospitalization, and opiate use.
- ***Pregnancy:*** Occasionally unsuspected by both patient and clinician. Point-of-care pregnancy testing should be considered in the assessment of all females of childbearing age.
- ***Obesity:*** Rarely a reason for acute presentation but may hinder recognition of other causes of abdominal swelling. A patient's subjective view on changes in their abdominal girth may be helpful.

? Box 11.47 Causes of abdominal swelling

Focal

- Organomegaly (e.g., massive hepatomegaly or splenomegaly, polycystic kidney disease, fibroid uterus, pancreatic pseudocyst)
- Distended hollow viscus (e.g., gastric outlet obstruction, urinary retention)
- Hernia
- Solid organ tumour
- Lymphadenopathy
- Haematoma (e.g., within the rectus sheath)

Generalized (sometimes remembered as the '5 Fs')

- Ascites ('Fluid')
- Small or large bowel obstruction ('Flatus')
- Pregnancy ('Foetus')
- Constipation ('Faeces')
- Obesity ('Fat')
- Ruptured aneurysm

? Box 11.48 Causes of ascites according to type of fluid

- **Straw-coloured**
  - Malignancy (most common cause)
  - Cirrhosis
  - Infective:
    - Tuberculosis
    - Following intra-abdominal perforation—any bacterium may be found (e.g., *Escherichia coli*)
    - Spontaneous in cirrhosis
  - Hepatic vein obstruction (Budd–Chiari syndrome)—protein level high in fluid
  - Chronic pancreatitis
  - Congestive cardiac failure
  - Constrictive pericarditis
  - Meigs' syndrome (ovarian tumour)
  - Hypoproteinaemia (e.g., nephrotic syndrome)
- **Chylous**
  - Obstruction of main lymphatic duct (e.g., by carcinoma)—chylomicrons are present
  - Cirrhosis
- **Haemorrhagic**
  - Malignancy
  - Ruptured ectopic pregnancy
  - Abdominal trauma
  - Acute pancreatitis

Ascites

Ascites, the accumulation of fluid within the peritoneal cavity, has a broad differential diagnosis (Box 11.48), with first presentation often to acute medical services.

Clinical assessment

- When did the patient first notice abdominal swelling and at what rate did this progress?

Box 11.49 Common abnormal blood results faced by the AIM physician

Blood test	Reference range	Mild derangement	Moderate derangement	Severe derangement
Potassium (mmol/L)	3.5–5.0	3.0–3.4 5.1–5.9	2.5–2.9 6.0–6.4	<2.5 ≥6.5
Sodium (mmol/L)	135–145	130–134 146–150	125–129 151–155	<125 >155
Calcium (corrected) (mmol/L)	2.20–2.60	2.10–2.19 2.61–2.75	1.90–2.09 2.76–3.00	<1.90 >3.00
Creatinine (μmol/L)	53–97	98–133	134–177	≥178
Platelets ($\times 10^9$/L)	150–400	100–149	50–99	<50

Note: Reference ranges may vary slightly between laboratories, and the values provided here are indicative. Interpretation of results should consider patient context, clinical history, and any specific guidelines from the relevant healthcare institution. This table serves as a general guide and is not exhaustive.

- Are there associated symptoms which may point towards underlying malignancy (appetite or weight loss or localizing symptoms such as rectal or abnormal vaginal bleeding or change in bowel habit) or infection (e.g., fevers or night sweats)?
- **Jaundice** may signify malignancy (commonly pancreatic cancer or liver metastases) or liver disease.
- Seek risk factors for CLD (alcohol and drug use, medications including herbal or traditional remedies, sexual history, travel history, vaccination history, and any known previous liver disease) and examine for signs of CLD and portal hypertension (see Chapter 23).
- **Abdominal pain** could signify Budd–Chiari syndrome (usually right upper quadrant accompanied by tender hepatomegaly and jaundice), an underlying malignancy (e.g., pancreatic or ovarian carcinoma) or the superadded complication of spontaneous bacterial peritonitis.
- **Generalized body swelling** may indicate nephrotic syndrome, congestive cardiac failure (most commonly right heart failure, for instance due to cor pulmonale), or cirrhosis.
- **Tense ascites** may splint the diaphragm and increase the patient's work of breathing and is one indication for paracentesis.

Investigations

- **Laboratory investigations** should include full blood count, 'liver function tests' (including serum albumin), renal function, and clotting profile.
- **Urinalysis** will show heavy proteinuria (3–4+) in ascites due to nephrotic syndrome.
- **Chest X-ray** may show cardiomegaly or pulmonary oedema (congestive cardiac failure), pulmonary metastases, or signs of pulmonary tuberculosis.
- **Ultrasound:** POCUS is commonly used to confirm the presence of ascites and mark sites for diagnostic aspiration or drainage. Departmental ultrasound will allow detailed organ assessment to seek evidence of hepatic cirrhosis or portal hypertension, hepatic vein thrombosis or liver metastases.
- **CT** (contrast enhanced): Request where a high index of suspicion exists for malignant ascites.
- **Ascitic fluid analysis:** This should be performed under ultrasound guidance and is essential for determining the underlying aetiology. An ascitic tap should be performed in all patients presenting with ascites, and the fluid should be sent for analysis, including cell count, total protein, albumin, culture (including acid-fast bacilli smear where appropriate), and cytology. The serum-ascites albumin gradient (SAAG) can be calculated to differentiate between portal hypertension-related ascites (SAAG ≥ 11 g/L) and non-portal hypertension causes (SAAG < 11 g/dL), see **Box 23.20**. A neutrophil count ≥250 cells/mm^3 in ascitic fluid is diagnostic for SBP and empirical antibiotic therapy should be initiated without waiting for culture results.

Management

Management of ascites should be directed towards the underlying causes and varies depending on aetiology.

Potential indications for therapeutic drainage of ascites include:

- Relief of symptoms (commonly performed in patients with decompensated cirrhosis or malignancy)
- Diaphragmatic splinting impairing ventilation
- Hepatorenal syndrome with tense ascites (relief of abdominal hypertension).

Large-volume paracentesis in patients with decompensated cirrhosis requires albumin replacement to support colloid oncotic pressure (usually 8 g albumin per litre of ascites removed, using a 20% solution). Albumin replacement is not required when draining non-cirrhotic ascites, although it is customary to do this in a staggered fashion to allow for assessment of haemodynamic tolerance.

Abnormal blood results

A frequent challenge faced by the AIM physician is the management of unexpected abnormal outpatient blood results in an out-of-hours setting. Examples include electrolyte disturbances, full blood count abnormalities, renal dysfunction, or prolonged clotting times. Gather as much relevant information regarding the patient as possible and then take steps to ensure patient safety. Information gathering can be challenging, as the laboratory may have very limited documentation related to the requested blood test. Search the available records of the patient for information on their medical history and medications and for clues as to the indication for the blood test concerned.

Box 11.49 offers some common examples of abnormal blood results which the AIM physician may encounter out-of-hours. Severely deranged results should trigger patient contact and emergency review.

A common example is an alert from the biochemistry laboratory regarding a significantly elevated potassium level. After reviewing available electronic records, the next step is to consider the abnormal result's severity: a serum potassium ≥6.5 mmol/L is considered severe and potentially life-threatening and requires immediate assessment. The patient should be contacted and advised to attend their local ED for a repeat blood test and an ECG to look for features of severe hyperkalaemia.

In the case of mild hyperkalaemia (serum potassium 5.1–5.9 mmol/L) it is usually unnecessary for the patient to attend emergency services immediately and instead they could be advised to attend ambulatory emergency care within 24 hours.

Box 11.50 Five key questions to ask when confronted by an abnormal creatinine result

1. Why was the blood test taken—routine monitoring or acutely unwell patient?
2. Do they have a previous creatinine result to compare with?
3. Does the patient have known chronic kidney disease (CKD)?
4. Have any new medications been commenced that could be responsible?
5. Has the patient recently completed a pregnancy? (creatinine falls below 80% of normal in second trimester)

For moderate hyperkalaemia (serum potassium 6.0–6.4 mmol/L) context becomes pivotal; for instance, a patient with chronic kidney disease (CKD) and hyperkalaemia may require different management than someone with apparently new hyperkalaemia of unclear origin. When uncertain, proceed cautiously and arrange urgent repeat testing and assessment.

The original requestor should be notified, outlining the abnormal result and the subsequent management undertaken. Follow-up investigations and consultations should be arranged as necessary to comprehensively address the patient's condition and ensure resolution.

Acute kidney injury

Acute kidney injury (AKI) refers to a sudden reduction in renal function, potentially impairing fluid, electrolyte, and acid–base homeostasis. It is associated with up to 20% of emergency hospital admissions and has been linked to overall poorer health outcomes. AKI is often seen on the acute medical take, either in conjunction with other pathology or as a standalone referral from primary care having been discovered on a blood test. As AKI is a clinico-biochemical syndrome, biochemistry results need to be interpreted with an understanding of the clinical context in which the blood test was taken (Box 11.50).

Definition and staging

AKI is recognized either by an acute rise in serum creatinine or fall in urine output and categorized into three stages of severity by the Kidney Disease Improving Global Outcomes (KDIGO) staging criteria (Box 11.51). Increasing severity of AKI correlates with worse clinical outcomes.

AKI may cause sudden, life-threatening metabolic and clinical disturbances such as hyperkalaemia or pulmonary oedema, particularly when a patient is oligo-anuric. While there is no explicit threshold for referral and admission to hospital, AKI stage 1 is usually investigated and managed via outpatient pathways such as ambulatory care or SDEC. Management of stage 2 and 3 AKI depends on clinical context, although most patients require referral to secondary care for inpatient management.

When assessing a patient with AKI, the parallel aims should be to identify and treat life-threatening complications and determine the cause—pre-renal, renal, or postrenal.

Causes of AKI

AKI may result from:

- ***Pre-renal*** causes—Falling kidney blood flow leads to a falling glomerular filtration rate (GFR). This might be due either to changes in the circulation, or to intrarenal vasomotor changes that drop glomerular perfusion pressures. Common causes of a falling effective circulating volume leading to renal hypoperfusion are outlined in Box 11.52. Common drugs interfering with intrarenal blood flow include non-steroidal anti-inflammatory drugs (NSAIDs) and ACE inhibitors/angiotensin receptor blockers (ARBs), which may contribute to AKI in patients with concomitant factors causing renal hypoperfusion (Fig. 11.26).
- ***Intrarenal*** parenchymal disorders (injury to glomerulus, tubule or vessels)—Most commonly acute tubular injury (ATI; formerly referred to as acute tubular necrosis), but with a broad differential diagnosis (see section on 'Intrarenal parenchymal AKI').
- ***Postrenal*** causes (urinary tract obstruction—functioning kidneys cannot excrete urine with back-pressure affecting function).

A useful mnemonic for common causes of AKI encountered on the medical take is 'STOP'—**S**epsis, **T**oxins, **O**bstruction, **P**arenchymal.

Box 11.51 KDIGO[a] classification of acute kidney injury

Stage	Serum creatinine concentration (SCr)	Urine output criteria
1	SCr 1.5–1.9 times baseline OR ≥ 26.5 μmol/L (0.3 mg/dL) increase	<0.5 mL/kg/h for 6–12 h
2	SCr 2–2.9 times baseline	<0.5 mL/kg/h for 6–12 h
3	SCr 3 times baseline OR initiation of kidney replacement therapy OR in patients <18 years, decrease in eGFR to <35 mL/min per 1.73 m^2	Anuria for ≥12 h

[a]Kidney Diseases: Improving Global Outcomes.

Box 11.52 Causes of a reduced effective circulating volume

- Dehydration
- Reduced intake (nil by mouth, confusion)
- Gut losses (vomiting, nasogastric tube losses, diarrhoea)
- Kidney losses (diuretics, hyperglycaemia)
- Burns, sweating
- Haemorrhage
- Third space losses (pancreatitis, peritonitis, bowel obstruction)
- Systemic vasodilation (septic shock, cirrhosis)
- Cardiac failure or shock (myocardial infarction, arrhythmias, cardiomyopathy, tamponade)

Clinical assessment

- What is the **clinical context** (Box 11.50)? Is the patient unwell?
  - A large proportion of AKI in acute medical patients is pre-renal and a cause (e.g., sepsis, hypovolaemia, recent surgery) will be readily apparent.
  - A patient with seizures or who has fallen with a 'long lie' is at risk of rhabdomyolysis.
  - Recent angiography may lead to contrast-induced AKI or cholesterol emboli.
- Does the patient have any common **risk factors for AKI** (Box 11.53)?
- Has any **new medication** been started?
  - ACE inhibitors/ARBs may provoke AKI when started in a patient with unrecognized bilateral renal artery stenosis.
  - NSAID use is often not volunteered and should be specifically sought.

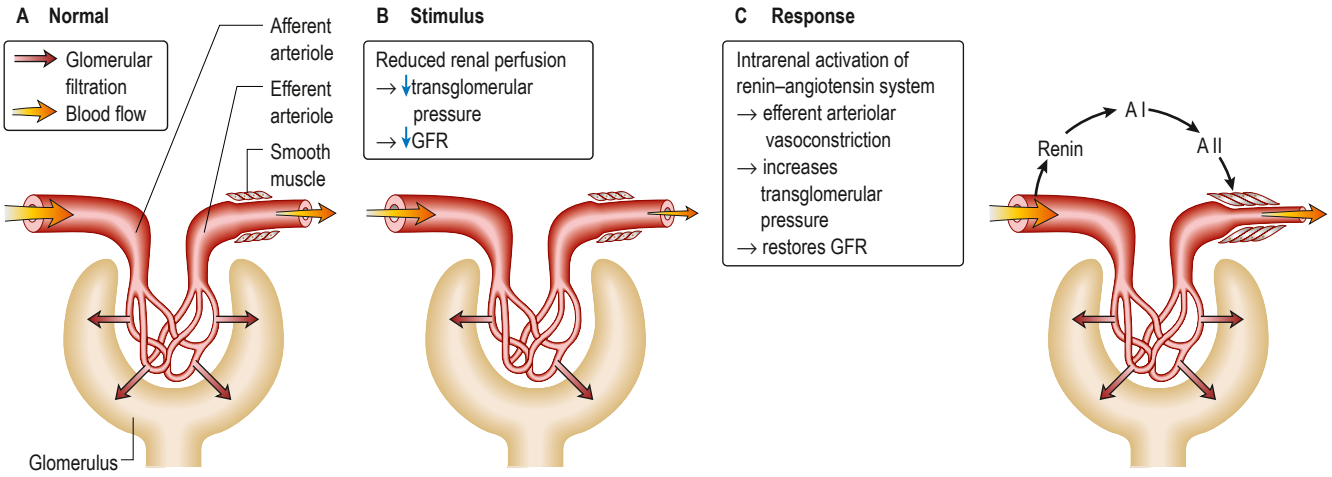


Fig. 11.26 Glomerular dynamics. **(A)** Glomerulus under 'resting' conditions; **(B)** reduced renal perfusion (e.g., due to systemic hypotension) results in a drop in transglomerular pressure and filtration; **(C)** intra-renal activation of the renin-angiotensin system restores GFR. ACE inhibitors/ARBs impair efferent arteriolar vasoconstriction and interrupt the normal glomerular haemodynamic response to a reduction in renal perfusion, compounding AKI. *AI,* Angiotensin I; *AII,* angiotensin II; *GFR,* glomerular filtration rate.

Box 11.53 Risk factors associated with development of AKI

Age >70
Immunosuppression or reduced immunity
CKD (eGFR < 60 mL/min/1.73 m^2)
Diabetes mellitus
Heart failure
Liver disease
Previous history of AKI
Neurological impairment (limiting access to fluids)
Nephrotoxic medication
Exposure to iodinated contrast agents in the prior week
Predisposing risk factors for urological obstruction such as enlarged prostate or previous history of hydronephrosis

- Recent antibiotic treatment may trigger tubulointerstitial nephritis.

- **Obstructive voiding symptoms** (e.g., poor stream, urinary hesitancy) raise suspicion of bladder outflow obstruction and postrenal AKI. A distended bladder may be palpable or detected with a bedside bladder scan.
- **Visible haematuria** may indicate bladder pathology (e.g., infection or malignancy) but should also prompt consideration of parenchymal renal disease (e.g., anti-GBM nephritis).
- **Systemic symptoms or signs** may point to a rarer multisystem disorder (e.g., purpuric rash or haemoptysis in small vessel vasculitis; bone pain and anaemia in multiple myeloma).
- **Fluid status assessment** to identify hypo- or hypervolaemia is critical (see Chapter 27). Reviewing fluid intake and losses, diuretic and antihypertensive medication, postural change in blood pressure, and general fluid status examination is paramount in any patient with an AKI. A fall in systolic blood pressure of 40 mmHg or more from a previous baseline, or a systolic pressure of ≤90 mmHg is a red flag. The most likely cause of shock can often be deduced from clinical examination (**Box 11.54**).

Box 11.54 Distinguishing causes of shock on clinical examination

	Hypovolaemia	Low cardiac output	Vasodilation
Hypotension	Postural hypotension	Hypotension	Hypotension
Peripheries	Cool	Cool	Warm
Jugular venous pressure	Low	Raised	Low

- Are any clinical features of **uraemia** present (e.g., encephalopathy, pericarditis)?

Investigations

- **Urinalysis:** Usually negative for blood and protein in pre-renal AKI or ATI. Blood and protein together should prompt consideration of a parenchymal cause (glomerulonephritis). Myoglobinuria in rhabdomyolysis causes a false-positive result for blood (due to cross-reaction with haemoglobin). Urinalysis is ideally performed prior to catheterization to avoid traumatic contamination.
- **Urine tests:** Urine sodium and markers of urine concentration may help to discriminate between pre-renal AKI and ATI (**Box 11.55**). Phase contrast microscopy may detect red cell casts or dysmorphic red cells indicating glomerulonephritis.
- **Arterial or venous blood gas** will provide a point-of-care potassium level and acid–base status.
- **Core lab tests:** These should include full blood count, urea and electrolytes (including potassium and calcium), and liver function tests. Serum CK should be checked in patients at risk of rhabdomyolysis and urate in patients at risk of tumour lysis syndrome.
- Depending on the clinical context, **additional laboratory tests** may be indicated (e.g., myeloma screening in a patient with hypercalcaemia or normocytic anaemia; autoantibody and blood-borne virus screening in a patient with possible glomerulonephritis). Indiscriminate use of advanced lab tests in AKI is costly

Box 11.55 Criteria for distinction between pre-renal and intrinsic causes of kidney dysfunction

Criteria	Pre-renal	Intrinsic
Urine specific gravity	>1.020	<1.010
Urine osmolality (mOsm/kg)	>500	<350
Urine sodium (mmol/L)	<10	>20
$Fe_{Na} = \frac{U_{Na}}{P_{Na}} + \frac{U_{Cr}}{P_{Cr}} \div 100$	<1%	>1%

With pre-renal AKI, the kidney avidly retains sodium and progressively concentrates urine in an effort to maintain renal blood flow, leading to increases in urine specific gravity and osmolality and a fall in urine sodium and fractional excretion of sodium. With established acute tubular injury ('intrinsic renal failure') these autoregulatory mechanisms cease to function. These criteria lack absolute specificity to discriminate pre-renal from intrinsic AKI and should be interpreted in parallel with clinical assessments of cause.
Cr, Creatinine; *Fe,* fractional excretion; *Na,* sodium; *P,* plasma; *U,* urine.

and may be misleading; where doubt exists the patient should be discussed with a nephrologist.

- An **ECG** may show electrical changes consistent with hyperkalaemia.
- **CXR** may reveal pulmonary oedema complicating oliguric AKI, or point towards an underlying cause (e.g., bihilar lymphadenopathy in sarcoidosis).
- **Ultrasound:** POCUS may be used to detect bladder outflow obstruction or hydronephrosis. Departmental ultrasound will provide more detailed assessment of renal size and architecture.
- **Kidney biopsy:** Usually requested and performed by a nephrologist, indications include unexplained or non-resolving AKI, or where a high index of suspicion exists for a parenchymal cause such as glomerulonephritis or tubulointerstitial nephritis.

While the finding of small kidneys (<8 cm in bipolar length) on ultrasound scan usually indicates significant CKD, the presence of anaemia or secondary hyperparathyroidism are not sensitive or specific discriminators between AKI and CKD. Conversely, normal kidney size on ultrasound does not exclude CKD (diabetic nephropathy and amyloidosis are both notably associated with preserved kidney size).

Management of AKI

Patients with severe AKI (including most with AKI stage 3) or adverse risk factors for ambulatory care management require admission. Early specialist review is advisable. Often unwell, patients with AKI should ideally be nursed in an ECU or high-dependency unit. Cardiac monitoring is required in patients with significant electrolyte disturbance, particularly hyperkalaemia.

General measures

Good nursing, infection control, and early physiotherapy are vital. Fluid balance monitoring, as intake and output (particularly urine output) and daily body weights, is key to monitoring progress and recovery and avoiding fluid overload. Medications require careful scrutiny, both to withhold nephrotoxins and adjust doses for current GFR, which may change on a daily basis. There is a particular risk with anticoagulants (oral and heparins; see Chapter 40). Patients with AKI stage 3 are at high risk of gastric stress ulceration and prophylactic treatment with an anti-acid (PPI or H_2 receptor antagonist) is usually indicated. Serial blood tests are required to monitor electrolytes (including serum potassium and calcium) and acid–base status.

Sepsis

Infection needs to be treated promptly, avoiding nephrotoxic antibiotics where possible. Use of prophylactic antibiotics or protective barrier nursing is not usually recommended. If there is loose stool, precautions will be necessary as appropriate.

Fluid and electrolyte balance

Replacement fluid may be a balanced crystalloid solution, 0.9% saline (e.g., burns, pancreatitis) or 4% albumin. There are no significant differences between these with respect to death rates, organ failure, need for RRT, or duration of hospitalization.

In general, once the patient is euvolaemic, daily fluid intake should equal urine output plus losses from the gut (including nasogastric losses) plus an allowance of 500 mL daily for insensible loss. Febrile patients will require an additional allowance. Most patients will meet this intake through oral fluids.

Large-volume fluid infusion or changes in daily weight (reflecting changes in fluid balance) make more regular reviews necessary. When overload is present, loop diuretics may be useful. Non-oliguric patients with AKI fare better than oliguric patients. However, conversion of oliguria to non-oliguria has not been shown to decrease mortality, and diuretics have not been demonstrated to prevent AKI or improve outcomes.

Nutrition

Salt and potassium should be restricted. Protein intake is sometimes limited to approximately 0.5 g/kg per day to try and delay progression to RRT, but this poses the risk of a negative nitrogen balance and is not standard practice in the UK. Patients on RRT are managed with 1–1.5 g/kg protein per day diet, though hypercatabolic individuals may require a greater nitrogen intake. Ideally, nutrition should be by mouth or nasogastric tube, or, as a last resort, by parenteral feeding.

Pre-renal AKI

Management largely depends on the underlying cause. Many cases of AKI are associated with hypovolaemia, so prompt fluid resuscitation is most often indicated. When there is uncertainty as to the volume status of a patient, a fluid challenge of 250 mL crystalloid will often prove whether hypotension and urine output are fluid responsive. Crystalloids and particularly balanced solutions such as Hartmann's solution and Ringer's lactate are the recommended resuscitative fluid, as 0.9% saline can lead to hyperchloraemic acidosis. Colloids such as hydroxyethyl starch are contraindicated, as they can lead to worsening AKI. Heart rate, blood pressure, and urine output will all guide response to resuscitation. For cardiogenic shock and septic shock, see Chapter 13.

Postrenal AKI

Obstruction of the urinary tract may occur at any point from the calyces to the external urethral orifice but commonly takes the form of bladder outflow obstruction (prostate disease in men) or bilateral ureteric obstruction (stones or tumours). The causes and presentation of urinary tract obstruction are dealt with in Chapter 26. Almost every case of unexplained AKI should result in an ultrasound to exclude obstruction, as once this is relieved (and if acute), kidney function will return to baseline.

Intrarenal parenchymal AKI

This is most commonly (80%–90%) due to ***acute tubular injury*** (ATI; see Chapter 26, **Box 26.28**). Almost any cause of pre-renal AKI, if prolonged to the point at which kidney autoregulation fails will lead to ischaemic ATI. If not ischaemic, then ATI usually results from

direct tubular toxins (often drugs). As a result, ATI is common in hospital practice. Other causes of parenchymal AKI include:

- Disease affecting the intrarenal arteries and arterioles, as well as glomerular capillaries, such as a vasculitis, accelerated hypertension, cholesterol embolism, haemolytic uraemic syndrome, pre-eclampsia, and crescentic glomerulonephritis.
- Acute TIN (see Chapter 26). Other mechanisms of tubulointerstitial injury include acute obstruction by crystals e.g., after rapid lysis of certain malignant tumours following chemotherapy (acute hyperuricaemic nephropathy), light chain casts (myeloma cast nephropathy), or pigment casts (rhabdomyolysis).
- Acute bilateral suppurative pyelonephritis or pyelonephritis of a single kidney, which can cause acute uraemia.

Conclusive diagnosis of intrarenal AKI often requires a diagnostic kidney biopsy and management is directed towards the underlying cause.

Renal replacement therapy—haemodialysis and haemofiltration

Regardless of cause, severe AKI may require supportive treatment with renal replacement therapy (RRT). The main indications are:

- symptomatic uraemia (including pericarditis or tamponade)
- hyperkalaemia not controlled by medical management
- pulmonary oedema not controlled with medical diuresis
- acidosis not controlled by medical management
- for removal of drugs causing toxicity e.g., gentamicin, lithium, aspirin overdose.

Contemporary RRT options include intermittent haemodialysis or haemodiafiltration (HDF), continuous venovenous haemofiltration (CVVHF) or peritoneal dialysis, the latter mainly in resource-poor countries. There is no data to favour haemodialysis or CVVHF as a superior mode of therapy in AKI. However, in haemodynamically unstable patients, particularly those on the ITU, continuous treatments like CVVHF are more widely used for systemically unwell patients with AKI.

Management of the recovery phase of AKI

Signs of recovery include an increase in urine output and a fall in serum creatinine (in the absence of kidney replacement). Dialysis or haemofiltration can be discontinued when kidney function is sufficient. A characteristic polyuric recovery phase develops as GFR recovers ahead of kidney tubular reabsorptive capacity for sodium, potassium, and water. Intravenous fluids with replacement sodium and potassium may be needed to prevent slowing of recovery due to volume depletion. Typically, the diuretic phase lasts for only a few days.

Key points in the management of AKI are summarized in **Box 11.56**.

Further discussion of specific causes of AKI can be found in Chapter 26.

Further reading

BMJ Best Practice: *Acute Kidney Injury*. Accessed at: https://bestpractice.bmj.com/topics/en-gb/3000117.
Kidney Disease Improving Global Outcomes (KDIGO) Clinical Practice Guidelines for Acute Kidney Injury. 2012.
NICE Guideline (NG) 148: Acute Kidney Injury: Prevention, Detection and Management. 2023.
UK Kidney Association: *Clinical Practice Guideline on Acute Kidney Injury*. 2019.

Hyperkalaemia

Hyperkalaemia can be classified as mild, moderate, and severe.
Mild: 5.5 to 5.9 mmol/L
Moderate: 6.0 to 6.4 mmol/L
Severe: ≥6.5 mmol/L

Box 11.56 Summary of key points (AKI)

- Clinically stable patients with stage 1 AKI can often be assessed and managed in an ambulatory care setting depending on local pathways and resources
- Patients identified with AKI should be reviewed early to ensure adequate volume resuscitation, prompt treatment of underlying infection, ensure obstruction is excluded, and nephrotoxic medication withheld ('STOP' AKI)
- Actively monitor for and treat complications of AKI including hyperkalaemia, acidosis, fluid overload, and manifestations of uraemia (e.g., pericarditis or encephalopathy)
- Development of complications or non-resolution with initial management should prompt referral to a nephrologist (or critical care unit) to consider the need for specialist investigations (e.g., kidney biopsy) and timely renal replacement therapy if needed

Box 11.57 Risk states and causes for hyperkalaemia

- Acute kidney injury
- Dialysis-dependent end-stage kidney disease (haemodialysis or peritoneal dialysis)
- Chronic kidney disease stages 4 and 5 (CKD, eGFR < 30 mL/min/1.73 m^2)
- Drugs (renin-angiotensin-aldosterone inhibitors, NSAIDs, trimethoprim)
- Cardiac failure (cardio-renal syndrome, RASi drugs, mineralocorticoid receptor antagonists)
- Diabetes mellitus (RASi drugs, diabetic ketoacidosis)
- Liver disease (mineralocorticoid receptor antagonists, hepatorenal failure)
- Addison's disease (primary adrenal insufficiency)
- Hyporeninaemic hypoaldosteronism (type 4 renal tubular acidosis)

Common risk states and causes for the development of hyperkalaemia are listed in **Box 11.57**.

Spurious causes for hyperkalaemia (e.g., delayed separation of a laboratory sample or resulting from thrombocytosis) can be excluded by measuring potassium on a point-of-care venous blood gas sample.

Management

The UK Kidney Association guidelines recommend that all patients with severe hyperkalaemia (potassium value of ≥6.5 mmol/L) irrespective of kidney function are referred to hospital for immediate assessment and treatment. Severe hyperkalaemia may be life-threatening, leading to cardiac dysrhythmias, particularly asystole or ventricular fibrillation. Mild to moderate hyperkalaemia (potassium values of 5.5–6.4 mmol/L) prompts urgent clinical and medication review but may sometimes be managed via ambulatory pathways depending on clinical context (see section on 'Abnormal blood results').

Treatment for severe hyperkalaemia requires emergency, acute, and subacute measures to stabilize the myocardium and prevent cardiac arrest, lower potassium, and prevent rebound, as summarized in **Box 11.58**. In parallel, the cause or contributors to development of hyperkalaemia should be sought and removed or treated. In patients with severe hyperkalaemia, patients should be managed in a higher dependency setting such as the resuscitation area of ED, enhanced care unit of AMU, or in an HDU setting with cardiac monitoring.

High potassium levels are cardiotoxic, as they inactivate sodium channels. Divalent cations, for example, **calcium**, restore the voltage

Box 11.58 Treatment of severe hyperkalaemia

This is a potentially life-threatening emergency. Management is divided into emergency, acute, and subacute management.

Emergency management

- ECG monitor and i.v. access

Protect the myocardium

- If there are ECG changes (tall T waves, widened QRS, prolonged PR interval, loss of P wave, or ultimately, sinusoidal waves), then 30 mL of 10% calcium gluconate i.v. over 10 min or 10 mL 10% calcium chloride i.v. over 5 min via a large vein
- Repeat dose after 5 min if ECG changes persist

Acute management

Drive K^+ into cells

- Insulin 10 units in 25 g glucose (either 50 mL of 50% glucose or 125 mL of 20% glucose) i.v. over 15–30 min, followed by regular checks of blood glucose and plasma K^+. Note: monitor for hypoglycaemia in the 6 h period following insulin/dextrose
- Consider correction of severe acidosis (pH < 6.9)—infuse $NaHCO_3$ (1.26%)
- Consider salbutamol 10–20 mg nebulized
- In a hyperkalaemic cardiac arrest, consider a sodium bicarbonate i.v. bolus (50 mL of 8.4% solution)

Subacute

Deplete body K^+ (to decrease plasma K^+ over the next 24 h)

- Novel potassium binders (sodium zirconium cyclosilicate, SZC, or patiromer) bind potassium in the GI tract in exchange for sodium and hydrogen ions. Consider SZC 10 g thrice daily for up to 72 h, or patiromer 8.4 g once daily
- Old potassium binders (e.g., polystyrene sulphonate resins such as calcium resonium) have been rendered obsolete since the advent of SZC and patiromer
- Renal replacement therapy, such as haemodialysis, haemofiltration, or peritoneal dialysis if the above fails

Monitoring and prevention

- Monitor serum potassium closely (1, 2, 4, 6, and 24 h) after emergency treatment to check effectiveness and exclude rebound
- Monitor blood glucose for 6 h after treatment with insulin-dextrose
- Consider cause of hyperkalaemia and prevent recurrence as best possible. This may include withdrawal of potassium supplements and potassium-sparing medications (e.g., ACEi/ARB or mineralocorticoid receptor antagonists) and dietary modifications

dependability of the channels. Calcium ions protect the cell membranes from the effects of hyperkalaemia but do not alter the potassium concentration.

Supraphysiological **intravenous insulin** (10 units) drives potassium into the cell and lowers plasma potassium by ~1 mmol in 60 min, but **must be accompanied by intravenous glucose** (25 g) to avoid hypoglycaemia. Regular measurements of blood glucose for at least 6 hours after use of insulin should be performed and extra glucose must be available for immediate use. The use of glucose alone in non-diabetic patients, to stimulate endogenous insulin release, does not produce the high levels of insulin required and therefore is not recommended.

Hypertonic (8.4%) **sodium bicarbonate** may reduce serum potassium through correction of systemic acidosis; potential unwanted effects include a paradoxical worsening of intracellular acidosis and reduction in cardiac output. As such, 8.4% sodium bicarbonate is only recommended for use in hyperkalaemic cardiac arrest.

Patiromer and ***sodium zirconium cyclosilicate*** (SZC) are new non-absorbed agents that bind potassium in the gastrointestinal (GI) tract to facilitate faecal excretion. Clinical trials of patiromer and SZC have demonstrated clear evidence of a dose-dependent potassium-lowering effect. Patiromer and SZC are effective and well tolerated and have now largely superseded ion-exchange resins for the treatment of hyperkalaemia. Potassium binders are now also prescribed for long-term use in patients with heart failure and CKD to attenuate ACE inhibitor/ARB-induced hyperkalaemia.

Ion-exchange resins (polystyrene sulphonate resins, SPS, such as calcium resonium) are an older form of gut-acting therapy to maintain normokalaemia after emergency treatment. Their use has now been superseded by the newer gut-acting potassium binders described above.

In general, all of these measures are simply ways of 'buying time' either to correct the underlying disorder (e.g., stopping a culprit drug) or to arrange removal of potassium by dialysis, which is the definitive treatment for hyperkalaemia in severe acute or chronic kidney failure.

Follow-up

Potassium should be monitored frequently during the 24-hour period following emergency treatment of severe hyperkalaemia both to confirm effectiveness and to exclude early rebound (e.g., 1, 2, 4, 6, and 24 h). Blood glucose should be monitored for 6 hours following insulin-dextrose treatment (risk of hypoglycaemia).

Measures should be taken to reduce the risk of recurrent hyperkalaemia: often withdrawal of potassium supplements, suspension of potassium-sparing medications, or dietary modifications. Handover to the patient's usual clinician is important: suspension of potassium-sparing antihypertensive medications warrants closer monitoring of blood pressure and pro-activity to prevent deterioration in control. Likewise, interruption of potassium-sparing medications prescribed for prognostic benefit in CKD or heart failure (e.g., ACEi/ARB) should trigger early reassessment and re-evaluation of treatment options. The novel potassium binders may now be prescribed long term in patients with CKD or heart failure to facilitate re-introduction and up-titration of ACEi/ARB under close supervision in patients where hyperkalaemia has previously proved limiting.

Further reading

The UK Kidney Association Clinical Practice Guidelines—Treatment of Acute Hyperkalaemia in Adults. 2023.

The unconscious patient

Consciousness is dependent on the functioning of two separate anatomical and physiological systems:

- The ascending reticular activating system (ARAS), projecting from brainstem to thalamus. This determines arousal (the level of consciousness).
- The cerebral cortex, which determines the content of consciousness.

Impaired functioning of either anatomical system may cause coma.

Disturbed consciousness: definitions

- ***Coma:*** a state of unrousable unresponsiveness. Level of consciousness represents a continuum between being alert and deeply comatose. It may be quantified using the Glasgow Coma Scale (GCS) with a score between 3 and 15 (**Box 11.59**). Coma may be conveniently defined as a GCS score of 8 or lower. Terms such as stupor and obtundation have been superseded by the GCS score and are no longer used.

Box 11.59 Glasgow Coma Scale (GCS)[a]

Parameter	Score
Eye opening *(E)*	
Spontaneous	4
To speech	3
To pain	2
No response	1
Verbal response *(V)*	
Orientated	5
Confused conversation	4
Inappropriate words	3
Incomprehensible sounds	2
No response	1
Motor response *(M)*	
Obeys	6
Localizes	5
Withdraws	4
Flexion	3
Extension	2
No response	1

[a]GCS score = *E* + *V* + *M* (GCS minimum = 3; maximum = 15).

- ***Delirium:*** a confusional state in which reduced attention is a cardinal feature, usually with altered behaviour, cognition, orientation, and a fluctuating level of consciousness (from agitation to hypoarousal; see **Box 11.80**).

Mechanisms and causes of coma

Altered consciousness is produced by four mechanisms affecting the ARAS in the brainstem or thalamus and/or widespread impairment of cortical function (**Box 11.60**)

- ***Diffuse brain dysfunction.*** Generalized severe metabolic or toxic disorders (e.g., alcohol, sedatives, uraemia, hypercapnia) depress cortical and ARAS function.
- ***Massive cortical damage.*** Unlike brainstem lesions, extensive damage of the cerebral cortex and cortical connections is required to cause coma e.g., meningitis or hypoxic–ischaemic damage after cardiac arrest (**Fig. 11.27**).
- ***Brainstem lesion.*** A discrete brainstem or thalamic lesion e.g., stroke, may damage the ARAS.
- ***Brainstem compression.*** A supratentorial mass lesion within the brain compresses the brainstem, inhibiting the ARAS e.g., 'coning' from a brain tumour or haemorrhage. Mass lesions within the posterior fossa are particularly prone to causing brainstem compression and hydrocephalus.

A single focal hemisphere (or cerebellar) lesion does not produce coma, unless it compresses the brainstem. Cerebral oedema frequently surrounds masses, increasing their pressure effects.

The most common causes of coma are:

- metabolic disorders—35%
- drugs and toxins—25%
- mass lesions—20%
- other—including trauma, stroke, and CNS infections.

Immediate assessment and management of the unconscious patient

- Check the ***a***irway, ***b***reathing, ***c***irculation, ***d***isability, and ***e***xposure/***e***xamination.

Box 11.60 Principal causes of coma: examples of mechanisms

Diffuse brain dysfunction
- Drug overdose (accidental or deliberate), including alcohol
- Carbon monoxide poisoning
- Traumatic brain injury
- Hypoglycaemia, hyperglycaemia
- Severe uraemia (see Chapter 26)
- Hepatic encephalopathy (see Chapter 23)
- Respiratory failure with CO_2 retention (see Chapter 20)
- Hypercalcaemia, hypocalcaemia
- Hypoadrenalism, hypopituitarism and hypothyroidism
- Hyponatraemia, hypernatraemia
- Metabolic acidosis
- Hypothermia, hyperpyrexia
- Seizures—post-ictal state and non-convulsive status
- Metabolic rarities e.g., porphyria

Extensive cortical damage
- Hypoxic–ischaemic brain injury e.g., cardiac arrest
- Encephalitis, meningitis, cerebral malaria
- Subarachnoid haemorrhage

Direct effect within brainstem
- Brainstem haemorrhage, infarction, or demyelination
- Brainstem neoplasm e.g., glioma
- Wernicke–Korsakoff's syndrome

Pressure effect on brainstem
- Tumour
- Massive hemisphere infarction with oedema
- Intracranial haematoma
- Cerebral abscess
- Cerebellar mass

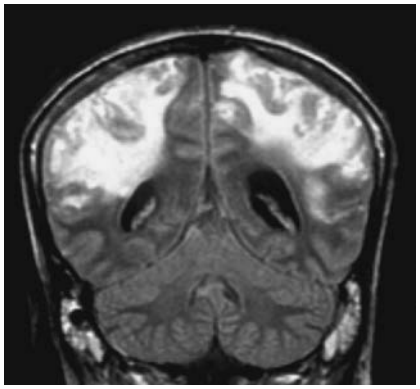


Fig. 11.27 Magnetic resonance image showing extensive cortical damage following meningitis. A permanent vegetative state resulted.

- Use a point-of-care testing device to measure blood glucose; if the patient is hypoglycaemic, give glucose (consider e.g., 50 mL 10% glucose in adults, repeating every minute until the patient has regained consciousness or a total of 250 mL has been administered). If intravenous access is not available, consider intraosseous administration or intramuscular glucagon 1 mg.
- Treat seizures with buccal midazolam (community) or intravenous lorazepam (hospital); if not terminated, use a second-line anticonvulsant (intravenous levetiracetam, phenytoin, or sodium valproate).
- If there is fever and meningism, give appropriate intravenous antibiotics.
- Give intravenous naloxone if opiate overdose is suspected or flumazenil (for *isolated* benzodiazepine overdose) and i.v.

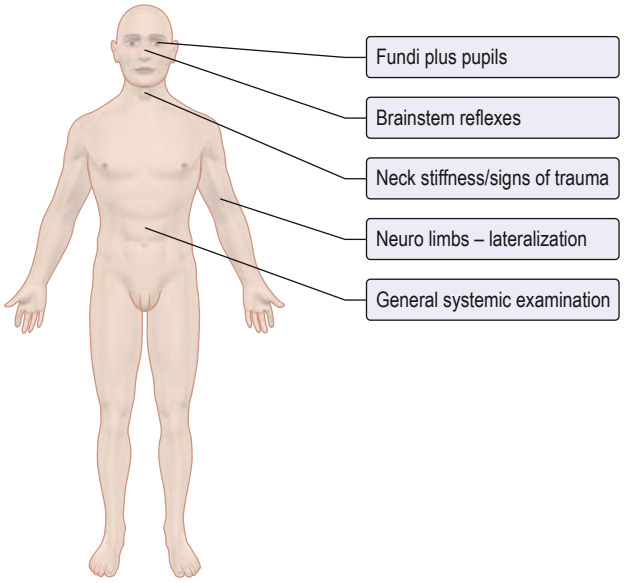


Fig. 11.28 **Assessing the unconscious patient.**

thiamine (in Wernicke's encephalopathy) in people who may be malnourished.

Obtain as much history as possible. A limited history is one of the problems in assessing the unconscious patient. What were the circumstances? Ask paramedics, police, and witnesses. Contact the patient's relatives, friends, and GP, and obtain past hospital notes. Look for drug details/bottles and identification data.

General and neurological examination (see Fig. 11.28)

General examination

- Measure the patient's temperature (with a low-reading rectal thermometer if hypothermic).
- Check for meningism.
- Sniff the patient's breath for ketones, alcohol, and hepatic fetor.
- Survey the skin for signs of trauma or spinal injury, rash (meningococcal sepsis), jaundice or stigmata of chronic liver disease (CLD), cyanosis, and injection marks.
- Assess the respiratory pattern e.g., Cheyne–Stokes (alternating hyperpnoea and periods of apnoea, indicating bilateral cerebral or upper brainstem dysfunction) or Kussmaul respiration (deep, sighing hyperventilation seen in metabolic acidosis including DKA and renal failure).

Neurological examination

Neurological examination aims to determine:

- depth of coma (GCS)
- brainstem function
- lateralization of pathology.

Depth of coma

GCS. Assessment of the GCS score is repeated regularly to determine whether the patient's level of consciousness is progressively declining. Use deep stimulation such as a trapezius squeeze or pressure over the supraorbital nerve, and record the *best* response. Shout commands.

Fundi. Look for papilloedema and subhyaloid retinal haemorrhage (sometimes seen in subarachnoid haemorrhage [SAH]).

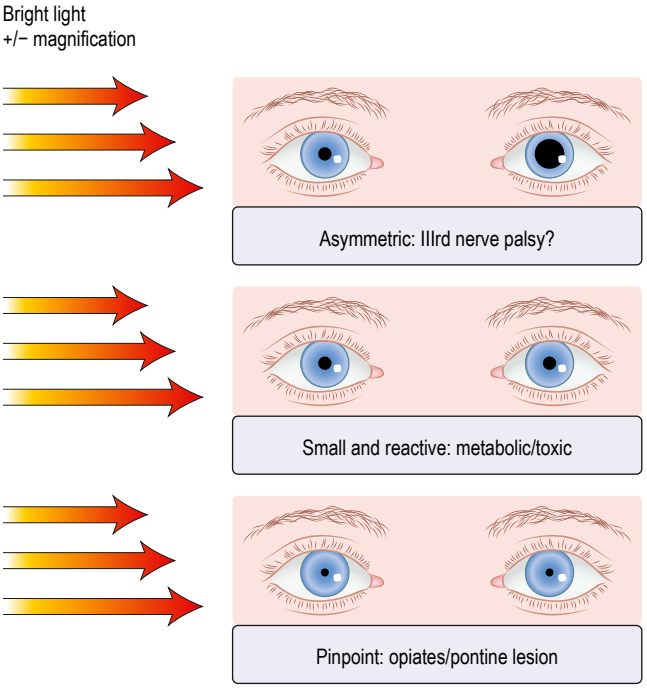


Fig. 11.29 **Examination of the pupils.**

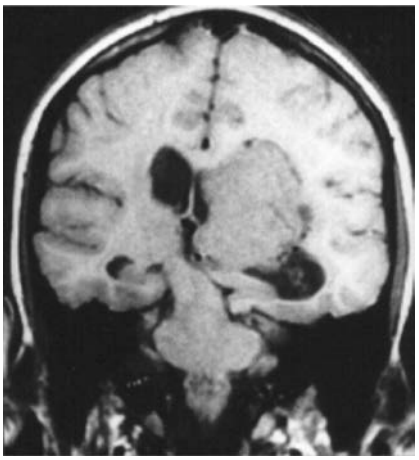


Fig. 11.30 **Medial temporal lobe (uncal) herniation causing third nerve compression.**

Brainstem function

Pupils

Record their size and reaction to light (Fig. 11.29):

- ***Dilation of one pupil*** that then becomes fixed to light indicates compression of the IIIrd nerve. This may indicate a potential neurosurgical emergency (Fig. 11.30).
- ***Bilateral midpoint reactive pupils*** (i.e., normal pupils) are characteristic of metabolic comas and secondary coma due to sedative drugs (except opiates).
- ***Bilateral light-fixed, dilated pupils*** are one cardinal sign of brain death. They can occur in deep coma of any cause, but particularly in barbiturate intoxication and hypothermia.
- ***Bilateral pinpoint, light-fixed pupils*** occur with pontine lesions (e.g., haemorrhage) and with opiates.

Mydriatic drugs and previous ophthalmic surgery can cause diagnostic difficulty.

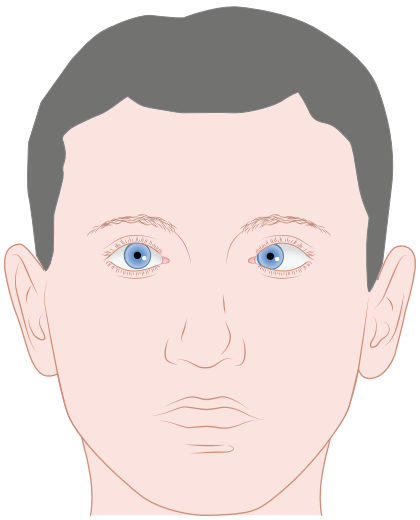
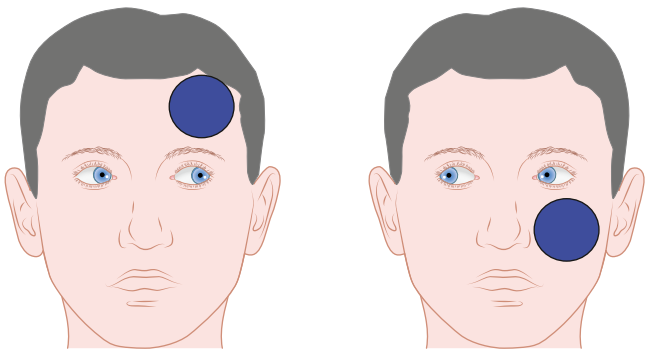


Fig. 11.31 Disconjugate eye position. This usually indicates brainstem lesion (the eyes may be mildly disconjugate in metabolic coma).



A B

Fig. 11.32 Conjugate gaze deviation. (A) Frontal lesion. (B) Brainstem lesion.

Eye movements and position

- Disconjugate eyes (divergent ocular axes) indicate a brainstem lesion e.g., ***skew deviation*** (one eye up, one eye down; Fig. 11.31).
- Conjugate gaze deviation means the eyes deviate towards the lesion in the frontal lobe and the normal limbs (unopposed activity of the intact frontal eye fields drives eyes to the opposite side); and away from the lesion in the brainstem and towards the weak limbs (the paramedian pontine reticular formation in the pons controls lateral gaze to the ipsilateral side; see Chapter 26) (Fig. 11.32).
- Vestibulo-ocular reflex means that passive head-turning produces conjugate ocular deviation away from the direction of rotation (doll's eye reflex). This reflex disappears in deep coma, brainstem lesions, and brain death (Fig. 11.33).
- In light coma, slow, side-to-side eye movements are seen ('windscreen wiper' eyes; Fig. 11.34). This is also seen with extensive cortical damage in deep coma.

Other brainstem reflexes

- Corneal reflex.
- Gag/cough reflex (via endotracheal tube if intubated).
- Respiratory drive.

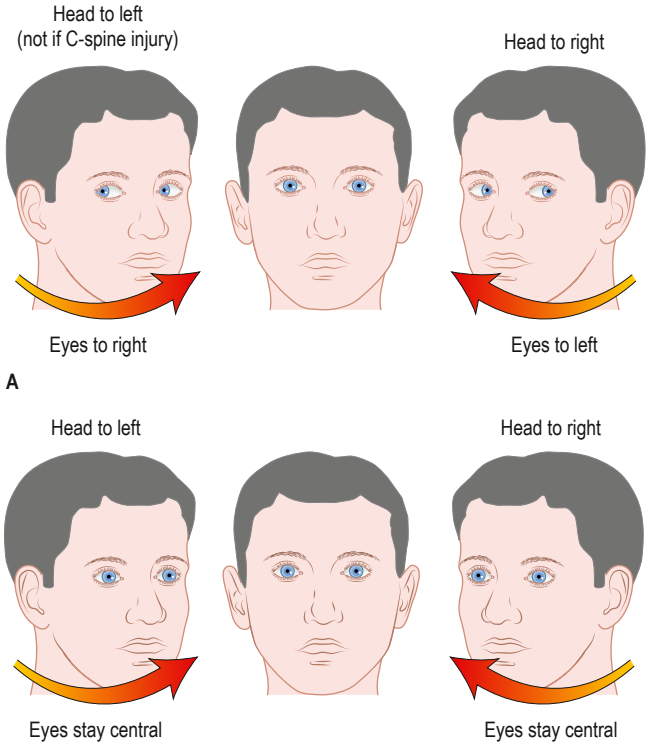


Fig. 11.33 Doll's eye movements. (A) Normal. (B) Abnormal.

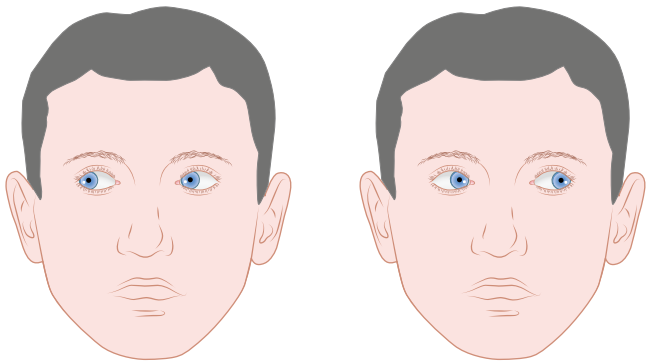


Fig. 11.34 'Windscreen wiper' eyes ('ping pong' eyes). Slow, side-to-side movements demonstrate diffuse cortical dysfunction. This is common in light coma and implies normal brainstem function. In deep coma it indicates extensive cortical damage.

Lateralizing signs

Coma makes it difficult to recognize lateralizing signs. These are helpful:

- ***Asymmetry of response to visual threat*** in a stuporous patient: suggests hemianopia.
- ***Asymmetry of face:*** drooping or dribbling on one side; blowing in and out of mouth when the paralyzed cheek does not move.
- ***Asymmetry of tone:*** unilateral flaccidity or spasticity—may be the only sign of hemiparesis.
- ***Asymmetry of decerebrate and decorticate posturing.***
- ***Asymmetrical response*** to painful stimuli.
- ***Asymmetry of tendon reflexes and plantar responses:*** both plantars are often extensor in deep coma.

Coma 'look-alikes'

- Psychogenic coma.
- 'Locked-in' syndrome—complete paralysis, except for vertical eye movements/blinking in ventral pontine infarction. Patients have a functioning cerebral cortex and are fully aware but unable to communicate, except through eye movements
- Severe paralysis e.g., myasthenic crisis or severe Guillain–Barré syndrome.

Diagnosis and investigations in coma

Often the cause is evident (e.g., head injury, metabolic disorder, overdose). Where lateralizing signs or brainstem pathology are found on examination, a mass lesion or infarction/haemorrhage is likely (note that hypoglycaemia may also cause focal signs). If no cause is obvious after clinical assessment, further investigations are essential.

Blood and urine

- ***Drug screen***—blood ethanol, paracetamol, and when indicated, salicylate levels, urine toxicology including screening for benzodiazepines, narcotics, amfetamines.
- ***Biochemistry***—urea, electrolytes, glucose, calcium, liver biochemistry. Consider ammonia level (needs to be on ice) in advanced liver disease or a history of ingestion of certain medications including valproate and topiramate.
- ***Metabolic and endocrine studies***—thyroid-stimulating hormone, cortisol.
- ***Arterial blood gases***—for acidosis or high CO_2 levels. Also check carboxyhaemoglobin (HbCO) and methaemoglobin (MetHb) levels.
- ***Other***—for example, cerebral malaria (request antigen test and *thick* blood film; see Chapter 42) and porphyria (Chapter 36).

Brain imaging

CT brain imaging is the most readily available and safest modality in the unconscious patient (MRI is useful where CT is normal but presents greater monitoring challenges in the unconscious patient). CT is quick and effective in demonstrating all types of brain haemorrhage and most mass lesions; infarcts may be missed in the early stages and where only the brainstem is affected.

Cerebrospinal fluid (CSF) examination

An LP should be performed in coma only after careful risk assessment. It is contraindicated when an intracranial mass lesion is a possibility; CT is essential to exclude this. Cerebrospinal fluid (CSF) examination is likely to alter therapy only if undiagnosed meningoencephalitis or other infection is present, or in SAH where CT may give a false-negative result, particularly after 24 hours.

Electroencephalography (EEG)

Electroencephalography (EEG) is of some value in the diagnosis of metabolic coma, encephalitis, and non-convulsive status epilepticus.

General management

Comatose patients need careful nursing, meticulous attention to the airway, and frequent monitoring of vital functions.

Longer-term essentials are:

- skin care—turning (to avoid pressure ulcers and pressure palsies)
- oral hygiene—mouthwashes, suction
- eye care—prevention of corneal damage (lid taping, irrigation)
- fluids—nasogastric or intravenous
- feeding—via a fine-bore nasogastric tube or via PEG
- sphincters—catheterization when essential (use a penile urinary sheath if possible in men); rectal evacuation.

The ongoing management and prognosis of longer-term disorders of consciousness is discussed in Chapter 28.

Box 11.61 Terms relating to dizziness and blackouts

Dizziness covers many symptoms, from a vague feeling of unsteadiness to severe, acute vertigo. It is frequently used to describe light-headedness (e.g., due to hypotension), panic, anxiety, palpitations, and chronic ill-health. The real nature of this symptom must be determined.

Vertigo means the illusion of movement, a sensation of rotation or tipping. The patient feels the surroundings are spinning or moving. This is distressing and often accompanied by nausea or vomiting.

Blackout, like dizziness, is simply descriptive, implying either altered consciousness (e.g., epilepsy or syncope), transient visual disturbance as part of presyncope or falling, or hypoglycaemia in a diabetic patient. A good history is essential.

Collapse is a vague term but often used. It should be avoided.

Fatigue is common. When it is an isolated symptom, it rarely indicates neurological disease, although it may be a symptom of many neurological disorders. General medical causes, such as anaemia and endocrine disorders, should be excluded.

Dizziness and blackouts

The potential causes for blackouts and dizziness are diverse and include vestibular, cardiovascular, neurological, and psychogenic disorders. Detailed history taking is essential, with particular focus on the precise nature of symptoms, time course of onset and offset, any identifiable triggers of symptoms, and a third-party eye-witness account if available.

Descriptive terms for blackouts and dizziness are often used inconsistently or incorrectly by both patients and doctors (**Box 11.61**); when probed, a patient complaining of 'blackout' may have actually experienced presyncopal symptoms without loss of consciousness, or even a simple fall. A careful history will usually allow the clinician to reach a 'syndromic diagnosis' (e.g., syncope, as opposed to seizure), with further focused history, clinical examination, and investigations then directed towards elucidating the underlying cause, although when the history is unclear a more extensive work-up may be required.

A helpful approach to the differential diagnosis of blackouts and dizziness is to consider causes associated with loss of consciousness and those without (see **Box 11.62**).

Clinical assessment

History taking is an essential skill to define the precise nature of the episode. A helpful structure to the history of the presenting problem is to adopt a 'before, during, and after' approach (**Box 11.63**). If the episode resulted in a fall, enquire about pain or injury, including whether the patient struck their head. If the patient describes dizziness, the character of this (particularly rotational dizziness vs. light-headedness) should be carefully established.

Examination

Examination should be focused both on eliciting signs that may aid in diagnosis but also those resulting from physical injury that may have been sustained if the patient lost consciousness.

Box 11.62 Differential diagnoses of dizziness, blackouts, and syncope

With loss of consciousness	Without loss of consciousness
Syncope • Vasovagal syncope • Cough or micturition syncope • Arrhythmia (e.g., complete heart block, ventricular tachycardia) • Other cardiovascular disorders (e.g., severe aortic stenosis, aortic dissection) • Massive pulmonary embolism	Presyncope • Orthostatic hypotension • Severe anaemia • Other causes for 'syncope'
Seizure • Generalized tonic–clonic • Complex partial • Absence ('petit mal')	Vertigo • Vestibular causes e.g., benign positional paroxysmal vertigo (BPPV), Ménière's disease, vestibular neuritis, labyrinthitis • Central causes e.g., cerebellar stroke
Other • Hypoglycaemia • Vertebrobasilar insufficiency • Migraine (severe basilar migraine) • Stroke (haemorrhagic or ischaemic), either with brainstem involvement or massive and resulting in 'coning' • Intoxication or substance misuse/withdrawal • Psychogenic e.g., dissociative seizures • Narcolepsy	Other • Hypoglycaemia • Migraine • Stroke or transient ischaemic attack • Focal seizures • Drop attacks • Panic attacks

- Signs of physical injury e.g., haematomas, open wounds, fractures, or dislocations.
- Inspection of the tongue and oral cavity may reveal signs of biting or trauma sustained during a seizure.
- Bedside observations: blood pressure (including **lying and standing blood pressure** to exclude orthostatic hypotension), heart rate and regularity (arrhythmias), temperature, respiratory rate, oxygen saturations, and blood glucose (to exclude hypo- or hyperglycaemia—either causative or concurrent).
- Peripheral and central pulse: irregularity may indicate arrhythmia, other signs may be present suggestive of valvular heart disease (e.g., the low volume, slow-rising pulse of severe aortic stenosis).
- Murmur e.g., ejection systolic murmur radiating to the carotid arteries in aortic stenosis.
- Physical features of substance misuse e.g., tremor for alcohol withdrawal, needle track marks in i.v. drug abuse.
- Neurological signs: nystagmus is expected in a patient with true vertigo; focal weakness may be present following a seizure (Todd's paresis) or resulting from a space-occupying lesion or intracranial bleed; cerebellar signs such as past-pointing and ataxia may be elicited in an 'unsteady' patient.

Investigations

Investigations should be tailored to the working diagnosis or differential diagnosis based on history and examination, although those commonly requested for a patient presenting with a transient loss of consciousness are listed below.

- **Laboratory investigations** including full blood count to exclude anaemia, 'liver function tests', bone profile including magnesium, renal function, blood glucose, and consideration of thyroid function tests.

Box 11.63 History taking in the patient experiencing a blackout or dizziness

Before	During	After
What is the last thing you remember before it occurred? *A preceding sensation of déjà vu or jamais vu suggests a seizure while sweating more likely suggests syncope*	Do you have any recollection during the event e.g., hearing people around you? *Maintained awareness during the event suggests it was not a generalized tonic–clonic seizure or complete loss of consciousness*	What is the next thing you remember after it occurred? *Patients experiencing a seizure typically have no recollection of abnormal behaviours witnessed before or after loss of consciousness*
When did it happen (time of day) or is there any pattern to episodes (if repeated)? Any triggers such as turning your head *(suggestive of vestibular disease if vertigo, or vertebrobasilar ischaemia if syncope)* or postural changes *(suggestive of vestibular disease or orthostatic hypotension)*?	Did you bite your tongue?—*may support the diagnosis of a seizure (but does not exclude syncope)*	Was there any evidence of injury afterwards? *E.g., facial injury suggestive of cardiac syncope*
How did you feel before it happened that day? E.g., at work, food and drink intake, sleep, stress, travel, social events *(considering fluid status, insensible losses, substance misuse)*	Did you experience faecal or urinary incontinence?—*This may support the diagnosis of a seizure (but does not exclude syncope)*	How long did it take for you to recover?—*May help differentiate causes, recovery of full consciousness is typically slower with seizures than syncope*
Who were you with? Did anyone witness the event?	Are there any witness descriptions of the event e.g., rhythmic limb jerking, snoring, unusual posturing, confusion? *Head-turning to one side during a loss of consciousness is strongly suggestive of seizure*	Do you remember how you got to the hospital?
Do you or did you get any warning symptoms? *(e.g., light-headedness, aura, palpitations)*		How are you feeling now? Do you feel any ongoing symptoms? *Recovery from a simple faint may be far quicker than other causes*
Has this ever happened before? If so, how frequent?		

- **Bedside tests:** a 12-lead ECG should always be performed after a syncopal episode to identify heart block, pre-excitation, or long QT syndrome.
- **Focused Acute Medical Ultrasound (FAMUS):** if presenting features are suspicious for cardiovascular causes such as aortic dissection or massive PE, or shock including sepsis.
- **CT brain** (plain or contrast enhanced): request where an intracranial/neurological cause is suspected or where concern exists for a traumatic complication of a blackout (e.g., subdural haematoma [SDH]).
- **Echocardiography and cardiac rhythm monitoring**—Holter monitoring and echocardiography are required where cardiac syncope is suspected. An implantable loop recorder (ILR) is occasionally needed for infrequent events with a possible cardiac origin (see Chapter 17, 'Ambulatory taped electrocardiography' and **Fig. 17.19**). Tilt-table testing (see Chapter 17, 'Tilt testing') is sometimes diagnostic in patients with recurrent unexplained syncope, but has low sensitivity.

Syncope or fainting

The simple faint that over half the population experiences at some time (particularly in childhood, youth, or pregnancy) is due to sudden reflex bradycardia with vasodilation of both peripheral and splanchnic vasculature *(neurocardiogenic or vasovagal syncope)*.

- ***Precipitants:*** a common response to prolonged standing, fear, venesection, or pain. Syncope almost never occurs in the recumbent posture.
- ***Prodrome:*** usually brief. There is dizziness and a light-headed feeling, often with nausea, sweating, a feeling of heat, and visual grey-out.
- ***Blackout:*** patients usually lie still but jerking and twitching movements can occur and are sometimes mistaken for a convulsion. Their appearance is pale. Incontinence of urine or faeces can occur and is not a good discriminator between seizure and syncope.
- ***Recovery:*** rapid, usually taking place over seconds, but may be followed by a feeling of general fatigue (as opposed to post-ictal drowsiness and confusion following a seizure).

Other types of syncope

- ***Cardiac syncope*** (Stokes–Adams attacks) is potentially serious and often treatable. Typically, there is little or no warning. Cardiac arrhythmias, such as those due to heart block, or left ventricular outflow tract obstruction may be the cause. Syncope during exercise is often cardiac in origin.
- ***Micturition syncope*** occurs during micturition in men, particularly at night.
- ***Cough syncope*** occurs when venous return to the heart is obstructed by bouts of severe coughing; it is also seen with laughter occasionally.
- ***Postural hypotension*** (see **Box 27.7**) can cause syncope and occurs in the elderly, in autonomic neuropathy, and with some drugs, such as antihypertensives.
- ***Carotid sinus syncope*** is due to a vagal response caused by pressure over the carotid sinus baroreceptors in the neck: for example, due to a tight collar.
- ***Convulsive syncope*** is when collapsing in a propped-up position following a syncope results in delayed restoration of cerebral blood flow and may lead to a secondary anoxic seizure following syncope.

Other conditions to consider

Seizures are discussed in **Box 11.78**.

Dissociative seizures (sometimes referred to as non-epileptic attack disorder or psychogenic seizures, see also Chapter 28) regularly cause difficulty in diagnosis as attacks may sometimes closely resemble generalized tonic–clonic seizures. Usually, there are bizarre thrashing, non-synchronous limb movements, but on occasion there can be extreme difficulty in separating these attacks from seizures. EEG video telemetry is valuable. People who are diagnosed with this condition can have periods of loss of control and/or awareness, that can appear like an epileptic seizure, and even be incorrectly labelled as status epilepticus. Blood gas analysis may resemble that of an epileptic seizure if prolonged (raised carbon dioxide, raised lactate, metabolic acidosis). An accurate bystander or personal witness account is paramount to try to elucidate the exact nature of an episode.

Serum prolactin level may rise (typically 3–4 times baseline) following an epileptic seizure; however, there is a significant false-positive rate (up to 25%) in patients with dissociative seizures and thus this test is of limited clinical value to discriminate between the two conditions.

Panic attacks trigger sudden sympathetic activation and often hyperventilation, leading to respiratory alkalosis. They cause some or all of the following symptoms: dizziness, chest pains or tightness, a feeling of choking or shortness of breath, tingling in face and extremities, palpitations, trembling, and a feeling of dissociation or of impending doom. Consciousness is usually preserved and attacks are easily recognized.

Drop attacks are instant, unexpected episodes of lower limb weakness with falling, largely in females over 60 years. Awareness is preserved. They are due to sudden change in lower limb tone, presumably of brainstem origin. Sudden attacks of leg weakness also occur in hydrocephalus. A specific cause for drop attacks often cannot be identified, but this diagnostic label should only be applied after a careful diagnostic work-up has excluded cardiac and vestibular disease and other neurological disorders.

Vertigo (see also Chapter 21)

Vertigo is defined as the false sense of motion of the body or its surroundings, which can be spinning, swaying, or tilting and can be subdivided into peripheral and central causes. Peripheral vertigo encompasses the inner ear apparatus and vestibular nerve and accounts for the majority of cases of vertigo. Central vertigo is caused by pathology in the brainstem or cerebellum. These present with distinct clinical features albeit with some overlap. **Box 11.64** summarizes the different causes of peripheral and central vertigo.

The majority of patients who present to the acute medical take with vertigo have continuous symptoms. Most cases, while debilitating for the patient, are of benign aetiology, with an acute peripheral vestibulopathy (APV, either vestibular neuronitis or labyrinthitis) being the most common cause. Vertigo can however indicate serious pathology within the brainstem, cerebellum, or inner ear and vertebrobasilar stroke in particular is an important differential diagnosis in sudden onset cases. It is therefore important to follow a systematic clinical approach focused on discriminating benign and self-limiting vertigo from serious underlying disorders. **Red flags** pointing to potentially life-threatening or disabling diagnoses are listed in **Box 11.65**.

Clinical assessment

The history must initially determine the symptom as true vertigo. The question *'Did you just feel lightheaded or did you see the world spin around as though you had just got off a playground roundabout?'* is validated for differentiating vertigo from non-rotatory dizziness.

Box 11.64 Causes of vertigo

Peripheral	Central
Benign paroxysmal positional vertigo (BPPV)	Cerebellar stroke (bleed or infarct)
Acute peripheral vestibulopathy • Vestibular neuronitis • Labyrinthitis	Brainstem stroke (bleed or infarct)
	Vestibular migraine
Ménière's disease	Multiple sclerosis
Ramsay Hunt syndrome[a]	Space-occupying lesion
	CNS infection

[a]NB—Lower motor neuron facial weakness present in Ramsay Hunt syndrome.

Box 11.65 Red flags for more serious causes of vertigo

	Red flags	Possible causes
Vertigo +	Abnormal neurology/ new onset headache	Vertebrobasilar stroke
		Tumour
		Multiple sclerosis
		Intracranial infection
Vertigo +	Tinnitus/deafness	Ramsay Hunt syndrome
		Ménière's disease
		AICA stroke
		Labyrinthitis
Vertigo +	Unable to tolerate oral fluids (risk of dehydration, AKI, etc.)	Multitude

AICA, Anterior inferior cerebellar artery.

The **timing of the vertigo episode**(s) is important. It is crucial to clarify whether vertigo is persistent or episodic, and if so, what exactly triggers the vertigo (e.g., head movement). Episodic vertigo lasting seconds is usually indicative of benign paroxysmal positional vertigo (BPPV). Onset of symptoms in APV (e.g., labyrinthitis) is often rapid and thus speed of onset is not a good discriminator between benign vertigo and cerebrovascular causes.

The patient should be questioned directly about the presence of **associated symptoms:**

- Neurological symptoms, including focal weakness, ataxia, speech disturbance, visual change, or headache, suggest a central cause of vertigo and require urgent investigation.
- Audiological symptoms, namely hearing loss, tinnitus, or ear pain, point towards peripheral causes other than APV.
- Severe nausea and vomiting points towards a peripheral cause, though it can also be present in central causes. Inability to keep fluids down poses a risk of dehydration and warrants effective symptom control before considering discharge.

Past medical history is an important part of the initial risk assessment. The presence of stroke risk factors such as atrial fibrillation, hypertension, and diabetes increases suspicion of a cerebrovascular cause. Personal or family history of multiple sclerosis or Ménière's disease would increase the likelihood of vertigo attributed to these causes prompting focused assessments and/or onward referral to a specialist clinic.

Clinical examination should focus on gait (when safe/possible to assess), neurological and otological assessment. **Box 11.66** outlines the important clinical features discriminating central from peripheral causes of vertigo elicited on examination.

The Hallpike manoeuvre is of particular use in diagnosing BPPV and is only appropriate when vertigo is paroxysmal (i.e., non-continuous) (**Box 11.67**)

The HINTS examination (Head impulse—Nystagmus—Test of skew) is used in patients with continuous vertigo and helps to differentiate peripheral from central causes. (**Box 11.67**)

Management

Further investigation and management depends on the underlying cause. Symptomatic treatment with prochlorperazine or certain antihistamines (e.g., cinnarizine) may be warranted, particularly in cases accompanied by severe nausea or vomiting. Both vertigo itself and side-effects of medication may have implications for fitness to drive or work, for example operating machinery, and patients should be counselled accordingly.

Although a careful clinical assessment will usually allow discrimination between benign causes of vertigo and more serious pathology, confident clinical separation of peripheral and central causes of vertigo by the general physician is not always possible and in cases of diagnostic uncertainty or the presence of red flag features, further assessment with brain imaging (most commonly MRI with diffusion-weighted sequences) is usually performed.

Increasingly, SDEC services can provide therapeutic options such as intravenous antiemetics and vestibular suppressants, or arrange admission if deemed unsafe to investigate or manage on an ambulatory basis, with access to expedited specialist opinions such as ENT, neurology, or stroke.

Further reading

Edlow JA, Gurley KL, Newman-Toker DE. A new diagnostic approach to the adult patient with acute dizziness. *J Emerg Med*. 2018;54(4):469–483.
European Society of Cardiology (ESC). *Clinical Practice Guidelines: Guidelines for Diagnosis/Management of Syncope.* 2018.
NICE Guideline CG109: *Transient Loss Of Consciousness ('blackouts') in over 16s.* 2023.

Headache

Headache is a common presenting problem on the acute medical take and is a symptom often referred to SDEC or ambulatory services for further investigation and management. History taking is vital and a broad range of primary and secondary pathology can result in a headache (see **Box. 11.68**). Many SDEC services now have robust pathways for the investigation and management of acute headache; these are quick and accessible for both patients and primary care physicians and allow for rapid diagnosis or exclusion of certain causes of secondary headache in the presence of red flag symptoms (**Box 11.69**). For example, temporal (giant cell) arteritis is now commonly managed through SDEC with a dedicated temporal arteritis pathway, with streamlined access to rheumatology services and temporal artery ultrasound ± biopsy.

Clinical assessment

History

A careful history of the presenting problem is essential and SOCRATES is a helpful mnemonic:

- **Site:**
  - Migraine is often felt as unilateral ('hemicranial') throbbing although may become generalized as an attack progresses

Box 11.66 Clinical features of central vs. peripheral vertigo

Clinical features	Peripheral	Central
Gait	Can walk, may be unstable and reluctant to mobilize	Complete inability to stand up or walk even with the eyes open
Hearing loss or tinnitus	Possible (e.g., Ménière's disease, labyrinthitis)	Rare e.g., AICA stroke or intracranial tumour
Otoscopy	Discharge, vesicular eruptions (suggestive of Herpes Zoster), or evidence of cholesteatoma	No relevant findings
Other neurological symptoms	No	Usually present e.g., cranial nerve dysfunction, visual disturbance, dysphasia, dysarthria, weakness, sensory changes
Nystagmus	Horizontal nystagmus (with a torsional component) Does not alter in direction when the gaze changes (unidirectional) Inhibited by gaze fixation, fatiguable	Pure vertical or torsional nystagmus Direction-changing (bidirectional) nystagmus on lateral gaze (right beating on right gaze, left beating on left gaze) Not inhibited by gaze fixation Non-fatiguable
Manoeuvres	**Box 11.67** describes expected results for Hallpike manoeuvre (paroxysmal vertigo) and HINTS test (continuous vertigo) for peripheral and central causes	

Box 11.67 Hallpike manoeuvre and Head Impulse—Nystagmus—Test of Skew (HINTS)

Name	Hallpike manoeuvre	HINTS
Aim	Diagnose *paroxysmal* vertigo.	Differentiate central from peripheral causes of *continuous* vertigo.
Mechanism	Positional manoeuvre to reproduce vertigo and identify canalithiasis.	Highly sensitive and specific triad predicting a central cause of continuous nystagmus.
Process	1. Ask the patient to sit up on the couch. Extend the neck and turn 45 degrees to one side. From this position rapidly lie the patient down, supporting the head until it is 20–30 degrees over the end of the couch. 2. Observe the eyes for 30 seconds for the presence of nystagmus. Note its direction, duration, and type. Ask the patient to report symptoms of vertigo. 3. Slowly sit the patient up and repeat with head facing to the opposite side.	1. **Head impulse test.** With the patient resting flat on a couch, ask them to fix on your nose. Gentle move their head to each side, each time returning to the neutral position. In a 'positive' test, you will see an obvious 'catch up' saccade when the head is turned, as the patient re-fixes on your nose. This is absent in a 'negative' test. 2. **Nystagmus.** Carefully observe the nystagmus: is it unidirectional or bidirectional? Is it purely horizontal or are there vertical or rotary components? Does nystagmus direction change with different directions of gaze? 3. **Test of skew.** Cover one eye for several seconds and then uncover. Look for vertical realignment of the eye back to the primary position (i.e., abnormal vertical 'skew').
Interpretation	In BPPV: Horizontal nystagmus is seen when turning the head to the affected side. Nystagmus takes a few seconds to evolve ('latency') and diminishes on repeated tests ('fatiguability'). In central causes of paroxysmal vertigo: Nystagmus onset is immediate (i.e., without latency) and does not fatigue with repeated tests; direction is often not classical (e.g., horizontal towards the downward ear).	Head impulse test: Positive if a visible 'catch-up' saccade is seen on head turning. This implies a peripheral vestibular cause. A negative test or bilateral findings indicate a central cause. Nystagmus: Unidirectional horizonal nystagmus is in keeping with a peripheral cause. Bidirectional, vertical, or gaze-changing nystagmus implies a central cause. Test of skew: Vertical realignment movements indicate a central cause.

- Excruciating, unilateral retro-orbital pain may suggest cluster headache
- Occipital pain may point to sub-arachnoid haemorrhage (SAH), vertebral artery dissection, or meningitis
- Pain and tenderness specific to one temporal region may suggest temporal arteritis, rare in patients under 50 years of age

- **Onset and duration of symptoms:**
  - What time of day did the headache start? Did they awake from sleep?
  - Was there a specific trigger e.g., a sudden neck movement or turning to look over their shoulder for instance while reversing their car, which could suggest vertebral artery dissection
  - Has there been a history of trauma? (Consider an intracerebral bleed or fracture)
  - Did the headache come on suddenly or gradually and if so, over weeks (i.e., subacute), days, hours, or minutes?
- **Character**
  - A headache described as a tight band across the forehead may suggest tension-type headache (TTH) while migraine is often described as 'throbbing'
  - Is the headache constant or intermittent in nature, or does it fluctuate in severity?
  - The sensation of being 'hit across the back of the head' is a common patient description of SAH
- **Radiation**
  - Does the headache radiate down the neck? (This may suggest meningism or vertebral artery dissection)
  - Does the headache start behind the eyes or across the forehead and then radiate to one side of the head (which may

Box 11.68 Causes of headache

Primary headache disorders	Secondary headache disorders
Migraine	Raised intracranial pressure (e.g., idiopathic intracranial hypertension, primary or secondary tumour)
Tension-type headache	Intracranial infections (e.g., meningitis, sinusitis)
Trigeminal autonomic cephalalgias (incl. cluster headache)	Cerebral venous sinus thrombosis
Other primary headache disorders (e.g., primary cough headache, primary exercise headache, primary headache associated with sexual activity)	Giant cell/temporal arteritis
	Intracranial bleed (e.g., subdural haematoma, extradural haematoma, and subarachnoid haemorrhage)
	Cervicogenic headache
	Acute glaucoma
	Low CSF volume (low-pressure) headache
	Medication overuse headache
	Facial pain syndromes (e.g., trigeminal neuralgia, temporomandibular joint disorders)

Box 11.69 Red flag features for secondary headache

- New onset or change in headache in patients who are aged >50 years
- Thunderclap: rapid time to peak headache intensity (seconds to 5 min)
- Focal neurological symptoms (e.g., limb weakness, aura <5 min or >1 hour)
- Non-focal neurological disturbance (e.g., cognitive disturbance)
- Change in headache frequency, characteristics, or associated symptoms
- Abnormal neurological examination
- Headache that changes with posture
- Headache waking up the patient
- Headache precipitated by physical exertion or Valsalva manoeuvre (e.g., coughing, laughing, straining)
- Patients with risk factors for cerebral venous sinus thrombosis
- Jaw claudication or visual disturbance
- Neck stiffness
- Fever
- New onset headache in a patient with a history of human immunodeficiency virus (HIV)
- New onset headache in a patient with a history of cancer

suggest migraine), or all over the head (more suggestive of tension-type headache)?

- **Associated symptoms**
  - Was the headache preceded by an aura? (suggestive of migraine)
  - Is there associated vomiting? (This may accompany headache of any cause, but occurs more commonly in migraine and raised ICP)
  - Are there any visual symptoms? (In one eye, this may suggest temporal arteritis; both eyes are more frequently affected in migraine. Transient disturbances of vision ('obscurations') may occur with raised ICP)
  - Is there associated neck stiffness, fever, loss of appetite, drowsiness, or a reported change in behaviour? (These features may suggest meningoencephalitis)
  - Is it painful to brush or wash their hair, or chew food? (This may be a feature of trigeminal neuralgia, although overlaps with symptoms also described by patients with temporal arteritis and temporomandibular joint dysfunction. The character of the pain in trigeminal neuralgia—usually shock-like and paroxysmal—usually allows it to be separated quite readily from these other possibilities)
  - Is there any excessive lacrimation or ptosis? (This may be reported in cluster headache)
  - A prodrome of weight loss, low-grade fever, malaise, and muscular aches and stiffness (particularly shoulder and hip girdle) may indicate temporal arteritis. Jaw claudication is present in ~50% of patients
- **Time/Duration**
  - Did the headache reach its maximum severity within the first 60 seconds from time of onset? (This is known as a 'thunderclap headache', a classical feature of SAH)
  - How long did the headache last from initial onset to resolution, and if intermittent in nature, how long between episodes?
- **Exacerbating factors**
  - Is the headache worse on lying down or first thing in the morning, or when you cough, sneeze, or bend down e.g., to tie shoelaces? (This may be suggestive of raised ICP)
  - Is the headache worse on sitting up/positional? (This may be the case in a low-pressure headache)
- **Severity**
  - Use a pain scale and ask the patient to grade the severity out of 10 (where 1 is mild, and 10 is the worst headache they have ever had). Does the intensity vary with time of day or when they are doing specific things? Does this change?
- **Past medical history**
  - Are there any relevant underlying medical conditions (e.g., adult polycystic kidney disease, posing risk of berry aneurysm and SAH)?
  - Do they have a history of hypertension, increasing risk of intracerebral bleed?
  - Do they have a history of obesity (associated with idiopathic intracranial hypertension)?
  - Are they pregnant (risk factor for venous sinus thrombosis)?
- **Medication history**
  - Are they taking antiplatelets or anticoagulants?
  - Are they taking any medication that may increase venous thromboembolic risk e.g., combined oral contraceptive pill?
  - Are they taking any known medication which can cause headaches as a side-effect? E.g., calcium channel blockers such as nifedipine.
  - Are they taking a high amount of over-the-counter analgesia (possibly suggestive of medication overuse headache)?
- **Social history**
  - What is their caffeine intake usually like?
  - Do they smoke?
  - Any recent travel or long-haul journeys?
  - Any history of alcohol or recreational drug use?
- Is there something very stressful occurring in their life currently?

Physical examination

- General bedside assessment: Does the patient appear distressed and in pain? Do you need to consider simple analgesia during your assessment?

- What is the Glasgow Coma Score (GCS)? Do you need to consider urgent imaging?
- Is there evidence of photophobia or phonophobia? (e.g., the patient preferring to lie in a darkened room with the door closed)—causes may include migraine and meningeal irritation (meningitis, SAH)
- Are there other signs of meningism (e.g., neck stiffness) suggesting SAH or meningitis?
- What is the blood pressure?
- Is there any evidence of alcohol withdrawal or intoxication?
- A full neurological examination—is there any focal neurological deficit?
- Ophthalmology assessment including fundoscopy—is there papilloedema indicating raised ICP? Are there any abnormalities of visual fields suggestive of space-occupying lesion?

Investigations

Investigations will be guided by history and clinical examination and most primary headache disorders are diagnosed by this alone. For secondary causes of headache, consider the following:

- **Laboratory investigations** including inflammatory markers such as erythrocyte sedimentation rate (ESR) and CRP if considering temporal arteritis
- **Brain imaging**
  - A non-contrast CT head is the first-line investigation for suspected SAH or other intracranial bleeds
  - Contrast-enhanced CT is usually required for identification and characterization of space-occupying lesions (e.g., tumour or abscess)
  - CT or MRI venography is required for diagnosis of cerebral venous sinus thrombosis
- **Lumbar puncture** is required where concern exists for intracranial infection. This is contraindicated with raised ICP resulting from a space-occupying lesion or hydrocephalus due to the attendant risk of tonsillar herniation (coning). While neuroimaging should not be used as a surrogate for the clinical identification of raised ICP, brain imaging is recommended in patients with symptoms and signs which could associate with the presence of structural CNS disease (e.g., decreased conscious level, focal neurological deficit, recent seizure or an immunocompromised state). LP for detection of xanthochromia is also indicated in patients with suspected SAH presenting late (>6 h) after symptom onset where CT does not confirm the diagnosis. In this setting, LP should be postponed until at least 12 h after symptom onset to maximize diagnostic yield.
- **Diagnosis of temporal arteritis** requires temporal artery ultrasound or biopsy, however, empirical treatment with prednisolone is usually commenced where a high index of suspicion exists for the diagnosis to reduce the risk of irreversible visual loss (discussed further in Chapter 32).

Migraine

Migraine is the most common primary headache disorder, affecting approximately 6 million people in the UK in the age range 16–65 years. In the Global Burden of Disease Study 2010 (GBD2010) migraine was ranked as the third most prevalent disorder in the world, and in GBD2015 it was ranked the third-highest cause of disability worldwide in both males and females under the age of 50. Migraine can be split into two types: with and without aura. A detailed description of pathogenesis and clinical features may be found in Chapter 28.

Box 11.70 POUND criteria for diagnosis of migraine

Pulsatile
One-day duration (4–72 h)
Unilateral
Nausea/vomiting
Disabling

Clinical assessment

The POUND criteria (**Box 11.70**) are useful in the diagnosis of migraine in an acute medicine setting. If four or more of these features are present, there is a 93% likelihood that this headache is a migraine. Sufferers often describe a 2- to 24-hour prodrome which may include myalgia, change in bowel habit, or mood change followed by an aura, and then progressing to the pain phase. It is crucial to intervene early when the pain is still mild, using adequate drug doses and appropriate routes of administration.

Management

Management can be divided into non-pharmacological (e.g., trigger avoidance) and pharmacological therapies. After acute management of migraine, it is also important to consider future prophylaxis. In the UK, both NICE and the British Association for the Study of Headache (BASH) advise that patients suffering from migraines should avoid over-the-counter remedies containing opiates as they will exacerbate nausea, impair gastric absorption of medication, and are strongly linked to the development of medication overuse headache: an intractable and often debilitating condition. Patients should be educated as to which over-the-counter treatments contain opiates.

Many patients with migraine find good relief with simple analgesics. These include paracetamol 1 g (up to four times daily), aspirin 900 mg (single dose, taken at onset of migraine), diclofenac 50 mg (up to three times daily), and naproxen 500 mg (up to twice daily).

There is mixed evidence for intravenous fluid as a treatment for acute migraine. While dehydration is a known trigger for migraine, and migraine-associated nausea often limits oral fluid intake, there is insufficient evidence to endorse admission to hospital or ambulatory care units for cannulation and fluid therapy. Liberal fluid therapy may be associated with adverse effects.

Oral medications should be administered with a prokinetic agent such as metoclopramide to facilitate enteral absorption; stomach emptying is greatly reduced in the pain phase of acute migraine. Triptans (5-$HT_{1B/1D}$ agonists) may also be effective in those not responsive to simple analgesics and parenteral administration (subcutaneous injection) may be effective where vomiting limits oral absorption.

Migraine suppression and recurrent migraine management are discussed in Chapter 28.

Tension-type headache

Tension-type headache (TTH) is the second most common disabling primary headache disorder and is characterized by a predisposition to attacks of mild to moderate headache with fewer associated symptoms. The diagnostic criteria are summarized in **Box 11.71**.

Management

For TTH; NICE guidelines advise simple analgesia such as paracetamol, aspirin, or NSAIDs in the first instance as soon as possible after the onset of the attack. Opioids should not be offered and, if so severe that they are needed, the diagnosis should be

Box 11.71 Diagnostic criteria for tension-type headache

(A) At least 10 episodes fulfilling the criteria B-D

(B) Headache lasting from 30 min to 7 days

(C) Headache has at least two of the following characteristics:
- Bilateral location
- Pressing/tightening (non-pulsatile) quality
- Mild to moderate intensity
- NOT aggravated by routine physical activity such as walking or climbing stairs

(D) Both of the following:
- No nausea or vomiting (anorexia may occur)
- No more than one episode of photophobia or phonophobia

(E) Not attributable to another disorder

reconsidered. Some co-morbidities that may be associated with TTH include stress, mood disorders, chronic pain, and sleep disorders. In the long term, there is no evidence for the use of SSRIs, botulinum toxin, homoeopathy, or beta-blockers. There is good evidence for the use of amitriptyline and regular exercise.

Further reading

Ahmed F, Bahra A, Tyagi A, and Weatherby S. *National Headache Management System for Adults, 2019*. British Association for the Study of Headache. 2019; http://www.bash.org.uk.

Chinthapalli K, Logan AM, Raj R, Nirmalananthan N. Assessment of acute headache in adults—what the general physician needs to know. *Clin Med*. 2018; 18 (5):422–427.

Clinical NICE guideline CG 150 Headaches in over 12's diagnosis and management. 2021.

International Headache Society (IHS). Headache classification committee of the International Headache Society (IHS) the international classification of headache disorders, 3rd edition. *Cephalgia*. 2018; 38(1):1–211.

Intracranial bleeds

This comprises:
- intracerebral and cerebellar haemorrhage (see Chapter 29)
- subarachnoid haemorrhage
- subdural and extradural haemorrhage/haematoma.

Subarachnoid haemorrhage

Subarachnoid haemorrhage (SAH) means spontaneous arterial bleeding into the subarachnoid space, and is usually clearly recognizable clinically from its dramatic onset. SAH accounts for some 5% of strokes and has an annual incidence of 6 per 100 000.

Aetiology

The causes of SAH are shown in **Box 11.72**; it is unusual to find any contributing disease.

Saccular (berry) aneurysms

Saccular aneurysms (**Fig. 11.35**) develop within the circle of Willis and adjacent arteries. Common sites are at arterial junctions:

a. between posterior communicating and internal carotid artery—posterior communicating artery aneurysm
b. between anterior communicating and anterior cerebral artery—anterior communicating and anterior cerebral artery aneurysm
c. at the trifurcation or a bifurcation of the middle cerebral artery—middle cerebral artery aneurysm.

Other aneurysm sites are on the basilar, posterior inferior cerebellar, intracavernous internal carotid, and ophthalmic arteries. Saccular aneurysms are an incidental finding in 1% of postmortems and can be multiple.

Box 11.72 Underlying causes of subarachnoid haemorrhage

Cause	(%)
Saccular (berry) aneurysms	70%
Arteriovenous malformation (AVM)	10%
No arterial lesion found	15%
Rare associations Bleeding disorders Mycotic aneurysms—endocarditis Acute bacterial meningitis Tumours (e.g., metastatic melanoma, oligodendroglioma) Arteritis (e.g., systemic lupus erythematosus) Spinal AVM → spinal subarachnoid haemorrhage Coarctation of the aorta Marfan or Ehlers–Danlos syndromes Adult polycystic kidney disease	<5%

Fig. 11.35 Digital subtraction angiogram: posterior communicating artery aneurysm.

Aneurysms cause symptoms either by spontaneous rupture, when there is usually no preceding history, or by direct pressure on surrounding structures; for example, an enlarging unruptured posterior communicating artery aneurysm is the most common cause of a painful IIIrd nerve palsy (see Chapter 28).

Arteriovenous malformation

Arteriovenous malformation (AVM) is a vascular developmental malformation, often with a fistula between arterial and venous systems, causing high flow through the AVM and high-pressure arterialization of draining veins. An AVM usually presents following a spontaneous intracerebral haemorrhage or with a seizure, often focal in onset. The risk of a first haemorrhage in unruptured AVMs (20% fatal and 30% resulting in permanent disability) is approximately 2%–3% per year. Once an AVM has caused a haemorrhage, the risk of rebleeds is increased to approximately 10% per year. AVMs may be ablated with endovascular treatment (catheter injection of glue into the nidus, usually), microsurgery, or stereotactic radiotherapy. An MDT approach with neurologist, interventional neuroradiologist, and neurosurgeon is required in deciding on treatment options. There is no clear consensus on the best treatment modality or, indeed, whether the considerable risk of intervention is lower than with a conservative approach.

Fig. 11.36 Magnetic resonance T2-weighted image: symptomless frontal cavernoma *(arrowed)*.

Fig. 11.37 Subarachnoid haemorrhage. Axial, non-contrast CT showing blood filling the basal cisterns around the brainstem.

Cavernous haemangiomas (cavernomas) are common (0.1%–0.5% prevalence) and consist of a tangle of low-pressure dilated vessels without a major feeding artery; they are frequently symptomless (Fig. 11.36) and seen incidentally on imaging. Multiple cavernomas often have a genetic basis. Cavernomas may cause seizures. Small haemorrhages may occur but are usually low-pressure bleeds and rarely cause severe deficits. Surgical resection is rarely needed, except where the cavernoma is gradually enlarging or causing significant neurological symptoms.

Clinical assessment

Patients with SAH classically present with a sudden, very severe headache, often occipital (mean time to peak headache 3 min). Headache is usually followed by vomiting and often by coma and death. Survivors of SAH may remain comatose or drowsy for hours, days, or longer. SAH is a *possible* diagnosis in any sudden headache.

Following major SAH, there is neck stiffness and a positive Kernig's sign. Papilloedema is sometimes present, with retinal and/or subhyaloid haemorrhage (tracking beneath the retinal hyaloid membrane). Minor bleeds cause few signs but almost invariably do cause headache (approximately 17% of patients have small 'sentinel bleeds' in the weeks before presenting with SAH).

Investigations

CT imaging (non-contrast) is the immediate investigation (Fig. 11.37). Subarachnoid and/or intraventricular blood is usually seen (sensitivity of CT to detect subarachnoid blood is 95% within 24 hours of onset but much lower over subsequent days).

If a CT scan performed *within 6 hours of symptom onset* (reported by a radiologist) shows no blood in the subarachnoid space, this can usually be considered to exclude the diagnosis, unless a very strong index of suspicion persists.

If a CT scan performed *more than 6 hours after symptom onset* shows no blood in the subarachnoid space, this should not be considered to exclude the diagnosis, and an LP should be considered.

A **lumbar puncture** is not necessary if SAH is confirmed by CT. In cases where doubt remains and a CT has not shown blood in the subarachnoid space, at least 12 hours from symptom onset should be allowed before conducting an LP (the time required for the yellow colour of xanthochromia to appear within the CSF, which subsequently remains detectable for 2 weeks). Visual inspection of supernatant CSF is usually sufficiently reliable for diagnosis. Spectrophotometry to estimate bilirubin in the CSF released from lysed cells is used to define SAH with certainty.

CT angiography or catheter angiography to identify the aneurysm or other source of bleeding is performed in patients potentially fit for surgery. In some, no aneurysm or source of bleeding is found, despite a definite SAH.

Differential diagnosis

SAH must be differentiated from migraine. This is sometimes difficult but a short time to maximal headache intensity and the presence of neck stiffness usually indicate SAH. ***Thunderclap headache*** is used (confusingly) to describe either SAH or a sudden (benign) headache for which no cause is ever found. The syndrome of reversible cerebral vasoconstriction (Call–Fleming's syndrome) presents with thunderclap headache. Acute bacterial meningitis occasionally causes a very abrupt headache, when a meningeal microabscess ruptures; SAH also occasionally occurs at the onset of acute bacterial meningitis. Cervical arterial dissection can present with a sudden headache.

Complications

Blood in the subarachnoid space can lead to obstructive hydrocephalus, seen on CT. Hydrocephalus can be asymptomatic but may cause deteriorating consciousness following SAH. Shunting may be necessary. Arterial spasm (visible on angiography and a cause of coma or hemiparesis) is a serious complication of SAH and a poor prognostic feature.

Management

Immediate treatment of SAH involves bed rest and supportive measures. Hypertension should be controlled. Nimodipine, a calcium-channel blocker given for 3 weeks, reduces mortality.

All SAH cases should be discussed urgently with a neurosurgical centre. Nearly half of SAH cases are either dead or moribund before reaching hospital. Of the remainder, a further 10%–20% rebleed

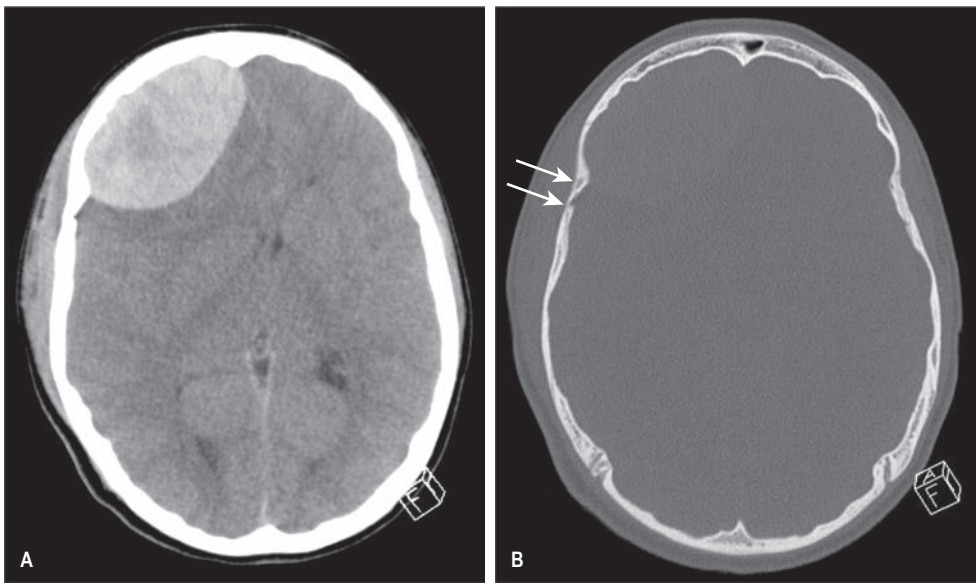


Fig. 11.38 Extradural haematoma. (A) Shown on hyperintensity on CT brain window. **(B)** Note the subtle linear non-depressed fracture of the right frontal bone *(arrows)*. This extended adjacent to the middle meningeal artery.

and die within weeks. Failure to diagnose SAH—for example, mistaking SAH for migraine—contributes to this mortality.

Where angiography demonstrates an aneurysm (the cause of the vast majority of SAHs), endovascular treatment by placing platinum coils via a catheter in the aneurysm sac, to promote thrombosis and ablation of the aneurysm, is now the first-line treatment. Endovascular coiling has a lower complication rate than surgery but direct surgical clipping of the aneurysm neck is still required in some selected cases. For asymptomatic (unruptured) aneurysms over 8 mm in diameter, the risk of treatment is less than the risk of haemorrhage if not treated. Patients who remain comatose or who have persistent severe deficits after SAH have a poor outlook.

Subdural and extradural bleeding

These conditions can cause death following head injuries unless treated promptly.

Subdural haematoma

Subdural haematoma (SDH) means accumulation of blood in the subdural space following rupture of a vein. This usually follows a head injury, sometimes a trivial one. The interval between injury and symptoms can be days, or may extend to weeks or months. Chronic, apparently spontaneous, SDH is common in the elderly, and also occurs with anticoagulants.

Headache, drowsiness, and confusion are common; symptoms are indolent and can fluctuate. Focal deficits, such as hemiparesis or sensory loss, develop. Epilepsy occasionally occurs. Stupor, coma, and coning may follow.

Extradural haemorrhage

Extradural haemorrhage (EDH) typically follows a linear skull vault fracture tearing a branch of the middle meningeal artery. Extradural blood accumulates rapidly over minutes or hours. A characteristic picture is that of a head injury with a brief duration of unconsciousness, followed by improvement (the lucid interval). The patient then becomes stuporose; there is an ipsilateral dilated pupil and contralateral hemiparesis, with rapid transtentorial coning. Bilateral fixed, dilated pupils, tetraplegia, and respiratory arrest follow. An acute progressive SDH presents similarly.

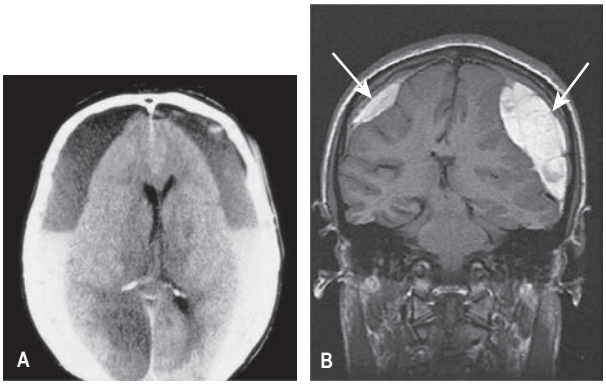


Fig. 11.39 Bilateral subdural haematomas. (A) CT scan. (B) Magnetic resonance T1-weighted image.

Management

Possible extradural or subdural bleeding needs immediate imaging. CT (Figs. 11.38 and 11.39A) is the most widely used investigation because of its immediate availability. MRI is more sensitive for the detection of small haematomas. T1-weighted MRI (Fig. 11.39B) shows bright images due to the presence of methaemoglobin.

EDHs *require urgent neurosurgery*; if it is performed early, the outlook is excellent. When the patient is far from a neurosurgeon, such as in wartime or at sea, drainage through skull burr holes has been lifesaving when an EDH has been diagnosed clinically.

Subdural bleeding usually needs less immediate attention but close neurosurgical liaison is necessary. Even large collections can resolve spontaneously without drainage. Serial imaging is required to assess progress.

Cortical venous thrombosis and dural venous sinus thrombosis

Intracranial venous thromboses are usually (>50%) associated with a prothrombotic risk factor, such as oral contraceptives, pregnancy, genetic or acquired prothrombotic states, and dehydration.

Head injury is also a cause. Infection, such as from a paranasal sinus, may be present. Venous thromboses can also arise spontaneously.

Cortical venous thrombosis

The venous infarct leads to headache, focal signs (e.g., hemiparesis), and/or seizures. Cortical haemorrhagic infarction is shown on MRI.

Dural venous sinus thromboses

Cavernous sinus thrombosis causes eye pain, fever, proptosis, and chemosis. External and internal ophthalmoplegia with papilloedema develops.

Thrombosis of the dural venous sinuses, such as when the sagittal sinus causes raised ICP with headache, papilloedema, and frequently seizures, may progress to coma.

Management

MRI and MR venography (MRV) show occluded sinuses and/or veins. Treatment is with heparin initially, followed by warfarin or other oral anticoagulants for 6 months. Anticonvulsants are given if necessary.

Meningitis

Meningitis is an infection of the meninges and bacterial meningitis is fatal unless treated. Microorganisms have usually travelled from mucosal surfaces such as the ears or nasopharynx via the bloodstream, or from a cranial injury or congenital meningeal defect. Immunocompromised patients are at risk of infection with unusual organisms. Non-infective causes of meningeal inflammation include malignant meningitis, intrathecal drugs, and blood following SAH. Infective causes are listed in (**Box 11.73**). Discussion of global trends in the incidence, causative agents, and role for vaccination against meningitis is found in Chapter 42.

? Box 11.73 Infective causes of meningitis in the UK[a]

Bacteria
- *Neisseria meningitidis*
- *Streptococcus pneumoniae*[a]
- *Staphylococcus aureus*
- Streptococcus group B
- *Listeria monocytogenes*
- Gram-negative bacilli e.g., *Escherichia coli*
- *Mycobacterium tuberculosis*
- *Treponema pallidum*

Viruses
- Enteroviruses:
  - ECHO
  - Coxsackie
- (Poliomyelitis—mainly eradicated worldwide)
- Mumps
- Herpes simplex
- HIV
- Epstein–Barr

Fungi
- *Cryptococcus neoformans*
- *Candida albicans*
- *Coccidioides immitis, Histoplasma capsulatum, Blastomyces dermatitidis* (USA)

[a]These organisms account for 70% of acute bacterial meningitis outside the neonatal period. A wide variety of infectious agents are responsible for the remaining 30% of cases. *Haemophilus influenzae b* (Hib) has been eliminated as a cause in many countries by immunization. Malaria often presents with cerebral symptoms and a fever.

In acute bacterial meningitis, the pia-arachnoid is congested with neutrophils. A layer of pus forms. This may organize to form adhesions, causing cranial nerve palsies and hydrocephalus. In chronic infection (e.g., tuberculosis), the brain is covered in a viscous grey–green exudate with numerous meningeal tubercles. Adhesions are invariable. Cerebral oedema occurs in any bacterial meningitis. In viral meningitis there is a predominantly lymphocytic inflammatory CSF reaction without pus formation, polymorphs, or adhesions; there is little or no cerebral oedema unless encephalitis develops.

Clinical assessment

Meningitis often presents with a triad of symptoms: headache, neck stiffness, and fever. Photophobia and vomiting are often present. In acute bacterial infection, there is usually intense malaise, fever, rigors, severe headache, photophobia, and vomiting, developing within hours or minutes. The patient is irritable and often prefers to lie still. Neck stiffness and a positive Kernig's sign usually appear within hours. Brudzinski's sign (severe neck stiffness causing a patient's hips and knees to flex when the neck is flexed) may be present.

Meningococcal meningitis may be part of a generalized meningococcal septicaemia: a non-blanching purpuric rash may be present and these patients may deteriorate rapidly with septic shock and peripheral vascular infarcts.

It is important to assess for signs suggesting raised ICP which should trigger urgent brain imaging and may impact the safety of LP:
- reduced or fluctuating level of consciousness
- relative bradycardia and hypertension
- focal neurological signs
- abnormal posture or posturing
- unequal, dilated, or poorly responsive pupils
- papilloedema
- abnormal 'doll's eye' movements (**Fig. 11.33**).

In uncomplicated meningitis, consciousness remains intact, although anyone with a high fever may be delirious. Progressive drowsiness, lateralizing signs, and cranial nerve lesions indicate complications such as venous sinus thrombosis, severe cerebral oedema or hydrocephalus, or an alternative diagnosis such as cerebral abscess or encephalitis.

It may be difficult to distinguish between the sudden headache of SAH, migraine, and acute meningitis. Meningitis should be considered seriously in anyone with headache and fever, and in any sudden headache. Neck stiffness should be looked for; it may not be obvious.

In less severe cases (e.g., many viral meningitides), there are less prominent meningitic signs. However, bacterial infection may also be indolent, with a deceptively mild onset.

Specific varieties of meningitis

Clinical clues point to the diagnosis (**Box 11.74**). If there is access to the subarachnoid space via skull fracture (recent or old) or occult spina bifida, bacterial meningitis can be recurrent, and the infecting organism is usually pneumococcus. **Box 11.75** summarizes typical CSF changes in different forms of meningitis.

Acute bacterial meningitis

Onset is typically sudden, with rigors and high fever. Meningococcal septicaemia is associated with a petechial rash, sometimes sparse (**Fig. 11.40**). The meningitis may be part of a generalized meningococcal septicaemia with septic shock and peripheral vascular

infarcts (see Chapter 42). Acute septicaemic shock may develop in any bacterial meningitis.

Viral meningitis

This is almost always a benign, self-limiting condition lasting 4–10 days. Headache may follow for some months. There are no serious sequelae, unless an encephalitis is present (see below). Causes are listed in **Box 42.34**.

Chronic meningitis

For further discussion on chronic meningitis, see Chapter 28.

Chronic meningitis sometimes resembles an intracranial mass lesion, with headache, epilepsy, and focal signs. *Cerebral malaria* (Chapter 42) can mimic bacterial meningitis.

Investigations

When meningococcal meningitis is diagnosed clinically by the presence of the typical petechial rash, immediate parenteral antibiotics should be given, ideally in the pre-hospital setting (see **Box 11.77**); in this situation LP is unnecessary. In other cases of meningitis, an LP is performed if there is no clinical suspicion of a mass lesion.

- **Laboratory investigations should** include full blood count, CRP, renal profile, liver function tests, and coagulation screening. Blood cultures are important, ideally prior to commencing intravenous antibiotics. Whole blood polymerase chain reaction (PCR) testing for meningococcus and pneumococcus should be sent. Be aware that a normal CRP and white blood cell count does not rule out bacterial meningitis.
- **CT/MRI brain** may be indicated if there are red flag features on presentation or examination including any signs of raised ICP, focal neurological signs, or risk factors for a space-occupying lesion (e.g., immunocompromised patient). Identification of a mass lesion on LP contraindicates LP, as coning of the cerebellar tonsils may follow LP. Importantly, normal brain imaging does not exclude raised ICP and clinical features of raised ICP also contraindicate LP.
- **Lumbar puncture and CSF analysis** (**Box 11.75**)—Guidance recommends obtaining CSF samples within 1 h of presentation to increase probability of a positive culture. Opening pressure should be measured and samples sent for white cell count, microscopy, culture and sensitivity, viral PCR, total protein, glucose, and cytology. CSF stains demonstrate organisms (e.g., Gram-positive intracellular diplococci—pneumococcus; Gram-negative cocci—meningococcus). Ziehl–Neelsen stain demonstrates acid-fast bacilli (tuberculosis), though TB organisms are rarely numerous. Indian ink stains fungi.
- Meticulous attention should focus on microbiological studies in suspected CNS infection with close liaison between clinician and microbiologist. Specific techniques, such as PCR for meningococci and viruses, or CSF bacterial antigen testing, are invaluable. Syphilis and HIV serology should always be carried out.

Management

Recognition and immediate treatment of acute bacterial meningitis are vital. Minutes save lives. Bacterial meningitis has a high mortality and morbidity. Even with optimal care, mortality is around 15%. The following applies to adult patients; management is similar in children. Immediate antibiotic treatment for suspected meningococcal meningitis is indicated at first contact before investigation (**Box 11.76**).

? Box 11.74 Clinical clues in meningitis

Clinical feature	Possible cause
Petechial rash	Meningococcal infection
Skull fracture Ear disease Congenital CNS lesion	Pneumococcal infection
Immunocompromised patients	HIV with opportunistic infection
Rash or pleuritic pain	Enterovirus infection
International travel	Malaria
Occupational: work in drains, canals, polluted water, recreational swimming Clinical: myalgia, conjunctivitis, jaundice	Leptospirosis

Fig. 11.40 Rash of meningococcal septicaemia with meningitis. Purpuric lesions.

i Box 11.75 Typical changes in cerebrospinal fluid in viral, pyogenic, and tuberculous meningitis

	Normal	Viral	Pyogenic	Tuberculous
Appearance	Crystal clear	Clear/turbid	Turbid/purulent	Turbid/viscous
Mononuclear cells	<5/mm^3	10–100/mm^3	<50/mm^3	100–300/mm^3
Polymorph cells	Nil	Nil[a]	200–300/mm^3	0–200/mm^3
Protein	0.2–0.4 g/L	0.4–0.8 g/L	0.5–2.0 g/L	0.5–3.0 g/L
Glucose	⅔–½ blood glucose	>½ blood glucose	<½ blood glucose	<½ blood glucose

[a]Some CSF polymorphs may be seen in the early stages of viral meningitis and encephalitis.

Antibiotics (Box 11.77)

- A third-generation cephalosporin (intravenous ceftriaxone or cefotaxime) provides adequate cover for the majority of important community-acquired bacterial meningitis pathogens.
- High-dose penicillin such as amoxicillin is indicated in *Listeria monocytogenes* meningitis and should be added empirically in high-risk groups.

Corticosteroids

- Adjunctive immediate high dose steroid (dexamethasone 0.15 mg/kg i.v. 6-hourly), commenced with or before the first dose of antibiotics, has been shown to reduce neurological complications including deafness in bacterial meningitis due to pneumococcus or Haemophilus influenza type B (Hib). Dexamethasone should be commenced in all patients with bacterial meningitis of unknown cause, continued for 4 days in those with pneumococcus/Hib and stopped for all other organisms. In cases where no causative organism is found, advice should be sought from an infectious diseases physician as to whether dexamethasone should be continued.

Box 11.76 Meningococcal meningitis and meningococcaemia: emergency treatment

Suspicion of meningococcal infection is a medical emergency requiring treatment immediately.

Clinical features

- Petechial or non-specific blotchy, red rash
- Fever, headache, neck stiffness

Not all these features may be present—and meningococcal infection may sometimes begin like any apparently non-serious infection.

Immediate treatment for suspected meningococcal meningitis at first contact before investigation (pre-hospital, with GP/paramedics).

- i.v. or i.m. benzylpenicillin unless history of anaphylaxis

In meningitis, minutes count—delay is unacceptable.

On arrival in hospital

- Blood tests, including blood cultures, immediately
- Monitoring for septic shock
- Third-generation cephalosporin e.g., cefotaxime or ceftriaxone, as empirical therapy (increasing rates of penicillin resistance)
- Dexamethasone is not indicated for meningococcal meningitis; in bacterial meningitis of unknown cause, commence dexamethasone 0.15 mg/kg i.v. 6-hourly, but continue only if pneumococcus or Haemophilus influenza type B is isolated and discontinue if meningococcus

Local infection (e.g., paranasal sinus) should be treated surgically if necessary. Repair of a depressed skull fracture or meningeal tear may be required.

Prophylaxis

Meningococcal infection should be notified to the public health authorities, and advice sought about immunization and prophylaxis of contacts. Chemoprophylaxis with rifampicin or ciprofloxacin should be offered to all close contacts. MenC vaccine is given in the UK and MenB, a meningococcal B vaccine, is now available for population immunization of infants and for use in outbreaks. A combined A and C meningococcal vaccine is sometimes used prior to travel from the UK to endemic regions, such as Africa or Asia, and there is a quadrivalent ACWY vaccine for specific events, such as the Hajj and Umrah in Mecca.

Pneumococcal conjugated vaccine is now given to infants in many countries and pneumococcal polysaccharide vaccine is offered to older adults and those with, for example, immunodeficiency or splenectomy. Pneumococcal immunization has reduced the incidence of pneumococcal meningitis.

Hib (Haemophilus influenzae) vaccine is given routinely in childhood in the UK and many other countries, virtually eliminating a common cause of fatal meningitis.

Contact tracing is important and identifying any at-risk contacts:

- Identify those at risk
- Be aware that meningitis and meningococcal disease are notifiable disease
- Inform close contacts that the risk to a contact is low with meningococcal disease. The risk is highest in the first 7 days after a case is diagnosed and falls sharply thereafter
- Prophylaxis or vaccination for close contacts is usually arranged by secondary care
- Prophylaxis against meningococcal disease should be considered for the following close contacts, regardless of meningococcal vaccination status:
  - People who have had prolonged contact with case in a household-type setting during the 7 days before onset of illness (e.g., household contacts, pupils in the same dormitory, partners, or university students sharing a kitchen in a hall of residence).
  - People who have had transient close contact with a case only if they have been directly exposed to large particle droplets/secretions from the respiratory tract of a case around the time of admission to hospital.
- Antibiotic prophylaxis should be given as soon as possible (ideally within 24 h) after the diagnosis of the index case.

Box 11.77 Antibiotics in acute bacterial meningitis

Organism	Antibiotic	Alternative (e.g., allergy)
Unknown pyogenic	Third-generation cephalosporin e.g., ceftriaxone/cefotaxime (+ vancomycin in areas of high pneumococcal penicillin/cephalosporin resistance)	Chloramphenicol
Age >50 or immunocompromised	As above but add amoxicillin to cover *Listeria*	Co-trimoxazole for *Listeria*
Meningococcus	Third-generation cephalosporin initially Switch to benzylpenicillin if confirmed sensitive	Cefotaxime
Pneumococcus	Third-generation cephalosporin	Chloramphenicol
Haemophilus	Third-generation cephalosporin	Chloramphenicol

Viral encephalitis

Viral encephalitis is an inflammation of the brain parenchyma caused by a virus. It is the most common type of encephalitis and may coexist with viral meningitis (see **Box 42.34**). The viruses isolated from adult UK cases are usually herpes simplex (HSV 1 and 2), varicella zoster (VZV) and other herpes group viruses, HHV-6, 7, enteroviruses, and adenovirus. HSV encephalitis typically affects the temporal lobes initially and is often asymmetric. Frequently, the virus is never identified. Outside the UK, in endemic regions, different pathogens cause encephalitis, including flaviviruses (Japanese encephalitis, West Nile virus, tick-borne encephalitis) and rabies.

Local epidemics can occur. For example, in New York in the 1990s, West Nile virus caused an epidemic and Venezuelan equine virus was isolated from encephalitis cases in South America.

Clinical features

- Flu-like symptoms such as headache, fever, myalgia, fatigue, or weakness
- Confusion
- Headache
- Personality changes or agitation
- Seizures
- Loss of consciousness including coma

In cases of autoimmune encephalitis (see Chapter 28), symptoms may develop more slowly over several weeks.

Investigations

- MRI classically shows areas of inflammation and swelling, generally in the temporal lobes in HSV encephalitis. Raised ICP and midline shift may occur, leading to coning.
- EEG shows periodic sharp and slow-wave complexes.
- CSF shows an elevated lymphocyte count (95%).
- Viral detection by CSF PCR is highly sensitive for several viruses, such as HSV and VZV. However, a false-negative result may occur within the first 48 h of symptom onset. Serology (blood and CSF) is also helpful.
- Brain biopsy is rarely required since the advent of MRI and PCR.

Management

Suspected HSV and VZV encephalitis is treated immediately with intravenous aciclovir (10 mg/kg three times a day for 14–21 days), even before investigation results are available. Early treatment significantly reduces both mortality and long-term neurological damage in survivors. Seizures are treated with anticonvulsants. Occasionally, decompressive craniectomy is required to prevent coning but coma confers a poor prognosis.

Long-term complications are common and include memory impairment, personality change, and epilepsy.

Seizures

Seizures are a common neurological emergency encountered in the acute medical setting. An epileptic seizure is defined as a sudden synchronous discharge of cerebral neurons causing symptoms or signs apparent to the patient or observer, with a broad spectrum of manifestations ranging from a subjective 'aura' noticeable only to the patient to a generalized tonic–clonic seizure with loss of awareness. Causes of seizures are summarized in **Box 11.78**.

Seizure types

Seizures may be generalized (involving both cerebral hemispheres with loss of awareness) or focal (arising within one hemisphere, with preservation of awareness if 'motor' only). Distinguishing between generalized and focal seizures is important, both with regard to potential causes (e.g., focal motor seizures may localize to an

? Box 11.78 Causes of seizures

Causes of seizures	Description
Primary generalized epilepsy	A chronic neurological disorder characterized by recurrent, seizures with a structurally normal brain, usually with onset in childhood or early adulthood. Thought usually to be polygenic, although taking a family history is important.
Febrile seizures	Seizures triggered by high fever in young children (typically under the age of 5). This diagnosis should not be made in adults, and the coexistence of seizures and fever should prompt consideration for serious pathology such as intracranial infection.
Metabolic disturbances	E.g., Electrolyte imbalances (e.g., hyponatraemia or hypocalcaemia), hypoglycaemia, or uraemia.
Structural brain lesions	E.g., Tumours, vascular malformations, or traumatic brain injury.
Infections	Meningitis, encephalitis, or brain abscess may present with seizures. Consider the possibility of immunocompromise and HIV infection.
Toxic exposures	Exposure to toxins, such as recreational drugs, alcohol (particularly in withdrawal), or environmental toxins.
Prescription medication	A number of medications may trigger seizures either in toxicity (e.g., tricyclic antidepressants, antipsychotics, lidocaine, ciclosporin) or withdrawal (e.g., benzodiazepines)
Vascular events	A cerebral infarct (stroke) can disrupt blood flow to the brain and lead to seizures, either at the time of an acute event, or subsequently as a result of a cerebral 'scar'. Cerebral venous sinus thrombosis may also result in seizures. Enquire about both cardiovascular and venous thromboembolic risk factors when taking a history.
Autoimmune encephalitis	An autoimmune response against brain tissue can cause seizures. Evaluate for autoimmune disorders and ask about cancer (which can cause paraneoplastic encephalitis).
Hypoxic-ischemic events (Anoxia)	Lack of oxygen to the brain due to conditions like drowning or cardiac arrest may result in seizures.
Genetic disorders (e.g., Dravet syndrome)	Certain genetic conditions predispose individuals to seizures. Enquire about family history and genetic testing.
Idiopathic (unknown)	In some cases, no underlying cause is identified, and seizures are categorized as idiopathic.
Dissociative seizures	Seizures without underlying neurological causes, often associated with psychological stressors.

underlying structural brain lesion) and management. Seizures may or may not lead to a diagnosis of epilepsy; classification is complex and the latest classification of seizures, epilepsy types, and epilepsy syndromes is discussed further in Chapter 28.

Clinical assessment

When a patient presents with an acute seizure, the primary focus is on ensuring safety and stabilizing their condition. This involves protecting the patient from injury during the seizure, evaluating their airway, breathing, and circulation, and providing necessary immediate interventions like administering oxygen and initiating cardiac monitoring.

History

Detailed history from the patient and **any witnesses** helps understand the nature of the seizure and identify any predisposing factors or underlying causes. Specifically, enquire about events preceding, during and after the episode (including any aura, alteration in awareness, or post-ictal confusion or headache), potential triggers or seizure risk factors (e.g., previous stroke or head injury), as well as alcohol or recreational drug use (see **Box 28.29**). Patients with a prior diagnosis of epilepsy should be asked about any recent changes to their prescription medications or new medicines which may alter the seizure threshold and concordance with treatment. A seizure is a clinical diagnosis and needs to be carefully separated from other causes of impaired consciousness (e.g., syncope) and dissociative seizures (**Box 11.63**).

Physical examination

Once the seizure has terminated, carry out a full neurological examination and assess for any signs of injury or trauma resulting from the seizure. Transient focal neurological symptoms and signs may be seen following a generalized tonic–clonic seizure (Todd's paralysis), but persistent abnormal findings raise concern for a vascular event or structural brain lesion.

Investigations

Blood tests, including full blood count, electrolyte panel, glucose, renal and liver function tests, toxicology screen if indicated, and antiepileptic drug levels, can provide information about potential underlying causes or contributing factors. A 12-lead ECG should be performed to assess rhythm, conduction abnormalities, and QT interval.

Brain imaging is almost always indicated for a first presentation with seizure although the optimal modality varies with the scenario: MRI (± venography or diffusion-weighted imaging) often provides the most data, although non-contrast CT is a rapid and sensitive test where intracranial bleeding is suspected. EEG may be useful in selected cases, such as suspected non-convulsive status epilepticus or to differentiate seizure types; it is frequently requested as part of the post-discharge 'first seizure' work-up although it is insensitive and often normal.

Management

Most seizures last only minutes and terminate spontaneously. A prolonged seizure (>5 min) or repeated seizures may be terminated with rectal diazepam, intravenous lorazepam, or buccal midazolam. Oxygen should be given and the airway monitored in the post-ictal phase. Additionally, it is important to protect the patient from injury during the post-ictal period.

Many hospitals run rapid-access first seizure clinics accepting referrals directly from ED and acute medical services. Where such services exist, patients who experience a single seizure with full recovery, and have a normal examination and first-line investigations (blood tests/ECG) may be discharged with outpatient follow-up (see **Box 28.31** for first seizure checklist).

For patients with recurrent seizures or suspected epilepsy, it is advisable to consult a neurologist for further evaluation and guidance on long-term management (see Chapter 28).

In all instances, patients and families should be provided with education and support regarding seizure recognition, management, and safety. Information about potential triggers of seizures and advice on lifestyle modifications, such as ensuring adequate sleep, avoiding excessive alcohol consumption, and managing stress, should be given. Patients should be educated about the importance of medication adherence and potential side-effects.

On discharge, patients should receive clear instructions regarding their condition and any necessary precautions. It is important to advise patients to avoid potentially hazardous activities (e.g., bathing or swimming alone) and to stop driving and in the UK, inform the Driver and Vehicle Licensing Agency (DVLA). Patients and their families should be encouraged to have a seizure action plan in place, including when to seek medical help and how to provide immediate assistance during a seizure.

Status epilepticus

This medical emergency (**Box 11.79**) means continuous seizures for 30 minutes or longer (or two or more seizures without recovery of consciousness between them over a similar period). Status epilepticus has a mortality of 10%–15%. The longer the duration of status, the greater the risk of permanent cerebral damage. Rhabdomyolysis may lead to AKI in convulsive status epilepticus.

Over 50% of cases occur without a previous history of epilepsy. Some 25% with apparent refractory status have dissociative (psychogenic or non-epileptic) seizures.

Not all status is convulsive. In non-convulsive status epilepsy; the patient is in a continuous, distant, stuporous state. ***Focal status*** also occurs. ***Epilepsia partialis continua*** is continuous seizure activity in one part of the body, such as a finger or a limb, without loss of consciousness. This is often due to a cortical neoplasm or, in the elderly, a cortical infarct.

Delirium

Delirium, otherwise referred to as 'acute confusional state', is characterized by an acute and fluctuating disturbance in attention and cognition and represents a frequent presentation to the acute medical service. This transient mental state alteration can arise from various underlying causes and demands a comprehensive approach for accurate assessment and management.

Delirium can be considered as an acute or subacute 'brain failure', in which impairment of attention is accompanied by abnormalities of perception and mood (**Box 11.80**). During the acute phase of delirium, thought and speech are incoherent, memory is impaired, and misperceptions occur. Patients may experience episodic visual hallucinations (or illusions) or persecutory delusions occur and as a consequence, the patient may be frightened, suspicious, restless, and possibly uncooperative.

Delirium is often subclassified as 'hyperactive' (agitated, restless, wandering if physically able) or 'hypoactive' (withdrawn, lethargic, immobile) although in practice patients may fluctuate between psychomotor states. Agitation is usually worse at night, with consequent sleep reversal, so that the patient may be asleep in the day and awake all night.

Box 11.79 Status epilepticus: management

Early status (up to 30 min)—first-line treatments

- Administer oxygen, monitor ECG and blood pressure, perform routine blood tests (include glucose, calcium, drug screen, anticonvulsant levels urgently)
- Give lorazepam i.v. 4 mg bolus. Repeat once if necessary. Buccal midazolam is an alternative

Established status (30–90 min)—second-line treatments

- Levetiracetam: give at a loading dose of 60 mg/kg (maximum 4500 mg) infused over 15 min in 100 mls of sodium chloride 0.9% (often preferred as first-line in view of quicker administration and fewest adverse effects)

or

- Phenytoin: give 20 mg/kg i.v. (max. 2 g) diluted to 10 mg/mL in saline into a large vein at 25–50 mg/min (ECG monitoring required)

or

- Fosphenytoin: this is a prodrug of phenytoin and can be given faster than phenytoin. Doses are expressed in phenytoin equivalents (PE): fosphenytoin 1.5 mg = 1 mg phenytoin. Give 15 mg/kg (PE) fosphenytoin (15 mg × 1.5 = 26.5 mg) diluted to 10 mg/mL in saline at 50–100 mg (PE)/min

or

- Valproate: given at a loading dose of 40 mg/kg at a rate of 10 mg/kg per min (maximum dose 3000 mg)

If ongoing seizures:

- Consider using an alternative second-line treatment under expert guidance

or

- Phenobarbital 10 mg/kg i.v. diluted 1 in 10 in water for injection at <100 mg/min

Refractory status (>90 min)—general anaesthesia

- Only in intensive care setting; intubation and ventilation usually required
- Propofol 1–2 mg/kg bolus, followed by a continuous infusion of up to 4 mg/kg per hour for a minimum of 24 hours
- Thiopental and midazolam infusions may also be used
- Use continuous EEG monitoring to assess efficacy of treatment—aim for EEG burst suppression pattern
- Reinstate previous AED medication via nasogastric tube
- Establish diagnosis: CT or MRI may reveal an underlying cause
- Remember: 25% of apparent status cases turn out to be pseudostatus

Box 11.80 Delirium: diagnostic criteria

- Disturbance of consciousness:
  - ↓ Clarity of awareness of environment
  - ↓ Ability to focus, sustain, or shift attention
- Change in cognition:
  - Memory deficit, disorientation, language disturbance, perceptual disturbance
- Disturbance develops over a short period (hours or days)
- Fluctuation over the course of a day

(Derived from American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, 5th edn. Text revision [DSM-5]. Washington, DC: APA; 2013.)

The differential diagnosis of delirium is laid out in **Box 11.81**. Patients with pre-existing brain disease (e.g., dementia, cerebrovascular disease), sensory impairment, or advanced age are at increased risk of delirium during periods of intercurrent illness (**Box 11.82**).

Clinical assessment

When confronted with a patient manifesting delirium, a thorough and systematic evaluation is essential to establish the cause and implement appropriate investigations/interventions.

Box 11.81 Common causes of delirium

- Infection:
  - Any systemic infection, particularly with high fever (e.g., malaria, septicaemia)
  - Intracranial infection (e.g., encephalitis, cerebral abscess)
- Metabolic disturbance:
  - Hepatic failure
  - Chronic kidney disease (uraemia)
  - Disorders of electrolyte balance, dehydration
  - Hypoxia
- Vitamin deficiency:
  - Thiamine (Wernicke–Korsakoff's syndrome, beriberi)
  - Nicotinic acid (pellagra)
  - Vitamin B_{12}
- Endocrine disease:
  - Hypothyroidism
  - Cushing's syndrome
- Intracranial causes:
  - Trauma
  - Tumour
  - Subarachnoid haemorrhage
  - Epilepsy
- Drug intoxication:
  - Anticonvulsants
  - Antimuscarinics
  - Anxiolytics/hypnotics
  - Tricyclic antidepressants
  - Dopamine agonists
  - Digoxin
- Drug/alcohol withdrawal
- Postoperative states

Box 11.82 Predisposing factors in delirium

Endogenous

- Extremes of age (developing or deteriorating brain)
- Damaged brain:
  - Any dementia (most common predisposition)
  - Previous head injury
  - Alcohol brain damage
  - Previous stroke
- Visual or hearing impairment

Exogenous

- Dislocation to an unfamiliar environment (e.g., hospital admission)
- Sleep deprivation
- Sensory extremes (overload or deprivation)
- Immobilization

History

A **collateral history** (e.g., from a relative or bystander) is a key component of this process, both in establishing a change in affect or behaviour where delirium is subtle, and as the patient themself may be unable to provide a coherent account of their illness.

- **Onset and duration:** Crucial both for discerning potential causes and discriminating from chronic forms of confusion (e.g., dementia).
- **Precipitating factors:** A history of preceding fall or trauma should raise suspicion of an intracranial injury (e.g., SDH). A detailed social history should include assessment of both alcohol and recreational substance use which may cause confusion both in toxicity or withdrawal, the latter sometimes manifesting several days after hospitalization in patients admitted for other reasons. Environmental changes (hospitalization, unfamiliar

surroundings) may provoke delirium in the elderly or those with underlying dementia (acute-on-chronic confusional state).

- **Associated symptoms:** Explore for symptoms of the underlying cause, such as fever or other focal symptoms indicating infection or headaches or focal neurological symptoms, suggesting an intracranial disorder. Visual hallucinations are more common in delirium associated with alcohol and drug withdrawal.
- **Medication review:** Conduct a thorough review of the patient's medications, including prescription and over-the-counter drugs, to identify potential culprits contributing to delirium. Overdose (either intentional or inadvertent) should be considered as should prescription drug withdrawal states.
- **Past medical history:** Scrutinize past medical history for pre-existing conditions that may predispose the patient to delirium (e.g., post-ictal delirium, anticonvulsant toxicity, or non-convulsive status epileptics in a patient with epilepsy). Some patients may wear medical alert bracelets with important information which may be diagnostically helpful (e.g., history of Addison's disease). Patients with a preceding history of dementia or brain injury are at heightened risk of developing delirium.

Physical examination

- Fever may indicate infection (and should prompt consideration of encephalitis, in the appropriate context).
- Bedside tests will usually identify hypoxia and hypoglycaemia.
- Look for external signs of head injury.
- Examine for signs of CLD and check for asterixis (the flapping hand tremor which may be seen in both hepatic and uraemic encephalopathy and with CO_2 retention).
- Look for signs of meningeal irritation (e.g., neck stiffness) but remember that meningism is uncommon in encephalitis (e.g., due to HSV-1).
- If possible, examine the optic fundi for papilloedema (raised intracranial pressure [ICP]).
- Ophthalmoplegia and ataxia may suggest Wernicke encephalopathy while focal motor weakness (e.g., hemiparesis) may result from intracranial bleeds or mass lesions.

The presence of delirium may not be immediately evident when a patient presents to an acute medical admission setting. It is important to screen for delirium in patients who are susceptible. The 4 'A's Test **(4AT)** is a rapid and systematic tool designed for the assessment of delirium (**Box 11.83**). This screening instrument is particularly useful in busy clinical settings to promptly identify individuals at risk of or currently experiencing delirium. The 4AT incorporates four key domains, each beginning with the letter 'A', to facilitate easy recall and application.

Box 11.83 The 4 'A's' Test (4AT) for identification of delirium

Alertness	Normal (0) Mild sleepiness for less than 10 seconds after waking, then normal (0) Clearly abnormal (4)
Acute Abbreviated Mental Test Score (AMT4) *Ask the patient to state their age, their date of birth, the current year, and the name of the current location.*	No mistakes (0) 1 mistake (1) 2 or more mistakes, or untestable (2)
Altered Attention *Instruct the patient to list the months of the year in reverse order, starting with December.*	Lists 7 or more months correctly (0) Starts but lists less than 7, or refuses to start (1) Untestable (2)
Acute Onset or Fluctuating Course *Evidence of significant change or fluctuation in mental status within the last 2 weeks and still persisting in the last 24 hours*	No (0) Yes (4)

A total score of 4 or more suggests a high likelihood of delirium, necessitating further assessment and intervention.

Investigations

Investigations required will vary depending on the clinical context and findings on clinical assessment.

- **Laboratory:** First-line blood tests include full blood count, renal and liver profiles, electrolytes including serum calcium, and thyroid function tests.
- **Urinalysis:** While urinary tract infection may provoke delirium particularly in older adults, the high incidence of asymptomatic bacteriuria in the elderly means urinalysis has a poor PPV for this diagnosis in patients over 65 and should not be relied upon in isolation.
- **Imaging:** Brain imaging (CT or MRI) is warranted if focal neurological symptoms or signs are present, a structural brain abnormality is suspected or LP is planned. In practice, brain imaging is often requested early in the diagnostic pathway where the cause of delirium is not readily apparent.
- **Toxicology:** This may include blood or urine testing for substances of abuse or prescription drug levels (e.g., digoxin, antiepileptic drugs).
- **Further testing:** LP is usually warranted where meningitis or encephalitis is suspected. In cases where the cause remains elusive, EEG can aid in detecting abnormal brain wave patterns associated with delirium and discriminate between metabolic encephalopathy and non-convulsive status epilepticus.

Management

The management of delirium hinges on identifying and addressing the underlying causes. Assessment of mental capacity should be undertaken when making treatment decisions (in the UK, a Deprivation of Liberty or 'DoLS' application may be made to offer treatment in line with the guiding principles of the Mental Capacity Act in patients unable to consent).

Key management principles encompass:

- **Treatment of underlying causes:** Address the specific cause(s) of delirium, whether it be infection, metabolic derangements, or substance withdrawal.
- Patients with a history of alcohol dependency should receive empirical treatment with thiamine (as Pabrinex) to diminish risk of permanent neurological sequelae. In patients with suspected encephalitis, treatment with acyclovir should be commenced immediately and even before definitive investigation results are available.
- **Environmental modification:** Optimize the patient's environment by minimizing stimuli, ensuring adequate lighting, and promoting a consistent daily routine.
- **Medication review and adjustment:** Re-evaluate and adjust medications, particularly those with sedative or anticholinergic properties.
- **Multidisciplinary team:** Engage an MDT, including nursing, therapies, and specialty involvement as appropriate e.g., care of the elderly, psychiatry, neurology.
- **Reassurance and support:** Offer reassurance and support to the patient and their caregivers, emphasizing the reversible nature of delirium with appropriate interventions.

- **Psychoactive drugs** should be avoided if possible and reserved for instances where a delirious patient is considered at risk to themselves or others and verbal and non-verbal de-escalation techniques have failed to mitigate this. When needed, short-term haloperidol is usually preferred and benzodiazepines avoided.

The longer-term inpatient and outpatient management of delirium in the context of underlying dementia is discussed in Chapter 49).

Alcohol withdrawal syndrome

Acute alcohol withdrawal syndrome, sometimes refer to as delirium tremens (DTs), is a serious withdrawal state and occurs 1–3 days after alcohol cessation. Patients may present to hospital with a symptomatic withdrawal state but may also develop this 1–2 days after admission to hospital for other, unrelated reasons.

Patients become disorientated and agitated and typically have a marked tremor and may experience visual hallucinations (e.g., insects or small animals). Signs include sweating, tachycardia, tachypnoea, and pyrexia. Complications can include dehydration, seizures, infection, or the Wernicke–Korsakoff's syndrome (see Chapter 30). Patients may have other alcohol-related health issues (e.g., CLD), concomitant psychiatric disease, or a history of dependence on other substances and need a careful global assessment to ensure their holistic needs are met as best possible.

Recognizing a patient at risk of alcohol withdrawal

In patients who drink alcohol on a regular basis, symptoms of alcohol dependence syndrome should be sought at the point of admission (see **Box 30.31**). Screening instruments like AUDIT (Alcohol Use Disorders Identification Test) are useful to detect alcohol dependence (https://auditscreen.org/). High AUDIT scores point to dependence and should prompt proactive measures to act to prevent withdrawal in patients requiring hospitalization. Additional tools such as SADQ (Severity of Alcohol Dependence Questionnaire) or LDQ (Leeds Dependence Questionnaire) may be used to further stratify severity of dependence.

Patients may conceal alcohol use for many different reasons (including employment, cultural, or religious reasons) and thus an index of suspicion should be maintained for alcohol withdrawal even in the absence of a positive alcohol history in the appropriate circumstances.

Management

The principles of management of a patient with acute alcohol withdrawal are summarized in **Box 11.84**. Most UK hospitals have well-developed guidelines for identification of, monitoring, and management of patients with alcohol withdrawal.

Intravenous thiamine (as Pabrinex) should be given to all patients with acute alcohol withdrawal, a severe dependence on alcohol, or additional risk factors for Wernicke-Korsakoff's syndrome (e.g., malnutrition), and oral thiamine replacement continued subsequently. Wherever possible, thiamine should be given before use of i.v. glucose, but should not delay emergency treatment of severe or symptomatic hypoglycaemia.

The CIWA-Ar scale (Clinical Institute Withdrawal Assessment for Alcohol) is commonly used to monitor clinical progress during treatment of withdrawal, or to trigger symptom-driven treatment in patients at risk of developing a withdrawal state. Benzodiazepine requirements can vary widely from patient to patient, and careful monitoring and titration is required to prevent complications of either over-treatment (sedation) or under-treatment (seizures or severe DTs).

All patients hospitalized with alcohol withdrawal syndrome should be offered referral to outpatient substance abuse services (see Chapter 30 for further discussion).

Box 11.84 Management of alcohol withdrawal syndrome

General measures

- Admit the patient to a medical bed
- Correct electrolyte abnormalities (including magnesium) and dehydration
- Treat any co-morbid disorder (e.g., infection)
- Give parenteral thiamine as Pabrinex slowly (1 pair of ampoules = 250 mg thiamine daily for 3–5 days). *Note:* Beware small risk of anaphylaxis. A higher dose regimen is indicated if clinical features of Wernicke-Korsakoff syndrome are present (usually, 2–3 pairs thrice daily for 3–5 days, followed by 1 pair once daily for a further 3–5 days or for as long as improvement continues)
- Control seizures with a fast-acting benzodiazepine (either buccal midazolam or i.v. lorazepam)

Specific drug treatment

- Give one of the following orally:
  - Chlordiazepoxide 20–40 mg 6-hourly (depending on severity of symptoms; preferred due to lower toxicity risk and lesser potential for abuse)
  - Diazepam 10–20 mg 6-hourly (longer half-life, sometimes preferred in patients requiring very large doses of benzodiazepines)

In cases of severe liver impairment, specialist advice should be sought; lorazepam and oxazepam are favoured due to shorter half-life and reduced risk of accumulation.

Monitoring and continued treatment

Patients should be monitored for:

- Over-sedation (sedation scale)
- Adequacy of symptom control (CIWA-Ar)

Initial response should be assessed at 1 hour—additional PRN benzodiazepine should be available if symptoms and signs are not adequately controlled.

Subsequent treatment may be either via a fixed-schedule taper (usually gradually reducing dose over 5–7 days) or using a protocolized, symptom-triggered regimen based on regular CIWA-Ar scoring (not in patients with withdrawal seizures, and only where sufficient nursing expertise and staffing is available to ensure timely assessments and treatment).

Functional neurological disorder

Functional neurological disorder (FND) refers to a group of disorders mediated via the central nervous system but without evidence of physical neurological disease (see Chapter 30). The precise underlying mechanisms remain the subject of ongoing research. Although motor symptoms are most common, memory, concentration, cognition, and sensation can also be affected.

Early diagnosis and treatment can help lessen symptoms, aid appropriate management, and possibly minimize any stigmatization of sufferers, as well as potential harm from over-investigation or inappropriate management.

Clinical features

FND causes real symptoms that significantly interfere with daily life. Unplanned movements and symptoms can occur subconsciously and are distinct from symptoms that are intentionally produced. Symptoms may appear suddenly, and intermittently increase in severity, but also potentially decrease when the patient is distracted.

FND can be difficult for patients, family members, and healthcare workers to understand.

FND has two primary categories:

1. Dissociative seizures (sometimes referred to as psychogenic non-epileptic attack disorder)
2. Functional movement disorder

There are many types of FND, manifesting with a diverse range of neurologic symptoms.

Dissociative seizures may resemble generalized tonic–clonic seizures, or other forms of epileptic seizures, but are not caused by obvious abnormal electrical signalling in the brain. There may be episodes of abnormal movement, sensation, or behaviour (including temporary loss of attention or memory lapse) similar to an epileptic seizure. Dissociative seizures may occur in patients with proven epilepsy, so delineating an accurate history of the acute episode is always important.

In some instances, dissociative seizures may occur in relation to psychological stress or in correlation to current or past life events. The disorder is more common in females and often begins in young adulthood. In severe cases, attacks may be frequent or prolonged although with treatment, attacks may cease or decrease in frequency.

Functional movement disorder affects movement of the body. Symptoms may include:

- Limb weakness or paralysis
- Tremor
- Sudden, brief, involuntary twitching or jerking of a muscle or group of muscles
- Involuntary muscle contractions that cause slow repetitive movements or abnormal postures (dystonia)
- Problems with walking motion (gait), posture, or balance
- Spasms and contractures (in which the tendons become fixed in awkward or uncomfortable positions)
- Muscle stiffness
- Tics

FND may manifest as symptoms relating to **other brain functions** that may include:

- Speech difficulties, such as sudden onset of stuttering or trouble speaking
- Problems with seeing or hearing
- Pain (including migraine)
- Extreme slowness and fatigue
- Numbness or perceived reduced sensation

Clinical assessment

Risk factors for FND include:

- Factors in the patient's personal history (such as childhood trauma, anxiety, being witness to violence, maltreatment, or sexual abuse).
- Sociological factors (including interpersonal relationships and stress).
- Psychological factors: FND may sometimes overlap with somatic symptoms disorder, typically with other multisystem symptoms resulting in distress to the patient.

A detailed medical and family history and physical examination is required to rule out organic neurological or systemic disease that may provide an alternative explanation for symptoms, or may coexist with FND neurology experts, preferably with experience of FND, should assist with timely accurate diagnosis and management.

Investigations

No single test can confirm a diagnosis of FND. Tests, if indicated, may include brain imaging, primarily to exclude other disorders. Other tests such as electromyography (EMG) and EEG can help identify a biological movement disorder. Video EEG can aid with diagnosis of dissociative seizures.

Management

There are no specific treatments for FND but interventions may be directed at some of its symptoms. Clear, empathic, and patient-centred communication about medically unexplained or functional symptoms is crucial in helping patients manage their condition, discussed further in Chapter 30. The patient's social support network should ideally also be involved to help understand FND and to support the patient in dealing with symptoms, treatment, and any stigma associated with the condition.

- **Medications:** Medications may be effective in treating associated pain, anxiety, depression, or insomnia. Unless a patient also suffers from epileptic seizures there is no role for anticonvulsant medications in the treatment of dissociative seizures.
- **Psychotherapy:** Cognitive behavioural therapy (CBT) can help modify thought patterns to address emotional or behavioural factors contributing to FND. Relaxation therapy or mindfulness activities may help alleviate psychological stress.
- **Physical, speech, and occupational therapy:** Therapy input can help with rehabilitation of muscular weakness or impaired movement and is a particularly important component of therapy in instances where symptoms are prolonged or repetitive.

Polysymptomatology

With improved healthcare, patients are living longer, often with one or more chronic health conditions and frequently presenting with multiple simultaneous symptoms or problems to the acute medical take. While the initial patient assessment should include a brief description of all active symptoms, an initial 'symptom triage' may be required to separate problems that require immediate attention (severe, progressive, or potentially life-threatening) from those more effectively dealt with at a later stage.

Formulation of a 'problem list' as part of the medical clerking, dividing problems into 'acute', 'chronic' or 'resolved' and indicating the proposed course of action for each, can help maintain an orderly and structured approach to a hospital admission and assist with criterion-based discharge. Importantly, the symptom deemed most impactful by a patient may not concord with that identified as a concern by the referrer or acute clinician and an initial exploration of a patient's ideas, concerns, and expectations can help with 'expectation setting' with regard to the agenda of an acute hospital admission. In an ideal medical system, every encounter would serve as an opportunity to provide holistic patient care, however, with limitations in time, resources, and capacity this might not always be attainable.

Swollen joint

The painful swollen joint is a common clinical presentation to acute medical services. A solitary hot, swollen, and tender joint (monoarthritis) should always raise concern for infection (septic arthritis) which may lead to rapid destruction of the joint and lasting disability unless diagnosed and treated promptly. Other potential causes include acute crystal arthropathies (gout or pseudogout), haemarthrosis (e.g., following trauma, or spontaneously in patients with a history of bleeding disorder or anticoagulation), or reactive arthritis. Systemic disorders (e.g., rheumatoid arthritis, seronegative spondyloarthropathies, SLE) more typically associated with multiple joint involvement (polyarthritis) may occasionally present acutely as a monoarthritis. Periarticular inflammation

(e.g., bursitis, enthesitis, or cellulitis) may mimic arthritis and be considered during the clinical assessment. The common causes of monoarthritis with key accompanying clinical features are outlined in Box 11.85.

Clinical assessment

History

- Take a description of the **onset and time course** of symptoms (truly acute, or following a longer history of pain, swelling, or stiffness in the joint?)
- Enquire regarding **prior episodes** and/or diagnoses (e.g., have they required a diagnostic joint aspiration in the past?)
- Describe the **pattern of joint involvement** in patients with multi-joint symptoms, including the presence of back or neck pain
- **Fever or chills** is a cardinal sign of infection but may be absent in the elderly or immunosuppressed patient with a septic joint
- **Prosthetic joints** are at higher risk for infection and diagnosis and management usually requires specialist orthopaedic involvement
- **Systemic anticoagulation** or a history of **trauma** should prompt consideration of haemarthrosis
- **Systems enquiry** should include seeking a history of recent GI symptoms (prompting consideration of reactive arthritis or IBD) or genitourinary symptoms (reactive arthritis or gonorrhoea), together with relevant travel and sexual history

Physical examination

- Assess for signs of sepsis
- Examine the affected joint: the inflamed joint (e.g., infection or crystal arthritis) will feel warm compared to the opposite side and a joint effusion usually present
- Other joints should be inspected for the presence of swelling and deformity pointing towards systemic disorders, including the joints of the feet and spine
- Skin should be inspected for the presence of rashes (e.g., erythema nodosum in sarcoidosis, pustular lesions in gonorrhoea), nail changes (e.g., pitting in psoriasis), rheumatoid nodules over extensor surfaces or tendons, or gouty tophi
- Examine the cardiovascular system for signs of endocarditis or more rarely, rheumatic fever
- Conjunctivitis should raise suspicion of reactive arthritis

Investigations

- **Laboratory investigations** should include:
  - Full blood count e.g., neutrophil leucocytosis in septic arthritis or crystal arthritis.
  - Basic metabolic tests including urea and electrolytes, liver function tests, and bone profile.
  - Random glucose may identify occult hyperglycaemia, a risk factor for infection.
  - Coagulation screen is required where joint aspiration is being considered, or if coagulopathy is suspected.
  - CRP is typically elevated in inflammatory arthritis of any cause, including septic and crystal arthritis
  - **Blood cultures** are essential if local or systemic infection is suspected.
  - Urate levels can be sent but may be normal or even low in the acute phases of gout.
  - Where appropriate, additional tests for systemic disorders (e.g., rheumatoid factor/anti-CCP antibody if suspecting rheumatoid arthritis, anti-nuclear antibodies, anti-dsDNA, and serum complement levels if suspecting SLE).
- **Plain radiographs** can be useful to rule out fractures or avascular necrosis and may identify chronic changes (e.g., those typical of osteo- or rheumatoid arthritis) but may only show soft tissue swelling in the acute phases of many arthritides. Look for the presence of chondrocalcinosis (often present in joints affected by pseudogout).
- **Joint aspiration** (Fig. 32.2) is key in diagnosis of septic or crystal-induced arthritis. Samples should be sent for cell count and Gram stain, microbiological culture testing and cytology, including examination under polarized light. Synovial fluid aspirated from an infected joint may appear turbid or purulent to the naked eye. In the absence of another identifiable cause, a high fluid WBC (neutrophils) remains suspicious for infection, even if cultures are negative; of note, only ~50% of joint aspirates in patients with gonococcal arthritis are culture positive. Consider synovial PCR or ribosomal 16S testing if culture negative and high index of suspicion for infection.
- **Other microbiological samples including** throat, cervical, rectal, or skin lesion swabs may be of use (e.g., haematogenous spread of infection from bacterial pharyngitis, urethritis triggering reactive arthritis, gonococcal arthritis).

Septic arthritis

Septic arthritis of native (non-prosthetic) joints occurs via direct injury or by blood-borne infection from an infected skin lesion or other site, with approximately 2 cases per 100 000 people per year. It is a medical emergency that requires prompt treatment to prevent irreversible joint destruction and has a significant mortality of up to 11%, reaching 50% in polyarticular sepsis.

The organism that most commonly causes septic arthritis is *Staphylococcus aureus*. Other organisms include streptococci, other species of staphylococcus, *Neisseria gonorrhoeae*, *Haemophilus influenzae* in children, and these and other Gram-negative organisms in the elderly or complicating RA.

Clinical features

In young and previously fit people, the joint is hot, red, swollen, and agonizingly painful; it is held immobile by muscle spasm. In contrast, the onset may be insidious with a lack of systemic symptoms in the elderly, the immunosuppressed, and patients with RA, and a high index of suspicion is needed. In 20% of patients, the infection affects more than one joint. Recent aspiration or steroid injection of the joint may lead to clinical suspicion of infection.

Investigations

- ***Aspirate*** the joint and send the fluid for urgent Gram-staining and culture. The fluid is usually frankly purulent. The culture techniques should include those for gonococci and anaerobes.
- ***Blood cultures*** are often positive.
- ***Leucocytosis*** is usual, unless the person is severely immunosuppressed.
- ***X-rays*** are of no value in diagnosis in acute septic arthritis, but may show other pathologies such as chondrocalcinosis in pseudogout.
- ***Skin wound swabs, sputum and throat swab, or urine*** may be positive and indicate the source of infection.

Management

There are no RCTs to guide management in adults. The prognosis is similar for patients with a proven microbiological diagnosis to those with a suspected but not proven diagnosis; therefore, it is vital to treat on the basis of clinical suspicion, probability (with subsequent

Box 11.85 Acute monoarthritis (and mimics): common causes and clinical features

Disease	History	Examination	Investigation	Management
Septic arthritis	Risk states include extremes of age, immunocompromise including diabetes and alcohol excess), pre-existing joint disease including prostheses, history of i.v. drug use, or recent intra-articular injection	Often large joints (knee, hips, wrist, ankle, and shoulder) Look for signs of endocarditis Pustular skin rash common with gonococcal infection	Blood cultures are essential Throat, cervical, and rectal swabs in selected cases Joint aspiration is required for definitive diagnosis	Systemic antibiotics Purulent arthritis usually requires joint washout
Gout	Rare in premenopausal females Risk factors include alcohol consumption, use of diuretics, and a prior history of gout	Multiple joints may be involved simultaneously, usually lower limb 1st MTP joint ('podagra') is classic site Gouty tophi if chronic	Raised inflammatory markers (WBC/CRP) Serum urate may be normal in acute flare X-ray may reveal erosions Joint aspiration—needle-shaped negatively birefringent urate crystals	NSAIDs, colchicine, steroids Offer lifestyle and dietary advice, review medications (e.g., diuretics)
Pseudogout (acute calcium pyrophosphate deposition disease)	Often older (>55), coexists with osteoarthritis	Multiple joints may be involved simultaneously, upper or lower limb joints Rarely syndrome of acute or recurrent neck and shoulder pain ('Crowned dens syndrome')	Raised inflammatory markers (WBC/CRP) X-rays may reveal chondrocalcinosis CT will identify Crowned dens syndrome Weakly positively birefringent crystals may be seen on microscopy	Consider NSAIDs, steroids Check ferritin (haemochromatosis) and thyroid function
Trauma	History of trauma may not be recalled if patient lost consciousness (e.g., seizure)	Signs of intra-articular trauma (e.g., meniscal tear, ligamentous rupture) may be difficult to detect if a large joint effusion is present	X-ray may reveal a peri-articular fracture MRI is usually required to delineate intra-articular injury	Acute joint effusion usually responds to conservative management Orthopaedic referral required if suspicion of damage to structures within the joint
Haemarthrosis	Recent trauma, bleeding diatheses, or anticoagulation	Bruising may be evident	Check coagulation and platelets	Consider clotting factors or reversal of anticoagulation
Reactive arthritis	Recent infection, usually diarrhoea or urethritis	Mono-, oligo-, or polyarthritis, usually asymmetrical. Knee commonly involved and low back pain (SIJ) common. Look for extra-articular signs: conjunctivitis, perianal or genital ulceration, keratoderma blennorrhagica (soles and palms)	Throat, cervical, and rectal swabs may be of use in identifying underlying infection	Usually self-limiting. NSAIDs first-line for symptomatic treatment although corticosteroids may be prescribed in severe or polyarticular forms. Genitourinary infection should be treated with appropriate antibiotics
Osteoarthritis	Older patients, pain worse with activity, relieved by rest	Single/multi-joint and generalized forms Gradual onset—an acute arthritis in an osteoarthritis joint should raise concern for infection or crystal arthritis PIP and DIP joints involved in generalized forms (Bouchard's and Heberden's nodes)	Inflammatory markers usually normal Radiology may show loss of cartilage, subchondral sclerosis, osteophytes, and subarticular cysts	Simple analgesia and exercise are mainstay of treatment

Box 11.85 Acute monoarthritis (and mimics): common causes and clinical features—cont'd

Disease	History	Examination	Investigation	Management
Systemic disorders	Rheumatoid arthritis (RhA), seronegative spondyloarthritides, sarcoidosis, and SLE may all rarely present with acute monoarthritis Chronic infections e.g., tuberculosis, Lyme disease, may occasionally simulate an acute, septic joint Serum sickness may complicate immune therapies (e.g., anti-thymocyte globulin) Rarely mono- or polyarthritis may be paraneoplastic	Multiple joint involvement more common Low back pain (SIJ) and enthesitis common in seronegative arthritis Look for rheumatoid nodules (RhA), erythema nodosum (sarcoidosis, TB), facial butterfly rash (SLE) Serum sickness/ immune complex arthritis often favours temporomandibular joint	Tailored to presentation but may include Rheumatoid factor/ anti-CCP Ab testing (RhA), anti-nuclear antibody, anti-dsDNA antibody and serum complement testing (SLE), and serum ACE (sarcoidosis) Joint aspirates should be sent for mycobacterial and fungal cultures in atypical or subacute presentations	Directed towards underlying cause
Bursitis	Often related to trauma or overuse injuries	Common sites include olecranon process and pre-patellar bursa Secondary infection may occur ('septic bursitis')	Diagnosis often clinical although ultrasound or MRI may be useful Blood cultures and bursal aspiration if concern of septic bursitis	Antibiotics towards causative agent if septic bursitis
Cellulitis	May mimic arthritis if skin overlies joint	Entry point may be evident Absence of joint effusion	Avoid aspirating a joint with overlying cellulitis (risk of bacterial innoculation)	Antibiotics towards causative agent
Tendonitis	Injury or repetitive exercise Quinolone antibiotic use may predispose	Quinolones typically cause Achilles tendon inflammation	Ultrasound or MRI may be of use in diagnosis	Rest, ice packs, and support. Analgesia and physiotherapy may be needed

DIP, Distal interphalangeal; *MTP,* metatarsophalangeal; *PIP,* proximal interphalangeal; *SIJ,* sacro-iliac joint; *SLE,* systemic lupus erythematosus.

confirmed identity) of the organism, local pattern of antibiotic sensitivity, and microbiological advice.

Therapy should be started immediately before culture results are available because joint destruction may occur rapidly.

- The joint should be aspirated and then immobilized initially, followed by early physiotherapy to prevent stiffness and muscle wasting.
- Intravenous antibiotics should be given for 2 weeks and then oral antibiotics for a further 4 weeks. Response is monitored clinically and with ESR or CRP.

Empirical treatment in septic arthritis

Empirical regimens vary according to local advice and index of suspicion. For suspected Gram-positive infection first-line therapy is vancomycin, with clindamycin or cephalosporin second-line. For suspected Gram-negative infection first-line therapy is a third-generation cephalosporin with i.v. ciprofloxacin second-line. In confirmed methicillin-sensitive *Staphylococcus aureus* (MSSA) primary options include flucloxacillin alone or combined with fusidic acid or gentamicin, while penicillin-allergic patients should receive vancomycin with clindamycin. In cases of methicillin-resistant *Staphylococcus aureus* (MRSA), vancomycin is the drug of choice. Antibiotics should be altered according to sensitivities. Drainage of the joint and/or arthroscopic joint washouts are required in all cases.

Management of infected prostheses

Long-term goals of prosthetic joint infection are to eradicate the infection, reduce pain, and restore function. Two-stage surgery is required to remove the prosthesis followed by treatment with appropriate i.v. antibiotics for 4–6 weeks before a new prosthesis is inserted. A 2–6-week course of pathogen-directed i.v. antibiotics followed by 3–6 months of oral antibiotics are also given.

The unwell diabetic patient

Patients with diabetes may present to acute medical services with an emergency directly related to their diabetes (e.g., hyperglycaemia, diabetic ketoacidosis (DKA), hyperosmolar hyperglycaemic state (HHS), hypoglycaemia, a foot emergency, continuous subcutaneous insulin infusion (CSII) pump failure) or may be admitted for many other emergencies where diabetes is a secondary diagnosis (e.g., sepsis, stroke, trauma). Patients presenting to emergency services may also have an elevated blood glucose and no previous diagnosis of diabetes; a proportion of these patients will have unrecognized diabetes, while some will have stress hyperglycaemia. This section will cover:

- Diabetic ketoacidosis
- Hyperosmolar hyperglycaemic state
- Hyperglycaemia
- Hypoglycaemia

Terms used with reference to the acutely unwell diabetic patient are defined in **Box 11.86**.

Diabetic ketoacidosis

Diabetic ketoacidosis (DKA) remains a frequent and life-threatening complication of diabetes, with a mortality rate of approximately 1%.

Box 11.86 Terms used in diabetes-related emergencies

Term	Definition
Ketonuria	Detectable ketones in the urine; it should be appreciated that ketonuria occurs after fasting in people without diabetes and may be found in people with relatively well-controlled type 1 diabetes
Ketosis	Elevated plasma ketone levels in the absence of acidosis
Diabetic ketoacidosis	A metabolic emergency in which hyperglycaemia is associated with a metabolic acidosis due to greatly raised (>3 mmol/L) ketone levels
Hyperosmolar hyperglycaemic state	A metabolic emergency in which uncontrolled hyperglycaemia induces a hyperosmolar state in the absence of significant ketosis
Lactic acidosis	A metabolic emergency in which elevated lactic acid levels induce a metabolic acidosis. In people with diabetes, it is rare and most commonly associated with biguanide therapy

Population studies suggest that there are 5–8 episodes annually per 1000 people with diabetes. It occurs more frequently in younger people, but the mortality is higher in older people. Although treatment has improved with better outcomes, the incidence is rising. As DKA results from marked insulin deficiency, it mostly occurs in people with type 1 diabetes but may occasionally present in people with type 2 diabetes. It is usually seen in the following circumstances:

- previously undiagnosed diabetes (10%–20% of cases)
- interruption of insulin therapy (15%–30% of cases)
- the stress of intercurrent illness and infection (30%–40%)
- sodium-glucose co-transporter-2 (SGLT2) inhibitor-related DKA.

Many cases of DKA can be prevented by earlier diagnoses, better communication between the person with diabetes and their healthcare professional, and better self-management education. A common error is to reduce or omit insulin because the individual feels unable to eat, owing to nausea or vomiting. This is a factor in at least 25% of all hospital admissions. Although insulin adjustment is necessary, it should never be stopped. The UK National Diabetes Inpatient Audit found that many people developed DKA while in hospital, suggesting that healthcare professionals also make the same mistakes.

Pathogenesis

Ketoacidosis is a state of uncontrolled catabolism associated with marked insulin deficiency and elevated counter-regulatory hormones, which accelerate the effects of insulin deficiency (Fig. 11.41). The insulin deficiency leads to hyperglycaemia secondary to increased hepatic glucose output and diminished insulin-mediated peripheral glucose uptake, which in turn causes an osmotic diuresis and profound dehydration and loss of electrolytes (Fig. 11.42). However, the most important biochemical abnormality in DKA is not the hyperglycaemia but the uncontrolled lipolysis in adipose tissue and uncontrolled ketogenesis in the liver. Marked insulin deficiency is a necessary precondition for DKA since very little insulin is needed to inhibit hepatic ketogenesis and the breakdown of adipose triglycerides to non-esterified fatty acids (NEFAs). NEFAs are transported to the liver where they are partially oxidized to acidic ketone bodies, such as acetoacetic acid and 3-hydroxybutyric acid, and acetone. This occurs because the absence of insulin impairs hepatic

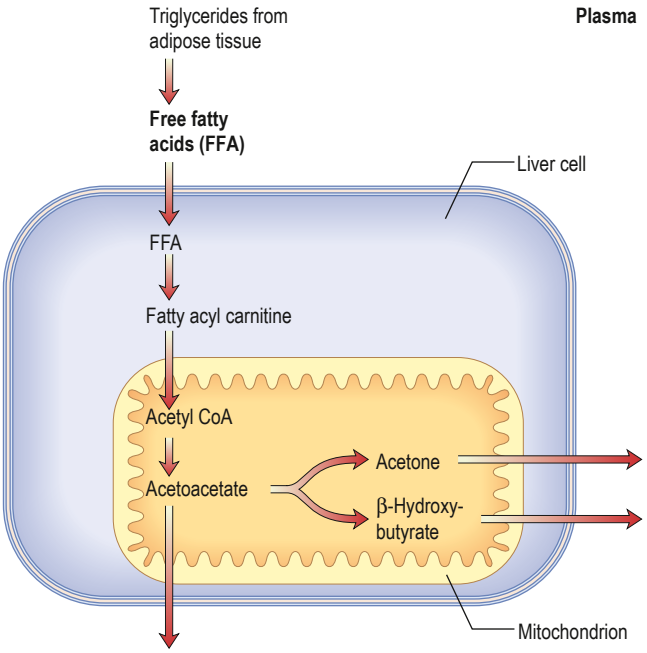


Fig. 11.41 Ketogenesis. During insulin deficiency, lipolysis accelerates and free fatty acids taken up by liver cells form the substrate for ketone formation (acetoacetate, acetone, and β-hydroxybutyrate) within the mitochondrion. These ketones pass into the blood, producing acidosis. *CoA,* Coenzyme A.

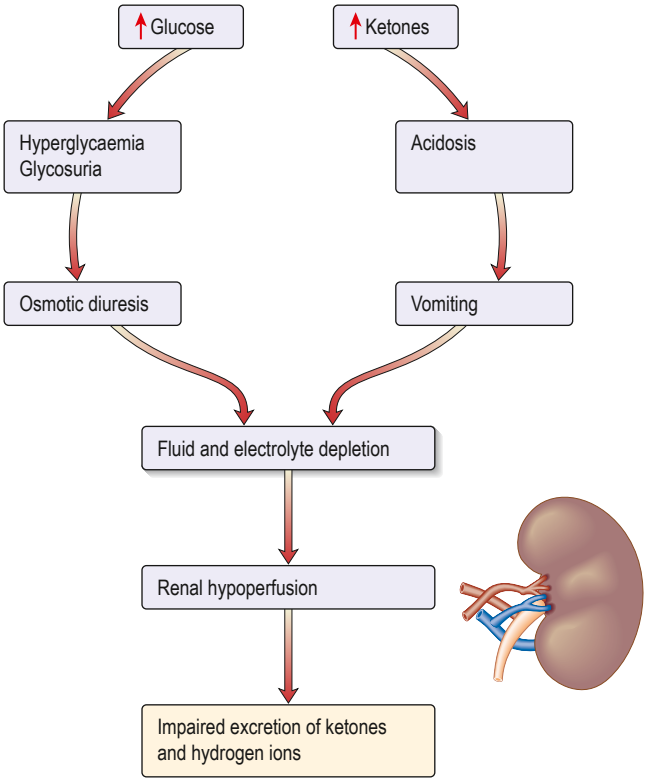


Fig. 11.42 Dehydration occurs during ketoacidosis as a consequence of two parallel processes. Hyperglycaemia results in osmotic diuresis, and hyperketonaemia results in acidosis and vomiting. Renal hypoperfusion then occurs and a vicious circle is established as the kidney becomes less able to compensate for the acidosis.

re-esterification of NEFA to triglyceride. The liver exports the ketone bodies as an alternative fuel supply, but these build up in the circulation because of impaired uptake into peripheral tissues, such as muscle. The excess ketones are excreted in the urine but also appear in the breath, producing a distinctive smell like that of acetone. Ketone bodies are strong organic acids that rapidly exceed the body's buffering capacity and cause severe metabolic acidosis. Ketone bodies are nauseating, and many people will vomit, worsening the dehydration and electrolyte loss further. Plasma osmolality rises and renal perfusion falls, which impairs renal excretion of hydrogen ions and ketones, aggravating the acidosis. Respiratory compensation for the acidosis leads to hyperventilation, graphically described as 'air hunger'.

As the pH falls below 7.0 ($[H^+]$ >100 nmol/L), pH-dependent enzyme systems in many cells function less effectively and, untreated, severe ketoacidosis is invariably fatal.

Clinical assessment

Patients presenting with DKA often have a variety of signs and symptoms on presentation and usually these develop rapidly within several hours of the precipitating event.

History

Symptoms of hyperglycaemia preceding presentation are most common, including polyuria, polydipsia, and unintentional weight loss. Other common symptoms of DKA include:

- Nausea and vomiting
- Abdominal pain—this is sometimes so severe that it can be confused with a surgical acute abdomen
- Weakness and fatigue
- Complaining of 'brain fog' or the inability to think straight.

Physical examination

Clinical examination signs include:

- Clinical features of dehydration: dry mucous membranes, reduced skin turgor, hypotension and tachycardia, reduced urine output, and in severe cases, features of shock
- Lethargy, confusion, and reduced GCS (up to 5% present in coma)
- Smell of ketones on breath (not all people are able to detect this odour)
- Hypothermia
- Increased respiratory rate as a result of acidosis, often referred to as Kussmaul breathing (rapid and deep respiratory effort in attempt to compensate for metabolic acidosis).

Investigations and diagnosis

- **Point-of-care bedside testing** including capillary blood glucose, blood ketones, and blood gas analysis. The **diagnosis of DKA** is confirmed by demonstrating hyperglycaemia with ketonaemia or heavy ketonuria, and acidosis as follows:
  - Ketonaemia ≥3 mmol/L (31 mg/dL) *or* significant ketonuria (≥2 on standard urine sticks)
  - Blood glucose >11 mmol/L (200 mg/dL) *or* known diabetes mellitus
  - Bicarbonate (HCO_3^-) below 15 mmol/L (15 mEq/L) and/or venous pH <7.3.
- **Laboratory investigations** including renal function and electrolytes in particular potassium. Acidosis causes hyperkalaemia as potassium is exchanged for hydrogen ions moving into the cell. As insulin promotes the co-transport of potassium along with glucose into cells, insulinopaenia further exacerbates hyperkalaemia. Although serum potassium may be elevated at presentation, there is a severe whole-body potassium *deficiency* as significant quantities of potassium are lost in vomit and urine and serum potassium often falls rapidly after initiation of treatment. Inflammatory markers such as CRP are also important.
- **Septic screen including** blood cultures, urine culture, and CXR if features suspicious of infection. Respiratory viral testing (including COVID-19, which has been associated with an increased risk of DKA) should be considered
- **12-lead ECG** if tachycardic, concern of arrhythmia (either related to infection or shock), significant electrolyte imbalance, or suspicion for cardiac ischaemia.

The indicators of severe DKA are shown in **Box 11.87**.

Management

Management is summarized in **Box 11.88**. If the patient has severe DKA (**Box 11.87**), they should be managed in an enhanced care area or high dependency unit.

Replacement of the fluid losses

Individuals with DKA have a typical water deficit of 100 mL/kg, equivalent to 7.5 L in a 75 kg adult. The replacement fluid of choice is 0.9% sodium chloride with additional potassium chloride (potassium cannot be added to Hartmann's solution). The aim of the first few litres of fluid is to correct any hypotension, replenish the intravascular deficit, and counteract the effects of the osmotic diuresis with correction of the electrolyte disturbance. Over-rapid fluid replacement may lead to cerebral oedema in children and young adults and so the remainder of the fluid should be replaced more cautiously over the next 2 days. The rate and volume of fluid replacement need to be modified in older people and in those with renal or heart failure.

Replacement of the electrolyte losses

Typical electrolyte deficits are sodium 7–10 mmol/kg, chloride 3–5 mmol/kg, and potassium 3–5 mmol/kg. After the initiation of treatment with insulin, potassium levels can fall rapidly as insulin promotes potassium uptake by the cells. Commercially, 0.9% sodium chloride is available with premixed potassium chloride (40 mmol/L (0.3%)) allowing the potassium to be replaced safely. Hypokalaemia and hyperkalaemia are potentially life-threatening conditions during the management of DKA and so potassium levels should be monitored frequently. Patients are frequently hyperkalaemic at presentation and so potassium is typically not prescribed with the initial fluid resuscitation (1 L 0.9% NaCl infused over 1 hour), but proactive replacement is subsequently required to prevent hypokalaemia and KCl supplementation only withheld if serum potassium level remains above 5.5 mmol/L.

Box 11.87 Indicators of severe diabetic ketoacidosis

- Blood ketones over 6 mmol/L
- Bicarbonate level below 5 mmol/L
- Venous/arterial pH below 7.0
- Hypokalaemia on admission (under 3.5 mmol/L)
- Glasgow Coma Scale less than 12 or abnormal AVPU scale
- Oxygen saturation below 92% on air (assuming normal baseline respiratory function)
- Systolic BP below 90 mmHg
- Pulse over 100 or below 60 beats/min
- Anion gap above 16 [anion gap = $(Na^+ + K^+) - (Cl^- + HCO_3^-)$]

Box 11.88 Management of diabetic ketoacidosis (DKA)

Immediate measures (0–60 min after diagnosis and initiation of intravenous fluids)

1. Carry out clinical assessment to determine the severity of the DKA episode
2. Perform initial investigations, including blood ketones, capillary and venous glucose, serum creatinine and electrolytes, venous blood gases, full blood count, cultures, ECG, CXR if indicated, and cardiac monitoring
3. Initiate the intravenous infusion with 0.9% sodium chloride
   a. Give 1 L over the first hour unless there is hypotension requiring more rapid correction of fluid deficit
   b. Serum potassium is often elevated at presentation and thus potassium is not added to the initial 1 L of 0.9% saline. Serum potassium will subsequently fall rapidly with rehydration and insulin treatment and needs proactive replacement with subsequent fluids
4. Commence a fixed-rate intravenous insulin infusion at a rate of 0.1 unit/kg per h using soluble human insulin. Estimate the weight if necessary
5. Continue the long-acting basal insulin at the usual dose and time

Management from 60 min to 6 h

1. Reassess the patient
   a. Consider urinary catheterization if the patient is incontinent or anuric after 2–4 h
   b. Consider nasogastric tube insertion if the patient has a reduced GCS or is persistently vomiting
2. Continue intravenous fluids
   a. 0.9% sodium chloride 1 L with potassium chloride over 2 h
   b. 0.9% sodium chloride 1 L with potassium chloride over 2 h
   c. 0.9% sodium chloride 1 L with potassium chloride over 4 h
   d. 0.9% sodium chloride 1 L with potassium chloride over 4 h
   e. 0.9% sodium chloride 1 L with potassium chloride over 6 h

   40 mmol KCl is typically added to each litre of fluid if serum potassium is in the range of 3.5–5.5 mmol/L. No KCl should be added if serum potassium is >5.5 mmol/L, and senior advice sought if serum potassium is <3.5 mmol/L.
3. Measure blood ketones and capillary glucose hourly
   a. If blood ketones are not falling by at least 0.5 mmol/L per h, increase the insulin infusion rate by 1.0 unit per h increments hourly until the ketones are falling at target rates. Seek senior review
   b. An alternative to blood ketones is venous bicarbonate which should rise by at least 3.0 mmol/L per h
   c. A further alternative is glucose that should fall by at least 3.0 mmol/L per hour (50 mg/dL)
   d. If the glucose falls below 14.0 mmol/L (250 mg/dL), commence 10% glucose given at 125 mL per h alongside the 0.9% sodium chloride solution
4. Measure venous blood gas for pH, bicarbonate and potassium at 60 min and hourly thereafter
   a. Adjust the potassium replacement accordingly
5. Identify and treat precipitating factors

Management from 6 to 12 h

1. Reassess the patient and review biochemical and metabolic measurement to check for resolution of DKA

Conversion to subcutaneous insulin

- Convert the patient to an appropriate subcutaneous regimen when biochemically stable (blood ketones less than 0.6 mmol/L, pH over 7.3) and ready and able to eat.

Hyperchloraemic acidosis may develop during treatment since a large variety of negatively charged electrolytes are lost in DKA, which are replaced with chloride. The kidneys usually correct this spontaneously within a few days.

Restoration of the acid–base balance

In most cases, once the circulating volume is restored, the metabolic acidosis will rapidly compensate. Supplemental i.v. bicarbonate is seldom indicated because it is implicated in the development of cerebral oedema and may cause a paradoxical increase in CSF acidosis. It is used only if the pH is <7.0 ($[H^+]$ >100 nmol/L) and is best given as an isotonic (1.26%) solution.

Insulin replacement

Insulin should be given by an intravenous infusion using a fixed rate 0.1 units/kg per hour (typically 6–10 units per h). This modest insulin dose is sufficient to suppress ketogenesis, lower the glucose, and correct the electrolyte disturbance. For those already treated with insulin, the long-acting basal subcutaneous insulin should be continued.

As the intravenous insulin should be continued until the ketosis has resolved, intravenous glucose alongside the 0.9% sodium chloride is often needed to prevent hypoglycaemia; 10% glucose is recommended once the blood glucose falls below 14.0 mmol/L (250 mg/dL) and should be continued until the person is eating and drinking normally.

Monitoring progress

The aim of treatment is lower the blood ketone concentration by 0.5 mmol/L per hour, increase the venous bicarbonate by 3.0 mmol/L per hour, and reduce capillary blood glucose by 3.0 mmol/L per hour (50 mg/dL per h) while maintaining potassium between 4.0 and 5.5 mmol/L. If these targets are not met, the insulin infusion rate should be increased.

Seeking the underlying cause

Physical examination may reveal a source of infection. Two common markers of infection are misleading: fever is unusual, even when infection is present; and polymorphic leucocytosis is present, even in the absence of infection as a result of demargination of circulating neutrophils due to metabolic acidosis. Relevant investigations include a CXR, urine and blood cultures, and an ECG (to exclude myocardial infarction). If infection is suspected, broad-spectrum antibiotics are started once the appropriate cultures have been taken.

Other measures

As patients are significantly dehydrated on admission, it is unlikely that they will pass urine for several hours after the initiation of fluid replacement. However, if no urine is passed within 2–4 hours of admission, the insertion of a urinary catheter should be considered.

A central venous cannula may be required to monitor fluid balance in older people and those with cardiac disease.

Aspiration of vomit may be fatal in people with an impaired conscious level and so a nasogastric tube should be considered.

The dehydration, increased blood viscosity, and coagulability of DKA increase the risk of a potentially fatal thromboembolism and so thromboprophylaxis with LMWH should be considered in older or high-risk individuals unless contraindicated.

Complications of diabetic ketoacidosis

Cerebral oedema is a rare but serious complication with high mortality and mostly happens in children or young adults. Severe hypothermia with a core temperature of more than 33°C may occur and can be overlooked unless a rectal temperature is taken with a low-reading thermometer.

Ongoing management

Once the person is able to eat and keep food down, subcutaneous insulin treatment should be resumed or initiated. A subcutaneous

bolus insulin is given with the meal and for those whose long-acting basal insulin was continued during the DKA episode, the insulin infusion and fluids can be discontinued 30 minutes after the meal. If the long-acting insulin was stopped, the intravenous insulin infusion and fluids should be continued until some form of background insulin has been given.

The inpatient stay provides an educational opportunity to assess why the DKA occurred and what steps are needed to prevent it from happening again. Any admission with DKA should prompt review of current insulin therapy, and consideration of whether an alternative therapy system such as a continuous glucose monitor or closed-loop insulin pump system may be appropriate (Chapter 35).

Sodium-glucose co-transporter-2 (SGLT2) inhibitor-related DKA

SGLT2 inhibitors (including dapagliflozin, empagliflozin, and canagliflozin) are a relatively new class of drugs for the treatment of type 2 diabetes and act by inhibition of renal glucose reabsorption (see Chapter 35). They are prescribed with increasing frequency due to a generally favourable side-effect profile, ease of administration, mild associated weight loss, and additional benefits for renal and cardiac health above and beyond their effect on serum glucose.

SGLT2 inhibitors are associated rarely with DKA, usually in the context of a normal blood glucose ('euglycaemic DKA'). DKA likely results from insulinopaenia (due, in turn, to suppression of insulin secretion as a result of induction of glycosuria) and risk is highest in patients with reduced pancreatic β cell mass (e.g., patients with type 2 diabetes and low C-peptide levels, or unrecognized latent autoimmune diabetes) or in clinical scenarios where insulin counter-hormone secretion is high driving lipolysis and ketogenesis. (e.g., infection or surgery). Other contributors may include increased renal ketone body resorption (driven by an increased concentration of sodium ions within renal tubules) and the diuretic effect of SGLT2 inhibitors, reducing overall blood volume and promoting ketoacidosis. Patients prescribed SGLT2 inhibitors should be counselled about 'sick day rules' and treatment interrupted in any patient hospitalized for major surgery or with a serious acute illness. SGLT2 inhibitors should be borne in mind as a risk-factor for euglycaemic DKA, and the clinician should have a low threshold to test plasma ketones and blood gases in an unwell patient who has continued to take this medication up until presentation.

Further reading

Buse JB, Wexler DJ, Tsapas A, et al. (2020) 2019 update to: Management of hyperglycemia in type 2 diabetes. A consensus report by the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). *Diabetes Care* 43(2), 487–493.

Dhatariya K, Corsino L, Umpierrez GE. Management of diabetes and hyperglycemia in hospitalized patients. [Updated 2020 Dec 30]. In: **Feingold KR, Anawalt B, Blackman MR, et al.**, editors. *Endotext [Internet]*. South Dartmouth (MA): MDText.com, Inc.; 2000. Available from: https://www.ncbi.nlm.nih.gov/books/NBK279093/.

The Management of Diabetic Ketoacidosis in Adults. Joint British Diabetes Societies Inpatient Care Group. Guideline. Revised March 2023. https://abcd.care/joint-british-diabetes-societies-jbds-inpatient-care-group.

Hyperosmolar hyperglycaemic state

Hyperosmolar hyperglycaemic state (HHS) is a medical emergency typically seen in patients with type 2 diabetes, characterized by severe hyperglycaemia (>30 mmol/L), marked dehydration, and elevated serum osmolality (usually >320 mOsm/kg). Unlike DKA, ketonaemia and metabolic acidosis are usually absent. People affected are often in middle or later life, with poorly controlled or previously undiagnosed diabetes. Common precipitating factors include infection (most commonly chest or urine), consumption of glucose-rich fluids, or new prescription medication (e.g., corticosteroids).

Box 11.89 Comparison of the biochemical changes in diabetic ketoacidosis and the hyperosmolar hyperglycaemic state

Examples of blood values	Severe ketoacidosis	Hyperosmolar hyperglycaemic state
Na^+ (mmol/L)	140	155
K^+ (mmol/L)	5	5
Cl^- (mmol/L)	100	110
HCO_3^- (mmol/L)	5	25
Urea (mmol/L)	8	15
Glucose (mmol/L)	30	50
Arterial pH	7.0[a]	7.35

HHS

Patients become severely hypernatraemic as a result of free water losses and inadequate water intake (contrast with uncomplicated hyperglycaemia, where the serum sodium is typically low as a result of osmotic water movement out of the intracellular compartment). The normal range of osmolality is 285–300 mOsm/kg. It can be measured directly, or can be calculated approximately from the formula $2(Na^+ + K^+)$ + glucose + urea. Serum osmolality is typically >320 mOsm/Kg in HHS (in this example 385 mosm/kg). Ketonaemia is not a feature and hence arterial pH and venous HCO_3^- are both within normal limits.

DKA

DKA is defined by an arterial pH of <7.30, although patients are often severely acidotic with very low serum bicarbonate. Anion gap can be calculated by $(Na^+ + K^+) - (Cl^- + HCO_3^-)$ and is elevated in DKA, due to ketone bodies acting as an unmeasured anion (in this instance anion gap is 40, compared to a normal upper limit of 16). Anion gap falls progressively with treatment, although resolution of acidosis and normalization of serum bicarbonate may 'lag behind' as a result of increasing serum chloride resulting from 0.9% NaCl infusion (which induces a normal anion gap metabolic acidosis).

[a]$[H^+]$ >100 nmol/L.

The cardinal biochemical differences between HHS and DKA are summarized in **Box 11.89**. The biochemical differences may partly be explained as follows:

- ***Age.*** Older people (the typical demographic developing HHS) experience thirst less acutely and become dehydrated more readily. In addition, the mild renal impairment associated with age results in increased urinary losses of fluid and electrolytes. Onset of HHS is also more gradual, usually over a few days, allowing for the development of more profound dehydration.
- ***The degree of insulin deficiency.*** This is less severe in the hyperosmolar hyperglycaemic state. Endogenous insulin levels are sufficient to inhibit hepatic ketogenesis but insufficient to inhibit hepatic glucose production.

Overlap presentations of HHS and DKA may sometimes occur, and it is important to test blood ketones in all patients at presentation.

Pathogenesis of HHS

- HHS is often triggered by an *acute stressor*, which increases levels of cortisol and catecholamines (thereby reducing insulin sensitivity).
- HHS occurs in patients with *enough* insulin to prevent ketoacidosis, but *not* enough insulin to control hyperglycaemia.
  - Higher endogenous levels of insulin are required to control hyperglycaemia, compared to the amount required to prevent ketogenesis (as seen in the way that many patients in the ICU

develop mild insulin resistance and *hyperglycaemia—without* developing ketoacidosis).
- Uncontrolled hyperglycaemia causes an *osmotic diuresis*, with loss of water.
- Patients *fail* to compensate adequately for water loss by increasing oral water intake (e.g., due to baseline debility, bed-bound status, or a relatively insensitive central drive to maintain normal tonicity).
- Over a period of several days, uncontrolled water loss leads to a hypertonic state and progressive hypernatraemia. This may lead to altered mental status (which exacerbates the patient's inability to drink an adequate amount of water).

Clinical assessment

The characteristic clinical features of HHS are severe dehydration, and stupor or coma. Impaired consciousness is directly related to the degree of hyperosmolality. Evidence of underlying illness, such as pneumonia or pyelonephritis, may be present, and the hyperosmolar state may predispose to stroke, myocardial infarction, or lower limb arterial insufficiency.

Investigations

Core investigations are as for DKA (see previous section). In HHS, the plasma osmolality is usually extremely high (>320 mOsm/kg). It can be measured directly or calculated as $(2(Na^+ + K^+)$+glucose+urea), all in mmol/L. Plasma osmolality should be calculated hourly during the early stages of treatment to ensure that over-rapid changes in osmolality are avoided which may otherwise rarely provoke neurological complications including cerebral oedema or central pontine myelinolysis.

Management

The most important aspect of management in HHS is **fluid replacement**, with the intent of correcting dehydration and gradually reducing blood glucose and serum osmolality. **Serum osmolality should be calculated hourly** during treatment, with the goal of fluid therapy being to achieve a sufficiently positive fluid balance to result in a fall in osmolality between 3.0 and 8.0 mOsm/kg/hour. Despite the majority of patients being hypernatraemic, 0.9% sodium chloride is the treatment of choice, and 0.45% sodium chloride should only be considered if the osmolality is not declining despite adequate positive fluid balance (usual target is to achieve a positive balance of ~2–3 L over the initial 6 hours of treatment). Serum sodium may be observed to rise initially with treatment, as a result of falling plasma glucose levels and consequent redistribution of water into the intracellular compartment; subsequently the rate of fall of plasma sodium should not exceed 10 mmol/L in 24 hours to reduce the risk of neurological complications.

Fluid replacement alone is often sufficient to lower the glucose and in the majority of patients insulin should *not* be started as part of initial therapy. Many patients are extremely sensitive to insulin and the glucose concentration may fall rapidly with initiation, potentially provoking haemodynamic instability or over-rapid changes in osmolality. Insulin (0.05 units/kg per h) may be started once adequate fluid replacement has been provided if the glucose is no longer falling with fluids alone or if the patient develops significant ketonaemia (in which case the diagnosis of HHS should be reconsidered). In order to reduce risk of cerebral damage, the fall in blood glucose should be no more than 5 mmol/L per hour (90 mg/dL per h). Patients fulfilling criteria for concomitant diabetic ketoacidosis (significant ketonaemia >3 mmol/L with pH < 7.3) require immediate insulin and should be treated in line with DKA guidance.

Serum potassium requires careful monitoring and often falls during treatment; local protocols for fluid infusion in HHS usually include protocolized potassium replacement.

Prophylactic LMWH should be given. Regular clinical assessment of fluid balance is vital given the frequent coexistence of cardiovascular disease and heart failure. Foot checks and pressure care (including heel off-loading) are also priority tasks.

While it may seem daunting to balance changes in serum osmolality (aiming for a fall between 3.0 and 8.0 mOsm/kg/h), serum sodium (limiting hourly fall to <10 mmol) and plasma glucose (limiting hourly fall to <5 mmol), in reality all three parameters closely co-vary during treatment and these criteria can usually be satisfied simultaneously with an attentive approach to monitoring and following the principles laid out above.

Prognosis

The reported mortality ranges as high as 15%–20%, mainly because of the more advanced age of the patients and the frequency of intercurrent illness. Unlike ketoacidosis, the hyperosmolar hyperglycaemic state is not an absolute indication for subsequent insulin therapy, and survivors may do well on diet and oral agents.

Metformin associated lactic acidosis

The risk of lactic acidosis in people taking metformin is extremely low, provided that the therapeutic dose is not exceeded and the drug is withheld in people with advanced hepatic or renal dysfunction. Current guidance is to discontinue metformin in patients with an estimated GFR < 30 mL/min/1.73 m^2. Severe lactic acidosis usually occurs in scenarios where patients have experienced an intercurrent illness (e.g., diarrhoea or vomiting), leading to dehydration and AKI and have continued to take metformin as usual during this period, leading to drug accumulation. Systemic upset is common at presentation (potentially also contributing to lactic acidosis) and AKI may also aggravate acidosis. Patients often have a severe metabolic acidosis with a high anion gap, usually without significant hyperglycaemia or ketosis. Treatment is by rehydration, discontinuation of metformin, and infusion of isotonic 1.26% bicarbonate. Critical care admission is often required, with haemofiltration if there is severe coexistent AKI. The mortality is in excess of 50%.

Further reading

The Management of The Hyperosmolar Hyperglycaemic State (HHS) in Adults With Diabetes. Joint British Diabetes Societies Inpatient Care Group. Guideline. 2022.
Accessed at: https://abcd.care/resource/jbds-06-management-hyperosmolar-hyperglycaemic-state-hhs-adults-diabetes.
https://cks.nice.org.uk/topics/diabetes-type-2/diagnosis/when-to-suspect-hyperglycaemic-emergencies-dka-hhs/.

Hyperglycaemia

Hyperglycaemia alone without the additional adverse features of DKA or HHS is often seen in patients in the ED or admitted to the AMU, either as a result of long-standing poor glycaemic control (as evidenced by HbA_{1c}) or stress counter-hormone secretion secondary to acute illness.

Patients with type 1 diabetes should be instructed **not to omit** their basal (long-acting) insulin, or ketosis will surely develop.

Plasma ketones should be checked in all patients with blood glucose >10 mmol/L (regardless of diabetes type, although risk of ketosis is substantially higher in type 1 diabetics). Those with

elevated ketones (>1.5 mmol/L) often warrant i.v. fluids (0.9% NaCl) and require more formal testing to exclude DKA.

In patients without ketosis, a blood glucose >13 mmol/L (type 1 diabetes) or >18 mmol/L (type 2 diabetes) is commonly regarded as the threshold where further immediate intervention is required. The most appropriate response will vary depending on whether the patient has a known diagnosis of diabetes (and if so, what type), recent control, whether the patient is already on insulin treatment and whether the patient otherwise requires admission to the hospital or may be suitable to discharge.

NHS England (London) guidance created for AMUs during the COVID-19 pandemic advocates the use of a stat dose of fast-acting insulin (e.g., Novorapid, dose adjusted based on body weight and prior insulin dose) paired with a bolus of i.v. 0.9% NaCl to achieve prompt glycaemic control, and enable a continuation plan to be devised based on the patient's individual characteristics and in conjunction with the local diabetes team. Stat insulin doses should not be repeated within 2 hours and variable rate insulin infusion may need to be started if multiple stat doses are required.

The approach to management for patients with newly diagnosed type 2 diabetes is discussed in Chapter 35).

Further reading

NHS London Clinical Networks, Guideline on Management of Diabetes in Emergency Departments During the Coronavirus Pandemic, 2020. Accessed at: https://www.england.nhs.uk/london/wp-content/uploads/sites/8/2020/04/Covid-19-Management-of-diabetes-in-emergency-department-crib-sheet-updated-150420.pdf.

Hypoglycaemia

Hypoglycaemia (low blood glucose) is the most common side-effect of insulin therapy and is the major limitation to what can be achieved with insulin treatment. A glucose concentration below 3.0 mmol/L (54 mg/dL) is defined as 'clinically significant hypoglycaemia' but it is recognized that any value below 4.0 mmol/L (70 mg/dL) represents an 'alert value' that requires treatment. 'Severe hypoglycaemia' is defined as hypoglycaemia requiring external help for recovery. Most people treated with insulin will experience several episodes of symptomatic hypoglycaemia per week and one to two severe episodes per year. Interestingly with the advent of continuous blood glucose monitoring, it is apparent that asymptomatic hypoglycaemia occurs much more frequently than previously thought. Hypoglycaemia is more common in people with type 1 diabetes, where there is absolute insulin deficiency, and with longer duration of diabetes. It occurs more frequently in young children and those trying to achieve tight glycaemic control.

Hypoglycaemia greatly impairs quality of life and induces fear and anxiety in the person with diabetes, their family, and carers. In an older study, the fear of severe hypoglycaemia was worse than the fear of developing blindness or end-stage renal disease. Nocturnal hypoglycaemia can be particularly disruptive. In addition to the psychological effects, hypoglycaemia is associated with a number of long-term medical consequences, including cardiovascular events and falls (**Box 11.90**).

Hypoglycaemia is caused when more insulin is injected than is needed. Irregular eating habits, unusual exertion, and alcohol excess may all precipitate hypoglycaemic episodes but insulin errors and variation in insulin absorption, for example as a result of lipohypertrophy, are also important. The times of greatest risk are before meals, during the night, and during and after exercise.

Box 11.90 Some of the consequences of hypoglycaemia

Acute symptoms
- Autonomic
  - Sweating, paraesthesia, feeling hot, shakiness, anxiety, palpitations and pallor
- Neuroglycopenia
  - Slurring of speech, altered behaviour, loss of concentration, drowsiness, low mood, dizziness, hemiplegia, fits, coma, and rarely, death
  - Loss of warning with recurrent hypoglycaemia and long duration of diabetes

Medical
- Increased risk of falls
- Increased risk of myocardial infarction
  - Increased thrombosis
  - Atherosclerotic plaque instability
  - Cardiac arrhythmias
- Increased risk of hospitalization
- Increased risk of sudden death

Psychological
- Poor quality of life
- Embarrassment
- Fear of hypoglycaemia
  - Reduced medication taking to avoid future episodes
  - Higher HbA_{1c}

Financial
- Inability to work effectively
- Increased out-of-pocket expenses
- Increased medical costs
  - Extra visits
  - More glucose testing

The mechanisms that protect against hypoglycaemia in people without diabetes are compromised because of abnormal islet cell function or defective counter-regulatory hormone secretion. The delivery of insulin after injection cannot be halted and glucagon secretion is impaired because its main regulator is a fall of intra-islet insulin. Even the third-line defence of counter-regulatory hormones and other sympathetic nervous system responses to hypoglycaemia may become diminished in people with longstanding diabetes or recurrent hypoglycaemia.

Clinical assessment

The symptoms of hypoglycaemia can be divided into two main categories, ***autonomic*** and ***neuroglycopenic***, although non-specific symptoms may also be experienced (see **Box 11.90**). Autonomic symptoms and signs occur because of activation of the adrenergic and cholinergic parts of the autonomic nervous system. As the brain is dependent on glucose as its primary fuel source, neuroglycopenic symptoms arise from inadequate glucose supply to the brain.

Usually, autonomic symptoms develop before cognitive impairment. This is clinically important as these symptoms alert the individual to hypoglycaemia and prompt them to take corrective action. In longstanding diabetes or after repeated episodes of hypoglycaemia, the autonomic symptoms may only develop at glucose levels below the threshold for neuroglycopenic symptoms. When this occurs, hypoglycaemia awareness becomes impaired and the person is at increased risk of severe hypoglycaemia. Awareness improves if hypoglycaemia is scrupulously avoided.

The symptoms and signs of hypoglycaemia resolve once the hypoglycaemia is treated and the glucose returns to within the normal range.

Management

Hypoglycaemia is an acute medical emergency that requires immediate treatment with 15–20 g of oral glucose. This should be repeated after 15 minutes if the glucose concentration has not risen above 4.0 mmol/L (70 mg/dL). The glucose can be taken either as a liquid or solid food containing fast-acting carbohydrate, such as glucose tablets or sugary sweets, for example, Jelly Babies. More carbohydrate than necessary should not be taken in order to prevent rebound hyperglycaemia.

The diagnosis of severe hypoglycaemia resulting in confusion or coma is simple and usually made on clinical grounds, backed by a bedside blood test. Many people with diabetes carry a card or wear a bracelet or necklace saying that they have diabetes, and these should be sought in unconscious patients. The treatment is either intramuscular glucagon or intravenous glucose. Glucagon acts by mobilizing hepatic glycogen and works almost as rapidly as glucose. It is simple to administer and can be given at home by relatives. It does not work when liver glycogen levels are low, for example, after a prolonged fast.

Once the individual revives, they should eat some longer-acting carbohydrate to replenish glycogen reserves (e.g., two biscuits or a slice of bread). After the acute event has been treated, it is important to ascertain why the hypoglycaemia occurred to try to prevent future episodes.

Deep vein thrombosis and pulmonary embolism

Definitions

Most often, venous thrombosis originates in the deep veins of the legs, hence the term ***deep vein thrombosis (DVT)***. It is thought that the process starts within the pocket of one of the valves that line the veins, where flow may be turbulent and localized hypoxia may develop, resulting in endothelial dysfunction. The thrombus may remain localized to the leg veins or may embolize through the circulation to result in a **pulmonary embolism (PE)**. A minority (about 10%) of episodes of venous thrombosis arise in other sites, such as the upper limb, the cerebral venous sinuses, and the splanchnic veins (hepatic, portal, and mesenteric veins). Apart from upper limb venous thrombosis, these ***unusual-site thromboses*** are described further in relevant specialty-specific chapters (see 'Cortical venous thrombosis' further on in this chapter as well as Chapters 23 and 26). Immunogenic thrombosis *in situ* is well recognized in some conditions, where the pulmonary clot has not embolized but developed locally in the lungs. This is relatively common in inflammatory pulmonary conditions including bronchopneumonia and COVID-19 infection.

If confined to the calf veins, the thrombus is called a ***calf, distal, or below-knee DVT***. Untreated, the thrombus may extend proximally and, when it reaches the popliteal vein or above, is called a ***proximal, or above-knee DVT***. Thrombi at this level are larger, and more likely to embolize and be transported with blood flow through the large veins of the pelvis and abdomen to the right atrium and ventricle. From there, they are pumped into the pulmonary arteries, which progressively divide into smaller arteries as they course through the lungs to supply the alveoli. The emboli stop in the pulmonary arteries, where they are no longer physically able to progress, and, in so doing, obstruct the flow of blood distally. It is thought to take at least several days for venous thrombi to become clinically evident.

Box 11.91 Risk factors for venous thromboembolism

Transient

- Surgery, especially major, lower limb/pelvis or cancer related
- Trauma, especially lower limb/pelvis
- Active cancer
- Acute medical admission
- Immobilization (bed rest >3 days)
- Plaster cast
- Pregnancy/puerperium
- Oestrogen administration (combined hormonal contraception, oral hormone therapy)
- Recent long-haul travel (>4 h)
- Central venous catheter
- Heparin-induced thrombocytopenia
- Superficial vein thrombosis

Persistent

- Increasing age
- Body mass index >30 kg/m^2
- Ethnicity
- Previous episode of venous thromboembolism
- Inflammatory conditions e.g., inflammatory bowel disease, systemic lupus erythematosus, Behçet's syndrome
- Nephrotic syndrome
- Lower limb paresis e.g., after stroke
- Heritable thrombophilia (factor V Leiden, prothrombin gene mutation, deficiencies of antithrombin, protein C or protein S)
- Antiphospholipid syndrome
- Myeloproliferative neoplasms

Epidemiology and risk factors

The annual incidence of VTE within the community is estimated at around 1 per 1000 per year overall and is age dependent. VTE is uncommon in childhood, affects around 1 per 10 000 young adults per year and 1 per 1000 middle-aged adults per year, and approaches an incidence of 1 in 100 elderly adults per year. Historical autopsy data suggest that about 10% of patients who died in hospital did so from PE. Over 5% of people will develop one or more episodes of VTE during their lifetime. Clinically, about 60% of episodes of VTE present with DVT and the other 40% with PE, with or without accompanying features of DVT.

Some conditions may predispose to immunothrombosis *in situ* in the pulmonary vasculature such as Behçet's disease and pneumonias or pneumonitis including COVID-19 infection. Steroids may be indicated in their therapy, with or without anticoagulation therapy (without usually in the case of Behçet's).

Other risk factors for VTE are shown in **Box 11.91**. They can be divided into transient provoking factors (such as major surgery) and longer-term predisposing factors (such as obesity or heritable thrombophilias).

- ***Stasis-related factors*** include surgery, bed rest, plaster cast immobilization, pregnancy, and long-haul travel.
- ***Hypercoagulability*** may be caused by cancer, surgery, pregnancy, and administration of oestrogens in combined hormonal contraception and hormone replacement, as well as the heritable thrombophilias.
- ***Vessel wall factors*** including previous thrombosis and invading malignancy.

Risk factors differ in the degree to which they increase the likelihood of VTE occurring:

- ***Strong risk factors*** increase the risk ten- to fiftyfold and include major surgery, trauma, and absolute bed rest.

Box 11.92 Differential diagnosis of venous thromboembolism

Deep vein thrombosis

- Ruptured Baker's cyst
- Musculo-tendinous—trauma, haematoma, myositis, tendonitis
- Superficial vein thrombosis
- Post-thrombotic syndrome
- Cellulitis
- Osteoarthritis, osteomyelitis, synovitis, fracture, tumour
- Acute arterial occlusion
- Lymphoedema
- Congestive cardiac failure and hypoalbuminaemia—usually cause bilateral leg swelling

Pulmonary embolism

- Chest infection/pneumonia
- Exacerbation of chronic obstructive pulmonary disease
- Asthma
- Pneumothorax
- Congestive cardiac failure
- Acute coronary syndrome
- Costochondritis
- Musculoskeletal pain or rib fracture
- Aortic dissection
- Pericardial tamponade
- Lung cancer
- Primary pulmonary hypertension
- Anxiety/hyperventilation

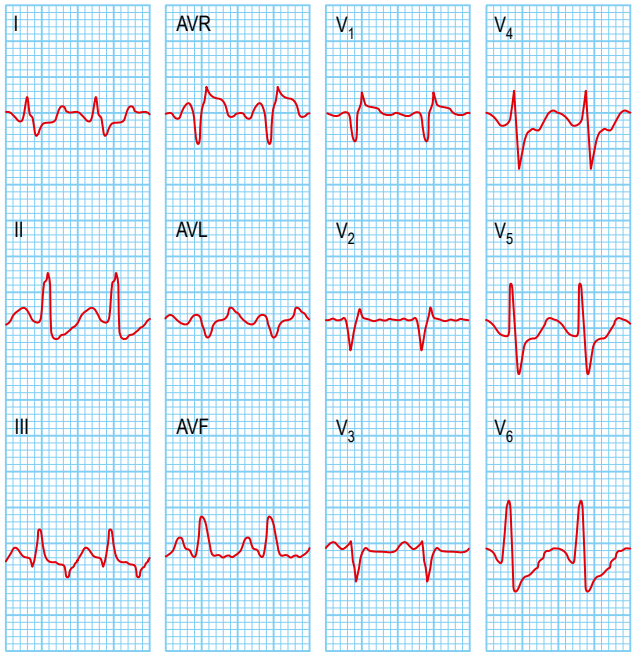


Fig. 11.43 Acute pulmonary embolism shown on a 12-lead ECG. There is an S wave in lead I, a Q wave in lead III, and an inverted T wave in lead III (the S1, Q3, T3 pattern). There is sinus tachycardia (160 beats/min) and an incomplete right bundle branch block pattern (an R wave in AVR and V_1 and an S wave in V_6).

- ***Moderate risk factors*** increase the risk three- to tenfold and include pregnancy, oestrogen therapy, and minor surgery under general anaesthesia. Most heritable thrombophilias also increase the risk three- to tenfold.
- ***Weak risk factors*** increase the risk up to threefold and include obesity and long-haul travel. For reasons that are not yet evident, the incidence of VTE is highest in people of African descent, intermediate in white people and lower in, for example, East Asian populations.

Clinical assessment

The differential diagnosis of suspected DVT or PE is very wide (Box 11.92), and most patients (>80%) who present with suspected VTE have the diagnosis excluded rather than confirmed. The clinical features are often non-specific and the majority of people who die from PE probably do so because earlier warning signs were missed rather than because of sudden collapse and death or failure of treatment. The challenge in the diagnostic work-up of suspected DVT and PE is therefore to identify rapidly and accurately, when urgent treatment is required, and to distinguish these patients from those who do not have the condition, in whom unnecessary investigation and treatment would be potentially harmful.

The typical features of ***DVT*** include pain and swelling in one leg. The leg may be red and warm to the touch. There may also be tenderness along the course of the deep veins, especially in the popliteal fossa, and dilation of the collateral superficial veins.

PE presents:

- In about 65% of cases, with pleuritic chest pain and breathlessness, sometimes but rarely accompanied by haemoptysis. Tachypnoea and tachycardia are typically present. Crackles and a pleural rub over a localized area of pulmonary infarction may be evident on auscultation.
- In another 25%, with isolated breathlessness, sometimes only evident on exertion.
- In the remaining 10%, with more severe features, including syncopal episodes, systolic hypotension or shock, and myocardial ischaemia with associated central chest pain. With this more severe presentation, the patient is tachypnoeic, and has a tachycardia with peripheral shutdown, a raised jugular venous pulse (JVP) with a prominent α-wave, right ventricular heave, gallop rhythm and a widely split second heart sound. Cardiac arrest may occur, typically with pulseless electrical activity. The severity of presentation depends on both the thrombus burden and the individual's cardiopulmonary reserve.

At least one-third of patients presenting with DVT have clinically silent pulmonary emboli. About 70% of people presenting with symptomatic PE have an associated DVT that is symptomatic in about a quarter of cases.

Investigations

The initial investigations that might be undertaken because of chest pain or breathlessness are often non-diagnostic in patients with PE:

- ***ECG*** (Fig. 11.43): A right ventricular strain pattern may be seen, with T wave inversion in the inferior (II, III, AVF) and right precordial (V1–V4) leads. The classical ECG is of an 'S1Q3T3' pattern, with a prominent S wave in lead I and a prominent Q wave and inverted T wave in lead III, but this is present in only a minority of patients. It is more common to see a sinus tachycardia or atrial fibrillation.
- ***CXR:*** This may be normal but more often shows non-specific abnormalities, including atelectasis, parenchymal abnormalities, cardiomegaly, elevation of the hemidiaphragm, or a small pleural effusion. Occasionally, there may be a lack of vessel markings distal to a large thrombosis, known as Westermark sign.

- ***Arterial blood gas analysis:*** Blood gases typically show hypoxia and hypocapnia, but again these are non-specific findings and are not always present.
- ***Biomarkers of cardiac injury:*** Plasma levels of brain natriuretic peptide (BNP) or its precursor, amino-terminal pro-BNP, may be elevated because of stretching of the right ventricle, and troponin may be increased because of right ventricular injury due to strain, but once again these findings are non-specific. They are useful indicators of severity and may guide therapeutic options including whether to admit the patient, or consider advanced therapies.
- ***POCUS/point-of-care echocardiography:*** Focused ultrasound can help diagnose DVTs (loss of colour flow, and loss of compressibility of deep veins). Standard point-of-care echocardiography (using a phased array (cardiac) probe, or a curvilinear (abdominal) probe utilizing the liver as an acoustic window for a subcostal view, may show signs of right ventricular (RV) strain including flattening of the intraventricular septum, or even RV bulging in to the left ventricular cavity (which should have a higher pressure). The acquisition of these images may help guide diagnosis and clinical assessment, but recognition of possible chronic pulmonary hypertension is important to avoid cognitive bias.

Diagnosis of DVT and PE

For patients in the UK, other than those with shock or hypotension due to suspected PE, the approach to the diagnosis of VTE recommended by the National Institute for Health and Care Excellence (NICE) is generally used.

Risk scoring

The diagnostic process starts with assessment of the clinical probability using a clinical prediction score that takes account of the individual patient's clinical features, the presence or absence of risk factors for VTE, and whether an alternative diagnosis is likely to explain the symptoms and signs. There are a number of clinical prediction scores available, and NICE recommends the modified two-level ***Wells score*** for DVT (Box 11.93) and the two-level **Wells score** for PE (Box 11.94); these categorize patients into groups that are likely or unlikely to have DVT or PE, respectively. Clinical prediction scores standardize clinical assessment and increase reproducibility among less experienced clinicians. However, in themselves, they do not confirm or exclude the diagnosis of VTE and further assessment is necessary, although experienced clinicians may use the pulmonary embolism rule-out criteria (the PERC rule) to help determine whether any further investigations for PE are needed in patients in whom the clinical suspicion of PE is low (estimated pretest likelihood <15%) (Box 11.95).

Measurement of D-dimer

In those considered unlikely to have VTE based on the Wells score, the next step is D-dimer testing. D-dimer is a fibrin degradation product that can be measured in plasma or whole blood by using highly sensitive, fully quantitative laboratory point-of-care tests. Raised levels indicate activation of the coagulation system but are not specific for VTE, as they are also seen, for example, with advanced age, infection, inflammation, postoperatively, in cancer and during pregnancy. As D-dimer can rise with age, some centres use an age-adjusted D-dimer test threshold for people aged over 50.

The value of D-dimer in VTE diagnosis is based on its high negative predictive value: that is, VTE is very unlikely in a patient who has a low pretest probability of DVT or PE by the Wells score and in whom D-dimer, measured by a sensitive assay, falls below a predefined cut-off (typically quoted as 'normal'). Such patients do not need further diagnostic testing for VTE and alternative diagnoses should be considered. With this approach, it must be borne in mind that there is a small failure rate (<2% within 3 months) and that it does not absolutely exclude VTE. The signs and symptoms of DVT and PE, and where to seek further medical help, should be discussed with the patient.

In contrast, patients whose Wells score indicates that VTE is unlikely but whose D-dimer is raised (lies above the cut-off value) require formal radiological imaging to confirm or exclude VTE.

Box 11.93 Wells score for deep vein thrombosis (DVT) (two-level)

Clinical feature	Points
Active cancer (treatment ongoing, within 6 months, or palliative)	1
Paralysis, paresis, or recent plaster mobilization of the lower extremities	1
Recently bedridden for ≥3 days or major surgery within 12 weeks requiring general or regional anaesthesia	1
Localized tenderness along the distribution of the deep venous system	1
Entire leg swollen	1
Calf swelling at least 3 cm larger than asymptomatic side	1
Pitting oedema confined to the symptomatic leg	1
Collateral superficial veins (non-varicose)	1
Previously documented DVT	1
Alternative diagnosis at least as likely as DVT	–2
Clinical probability simplified score	
DVT *likely*	≥2
DVT *unlikely*	≤1

(From Wells PS, Anderson DR, Rodger M et al. Evaluation of D-dimer in the diagnosis of suspected deep-vein thrombosis. *N Engl J Med* 2003; 349:1227–1235.)

Box 11.94 Wells score for pulmonary embolism (PE) (two-level)

Clinical feature	Points
Clinical signs and symptoms of deep vein thrombosis (DVT) (minimum of leg swelling and pain with palpation of the deep veins)	3
Alternative diagnosis less likely than PE	3
Heart rate >100 beats/min	1.5
Immobilization (for >3 days) or surgery in the previous 4 weeks	1.5
Previous DVT/PE	1.5
Haemoptysis	1
Malignancy (on treatment, treated in the last 6 months, or palliative)	1
Clinical probability simplified score	
PE *likely*	>4
PE *unlikely*	≤4

(From Wells PS, Anderson DR, Rodger M et al. Derivation of a simple clinical model to categorise patients probability of pulmonary embolism: increasing the model's utility with the SimpliRED D-dimer. *Thromb Haemost* 2000; 83:416–420.)

Box 11.95 The Pulmonary Embolism Rule-out Criteria (PERC) rule

If any items are present in a patient with a low likelihood of PE, the PERC test is positive and further assessment is required

- Age ≥50 years
- Heart rate ≥100 beats/min
- Oxygen saturation <95% while patient is breathing room air
- Swelling in one leg
- Haemoptysis
- Surgery or trauma within past 4 weeks
- Previous DVT or PE
- Hormone use

Similarly, imaging is also required for all patients whose Wells score indicates that VTE is likely. There is no evidence to support the use of D-dimer in pregnancy (see Chapter 48).

Imaging for DVT

The diagnosis of DVT is generally confirmed by ***ultrasonography*** of the deep venous system, which has long replaced the historical gold standard of venography, as it is quicker and non-invasive. At a minimum, ultrasonography involves examination of the proximal venous system. Typically, this includes compression of the popliteal and femoral veins by the ultrasound probe to determine whether the vein is compressible or not (see Fig. 8.33). This is usually supplemented by direct thrombus imaging with vein enlargement and assessment of blood flow by colour Doppler. Ultrasonography is very sensitive for proximal DVTs (>95%) but less so for distal DVTs (70%). In general terms, a positive scan will confirm the diagnosis. However, in those with a Wells score indicating that a DVT is likely, a negative scan, if limited to the proximal venous system, does not exclude the diagnosis, as it will not identify a proportion of patients who have a distal DVT. Further assessment of these patients is required (**Fig. 11.44**).

Alternatively, whole-leg ultrasonography can be performed. This may avoid the need for repeat scanning but is more time-consuming and requires greater expertise.

The diagnosis of DVT in those with a prior event is often problematic because of the presence of residual vein occlusion, and expert radiological advice should be sought.

Imaging for PE

The approach to the diagnosis of PE is similar in principle (**Fig. 11.45**). When the Wells score indicates that a PE is unlikely, a negative D-dimer excludes the diagnosis without further investigation. However, imaging is essential in those who either have a raised D-dimer level or are categorized as likely to have a PE based on their Wells score.

The most common imaging technique is ***computed tomographic pulmonary angiography (CTPA)*** (**Fig. 11.46**), which is widely available and sensitive, and can provide an alternative diagnosis when PE is excluded. The alternative is isotope lung scanning. Historically, planar ***ventilation–perfusion ($\dot{V}/\dot{Q}$) isotope lung scanning***, has been used (**Fig. 11.47**). This test is performed in two stages:

- ***a perfusion phase***, in which technetium-labelled albumin aggregates are injected intravenously and blood flow to the lungs is assessed
- ***a ventilation phase***, in which patients inhale radiolabelled xenon or technetium to assess air delivery to the lungs (sometimes this phase is not performed in pregnancy, to minimize radiation exposure, as assume ventilation will be normal).

A diagnosis of PE is made if there are mismatched defects in the perfusion scan, indicating impaired blood flow to the lungs, but normal ventilation because air entry to the lungs is normal. A normal $\dot{V}/\dot{Q}$ scan excludes the diagnosis of PE. This technique has the advantage of a lower radiation dose and is preferred in those with severe renal impairment (estimated GFR < 30 mL/min/1.73 m^2) or allergies to intravenous contrast agents. The disadvantages are that it is less readily available and the results are often non-diagnostic because other lung pathologies can cause matched or smaller mismatched defects. Increasingly centres offer ventilation/perfusion single photon emission computed tomography (V/Q SPECT) as a superior alternative to planar V/Q scanning.

Timescale of investigation

Diagnostic testing for VTE should be performed urgently and completed within 24 hours of initial presentation. Interim anticoagulation should be given if imaging for a PE that is likely by Wells score cannot be done immediately or if it will take more than 4 hours to investigate either a suspected PE that is considered unlikely by Wells score or a suspected DVT. Interim anticoagulation should be stopped after 24 hours if diagnostic testing is negative.

The approaches described apply only to patients presenting to primary care, EDs, or other acute medical services including SDEC. Inpatients (especially surgical, if bleeding risk) and pregnant patients should be regarded as high-risk and require appropriate expedited imaging if DVT or PE is suspected.

Emergency presentations

In the minority of patients with suspected PE who present with hypotension, mortality is high and prompt diagnosis is essential. If an urgent CTPA is not possible, a bedside transthoracic echocardiogram may be diagnostic, showing acute pulmonary hypertension and right ventricular dysfunction (see **Fig. 11.48**).

Upper limb DVT

The diagnostic algorithm for suspected upper limb venous thrombosis is less well established and patients with suggestive clinical features of arm pain and swelling typically require ultrasound imaging.

Further reading

Kahn S and de Wit K. Pulmonary Embolism. *N Engl J Med*. 2022; 387:45–57.
Khan F, Tritschler T, Kahn SR, Rodger MA. Venous thromboembolism. *Lancet*. 2021; 398:64–77.
Konstantinides S, Meyer G, Becattini G et al. 2019 ESC guidelines for the diagnosis and management of acute pulmonary embolism. *Eur Heart J*. 2020; 41:543–603.
National Institute for Health and Care Excellence. *NICE Clinical Guideline ng158: Venous Thromboembolic Diseases: Diagnosis, Management and Thrombophilia Testing*. NICE 2020; https://www.nice.org.uk/guidance/ng158.

Management

Anticoagulant therapy is the standard treatment for VTE. It is traditionally divided into three phases:

- the acute phase, lasting 5–14 days
- a maintenance primary treatment phase, lasting 6 weeks to 3 months
- a long-term secondary prevention phase beyond this.

The aim of treatment in the initial phase is to prevent thrombus extension and hence reduce the risk of embolization as well as post-thrombotic syndrome; the goal thereafter is to prevent thrombus recurrence.

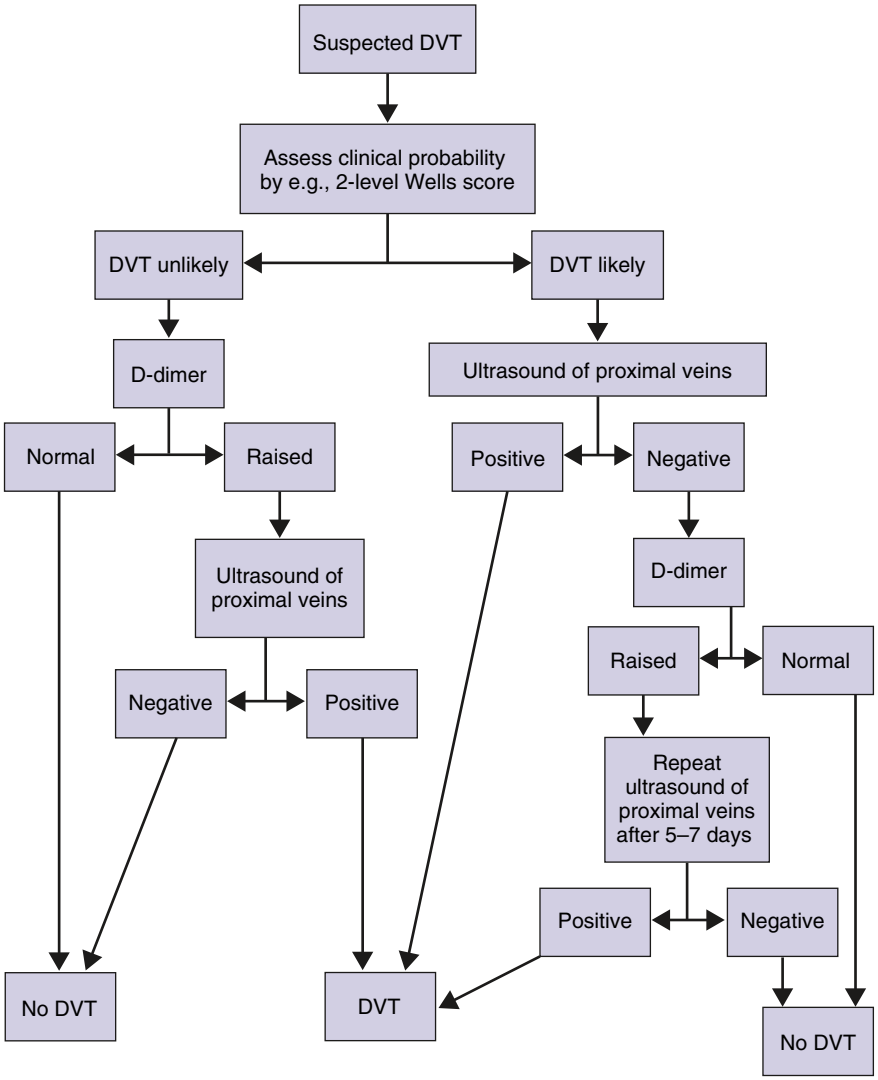


Fig. 11.44 **Investigation of suspected DVT.**

Primary treatment initiation

The long-established initial treatment of VTE involves a ***parenteral anticoagulant***, most commonly subcutaneous LMWH; alternatively, intravenous unfractionated heparin (UFH) or subcutaneous fondaparinux may be used. These provide an almost immediate anticoagulant effect.

Effective oral anticoagulant alternatives, the oral factor Xa and thrombin inhibitors (previously known as direct oral anticoagulants (DOACs) or non-vitamin K antagonist oral anticoagulants (NOACs)), are licensed for VTE treatment. Two of these, apixaban and rivaroxaban, can be used from the outset without any initial parenteral anticoagulant therapy. The initial dose is higher for the first 7 days for apixaban and 21 days for rivaroxaban.

For two others, edoxaban and dabigatran, the initial treatment consists of at least 5 days of parenteral therapy. The parenteral therapy (such as LMWH) is then stopped and on day 6, the edoxaban or dabigatran is started instead with no overlap.

Traditionally vitamin K antagonists such as warfarin were used as oral anticoagulant therapy and are still preferred in some circumstances such as the antiphospholipid syndrome and end-stage kidney failure. An overlap period in which they are used alongside the parenteral drug is essential to ensure effective management of the VTE because it takes at least 5 days before they provide therapeutic anticoagulation as judged by their effect on the ***International Normalized Ratio (INR)***. Heparin or fondaparinux can be stopped and warfarin continued alone once the INR is 2.0 or more on 2 consecutive days, indicating that the vitamin K antagonist is now providing sufficient anticoagulation.

The different treatment models are shown in **Fig. 11.49** and the drugs are described further in Chapter 40.

Special circumstances

A minority of patients warrant a different initial approach.

Pregnancy and lactation

LMWH is the treatment of choice. Warfarin, oral factor Xa, and thrombin inhibitors cross the placenta and should not be used in pregnancy. Warfarin is teratogenic and causes embryopathy between 6 and 12 weeks' of pregnancy, with skeletal abnormalities including nasal hypoplasia and stippled epiphyses. Later in pregnancy it causes bleeding in the fetus, particularly intracerebral bleeding, and a significant risk of intrauterine death. However, warfarin is safe until the 6th week of pregnancy. It is therefore important to recognize pregnancy early and to switch by the 6th week to LMWH, which does not cross the placenta and is safe for the fetus.

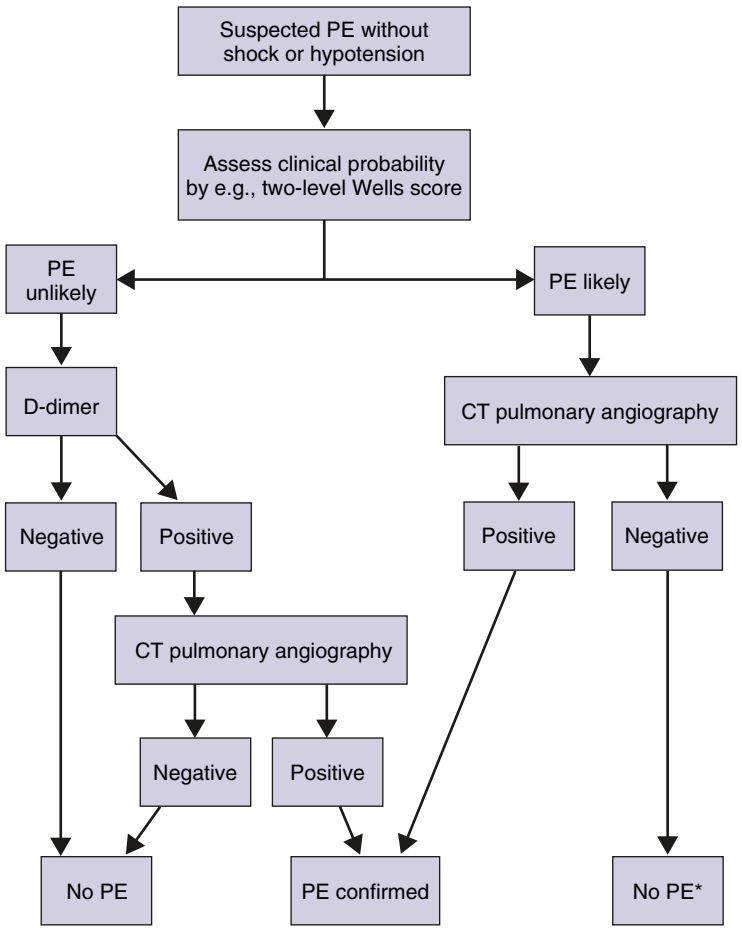


Fig. 11.45 Investigation of suspected PE without haemodynamic compromise. *In selected cases where clinical suspicion remains high, it may be necessary to consider either bilateral lower limb ultrasonography or V/Q scan.

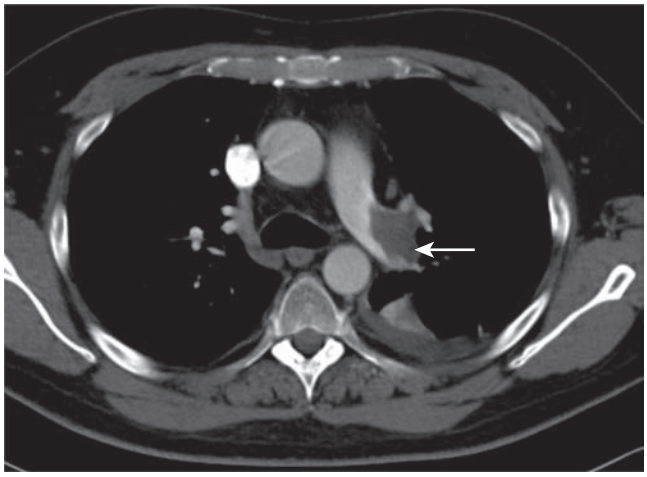


Fig. 11.46 CT pulmonary angiography in pulmonary embolism. The image is taken at the level of the main right and left pulmonary arteries, and shows a large thrombus *(arrowed)* in the left pulmonary artery.

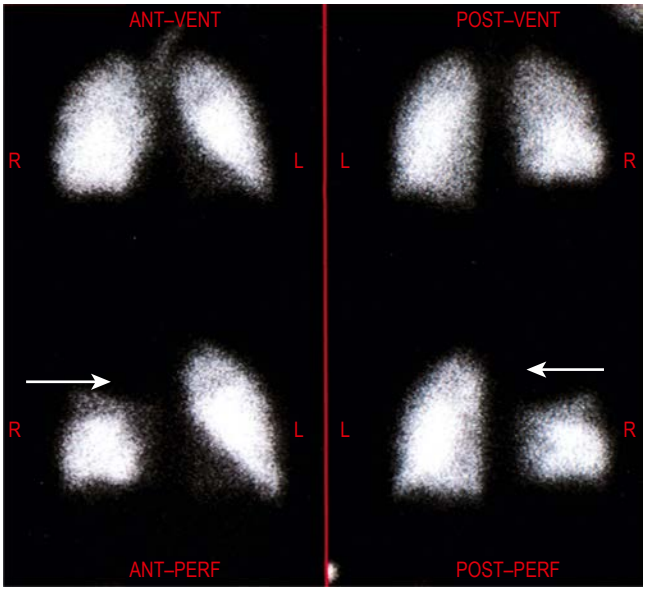


Fig. 11.47 Radionuclide ventilation/perfusion scanning. Ventilation *(top)* and perfusion *(bottom)* lung scans demonstrate an absence of perfusion in the right upper lobe: that is, probable pulmonary embolism *(arrowed)*.

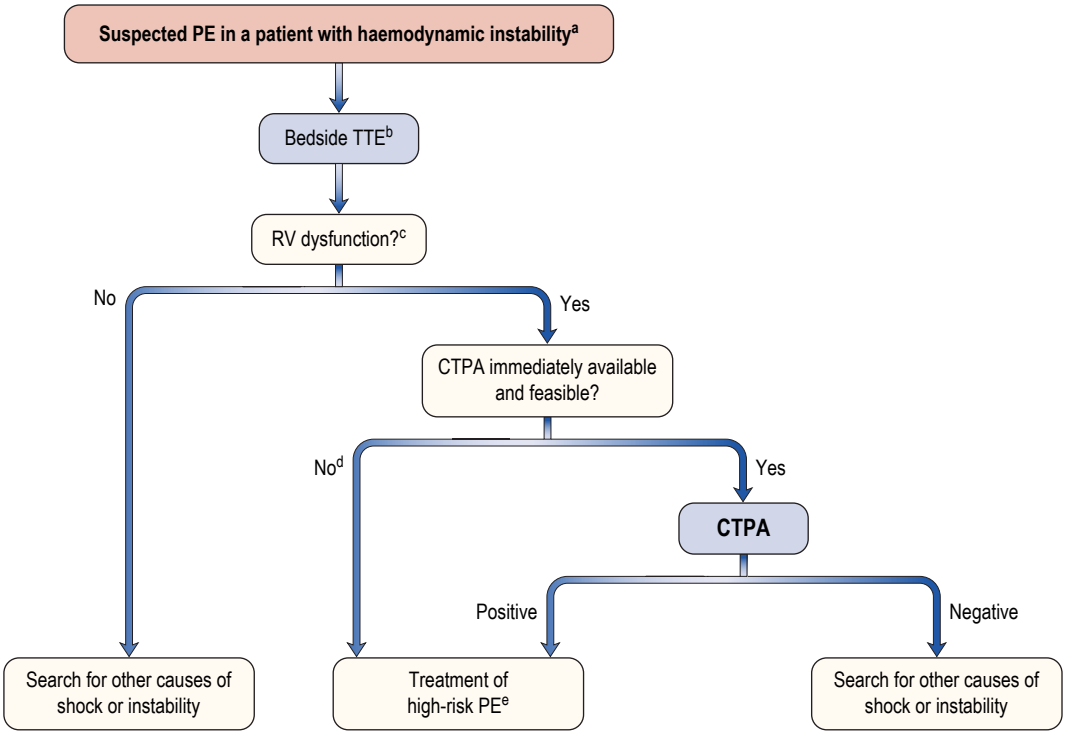


Fig. 11.48 Investigation of patients with suspected PE and shock or hypotension. Key: a = systolic blood pressure <90 mmHg (or vasopressors) with end-organ hypoperfusion OR need for resuscitation OR sustained systolic blood pressure <90 mmHg or fall in blood pressure > = 40 mmHg not due to another cause. b = or transoesophageal echocardiography or compression ultrasonography confirming VTE. c = right ventricular dysfunction with pressure overload. d = critical condition allowing only bedside diagnostic tests. *(Adapted from 2019 ESC guidelines on diagnosis and management of acute PE, Konstantinides et al. DOI: 10.1183/13993003.01647-2019. From Konstantinides S, Meyer G, Becattini C et al. 2019 ESC guidelines for the diagnosis and management of acute pulmonary embolism.* Eur Heart J *2020; 41:543–603.)*

Regular pregnancy testing should be performed by women taking warfarin who are trying to conceive. The consequences of fetal exposure to oral thrombin and factor Xa inhibitors are uncertain and these drugs are therefore avoided in pregnancy. Anticoagulation in pregnancy in those at very high risk of thrombotic complications; for example, metal heart valves should be done with expert counselling and surveillance. See Chapter 48 for further details of anticoagulation in pregnancy.

Both warfarin and LMWH are safe in breast-feeding but this has not been established for the oral thrombin and factor Xa inhibitors, which are therefore avoided (see also Chapter 48).

Patients with cancer

For many years LMWH has been the treatment of choice for cancer-associated thrombosis because of demonstrated superiority over warfarin. Recent trial data suggest that oral factor Xa inhibitors are as effective as LMWH in preventing recurrent cancer-associated DVT and PE, avoiding the need for daily subcutaneous injections, though they may increase the risk of bleeding particularly from the GI tract. Guidelines now support use of these agents for first-line treatment of cancer-associated DVT or PE after taking into account tumour site, interactions with other drugs including those used to treat the cancer, and bleeding risk.

Ambulatory care

Many patients with DVT can be managed on an outpatient basis, admission being reserved for those with a complex presentation or significant co-morbidities. Increasingly too, low-risk PE can be managed on an outpatient basis. Low-risk patients can be identified using the ***Pulmonary Embolism Severity Index (PESI) or simplified PESI*** (**Box 11.96**); management as an outpatient or with early discharge might be considered for this group. Higher-risk patients need admission for close observation or administration of oxygen therapy.

Role of thrombolysis

Anticoagulants help prevent thrombus extension and recurrence but do not dissolve blood clots, in contrast to thrombolytic agents. The latter are seldom used in the treatment of VTE because they carry a higher risk of major bleeding than anticoagulation, including a 2% risk of intracranial haemorrhage. However, in patients presenting with massive PE characterized by systolic hypotension (blood pressure ≤90 mmHg), there is a high risk of early death, and systemic thrombolysis administered intravenously, or occasionally by catheter infusion directly into the thrombus, may be lifesaving by rapidly restoring pulmonary perfusion.

The role of thrombolysis in the management of intermediate-risk PE—that is, without systolic hypotension but with evidence of right ventricular dysfunction and raised pro-BNP or troponin levels—remains controversial. Thrombolysis—either systemic, catheter-directed or pharmaco-mechanical—is occasionally used in the setting of management of limb-threatening ilio-femoral vein thrombosis. Local thrombolysis is also sometimes used in non-limb-threatening ilio-femoral vein thrombosis in an attempt to reduce

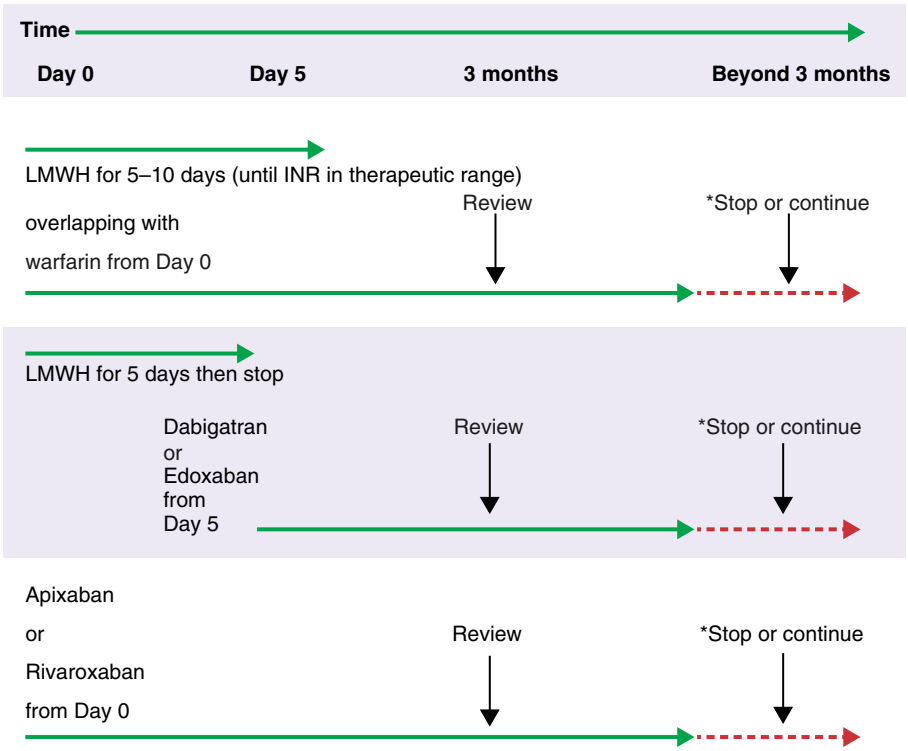


Fig. 11.49 Models for anticoagulant management of newly diagnosed venous thromboembolism (VTE). *The decision to stop or continue anticoagulation beyond 3 months is based on the balance of recurrence risk if anticoagulants are stopped, bleeding risk if anticoagulants continue, and patient preference. *INR,* International Normalized Ratio; *LMWH,* low-molecular-weight heparin.

Box 11.96 Original and simplified PESI scores

Parameter	Original version	Simplified version
Age	Age in years	1 point (if age >80 years)
Male sex	+10 points	–
Cancer	+30 points	1 point
Chronic heart failure	+10 points	1 point
Chronic pulmonary disease	+10 points	
Pulse rate ≥110 beats/min	+20 points	1 point
Systolic blood pressure <100 mmHg	+30 points	1 point
Respiratory rate >30 breaths/min	+20 points	–
Temperature <36°C	+20 points	–
Altered mental status	+60 points	–
Arterial oxyhaemoglobin saturation <90%	+20 points	1 point
Risk strata[a]		
	Class I: ≤65 points Very low 30-day mortality risk (0%–1.6%) **Class II: 66–85 points** Low mortality risk (1.7%–3.5%)	**0 points** = 30-day mortality risk 1.0% (95% CI 0.0%–2.1%)
	Class III: 86–105 points Moderate mortality risk (3.2%–7.1%) **Class IV: 106–125 points** High mortality risk (4.0–11.4%) **Class V: >125 points** Very high mortality risk (10.0%–24.5%)	**≥1 point** = 30-day mortality risk 10.9% (95% CI 8.5%–30.2%)

[a]Based on the sum of points.

CI, Confidence interval; *PESI,* Pulmonary Embolism Severity Index.

From Konstantinides S, Meyer G, Becattini C et al. 2019 ESC guidelines for the diagnosis and management of acute pulmonary embolism. *Eur Heart J* 2020; 41:543–603.

symptoms and prevent the post-thrombotic syndrome, although there is variable evidence of long-term benefit.

Interventional approaches

Embolectomy

When patients present with massive PE and thrombolysis is contraindicated, emergency pulmonary embolectomy can be lifesaving. This can be done surgically or percutaneously via a catheter inserted via the deep venous system.

Inferior vena cava filters

Occasionally, patients with newly diagnosed VTE have a contraindication to anticoagulation—for example, active bleeding—or a major bleeding risk, such as the need for urgent surgery. In such settings an inferior vena cava (IVC) filter can be inserted by an interventional radiologist to prevent emboli from the deep veins in the leg reaching the lungs. Contraindications to anticoagulation are usually temporary, and anticoagulant treatment should be started as soon as it is safe to do so because filters do not entirely prevent pulmonary emboli and are independently associated with an increased risk of DVT. IVC filters do not reduce the risk of recurrent PE compared to anticoagulation alone. As IVC filters can give rise to complications that include thrombosis, migration, and embolization, retrievable filters are preferred over permanent ones and they should be removed as soon as anticoagulation has been safely established.

For a discussion of treatment duration, complications, investigation of underlying cause, and prevention of venous thromboembolic disease, see Chapter 40.

Further reading

Howard LSGE, Barden S, Condliffe R et al. British Thoracic Society Guideline for the initial outpatient management of pulmonary embolism. *Thorax.* 2018; 73:ii1–ii29.
Kearon C, Akl EA, Ornelas J et al. Antithrombotic therapy for VTE disease. *Chest.* 2016; 149:315–352.
Konstantinides S, Meyer G, Becattini C et al. 2019 ESC guidelines for the diagnosis and management of acute pulmonary embolism. *Eur Heart J.* 2020; 41:543–603.
National Institute for Health and Care Excellence. *NICE Clinical Guideline 158: Venous Thromboembolic Diseases: Diagnosis, Management and the Thrombophilia Testing*. NICE 2020; www.nice.org.uk/guidance/ng158.

The unwell cancer patient

Patients living with cancer may present to the acute medical team for a variety of reasons including illness due to disease, complications of anticancer treatment, as a first presentation of malignancy, or due to an unrelated illness. This section details some of the more common medical emergencies which the acute medicine physician is expected to recognize and manage in this population of patients.

Many hospitals have developed an 'acute oncology service' (AOS) offering rapid specialist input for patients being treated for cancer who present to hospital via acute medical services. They provide a vital link between outpatient cancer services and the acutely unwell patient, and may have additional specialist information about the treatment patients have already received, as well offering a route to accessing urgent treatment modalities such as radiotherapy for metastases causing pain or spinal cord compression.

Neutropaenic sepsis

Neutropaenic sepsis is defined as the coexistence of a neutrophil count $<0.5 \times 10^9$/L accompanied by fever (>38°C) or other signs or symptoms consistent with sepsis in a patient undergoing anticancer therapy. This is the most common cause of attendance in the ED for any cancer patient and must always be considered in anyone who is unwell within a month of chemotherapy. Neutropenic patients are at high risk of bacterial and fungal infections, most often from enteric bowel flora; hence, some units will advocate a specific neutropenic diet that is low in bacterial and fungal contamination (**Box 11.97**). Patients must be warned of the possibility of neutropenic fever occurring.

Assessment should be focused on identifying potential sources of infection and establishing severity of infection. Common sources of infection in this population include the respiratory, GI, and urinary tracts. Ask specifically about cough, vomiting, diarrhoea, and dysuria. It will not always be possible to identify a source, and non-specific symptoms are also common, such as nausea, breathlessness, and drowsiness. A fever may not always be present.

The critical test is the full blood count. Although by strict criteria, neutropaenia is defined as a neutrophil count $<0.5 \times 10^9$/L, patients with falling neutrophils $<1.0 \times 10^9$/L after chemotherapy and fulfilling other criteria for sepsis are also managed as neutropaenic sepsis. Immediate introduction of broad-spectrum antibiotics is indicated. Signs of systemic illness, such as tachycardia, hypotension, or oliguria, mandate urgent admission and resuscitation with intravenous treatment (**Box 11.98**). Initial empirical therapy should be reviewed following microbiological results.

All such patients need to be discussed with the appropriate specialist oncology team and hospitals should have clear protocols for the rapid institution of antibiotics within an hour of arrival in the ED. In units practiced in the assessment of febrile neutropenia, it is possible to follow a more risk-stratified antibiotic policy and to avoid or curtail admission with oral co-amoxiclav plus ciprofloxacin when low-risk features are present: that is, absence of tachycardia, hypotension, hypoxia, and mucositis, and an expected short duration of myelosuppression.

Complications of immune checkpoint inhibitors

These relatively new agents act directly on T cells to reverse immune tolerance and encourage anti-tumour activity of the immune system (see Chapter 41). Because of their mechanism of action, side-effects include a variety of autoimmune phenomena which may cause disease affecting a wide range of organ systems (such as acute colitis, AKI secondary to tubulointerstitial nephritis, a variety of skin rashes, acute hepatitis). A high index of suspicion for drug side-effects is required in any patients receiving these agents, with close liaison with oncology teams.

Box 11.97 Neutropenic diet

Foods and drinks that should be avoided:
- All uncooked vegetables and most fruits—peel off thick skin from foods like oranges and bananas
- Raw or rare meat
- Fish
- Uncooked or undercooked eggs
- Foods from salad bars and delicatessen counters
- Soft blue/mould cheeses e.g., brie, camembert, stilton
- 'Well' water—or boil it for 1 min before drinking

Foods and drinks that are safe:
- Cooked vegetables
- Canned fruits and juices
- Well-done meat
- Hard-boiled eggs (no runny yolks)
- Vacuum-packed lunch meats (not freshly sliced meats)
- Pasteurized milk, cheese, yoghurt, and other dairy products only
- Tap water if the source is known—or boil it; bottled water

Box 11.98 Management of neutropaenic sepsis

Immediate intervention is essential

- Resuscitate with intravenous fluids to restore circulatory function; monitor vital signs, urine output, and Glasgow Coma Scale score
- Take cultures of peripheral blood, urine, sputum and stool and blood cultures drawn through any indwelling venous catheters
- Consider the need for admission to critical care, invasive monitoring (e.g., central venous pressure), and consider inotropic support at an early stage
- Give empirical antibiotics as per local policy and sensitivities:
  - Commonly used antibiotics should include activity against enteric Gram-negative bacteria and *Pseudomonas* e.g., ceftazidime or piperacillin–tazobactam with gentamicin; or meropenem monotherapy
  - Antibiotics against staphylococci may be needed e.g., vancomycin, especially with indwelling venous access lines
- Change antibiotics according to culture results, or empirically broaden Gram-negative cover and consider adding Gram-positive cover, if the patient deteriorates clinically and/or temperature is still elevated after 48 h; discuss with the microbiology department
- Consider imaging e.g., thoracic CT, if fever is not responding to broad-spectrum antibiotics, to detect an occult source of fever; consider adding treatment for opportunistic infections:
  - Liposomal amphotericin B or voriconazole—*Candida* and *Aspergillus*
  - High-dose co-trimoxazole—*Pneumocystis*
  - Clarithromycin—*Mycoplasma* and *Legionella*
  - Anti-tuberculous therapy—*Mycobacterium tuberculosis*
- Granulocyte colony-stimulating factor (G-CSF) is usually reserved for cases with profound neutropaenia (neutrophil count $<0.1 \times 10^9$/L) accompanied by adverse prognostic features (e.g., septic shock, prolonged neutropaenia, fungal disease)

Pulmonary embolus

This is a common complication of the coagulopathy of cancer and may be a side-effect of chemotherapy. In the context of malignancy, pulmonary embolus (PE) often presents with unexplained breathlessness and episodic exacerbations from multiple small emboli, rather than chest pain. A high level of suspicion should be maintained in any cancer patient with breathlessness, hypoxia, or chest pain. CT pulmonary angiography is the investigation of choice (**Fig. 11.46**). Prophylactic anticoagulation should be given to all immobilized patients. Treatment is discussed earlier in this chapter in the section 'Deep vein thrombosis and pulmonary embolism (special circumstances)'.

Spinal cord compression

Spinal cord compression (SCC; see Chapter 28), usually due to bony metastatic disease, needs to be rapidly diagnosed, and urgent treatment arranged within 24 hours of onset of paresis, to salvage as much functional capacity as possible. Symptoms include neck or back pain, which is frequently progressive and may be severe and unremitting, bladder or bowel dysfunction, numbness or paraesthesia, limb weakness, or difficulty walking. Neurological examination findings include signs of spastic (upper motor neuron) paraparesis or tetraparesis and sensory loss below the compression with a 'sensory level' usually 2–3 dermatome levels below the anatomical compression, although early neurological clinical features may be incomplete, and more subjective than objective in nature. Whole spine MRI scanning, including both T1-weighted images to identify spinal metastases and T2-weighted images to delineate compression of the cord is the investigation of choice. Treatment should begin with high-dose corticosteroids (usually dexamethasone 16 mg p.o. daily) and analgesia and both neurosurgical and oncological opinions sought. A neurosurgeon should advise on spinal stability, the need for spinal precautions for movement, and the role of decompressive surgery and external bracing for limited-extent disease although patients should be immobilized pending this input and nursed in a supine position to minimize axial weight-bearing. Where surgery is not considered, radiotherapy alone may be used.

Fig. 11.50 Superior vena caval obstruction resulting from a bulky anterior mediastinal mass in a patient presenting with T-lymphoblastic lymphoma.

The cauda equina syndrome

Lumbosacral bony metastatic deposits may result in the cauda equina syndrome: that is, bilateral flaccid (compare spasticity in spinal cord lesions), lower limb weakness, sacral numbness, retention of urine, erectile dysfunction, and areflexia, usually with back pain. Multiple lumbosacral nerve roots are involved. Onset is either acute—an acute flaccid paraparesis—or chronic, sometimes with intermittent claudication (usually with prominent exercise-induced sensory symptoms in addition to pain). Management is the same as for malignant spinal cord compression, with urgent imaging (whole spine MRI), analgesia, corticosteroids, and neurosurgical involvement with a view to surgical decompression.

Superior vena cava obstruction

Superior vena caval obstruction (SVCO; **Fig. 11.50**) can arise from any upper mediastinal mass but is most commonly associated with extrinsic compression from lung cancer and lymphoma. The patient presents with difficulty breathing and/or swallowing, stridor, a swollen, oedematous face and arms, venous congestion in the neck, and dilated veins in the upper chest and arms. CXR may show an abnormal mediastinal contour or widening although diagnosis requires contrast-enhanced CT scanning.

SVCO is a medical emergency and requires immediate assessment and treatment. The patient should be nursed sitting upright and supplemental oxygen administered. Initial treatment is with corticosteroids (usually dexamethasone 16 mg p.o. daily) followed by chemotherapy where the tumour is expected to respond (e.g., lymphomas, small-cell lung cancers, and germ cell tumours), and possibly mediastinal radiotherapy. Sometimes, vascular stents and anticoagulation are also required. Critical care team input and ITU supportive care may be required, pending a response to treatment.

Raised intracranial pressure

Raised ICP due to intracerebral metastases presents classically with headache, nausea and vomiting. There are often no localizing

neurological signs and almost never papilloedema until very late in the disease. However, for many, there is a slower onset with non-specific symptoms such as drowsiness or mental deterioration. Treatment is by high-dose steroids and investigation by MRI. Surgery is appropriate if the condition is unifocal and/or threatening the fourth ventricle; otherwise, whole-brain or local stereotactic CyberKnife radiotherapy is required.

Tumour lysis syndrome

This occurs if treatment triggers a massive breakdown of tumour cells, leading to increased serum levels of urate, potassium, and phosphate, and a secondary hypocalcaemia. These biochemical changes can give rise to cardiac arrhythmias and seizures. Urate deposition in the renal tubules can cause AKI (hyperuricaemic nephropathy). Vigorous hydration, combined with loop diuretics if needed, is crucial to maintain high urine outputs in such patients; however, a proportion will require dialysis for uraemia, oliguria, or severe electrolyte disturbances. The xanthine oxidase inhibitor allopurinol should be given before treatment is started in low-risk patients. Intravenous rasburicase, a recombinant urate oxidase, is used for prophylaxis in high-risk patients and in the treatment of tumour lysis syndrome.

Acute hypercalcaemia

Acute hypercalcaemia of malignancy often presents with dehydration, nausea and vomiting, nocturia and polyuria, drowsiness, and altered consciousness. The serum Ca^{2+} is >3 mmol/L and sometimes as high as 5 mmol/L. While investigation of the cause is under way, immediate treatment is mandatory if the patient is seriously ill or if the Ca^{2+} is >3.5 mmol/L.

Treatment is by resuscitation with intravenous fluids first, to establish a saline diuresis, and then an intravenous bisphosphonate, such as pamidronate or the more potent zoledronic acid (**Box 11.99**). Treating the cause is crucial. Denosumab and calcitonin can be used in intractable cases.

Hypercalcaemia of malignancy is a poor prognostic indicator and patients should have a full holistic assessment in collaboration with their parent oncology team.

Malignant bile duct obstruction

This presents with cholestatic jaundice. Lymphomatous obstruction responds very well to prompt initiation of therapy. A small proportion of pancreatic and bile duct tumours are surgically resectable, more commonly those in the distal bile duct as compared to those in the hilar region. However, in the greater proportion of patients, treatment is palliative. In recent years, endoscopic techniques have allowed the insertion of stents into the biliary tree to re-establish bile flow. Self-expanding metal stents have long periods of patency but are at risk of ascending infection. In the small proportion of patients in whom bile duct drainage is not possible endoscopically, the percutaneous route offers an alternative. Endoscopic photodynamic laser therapy with a photoporphyrin sensitizer can also prolong patency.

Further reading

European Society for Medical Oncology Clinical Practice Guidelines. Accessed at: https://www.esmo.org/guidelines.
Treatment of hypercalcemia of malignancy in adults: An endocrine society clinical practice guideline. *J Clin Endocrinol Metab*. 2023; 108:507–528.

Box 11.99 Management of acute severe hypercalcaemia

- **Rehydrate** using at least 4–6 L of 0.9% saline on day 1, and 3–4 L for several days thereafter, aiming to maintain a urine output of 100–150 mL/h. Central venous pressure (CVP) may need to be monitored to control the hydration rate. Fluid infused will correct dehydration and increase urinary calcium excretion, usually lowering serum calcium by 0.25–0.375 mmol/L over the first 24 h.
- **Intravenous bisphosphonates** are the treatment of choice for hypercalcaemia of malignancy or of undiagnosed cause. Pamidronate is preferred (15–60 mg, either as a single intravenous infusion in 0.9% saline or glucose over 2–4 h or in divided doses over 2–4 days with dose adjustment for serum calcium concentration; maximum 90 mg per course). Levels fall after 24–72 h, lasting for approximately 2 weeks. Zoledronic acid is an alternative. Renal function should be assessed prior to treatment and dose adjusted according to creatinine clearance.
- **Denosumab** (a monoclonal antibody against RANK ligand, inhibiting bone resorption) may be used as an alternative to bisphosphonates depending on local protocols, or under specialist supervision in circumstances where response to other therapies has been poor.
- **Prednisolone** (30–60 mg daily) is effective in some instances (e.g., in myeloma, sarcoidosis, and vitamin D excess) but in most cases is ineffective.
- **Calcitonin** (4–8 units/kg i.m. or s.c. every 6–12 h) rapidly lowers serum calcium by 0.25–0.50 mmol/L and may be useful as bridging treatment while awaiting a response to bisphosphonates/denosumab. Calcitonin inhibits bone resorption through interference with osteoclast function, although tachyphylaxis occurs within 48–72 h.
- **Oral phosphate** (sodium cellulose phosphate 5 g three times daily) produces diarrhoea and is now rarely used
- **Haemodialysis** may be required in refractory cases, usually where hypercalcaemia is accompanied by established renal failure.

Rash

Rashes constitute a frequently encountered clinical presentation, encompassing a myriad of underlying aetiologies. Managing patients presenting with a rash revolves around identification of the specific cause to tailor an effective treatment strategy. The origins of rashes are diverse and may stem from infectious, allergic, autoimmune, or systemic conditions. The clinical assessment and stepwise approach to assessing a skin rash is suggested in **Fig. 45.4**.

Important dermatological conditions which may present to the acute physician include skin infections (including cellulitis, necrotizing fasciitis, and toxic shock syndrome), erythroderma, and severe cutaneous adverse drug reactions. The sections below detail information gathering required to enable first-line investigation and treatment and, when required, enable effective discussion and referral to a dermatologist.

Clinical assessment

History

- **Onset and duration:** Determine when the rash first appeared and its evolution over time, noting any changes in size, shape, or associated symptoms.
- **Symptoms:** Enquire about associated symptoms such as pruritus, pain, or systemic manifestations like fever. This aids in discerning infectious or systemic causes.
- **Triggers and exacerbating factors:** Investigate potential triggers or exacerbating factors, including exposure to new substances, recent infections, or changes in environment (e.g., heat or sunlight exposure).
- **Medical and drug history:** Gather information on the patient's medical history, including any history of allergy, atopy, or autoimmune disease, focusing on conditions linked to dermatological manifestations. Explore recent changes in medications and enquire about supplements and recreational substances. Enquire about family history of skin disease.

- **Occupational and environmental exposures:** Enquire about the patient's occupation and environmental exposures, as these may contribute to certain types of rashes.
- **Review of systems:** Perform a review of systems to identify any concurrent or preceding symptoms that may be linked to the rash (e.g., respiratory symptoms, joint pain).

Physical examination

- **Rash characteristics:** Scrutinize the morphology, distribution, and configuration of the rash. Consider whether it presents as macules, papules, vesicles, pustules, or other distinctive patterns. Does the rash blanch with pressure? A non-blanching (petechial or vasculitic) rash should always prompt consideration of meningococcal sepsis, although the differential diagnosis is contextual and broad.
- **Distribution:** Assess the distribution of the rash across the body, considering localized versus generalized patterns. Note if it follows a specific dermatomal or anatomical distribution. Ask specifically about involvement of mucosal surfaces (e.g., eyes, mouth, genital area) and hair and nail disease.

Investigations

Core laboratory tests (including full blood count, urea and electrolytes, liver function tests, and CRP) will be warranted in all but the most mild presentations. Targeted investigations will vary widely depending on the nature of rash identified, although appropriate investigations may include:

- blood cultures (for the febrile patient or where suspicion exists for bacterial sepsis)
- pus swabs or viral nucleic acid testing (for pustular or vesicular eruptions)
- skin scrapings (to confirm a possible fungal rash)
- additional investigations to investigate for underlying cause (e.g., the finding of a vasculitic rash will prompt broad investigation for underlying infectious or autoimmune disorders).

Where the causes remain unclear, or the rash fails to resolve spontaneously or with treatment, skin biopsy by a dermatologist may be indicated.

Management

Tailor the management plan to the identified cause of the rash. Any culprit medication should be withdrawn. ***Erythroderma*** (inflammation involving >90% skin area, irrespective of cause, see 'Erythroderma and loss of skin function') is a medical emergency, threatening autoregulation of body temperature and fluid and electrolyte balance, and requires prompt clinical assessment (ABCDE approach) and resuscitation and may warrant critical care admission. Topical emollients may bring symptomatic relief in a broad range of dermatoses, while more definitive treatment is planned. Dermatology referral should be considered for any patient where the cause of rash is unclear or symptoms are severe (e.g., blistering eruptions or mucosal involvement) and is mandatory in erythrodermic eruptions.

Acute skin infections

Cellulitis and erysipelas

Cellulitis and erysipelas are caused by superficial and deeper infection of the dermis and subcutaneous tissues, respectively (Fig. 11.51). It is not always possible to make a clear distinction. Both present with tender confluent areas of inflamed skin and are often associated with fever and malaise. Cellulitis typically affects the lower leg or arm and may spread proximally. Other sites that may be affected include the abdomen and perianal and periorbital areas. Erysipelas is more common on the face and is more sharply demarcated. Localized blistering (clear or blood-stained), necrosis, abscess formation,

Fig. 11.51 Acute skin infections. **(A)** Cellulitis extending from the dorsum of the left foot to the shin. **(B)** Facial erysipelas.

lymphangitis, and lymphadenopathy may occur. The most common infective organisms are β-haemolytic streptococcus and *S. aureus*. Gram-negative or anaerobic bacteria may cause infection in immunocompromised people and those with diabetes.

Lymphoedema, obesity, venous stasis, and toe-web fungal infection are risk factors for lower limb cellulitis. The skin should be carefully examined to look for fissures or erosions that can allow entry of pathogens. Skin swabs are usually negative unless taken from broken skin. Serological tests can confirm a streptococcal infection (antistreptolysin O titre (ASOT) and anti-DNAse B titre).

Management

Systemic antibiotics (flucloxacillin or erythromycin) are indicated; they may be given as high-dose oral therapy, or intravenously via ambulatory care or as an inpatient, depending on the patient's age, disease severity, and co-morbidities. Single-dose dalbavancin may be an alternative to conventional antibiotics in the treatment of acute bacterial skin infections in adults, avoiding the need for hospitalization or repeat attendance.

Up to 50% of people with cellulitis experience repeat episodes. Prophylactic antibiotic therapy reduces the risk of further attacks during treatment but the effect wears off after discontinuation.

Necrotizing fasciitis

This fulminant, rapidly spreading, life-threatening bacterial infection of the subcutaneous and deep soft tissues causes severe tissue destruction and is associated with a high mortality. A number of organisms can cause necrotizing fasciitis: spontaneous disease that arises in previously healthy people is often caused by group A streptococcus (GAS), while disease related to abdominal surgery or immunosuppression is often caused by mixed aerobic and anaerobic populations.

Necrotizing fasciitis is characterized by severe pain that is out of proportion to the apparent signs of skin inflammation. Infection tracks rapidly along tissue planes, causing spreading erythema, swelling, pain, and sometimes tissue crepitus. Patients are typically febrile and severely unwell, and urgent surgical intervention (with extensive debridement or amputation) and intravenous antibiotic therapy are required. High-dose benzylpenicillin and clindamycin are used in confirmed GAS infection, and a broad-spectrum antibiotic combination with additional metronidazole is used where the pathogen is unknown. Typically, the CRP and white cell count are significantly raised. Imaging is sometimes helpful (e.g., by revealing gas in deep tissues) but should not delay urgent surgical intervention. Multiorgan failure is common and mortality is high.

Gas gangrene

Gas gangrene is caused by deep tissue infection with *Clostridium* spp. (especially *C. perfringens*), following contaminated penetrating injuries—historically, battlefield wounds but now commonly found in intravenous drug users and patients who have undergone GI tract surgery. Initial infection develops in damaged, necrotic tissue; toxins secreted by the bacteria destroy surrounding tissue and permit rapid spread of infection, contributing to severe systemic effects. Treatment is by surgical debridement of necrotic tissue and intravenous benzylpenicillin and clindamycin.

Toxin-mediated skin disease

Staphylococcal scalded skin syndrome (SSSS) and toxic shock syndrome (TSS) are diseases caused by bacterial secretion of toxins, the latter representing a life-threatening illness requiring urgent medical treatment. Blistering lesions seen in SSSS can resemble toxic epidermal necrolysis (TEN, see later) but the mucous membranes are not involved in SSSS and blistering occurs higher in the epidermis. The intraepidermal split of SSSS can be demonstrated by histology of a frozen section of skin.

Potential sources of infection and toxin production include tampons and menstrual cups (particularly if left *in situ* for longer than recommended), and cutaneous wounds or abscesses. Treatment revolves around resuscitation, antibiotics, and identification and elimination of the bacterial 'driver' (e.g., removal of a retained tampon, surgical debridement of an abscess). Critical care management may be required.

Erythroderma and loss of skin function ('skin failure')

Erythroderma ('red skin') is a term used to describe widespread inflammation affecting 90% or more of the body surface. There are several causes (**Box 11.100**), of which underlying skin disease and drug hypersensitivity are the most common. Extensive skin inflammation causes pain, soreness, and irritation, and leads to loss of normal skin function, especially fluid and temperature regulation. Symptoms include thirst, fever and chills, malaise, and dizziness. Erythroderma requires urgent medical treatment to avoid life-threatening complications (see next section). Examination should seek features of an underlying dermatosis, such as scalp or nail psoriasis. There may be associated generalized lymphadenopathy and a skin biopsy may be indicated, especially in suspected cutaneous lymphoma. In non-malignant disease, lymph nodes are often mildly enlarged and histology shows non-specific, reactive changes.

Complications

These include:

- high-output cardiac failure from increased blood flow
- hypothermia from heat loss
- pre-renal AKI from fluid depletion
- hypoalbuminaemia
- catabolism and increased basal metabolic rate
- secondary bacterial infection
- capillary leak syndrome: the most severe complication and potentially fatal, though rare. Release of inflammatory mediators from the skin causes extensive vascular leakage with oedema, hypovolaemic shock, and acute respiratory distress syndrome (ARDS) (see Chapter 13).

? Box 11.100 Causes of erythroderma

Common

- Atopic eczema
- Psoriasis
- Drugs (e.g., sulphonamides, gold, sulphonylureas, penicillin, allopurinol, captopril)
- Seborrhoeic eczema
- Idiopathic

Rare

- Chronic actinic dermatitis
- Cutaneous T cell lymphoma (Sézary's syndrome)
- Malignancy (especially leukaemias)
- Pemphigus foliaceus
- Pityriasis rubra pilaris (a hereditary disorder of keratinization)
- HIV infection
- Toxic shock syndrome

Box 11.101 Severe cutaneous adverse drug reactions

- Stevens–Johnson syndrome (SJS) and toxic epidermal necrolysis (TEN)
- Drug reaction with eosinophilia and systemic symptoms (DRESS)
- Acute generalized exanthematous pustulosis (AGEP)
- Erythroderma/exfoliative dermatitis
- Serum sickness-like reactions

Management

Intensive supportive treatment includes maintaining body temperature (with space blankets and heaters) and close monitoring of vital signs and fluid and electrolyte balance. Skin swabs should be taken and antibiotics commenced if secondary infection is suspected. All non-essential medication should be stopped. Topical therapy with a bland emollient or mild topical steroid can be initiated while the underlying cause is identified, at which point systemic therapy may be indicated under specialist guidance: for example, ciclosporin for psoriasis (see Chapter 45).

Severe cutaneous adverse drug reactions

Severe drug rashes are listed in **Box 11.101**. Common causes include antibiotics, NSAIDs, anticonvulsants, allopurinol, dapsone, and nevirapine. Early recognition and drug withdrawal can minimize morbidity and mortality. Recent advances in pharmacogenetics have identified human leucocyte antigen (HLA) associations, and cytochrome P450 polymorphisms that are associated with an increased risk of reaction to certain drugs.

Stevens–Johnson syndrome and toxic epidermal necrolysis

Stevens–Johnson syndrome (SJS) and toxic epidermal necrolysis (TEN) are severe mucocutaneous disorders that are considered variants of a disease spectrum. They are characterized by varying extents of blistering/epidermal detachment and mucosal ulceration:

- SJS: <10% skin detachment; one or two mucosal sites involved (oral, genital, ocular)
- SJS–TEN overlap: 10%–30% skin detachment
- TEN: >30% skin detachment; all mucosal sites involved in most cases.

The onset is usually 1–2 weeks after drug exposure. Initial symptoms are non-specific (malaise, myalgia, fever, and cough). These are followed by tender maculopapular erythematous lesions on the torso and inflamed mucosal surfaces. 'Target lesions' (as seen in erythema multiforme, see Chapter 45) usually affect the hands and feet in SJS. In TEN there is widespread flaccid blistering with skin that wrinkles like wet wallpaper on gentle pressure (Nikolsky's sign). Features of skin failure can ensue (see above). Respiratory mucosal and pulmonary involvement may require ventilation, and upper GI involvement can cause haemorrhage. Patients are at high risk of sepsis and require expert intensive supportive care, as for extensive burns. Multiorgan failure may occur and the mortality for TEN ranges from 30% to 50%. 'SCORTEN' is a clinical severity score that can help assess prognosis (**Box 11.102**). All potential causative drugs should be stopped.

Box 11.102 SCORTEN prognostic score in toxic epidermal necrolysis

Risk factor	0	1
Age	<40 years	>40 years
Associated malignancy	No	Yes
Heart rate (b.p.m.)	<120	>120
Serum urea (mmol/L)	<9.6	>9.6
Detached or compromised body surface	<10%	>10%
Serum bicarbonate (mmol/L)	>20	<20
Serum glucose (mmol/L)	<13.9	>13.9

The more risk factors present, the higher the SCORTEN score, and the higher the mortality rate:
0–1 risk factors: 3.2% mortality
2: 12.1%
3: 35.3%
4: 58.3%
≥5: >90%

(From Bastuji-Garin S, Fouchard N, Bertocchi M et al. SCORTEN: a severity-of-illness score for toxic epidermal necrolysis. J Invest Dermatol 2000; 115:149–153.)

Specialist dermatology input is essential. Treatment with high-dose steroids, ciclosporin, and intravenous immunoglobulin therapy remains controversial, as high-quality clinical trials are lacking.

Drug reaction with eosinophilia and systemic symptoms/drug hypersensitivity syndrome

Drug reaction with eosinophilia and systemic symptoms (DRESS)/drug hypersensitivity syndrome usually starts 2–6 weeks after initial exposure and is characterized by widespread erythema, facial oedema, fever, lymphadenopathy, and hepatosplenomegaly. Blood eosinophilia is usual with elevated hepatic transferases. Pneumonia and nephritis can develop and a mortality rate of 10% has been reported. Treatment is with systemic steroids tapered over at least 3 months. Aromatic anticonvulsants are one of the most common causes of DRESS and, as they cross-react, all drugs of this group must be avoided in the future. Sodium valproate is a suitable alternative.

Acute generalized exanthematous pustulosis/toxic pustuloderma

Acute generalized exanthematous pustulosis (AGEP)/toxic pustuloderma is an exanthem with numerous small, non-follicular, sterile pustules around the neck, axillae, and groin. It usually starts a few days after drug exposure and resolves with peeling. There may be mild systemic upset but internal organs are not involved. Localized forms can occur. Florid cases can resemble pustular psoriasis. Topical steroids and emollients can be used to relieve symptoms.

Questions for this chapter can be found on Elsevier eBooks+.

12 Sepsis and the treatment of bacterial infection

Mark Melzer, Andrew W. Leitch, and David W. Wareham

Clinical skills 278
History 278
Management of sepsis 278
Examination 280
Background 281
Pathophysiology 281
Organ dysfunction 281
Consensus definition (Sepsis-3) 281
Septic shock 281
Recognition of sepsis 282
At-risk groups 282
In-hospital surveillance 282
Origins of sepsis 283
Clinical features 283
Sepsis in special situations 284
Management of sepsis 284
Surviving Sepsis Campaign guidelines 284
Antimicrobial stewardship 286
UK-based targets 286
Considerations before starting empirical antibiotics 286
Antimicrobial decision-making at 72 hours 286
Outpatient parenteral antimicrobial treatment 287
Assessment of allergy risk 287
Antibiotic chemoprophylaxis 288
Antibiotic therapies 289
Antibiotics commonly used in the treatment of sepsis 289
Treatment of infection caused by multidrug-resistant organisms 295

CORE SKILLS AND KNOWLEDGE

Sepsis is a serious complication of infection in which a dysregulated immune response leads to organ dysfunction and sometimes death. It is the final common pathophysiological pathway in patients who die from infection. It is estimated that, worldwide, 49 million people develop sepsis every year, and 11 million die.

All healthcare staff should be aware of sepsis, which may present via the hospital emergency department having started in the community or develop on general medical or surgical wards. Specialist care for patients with severe complications of sepsis may be provided in a critical care unit. Infectious diseases (ID) physicians become involved in managing patients with complex or disseminated infections and may run outpatient antibiotic treatment (OPAT) services and specialist ID clinics.

Key skills in this chapter area include:

- Identifying sepsis and instituting emergency management using the Sepsis Six care bundle
- Appreciating the potential complications of sepsis and recognizing the need for appropriate escalation of care in acutely unwell or deteriorating patients
- Becoming familiar with the appropriate use of antibiotics to treat commonly encountered infections.

Opportunities for students to learn about sepsis and infection are gained by reviewing patients with suspected infection in the emergency department or acute assessment units, ID wards or critical care settings, or those deteriorating on general wards who may be identified by attending critical care outreach rounds. Observing specialist ID and OPAT clinics (including multidisciplinary team meetings) provides an insight into the longer-term follow-up of patients with infection.

CLINICAL SKILLS

History

When taking a history, this should be tailored to cover the points in **Box 12.1**, which are highly relevant to a presentation with suspected sepsis or infection.

Management of sepsis

Sepsis is a medical emergency, and an approach based on the Sepsis Six care bundle ensures that management and investigation proceed hand in hand, and patients are cared for in the most appropriate location for their needs (see **Fig. 12.1**).

Box 12.1 Features in the history relevant to sepsis

Symptoms
- Duration
- Symptoms of inflammatory response: fevers, chills, rigors, malaise, confusion, vomiting
- Localizing symptoms indicating site of infection:
  - Chest: cough, sputum, shortness of breath
  - Urinary tract: dysuria, urgency, loin pain
  - Central nervous system: headache, meningism
  - Biliary tract: upper abdominal pain, jaundice, vomiting
  - Skin and soft tissue: localized heat, erythema, and pain

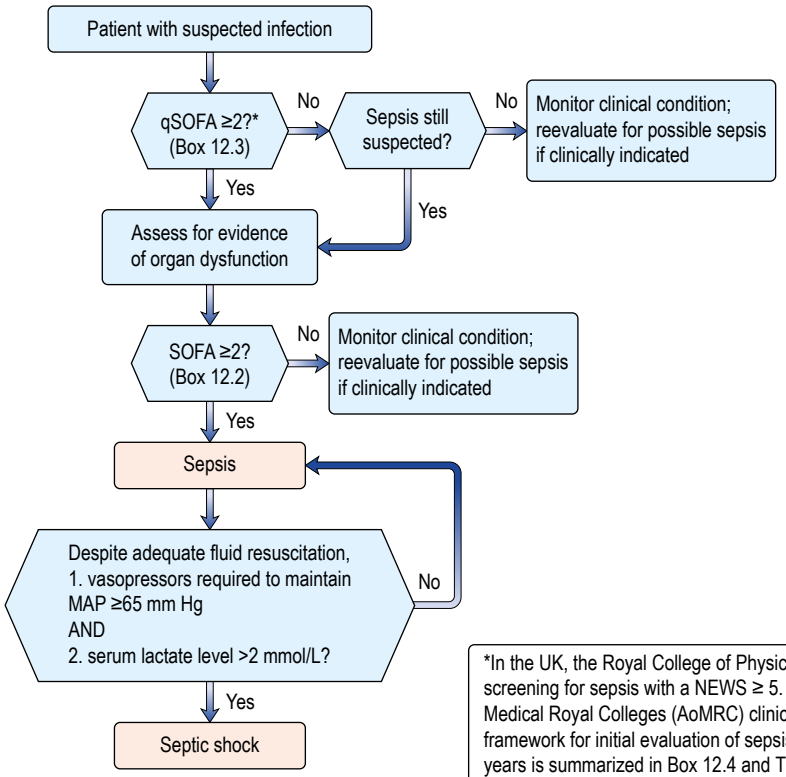
Past medical history
- Immunosuppression (diabetes, asplenism, human immunodeficiency virus, etc.)
- Pregnancy or patient in the puerperium
- Recent surgery or indwelling devices
- Previous sepsis

Drug history
- Immunosuppressant medication
- Recent courses of antibiotics

Risk factors for specific pathogen types
- Recent contact with medical institutions
- Previous infection with antibiotic-resistant organisms
- Travel history

General investigations

Bloods	FBC U&E LFTs CRP Lactate Glucose Coagulation
Urine	Nitrites Leucocyte esterase Culture
Arterial blood gas	
Imaging	X-ray Ultrasound CT/MRI Echo PET
Bacteriological	Blood culture x2 Specific site culture (e.g. CSF) PCR



*In the UK, the Royal College of Physicians recommends screening for sepsis with a NEWS ≥ 5. The Academy of Medical Royal Colleges (AoMRC) clinical decision support framework for initial evaluation of sepsis in adults ≥ 16 years is summarized in Box 12.4 and The Sepsis Six care bundle for initial treatment and resuscitation of suspected sepsis in Box 12.5.

End-organ failure
- Systolic blood pressure (SBP) <90 mmHg or >40 mmHg fall from baseline or mean arterial pressure (MAP) <65 mmHg
- Bilateral pulmonary infiltrates with no new need for oxygen to maintain saturations >90% or with PaO_2/FiO_2 ratio <300 mmHg or 39.9 kPa
- Serum lactate >2.0 mmol/L
- Serum creatinine >170 mol/L or urine output <0.5 mL/kg per hour for 2 successive hours
- INR >1.5 or an (activated) partial thromboplastin time (PTT) >60 sec
- Platelet count <100 10^9/L
- Bilirubin >32 mol

Fig. 12.1 Management of suspected sepsis. *CCU,* Critical care unit; *CRP,* C-reactive protein; *CT,* computed tomography; *ESR,* erythrocyte sedimentation rate; *FBC,* full blood count; FiO_2, fraction of inspired oxygen; *GCS,* Glasgow Coma Scale score; *Hb,* haemoglobin; *INR,* international normalized ratio; *LFTs,* liver function tests; *MAP,* mean arterial pressure; *MRI,* magnetic resonance imaging; *MSU,* midstream specimen of urine; *NAAT,* nucleic acid amplification technique; PaO_2, arterial oxygen tension; *PCR,* polymerase chain reaction; *PET,* positron emission tomography; *SOFA,* sequential organ failure assessment; *WBC,* white cell count. (*Modified from Singer M, Deutschman CS, Seymour CW et al. The Third International Consensus Definitions for Sepsis and Septic Shock [Sepsis-3].* JAMA *2016; 315:801–810.*)

Examination

Highlighted in red are sites of infection that are common causes of sepsis, with associated clinical signs. This list is not exhaustive and, in a patient with the sepsis syndrome but no clear source, careful examination is vital to establish the underlying diagnosis.

SITES AND SIGNS OF INFECTION

Observation chart
Fever
Tachypnoea (a key sign of clinical deterioration)
Tachycardia and hypotension
Hypoxia (may suggest lower respiratory tract infection)

General observation
Appears 'unwell'
Hot to touch
Confused or reduced consciousness

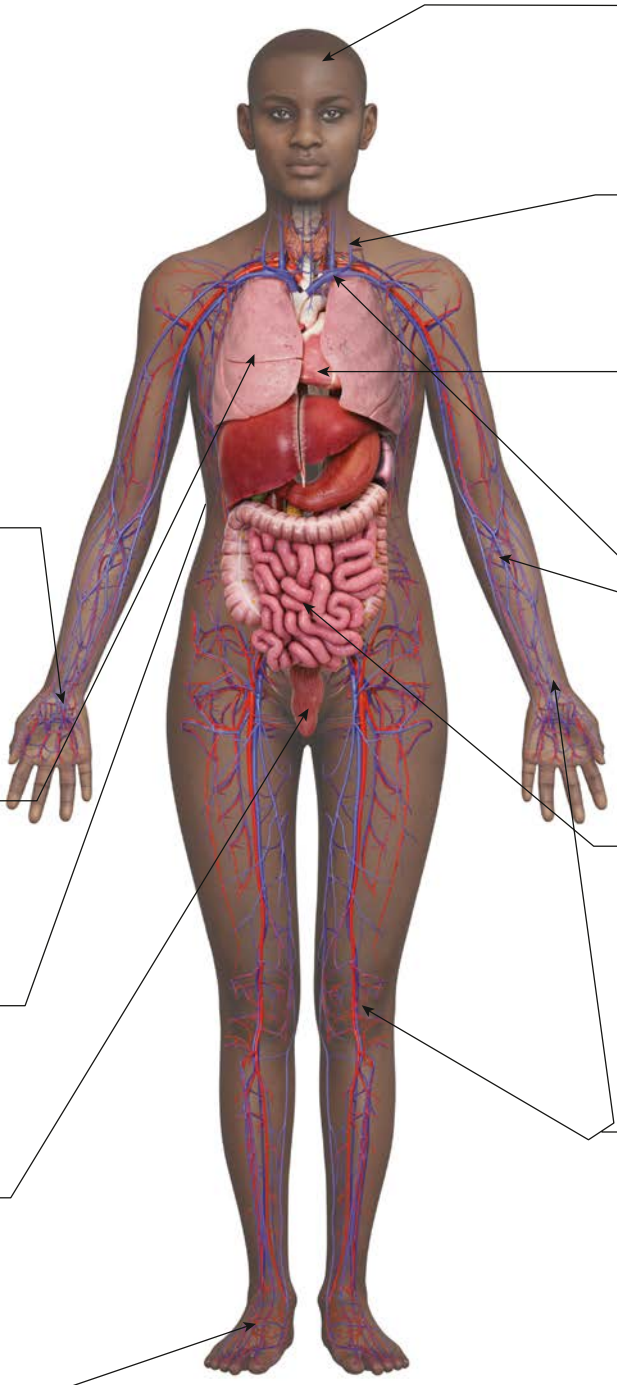
Pulse
Tachycardia common; acute atrial fibrillation possible
May be high-volume ('bounding') due to vasodilation
May be low-volume if 'peripherally shut down' due to relative hypovolaemia

Chest
Tachypnoea
Consolidation may be present in lower respiratory tract infection

Back
Localized tenderness may indicate discitis or spinal abscess
Loin tenderness common in pyelonephritis

Genital examination
Males – signs of epididymo-orchitis
Females – pelvic inflammatory disease; consider retained tampons causing toxic-shock syndrome

Feet
Infected ulcers (especially in patients with diabetes)



Central nervous system
Drowsiness and confusion common signs of sepsis in the elderly
Examine for signs of meningitis

Jugular venous pressure
May be reduced in dehydration/volume depletion

Heart
New or changed murmur may indicate infective endocarditis

Indwelling devices (e.g., Hickman lines)
Examine for signs of intravenous device infection such as discharge or induration
Obtain additional culture from line where possible

Abdomen
Examine for signs of peritonitis
Right upper quadrant tenderness common in biliary sepsis
Right lower quadrant tenderness seen in appendicitis, pelvic inflammatory disease and cystitis
Left-sided tenderness common in diverticulitis

Extremities
Coolness (may indicate hypotension or intravascular volume depletion)
Cellulitis or soft tissue infection (consider devastating infection such as necrotizing fasciitis)
High index of suspicion for septic arthritis if any joints are swollen

BACKGROUND

Sepsis is a rare but serious complication of infection. The syndrome arises when the adaptive immune response to infection becomes maladaptive, leading to organ dysfunction. Multiorgan failure in septic shock is the final common pathway for most patients who die from infection.

PATHOPHYSIOLOGY

The pathophysiology of sepsis is complex and includes the activation of both pro- and anti-inflammatory mechanisms, which can both contribute to clearance of infection but also organ injury. Pathogen-associated molecular patterns (PAMPs), which include components of bacterial, fungal, and viral pathogens, activate the innate immune system, which is chiefly comprised of macrophages, monocytes, granulocytes, natural killer cells, and dendritic cells (see Chapter 6). This activation occurs through Toll-like and C-type lectin receptors on cell surfaces and nucleotide-binding oligomerization domain (NOD)-like and retinoic acid-inducible gene-I (RIG-I)-like receptors in the cytosol. The downstream effects of these activation events, mediated through the upregulation of multiple genes, include the production of pro-inflammatory cytokines such as tumour necrosis factor-alpha (TNF-α) and interleukins 1 (IL-1) and 6 (IL-6). Further effects include increased innate immune activity, production of hepatic acute phase proteins (e.g., complement, fibrinogen, and C-reactive protein [CRP]), release of microparticles containing inflammatory, pro-oxidant and procoagulant lipids and proteins, and elaboration of neutrophil extracellular traps (NETs), which are scaffolds of DNA and antimicrobial proteins and enzymes. The release of microparticles and NETs, combined with increased intravascular tissue factor expression, leads to the phenomenon of 'immunothrombosis', whereby microbes are trapped in microthrombi, which, in turn, attract and further activate leucocytes.

This highly conserved innate response is usually very effective at controlling and neutralizing local infection. When the response exceeds a certain threshold, damage to the host becomes clinically evident. This may be mediated through mitochondrial dysfunction, leading to deficient adenosine triphosphate (ATP) production, and through widespread immunothrombosis as complement activation increases vascular permeability and causes disseminated intravascular coagulation (DIC); this leads to further inflammation, impaired microvascular function and oxygen delivery, and subsequent organ dysfunction. Sepsis becomes a self-reinforcing pathophysiological phenomenon, associated with fierce catabolism to fuel it.

Organ dysfunction

At organ level, sepsis represents a systemic loss of homeostasis, the dysfunction of each organ system adversely affecting the function of every other. The physiological demand requires an increased cardiac output, which most patients achieve after fluid resuscitation. However, the syndrome itself, and associated acidosis, can exert a negatively inotropic influence on the heart (septic cardiomyopathy). Vasoplegia leads to reduced total peripheral resistance, and capillary leak of plasma and protein-rich fluid causes tissue oedema, which in turn impairs oxygen delivery to cells. In the lungs, it can cause acute lung injury and adult respiratory distress syndrome (ARDS). Epithelial disruption in the gut can lead not only to malabsorption but also to translocation of bacteria into the bloodstream, further amplifying the septic insult. Acute kidney injury (AKI) is common and associated with an increased risk of death. Brain dysfunction manifests as septic encephalopathy, ranging from mild drowsiness and confusion to coma.

Consensus definition (Sepsis-3)

The complexity of the sepsis syndrome does not lend itself to a simple diagnostic test. Many biomarkers have been proposed but none is specific enough to be used to define the condition.

In 1991, the first consensus conference to agree a definition of sepsis allied it to the concept of the systemic inflammatory response syndrome (SIRS). This definition was modestly revised in 2001. The latest definition was published in 2016. In anticipation of further updates, as our understanding of the syndrome evolves, it was given the name 'Sepsis-3'.

Sepsis is currently defined as '***life-threatening organ dysfunction caused by a dysregulated host response to infection***'. SIRS is no longer used to define sepsis, although components of this syndrome (e.g., pyrexia, neutrophil leucocytosis) remain important in recognizing the presence of infection; the term 'sepsis' is reserved the syndrome of organ dysfunction attributable to the dysregulated host response to infection in contradistinction to simple infection. Clinicians must be clear about the diagnosis. Inappropriate use of the term 'urosepsis' to describe a simple urinary tract infection (UTI), for example, is misleading and distorts healthcare statistics.

Examination of a large dataset of electronic patient records enabled the authors of the Sepsis-3 definition to construct clinical criteria that best predicted which patients with infection developed life-threatening organ dysfunction, for which the surrogate measure of in-hospital mortality or a requirement for 3 or more days of intensive care treatment was used. In intensive care this consisted of changes in the Sequential Organ Failure Assessment (SOFA) score (**Box 12.2**), where are an increase of SOFA score by ≥2 points from baseline consequent on infection satisfies the Sepsis-3 criteria for organ dysfunction.

The best predictor of death or a requirement for 3 days or more of intensive care for ward patients was a new construct, the 'quick SOFA' (qSOFA) score, consisting of three simple parameters (GCS, respiratory rate, systolic blood pressure); **Box 12.3**. qSOFA is an effective bedside clinical tool to rapidly identify patients at high risk of mortality and who may require admission to critical care. In the UK the Royal College of Physicians recommends using a National Early Warning Score (NEWS 2) of 5 or more as an indicator that a patient is at higher risk of having sepsis (see Fig. 13.1). Remember that neither the qSOFA nor the NEWS 2 are themselves diagnostic of sepsis; they merely indicate that a patient is at higher risk of having sepsis and should be assessed for the condition (**Fig. 12.1**).

Septic shock

The term 'septic shock' is reserved for the subset of patients with sepsis whose circulatory and metabolic dysfunction is such that their risk of death is significantly increased. Clinically, patients are identified as in septic shock if they require vasopressor support to maintain a mean arterial pressure of 65 mmHg or more and have an elevated serum lactate concentration (>2 mmol/L) despite adequate volume resuscitation.

Further reading

Cecconi M, Evans L, Levy M et al. Sepsis and septic shock. *Lancet* 2018; 392:75–87.
Singer M, Deutschman CS, Seymour CW et al. The third international consensus definitions for sepsis and septic shock (Sepsis-3). *JAMA* 2016; 315:801–810.

Box 12.2 Sequential Organ Failure Assessment (SOFA) score[a]

	0	1	2	3	4
Respiratory P/F (kPa[b])	≥53.3	<53.3	<40	<26.7 and MV	<13.3 and MV
Coagulation Platelet count ($\times 10^9$/L)	≥150	<150	<100	<50	<20
Liver Bilirubin (μmol/L)	<20	20–32	33–101	102–204	>204
Cardiovascular	MAP ≥70 mmHg	MAP <70 mmHg	Dopamine <5 μg/kg per min or dobutamine (any dose)	Dopamine 5.1–15 μg/kg per min or adrenaline (epinephrine) or noradrenaline (norepinephrine) ≤0.1 μg/kg per min	Dopamine >15 μg/kg per min or adrenaline (epinephrine) or noradrenaline (norepinephrine) >0.1 μg/kg per min
Central nervous system Glasgow Coma Scale score	15	13–14	10–12	6–9	<6
Renal Creatinine (μmol/L)	<110	110–170	171–299	300–440	>440
Urine output (mL/day)	—	—	—	<500	<200

[a]An increase in SOFA score of ≥2 points identifies patients with organ dysfunction for the purposes of the Sepsis-3 definition. Any such patients should be considered for admission to the critical care unit.
[b]PaO_2/FiO_2 (FiO_2 as a fraction ≤1).
MAP, Mean arterial pressure; *MV*, mechanically ventilated.

Box 12.3 The 'quick SOFA' (qSOFA) score

	Score = 1 for each
Glasgow Coma Score (see Chapter 11)	<15
Respiratory Rate (breaths/min)	≥22
Systolic blood pressure (mmHg)	≤100
Patients are 'positive' for qSOFA if score is ≥2	

RECOGNITION OF SEPSIS

At-risk groups

Whether a simple infection will progress to sepsis or not depends on a range of factors, including the virulence of the pathogen, the microbial load, the site of infection, and the host response. The latter is influenced by genetic make-up, co-morbidities, past medical history, and chronic treatments. Groups at higher risk of developing sepsis include:

- older people (>65 years of age) and the very young (neonates)
- people who have previously had sepsis
- people with immunosuppressive medical conditions (such as HIV, asplenism, cirrhosis, autoimmune diseases, cancer)
- people who are iatrogenically immunosuppressed (those on immunosuppressive drugs, including systemic corticosteroids)
- patients with indwelling devices, especially if they breach normal barriers against infection
- pregnant (and recently pregnant) women
- people who abuse alcohol or intravenous drugs.

In-hospital surveillance

Inpatients have a higher risk of developing sepsis than patients in the community. Hospitals should therefore have systems in place to identify deteriorating patients early, aligned with processes that enable a prompt medical response. In the UK the NEWS 2 score (see earlier) assigns a numerical score to patients' observations, weighted according to the degree of deviation from normal values. It has long been known that cardiac arrest is usually preceded by several hours of increasingly deranged physiology, reflected by an increasing NEWS 2 score.

Most hospitals empower ward staff to call for help at certain NEWS 2 thresholds. In many instances this will be on the basis of a total NEWS 2 of 5 or more or on clinical concern even if the numerical score is lower. The request for assistance (conveyed using the SBAR approach, e.g., see Chapter 2) should prompt an urgent medical review, including screening for sepsis.

There is no binary test for sepsis. Screening consists of a risk assessment, carried out by an experienced and informed clinician, of the likelihood that a deteriorating patient is developing sepsis. That likelihood is increased in the presence of extreme physiological derangement, highly suspicious elements of recent medical history, AKI, or raised serum lactate, as incorporated into many guidelines. The diagnosis should be made on the basis of bedside observations and limited point-of-care testing, in order to minimize the time to treatment. Given that early antimicrobial therapy may alter the natural history of the syndrome, waiting for laboratory blood tests and other investigations to confirm the presence of organ dysfunction is counterproductive. Where available, point-of-care lactate measurement can be especially useful. Lactate level is prognostic in patients with sepsis; a raised lactate should never be ignored and is usually a sign of organ dysfunction, even in patients who might otherwise seem relatively well.

Recent guidance from the Academy of Medical Royal Colleges suggests tailoring the response to deteriorating patients with possible sepsis to the degree and rapidity of that deterioration. Early initiation of broad-spectrum antibiotics should be reserved for the sickest (NEWS 2 score of 7 or above, or NEWS 2 of 5 to 6 with other adverse signs), while less unwell patients can tolerate a delay to allow for further investigations to be undertaken (**Box 12.4**).

Box 12.4 The Academy of Medical Royal Colleges (AoMRC) clinical decision support framework for initial evaluation of sepsis in adults ≥ 16 years

Vital signs	Vital signs: NEWS-2 'Physiology first'	0	1–4	5–6	>7
Initial assessment	**History, examination, lab results**	*If clinical or carer concern, continuing deterioration, surgically remediable sepsis, neutropaenia, or blood gas/lab evidence of organ dysfunction, including elevated serum lactate, upgrade actions at least to next NEWS-2 level*			
	Comorbid disease, frailty, patient preferences?	*Consider influence of comorbid disease, frailty, and ethnicity on NEWS-2, and patient preferences for treatment intensity, limits, end-of-life care*			
Initial (generic) actions	**Monitoring and escalation plan**	Standard observations	• Registered nurse review <1 hour • Obs 4–5 hourly if stable • Escalate if no improvement	• Obs hourly • Review <1 hour by clinician competent in acute illness assessment • Escalate if no improvement	• Obs every 30 minutes • Review <30 minutes by clinician competent in acute illness assessment • Senior doctor review <1 hour if no improvement: refer to Outreach or ICU
	Initial treatment of precipitating condition	Standard care	<6 hours	<3 hours	<1 hours
Likelihood of infection and specific actions	Unlikely	Standard care	Review daily and reconsider infection if diagnosis remains uncertain		
	Possible	Review at least daily	**<6 hours:** • Source identification and control plan documented	**<3 hours:** • Microbiology tests • Antimicrobials: administer or revise • Source identification and control plan documented **<6 hours:** • Source control initiated **48–72 hours:** • Review antimicrobials with ID/micro/ senior clinician	**<1 hour:** • Microbiology tests • Antimicrobials: administer or revise (broad-spectrum if causative organism uncertain) **<3 hours:** • Source identification **3–6 hours:** • Source control initiated according to clinical urgency **48–72 hours:** • Review antimicrobials with ID/micro/senior clinician
	Probable or definite	**<6 hours:** • Diagnostic tests • Resuscitation plan	**<6 hours:** • Microbiology tests • Antimicrobials: administer or revise • Source identification and control plan • D/w ID/micro if uncertain, and review		

Bion J, Barton G, Boyle A, Carlson G, Carrol E, Christian W, Crossland S, Faust S, Gabbay F, Gent N, Gibbs P, Harden J, Howard P, Inada-Kim M, Lanzman M, Leanord A, Lees N, Noursadeghi M, Rajesh P, Shrestha S, Sriskandan S, Stockley S, Waldmann C, Walsh T, Wilson P, Singer M. Academy of Medical Royal Colleges Statement on the Initial Antimicrobial Treatment of Sepsis. Academy of Medical Royal Colleges 2022.

Origins of sepsis

Sepsis may be either community acquired, healthcare associated, or hospital acquired. Community-acquired disease is defined as sepsis occurring within 48 hours of hospital admission, whereas hospital-acquired sepsis occurs after 48 hours. Healthcare-associated sepsis, a subset of community-onset sepsis, is defined as sepsis occurring in:

- patients within 30 days of hospital discharge
- nursing home residents but not those living in residential homes
- patients accessing medical treatment in the community (e.g., patients on haemodialysis).

Community-acquired infection tends to be severe and is likely to be caused by virulent but more antibiotic-sensitive organisms. Examples include *Staphylococcus aureus*, *Streptococcus pneumoniae*, *Escherichia coli*, *Klebsiella pneumoniae*, and *Neisseria meningitidis*. The most common sites of infection are the urinary tract, biliary tract, and lower respiratory tract (pneumonia); rarely, there may be infective endocarditis or meningitis.

In contrast, organisms causing ***hospital-acquired infection*** are less virulent but often multidrug resistant; they are therefore harder to treat. Examples include *Stenotrophomonas maltophilia* and *Acinetobacter baumannii*. Infection is often procedure or medical device related. The risk of procedure-related sepsis (e.g., following surgical procedures or tissue biopsy) can be minimized by appropriate antibiotic prophylaxis, and where sepsis is associated with a medical device (e.g., a blocked urinary catheter or infected intravascular catheter), this should be removed (termed 'source control').

Clinical features

Early sepsis can be difficult to diagnose. Signs and symptoms evolve in variable fashion along a continuum from those of localized infection (e.g., cough, dysuria) to more systemic evidence of organ dysfunction (e.g., confusion, oliguria, petechial haemorrhage). Some patients' organ dysfunction evolves gradually, while others may exhibit a precipitous deterioration from minor symptoms to multiorgan failure in a matter of hours. No symptom is pathognomonic but patients typically complain of rigors, breathlessness, muscle pain, vomiting, mottled skin, and a feeling of being extremely unwell, in addition to the symptoms relating to

their underlying infection. Signs correspond to those of the infective source, together with those of systemic inflammation, such as tachycardia, tachypnoea, hypotension, warm peripheries, oliguria, and confusion or drowsiness. Immunosuppressed patients may display relatively few signs and symptoms of inflammation, although those relating to the physiological response should still be present.

Investigations reveal non-specific evidence of inflammation and organ dysfunction, such as raised white cell count and CRP, rising creatinine and bilirubin, falling platelet count, possible infiltrates on chest X-ray, and dysrhythmias on the electrocardiogram (ECG). Imaging may help identify the site of infection.

Sepsis in special situations

Returning travellers

A detailed history is required, with dates of travel, areas visited, and onset of symptoms. Exposures that put travellers at risk of tropical infections (e.g., insect bites or freshwater swimming) should be elicited, and enquiry should be made as to whether patients took malaria prophylaxis or were immunized prior to travel. Assessments should be carried out for viral haemorrhagic fever (in patients with fever within 21 days of return from an endemic region) and Middle East respiratory syndrome coronavirus (MERS-CoV) (in patients with onset of respiratory symptoms within 14 days of arrival from the Middle East). Minimum investigations include blood cultures and a malaria screen, in addition to routine blood tests. The approach to a returned traveller is covered more fully in Chapter 42.

People who inject drugs

Most episodes of sepsis are related to infected thrombophlebitis or local abscesses at or around injection sites. The most common causative bacteria are meticillin-sensitive and meticillin-resistant *S. aureus* (MSSA and MRSA, respectively) and *Streptococcus pyogenes*. Occasionally, staphylococcal infections can disseminate to distant sites, including heart valves, the vertebral column, and joints (see also Chapter 42). If the tricuspid valve becomes infected, embolization to the lungs may occur, resulting in multiple lung abscesses. In addition to source control (e.g., drainage of paraspinal collections, washout of peripheral joints, and, rarely, tricuspid valve replacement), patients often require weeks of antibiotic treatment.

People who inject drugs are more prone to aspiration pneumonia or empyema because of diminished cough reflex and fluctuating levels of consciousness. Rarely, when lemon juice is used as a solvent, yeasts in the blood may cause ocular endophthalmitis. Optimal management requires vitreal amphotericin implants and prolonged antifungal treatment.

Neutropenic patients

Most commonly, sepsis occurs in haematological oncology patients post chemotherapy (see Chapter 11). Due to the risk of overwhelming pseudomonal infection, most empirical treatment includes piperacillin/tazobactam (Tazocin) with or without an aminoglycoside.

Paralysed patients

Acutely, these patients are most commonly cared for in a critical care setting. The greatest risk is that posed by medical devices (e.g., central venous catheters, urinary catheters, and endotracheal tubes). Implementation of care bundles around the placement and aftercare of such devices minimizes the risk of infection, which may be caused by multidrug-resistant organisms that, consequently, are often harder to treat. Pressure sores can be prevented by use of appropriate mattresses and regularly turning the patient. Individuals with spinal cord injuries may present with signs and symptoms of autonomic dysreflexia rather than infection if the site is below the level of injury.

Further reading

Academy of Medical Royal Colleges. *Statement on The Initial Antimicrobial Treatment of Sepsis*. London: AoMRC, 2022.
National Institute for Health and Care Excellence. *NICE Guideline 51: Sepsis: Recognition, Diagnosis and Early Management*. NICE 2016; https://www.nice.org.uk/guidance/NG51.
Royal College of Physicians. *National Early Warning Score (NEWS 2): Standardising the Assessment of Acute-illness Severity in the NHS.* Updated Report of a Working Party. London: RCP, 2017.

MANAGEMENT OF SEPSIS

Surviving Sepsis Campaign guidelines

The cornerstones of management for sepsis are urgent treatment of the infection and support of failing organs. It is likely the patient will need to be admitted to a critical care unit in order to access specialized organ support. Suspected sepsis is a medical emergency, however, so initial treatment and resuscitation should start without delay. The Sepsis Six is a suggested bundle of care, which includes early broad-spectrum antibiotics (**Box 12.5**).

The Surviving Sepsis Campaign (SSC) produces regularly updated international guidelines on the management of sepsis and septic shock. The various recommendations and suggestions are graded according to the quality of research evidence available. The key recommendations are given in **Box 12.6**.

Supporting a failing circulation: fluid resuscitation

Fluid resuscitation is the first step in supporting organ perfusion. The SSC recommends infusing at least 30 mL/kg of crystalloid solution within 3 hours of diagnosis. Such early high-volume resuscitation is common practice in well-resourced healthcare systems. However, trial evidence is emerging that bolus resuscitation is not beneficial in all populations. Two high-quality trials conducted in low-income countries demonstrated that fluid resuscitation in patients with sepsis increased mortality, despite early improvement in indices of perfusion. Proposed explanations for these unexpected findings include lack of critical care support to counteract harmful side-effects of high-volume fluid resuscitation, and excessive reperfusion injury, which may be worse in a population that presents later in the natural history of the condition. Whether earlier initiation of vasopressor therapy (e.g., by administration of lower concentrations of noradrenaline through a peripheral cannula) while reducing fluid volumes may prove beneficial is the subject of ongoing research.

Box 12.5 The Sepsis Six

1. Give oxygen to maintain SpO_2 >94%
2. Take blood cultures
3. Give broad-spectrum intravenous antibiotics
4. Give a fluid challenge
5. Measure lactate
6. Measure urine output and chart hourly fluid balance

Box 12.6 Surviving Sepsis Campaign: main recommendations for early management

Screening
- Have systems in place to screen acutely ill high-risk patients for sepsis

Initial resuscitation
- Instigate i.v. fluid resuscitation with 30 mL/kg crystalloid within the first 3 hours
- Give further fluid guided by haemodynamic assessment and use of dynamic variables to ascertain fluid responsiveness
- Target mean arterial pressure to 65 mmHg
- Guide resuscitation to normalize lactate in patients with raised lactate

Diagnosis
- Obtain cultures before starting antibiotic treatment

Antimicrobial therapy
- Give i.v. antibiotics within 1 hour of recognition of septic shock or a high likelihood of sepsis; for patients at lower risk, a delay of up to 3 hours to enable further investigation is appropriate
- Start with broad-spectrum empirical therapy, with the aim of de-escalating once the pathogen and its sensitivities are known; assess for de-escalation daily

Source control
- Establish anatomical diagnosis and arrangements for emergency source control as soon as possible
- Remove potentially infected vascular access devices as soon as new access is obtained

Fluid therapy
- Apply a fluid challenge technique when on-going fluid administration is required
- Use balanced crystalloids (as opposed to normal saline) if available
- Consider human albumin solution in addition, when large volumes of fluid are required
- Avoid hydroxyethyl starch and gelatin-based solutions

Treating the underlying infection: antimicrobials and source control

A retrospective study of patients with sepsis found that, after the onset of hypotension, for every hour's delay in the administration of broad-spectrum antibiotics, the risk of death rose by 7.6%. More recent data from 49331 emergency department patients in New York State confirm a rise in mortality in direct proportion to the time to first antibiotic administration. This finding is biologically plausible, given that it is the microbial burden that is driving the immune response, which is, in turn, giving rise to the sepsis syndrome; however, the research underpinning these findings is fraught with difficulty (e.g., defining time zero, linearising non-linear changes). The evidence is strongest in the most unwell patients (e.g., those with septic shock), which, along with concerns about rising antimicrobial use and resistance, is the reason for increasingly nuanced recommendations for the timing of antibiotic administration in the guidelines (see above).

Not every infection can be cured with antimicrobials alone. Infected collections with an inadequate or absent blood supply, such as deep abscesses, infected prostheses, or intravascular devices, need physical removal. This process, whether it requires open surgery or percutaneous drainage, is termed source control (see later).

Useful diagnostic samples and laboratory processing

There is always time to obtain blood cultures before starting empirical antibiotic treatment. This is critical, even in an acutely septic or unstable patient. In adults a minimum of 5 mL of blood should be inoculated into both aerobic and anaerobic blood culture bottles. The one exception is in suspected meningococcal septicaemia, where antibiotic administration should never be delayed. Administration of antibiotics prior to sample collection may result in sterile blood cultures, but polymerase chain reaction (PCR) testing for bacterial DNA in whole blood may still enable detection of the causative bacterial pathogen (**Box 12.7**).

For other acute infections the ability to take further samples before antibiotics will depend on the stability of the patient. Examples include midstream urine (MSU), cerebrospinal fluid (CSF), stool, intravenous line tips, and pus. Where bacteria are cultured in pure growth, these samples can help to define sites of infection. The laboratory can also provide susceptibility data for bacteria, particularly when resistance to commonly used antibiotics is unpredictable. An example is a UTI associated with Gram-negative bacteraemia, typically with *E. coli* or *K. pneumoniae*, in which susceptibility to beta-lactams, quinolones, and aminoglycosides, often used as empirical therapy, is highly unpredictable. In stable patients whose site of infection remains undefined, imaging may reveal collections that require drainage or surgical intervention.

Culture-independent molecular methods of identifying pathogens may improve the speed and sensitivity of microbiological diagnosis, although the use of these technologies is not yet widespread (see **Box 12.7**). For example, if samples from a tissue collection are culture-negative post antibiotic exposure, they should be processed additionally for 16S PCR, a molecular test capable of identifying bacterial DNA.

In established chronic infections (e.g., diabetic foot infection, osteomyelitis, and chronic respiratory disorders), where patients rarely exhibit signs of sepsis, empirical treatment can be delayed while deep-tissue, pus, or sputum samples are obtained. Superficial swabs are rarely useful and may confuse management. Samples should always be transported to the laboratory in a timely fashion. As most results are available within 72 hours, this is the point when they may help with rationalizing antibiotic choices.

Source control

Management of infection requires not only appropriate antibiotic treatment but also an understanding of the site of infection. Occasionally, interventions at these sites are required and this is termed 'source control'. For example:
- Biomedical devices are commonly associated with infection and source control necessitates the removal of infected intravascular and urinary catheters, particularly when these are blocked.
- Obstructed biliary and urinary tract systems are common causes of sepsis and, under these circumstances, placement of a common bile duct stent, urinary catheter, or nephrostomy tube may be required.
- Intra-abdominal collections require drainage, as do empyemas and paraspinal collections.
- If a native peripheral joint is infected, an arthroscopic washout is required, and an infected prosthetic joint may require debridement and removal of the prosthesis.
- Valvectomy is increasingly used to optimize management of infective endocarditis, particularly for acute infections caused by *S. aureus.*
- Rapid surgical debridement of infected tissue is required for optimal management of monomicrobial necrotizing fasciitis caused by *S. pyogenes* or Fournier gangrene.

Box 12.7 Molecular identification of infectious agents causing sepsis

While traditional culture-based methods remain the gold standard, their disadvantages include:

- delay in obtaining results (it generally takes around 24 hours for cultures to become positive, and 48 hour for a firm bacterial identification to be made)
- the fact that cultures may be negative if antibiotics were administered before samples were taken.

Two methods have become widely accepted into clinical practice:

16S polymerase chain reaction (PCR) sequencing

- 16S ribosomal RNA is a component of the 30S small subunit of all prokaryotic ribosomes
- Due to the slow evolution of 'conserved' areas within the gene, to which DNA primers can anneal (allowing PCR amplification), sequencing of the hypervariable regions can be used to detect and speciate bacteria from a variety of clinical specimens (e.g., pus and deep tissue samples)
- Unlike culture, this diagnostic test does not provide susceptibility data but is of particular value for clinical samples that are culture-negative due to antibiotic exposure

Matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) mass spectroscopy

- Bacterial isolates are fixed in a matrix, which is then fragmented and subjected to mass spectroscopy
- Each protein-containing fragment of matrix has a different mass/charge profile, allowing a spectral profile to be generated that is unique to individual bacterial species
- Sometimes, genes conferring antibiotic resistance are detectable by MALDI-TOF, e.g., meticillin-resistant *Staphylococcus aureus* (MRSA) produces a different spectral profile to that of meticillin-sensitive *Staphylococcus aureus* (MSSA)

Further reading

Andrews B, Semler MW, Muchemwa L et al. Effect of an early resuscitation protocol on in-hospital mortality among adults with sepsis and hypotension: a randomized clinical trial. *JAMA* 2017; 318:1233–1240.
De Waele J, Girardis M, Martin-Loeches I. Source control in the management of sepsis and septic shock. *Intensive Care Med* 2022; 48:1799–1802.
Evans L, Rhodes A, Alhazzani W et al. Surviving sepsis campaign: international guidelines for management of sepsis and septic shock 2021. *Crit Care Med* 2021; 49(11):e1063–e1143.
Kumar A, Roberts D, Wood KE et al. Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock. *Crit Care Med* 2006; 34:1589–1596.
Maitland K, Kiguli S, Opoka RO et al. Mortality after fluid bolus in African children with severe infection. *N Engl J Med* 2011; 364:2483–2495.

ANTIMICROBIAL STEWARDSHIP

Antimicrobial stewardship, or the prudent use of antibiotics, is a key component in the fight against the emergence of multidrug-resistant organisms. It is both a national and an international public health priority. In lay terms, it can be summarized as administering the right antibiotic for the right condition, given at the right dose, for the right duration.

UK-based targets

In 2017, Commissioning for Quality and Innovation (CQUIN) targets were introduced in the UK for hospital-based National Health Service (NHS) trusts. Key performance indicators (KPIs) were a reduction in total antibiotic prescribing, particularly of carbapenems and tazobactam–piperacillin (Tazocin), and a documented antibiotic review at 72 hours. In subsequent years the reduction in total prescribing and carbapenem usage targets has continued but hospital doctors are now required to specify different prescribing options at 72 hours (see later). In general practice, KPIs are based on reductions in ciprofloxacin and amoxicillin/clavulanate usage and, for uncomplicated UTIs, prescribing nitrofurantoin rather than trimethoprim (as it is more effective against common urinary pathogens, including multidrug-resistant *E. coli*).

Considerations before starting empirical antibiotics

Most hospitals will have an empirical antimicrobial policy. Policies differ, depending on the local prevalence of antimicrobial resistance to key surveillance organisms. To prescribe appropriately likeliest site of infection must be determined, based on clinical assessment and other investigations. Previous microbiology results, such as previous infection/colonization with extended-spectrum beta-lactamase (ESBL)-producing Enterobacteriaceae, may help to determine appropriate choices of antimicrobial: for example, whether or not to prescribe a carbapenem.

Considerations that should be documented when choosing an initial antibiotic regimen include:

- route of administration
- frequency
- duration of treatment
- monitoring for potential toxicity
- where appropriate, drug levels (e.g., pre-dose gentamicin or amikacin levels)
- dose adjustment in renal/hepatic failure
- need for adjuvant therapy (e.g., rifampicin or fusidic acid for severe *S. aureus* infection)
- alternative antibiotics for severe or non-severe penicillin allergy.

Antimicrobial decision-making at 72 hours

At 72 hours, when most culture results are available, one of the following five decisions should be made:

- ***Stop antibiotic treatment.*** Symptoms of sepsis, such as confusion, may be caused by many different factors. After a period of observation and review of investigations, cessation of antibiotics is the appropriate action when patients are thought not to have had an infection after all.
- ***Step down to an oral alternative.*** For uncomplicated infections, e.g., pneumonia or pyelonephritis, treatment can normally be switched from intravenous to oral after 2 to 3 days if the patient is clinically stable and is showing signs of clinical improvement.
- ***Switch treatment.*** This may be necessary because of an unanticipated site of infection (e.g., infective endocarditis requiring prolonged intravenous antibiotic treatment) or unanticipated resistance (such as urosepsis and bacteraemia caused by an ESBL-producing *E. coli* requiring treatment with an intravenous carbapenem).
- ***Continue with intravenous treatment.*** The patient has a more complicated or difficult-to-treat infection, such as meningitis or septic arthritis.
- ***Discharge on outpatient parenteral antibiotic treatment.*** This is defined as the provision of intravenous antibiotics to patients out of hospital in either the community or an ambulatory care setting. For conditions such as skin and soft tissue infection, UTIs, and bone and joint infection, where the patient is otherwise well, treatment may be continued in the community, provided proper governance arrangements are in place (see later).

Where a switch or step down to an oral alternative is made, a key component of good antimicrobial stewardship is the choice of a narrow-spectrum agent when culture results are available. This reduces the risk of *Clostridium difficile* infection and colonization with multidrug-resistant organisms and is therefore both safer for patients and cost-effective.

Outpatient parenteral antimicrobial treatment

Outpatient parenteral antimicrobial treatment (OPAT) has been cited as one of five antimicrobial prescribing decision options in UK Department of Health guidance on antibiotic stewardship after 72 hours of treatment. It may be accessed via several routes:

- through the hospital emergency or ambulatory care department, where patients can be clinically assessed, investigated, and commenced on intravenous antibiotics for conditions such as lower limb cellulitis without hospital admission; OPAT is increasingly used for admission avoidance
- following a hospital admission for infections requiring inpatient admission (such as severe community-acquired pneumonia requiring supplementary oxygen), after the patient improves clinically but still requires intravenous antibiotics
- following a hospital admission required for source control (e.g., surgical removal of an infected joint prosthesis); OPAT is an option for carefully selected groups of patients, as it allows hospital discharge earlier than expected.

OPAT is associated with high levels of patient satisfaction, as most prefer to be treated out of hospital. Because there is less clinical supervision, governance arrangements around patient care are important, and clinical guidelines, published in both the UK and USA, should be complied with.

Conditions suitable for OPAT services

Skin and soft tissue infections—in particular, lower limb cellulitis—are the most common medical conditions referred to OPAT services. Patients are typically treated for 3 to 5 days with intravenous antibiotics but those with lymphoedema or underlying skin conditions typically require longer courses. Increasingly, multidrug-resistant UTIs may be treated in the community with intravenous antibiotics, and patients may be recruited for OPAT by direct laboratory referral after a multi-resistant organism is grown. Those with bone and joint infection, such as vertebral osteomyelitis or native or prosthetic joint infection, invariably require prolonged intravenous antibiotic courses. Other conditions suitable for OPAT include infected diabetic foot ulcers (with or without osteomyelitis), infective endocarditis, empyema, and abscesses of the brain and liver, once patients are deemed clinically stable.

The suitability of a patient to receive OPAT needs careful assessment and is dependent on age, co-morbidities, and severity of infection. OPAT also requires patients to engage reliably with therapy. Therefore intravenous drug users and patients with serious mental health problems are generally not suitable.

Delivery of OPAT services

Commonly used antibiotics are given once daily and reduce nursing time, although some teams can administer intravenous antibiotics up to four times per day. Examples of antibiotics suitable for once-daily dosing include ceftriaxone, ertapenem, amikacin, teicoplanin, and daptomycin. Some services deliver the first dose of antibiotic in hospital where resuscitation equipment is at hand in the event of allergic reaction, but many now give the first dose of intravenous antibiotics in the community to low-risk individuals. OPAT teams may need to monitor drug levels (e.g., pre-dose teicoplanin levels for efficacy, or pre-dose amikacin levels for nephrotoxicity). For most intravenous antibiotics, full blood count, urea and electrolytes, CRP, and liver function tests should be monitored on a weekly basis at least.

Antibiotics can be administered through a peripheral intravenous cannula, but for longer durations of treatment (>7 days), administration through a Hickman line or a peripherally inserted central catheter (PICC) is preferable. Insertion by a suitably trained healthcare professional or interventional radiologist reduces the risk of line infection. Patients can also be taught to self-administer or attach 'infusion devices', commonly known as elastomeric devices, which save district nurse time.

There are different models for delivering OPAT services. These are ***community-based*** (nurses deliver treatment within the patient's home) or ***hospital-based*** (patients attend hospital or an ambulatory care centre on a daily basis for intravenous antibiotic treatment). Ideally, an OPAT multidisciplinary team should consist of doctors, hospital-based nurses, community nurses, and a pharmacist.

Monitoring patients in OPAT services

Patients must be appropriately monitored with clinical assessments, blood tests, and, where relevant, imaging at appropriate intervals. In the UK, some NHS trusts capture patients on 'virtual wards', which facilitates clinical ownership and follow-up. Good practice includes regular multidisciplinary team meetings (where cases are discussed with an infection specialist), the taking of blood tests, and organization of other investigations. While many conditions can be managed exclusively by a clinical microbiologist or ID physician, joint care is required for orthopaedic or neurosurgical cases. The treatment of lower limb cellulitis can be managed by nurses and requires minimal involvement from medical staff.

Strict governance arrangements are required for OPAT services to help mitigate risk from the reduced level of clinical supervision. Adverse drug reactions are not uncommonly reported and central venous access can be associated with access site infections and thrombophlebitis. Overuse of broad-spectrum intravenous antibiotics can be associated with *C. difficile* infection and, depending on the condition being treated, relapse and hospital re-admission can occur. There should therefore be formal re-admission pathways to secondary care. Evidence suggests that OPAT is safe, provided it is administered through a formal service, designed to minimize risk, with doctors and nurses working together across primary and secondary care. OPAT services are expanding across the UK, driven by good safety data, high levels of patient satisfaction, and healthcare efficiencies.

Assessment of allergy risk

Around 10% of the population report a history of penicillin allergy but data show that the true figure is only around 1%. In some infections, penicillins are the first-choice antibiotic and other agents are associated with a worse outcome.

Anaphylaxis to penicillins or any antibiotic can be fatal, so careful evaluation is required. The timing of the reaction is of paramount importance:

- ***Immediate hypersensitivity reactions***, which include anaphylaxis, are immunoglobulin E (IgE)-mediated and classically begin within 1 hour of the dose, and often within minutes. Typically, they are characterized by facial swelling, rash, and severe shortness of breath.

- ***Delayed reactions*** appear after multiple doses of treatment, typically after days or weeks. While they may be immune mediated, they are not associated with anaphylaxis, although in some rare cases they can lead to severe or life-threatening conditions such as Stevens–Johnson syndrome and toxic epidermal necrolysis (see Chapter 11).

When a patient reports an antibiotic allergy the following information should be ascertained:

- the name(s) of the antibiotic(s) they report an allergy to
- the timing between starting the antibiotic and the symptoms or signs
- the nature of the symptoms or signs, asking specifically about rash, wheeze, swelling, and loss of consciousness
- time to resolution of the symptoms
- whether they tried the same or similar antibiotics since.

Severe allergy is typically a hypersensitivity reaction but occasionally can also be a delayed reaction. In type 1 (IgE-mediated allergy, Chapter 6) and other severe penicillin allergies, cephalosporins and carbapenems should also be avoided.

Non-severe allergy is typically non-IgE mediated and is commonly described as a mild rash. In these circumstances, cephalosporins, carbapenems, and monobactams can be cautiously used. Other antibiotic classes are safe alternatives.

Selected patients reporting a penicillin allergy should be referred to allergy services for formal testing. These include:

- patients with a history of an allergic reaction when on multiple drugs, e.g., during general anaesthesia
- patients allergic to multiple antibiotics
- patients for whom there is no reliable alternative antibiotic
- patients in whom there is a current requirement for a prolonged course of antibiotics and an allergy to 'gold standard' treatment
- patients with an absolute requirement for penicillin, e.g., those with central nervous system syphilis, immunodeficiency, or cardiac valve disorders requiring prophylaxis, or post splenectomy
- patients likely to need repeated courses of antibiotics, e.g., those with haematological malignancies, cystic fibrosis, or inflammatory bowel disease.

It is possible to desensitize patients with carefully introduced, gradually increasing doses of specific antibiotics. For example, a well-recognized protocol for co-trimoxazole sensitization has been used since the late 1990s.

Antibiotic chemoprophylaxis

The value of antibiotic chemoprophylaxis has been questioned, as there are relatively few controlled trials to prove efficacy. The evidence for chemoprophylaxis against infective endocarditis is an example. National Institute for Health and Care Excellence (NICE) guidelines recognize that procedures can cause bacteraemia but without a significant risk of infective endocarditis. Even patients at 'high risk', such as those with previous infectious endocarditis, prosthetic heart valves, and surgical shunts, do not always require prophylaxis (**Box 12.8**). However, there are a number of indications for which the prophylactic use of antibiotics is still advised. These

i Box 12.8 Antibiotic chemoprophylaxis

Clinical problem	Aim	Drug regimen[a]
General		
Splenectomy/spleen malfunction	To prevent serious pneumococcal sepsis	Phenoxymethylpenicillin 500 mg twice daily
Rheumatic fever	To prevent recurrence and further cardiac damage	Phenoxymethylpenicillin 250 mg twice daily or sulfadiazine 1 g if allergic to penicillin
Meningitis:		
Due to meningococci	To prevent infection in close contacts	Adults: rifampicin 600 mg twice daily for 2 days Children: <1 year: 5 mg/kg; >1 year: 10 mg/kg Alternative (single-dose) ciprofloxacin 500 mg (p.o.) or ceftriaxone 250 mg (i.m.)
Due to *Haemophilus influenzae* type b	To reduce nasopharyngeal carriage and prevent infection in close contacts	Adults: rifampicin 600 mg daily for 4 days Children: <3 months 10 mg/kg; >3 months 20 mg/kg
Tuberculosis	To prevent infection in exposed (close contacts) tuberculin-negative individuals, infants of infected mothers, and immunosuppressed patients	Oral isoniazid 300 mg daily for 6 months Children: 5–10 mg/kg daily
Endocarditis[b]		
Antibacterial prophylaxis and chlorhexidine mouthwash are *not* recommended for the prevention of endocarditis in patients undergoing dental procedures		
Antibacterial prophylaxis is *not* recommended for the prevention of endocarditis in patients undergoing procedures of the: upper and lower respiratory tract (including ear, nose, and throat procedures and bronchoscopy) genitourinary tract (including urological, gynaecological, and obstetric procedures) upper and lower gastrointestinal tract		
Any infection in patients at risk of endocarditis should be investigated promptly and treated appropriately to reduce the risk of endocarditis		
If patients at risk of endocarditis are undergoing a gastrointestinal or genitourinary tract procedure at a site where infection is suspected, they should receive appropriate antibacterial therapy that includes cover against organisms that cause endocarditis		
Patients at risk of endocarditis should be: advised to maintain good oral hygiene told how to recognize signs of infective endocarditis and advised when to seek expert advice		

[a]Unless stated, doses are those recommended in adults. For surgical procedure, see individual procedures in text.
[b]NICE guidelines for adults and children undergoing interventional procedures, March 2008, updated in 2016.
(Adapted from Joint Formulary Committee. British National Formulary, *70th edn. London: BMJ Group and RPS Publishing; 2015.)*

include surgical procedures carrying a high risk of infection (i.e., colon surgery) or with potentially serious consequences of infection (organ transplantation, post-splenectomy sepsis). The choice of agent(s) is determined by the likely infectious risk and the established efficacy and safety of the regimen. Antibiotic prophylaxis for the vast majority of surgical or radiological procedures should not extend for more than 24 hours post procedure and, for most operations, a single dose at induction is all that is required.

Further reading

National Institute for Health and Care Excellence. *NICE Guideline 15: Antimicrobial Stewardship: Systems and Processes for Effective Antimicrobial Medicine Use*. NICE 2015; https://www.nice.org.uk/guidance/ng15.

ANTIBIOTIC THERAPIES

Timely administration of antibiotics that are active against the infecting pathogen is critical in the effective management of sepsis. Delays in administration of antimicrobial therapy are well documented as having an impact on the survival and outcome of patients with septic shock.

The selection of an antibiotic requires knowledge of its spectrum of antimicrobial activity (**Fig. 12.2**) as well as standard pharmacokinetic considerations of absorption, distribution, metabolism, and excretion (see Chapter 7). Empirical antibiotic prescribing (prior to the availability of microbiological cultures with antibiotic sensitivities) also requires knowledge of the likely culprit organisms determined by infection site and the corresponding recognized antibiotic options (**Box 12.9**). Parenteral administration is usually indicated in the severely ill patient to ensure rapid, high, and consistent blood and tissue concentrations. Whether an antibiotic is able to kill bacteria (bactericidal) or can only inhibit the growth of the organism (bacteriostatic) at the infection site is another factor used for antibiotic selection.

Antibiotics commonly used in the treatment of sepsis

Beta-lactams (penicillins, cephalosporins, monobactams) and carbapenems

Beta-lactams are the most widely used antimicrobials in human medicine. They are naturally derived compounds that share a common ring structure. Modifications to the β-lactam ring and/or the side-chain may expand the antimicrobial spectrum to include many Gram-negative and positive organisms. Following the discovery of penicillin, dozens of β-lactam-based drugs (first to fifth generations) have been developed for clinical use, either from bacterial or fungal sources (e.g., cephalosporins, carbapenems) or by chemical engineering of the β-lactam ring. All β-lactams block bacterial cell-wall synthesis by binding to and inactivating specific penicillin-binding proteins, peptidases, which are involved in the final stages of cell-wall assembly and division. They are generally bactericidal to susceptible bacterial cells in a time-dependent manner.

Many bacteria produce β-lactamase enzymes, which inactivate antibiotics of this class. The emergence of Gram-negative organisms producing extended-spectrum β-lactamases (ESBLs) and carbapenemases (see below) has rendered some bacteria resistant to all β-lactams.

Penicillins

Benzylpenicillin can only be given parenterally and is still the drug of choice for some serious infections. However, due to increasing antimicrobial resistance, it should not be used empirically in serious infections without laboratory confirmation that the organism is penicillin sensitive. Uses include serious streptococcal infections (including infective endocarditis), group A streptococcus *(S. pyogenes)* necrotizing fasciitis, and gas gangrene (usually combined with other antibiotics).

Flucloxacillin is used in infections caused by penicillinase-producing staphylococci and remains the drug of choice for serious infections caused by MSSA.

Amoxicillin is susceptible to β-lactamase but its antimicrobial activity includes streptococci, pneumococci, and enterococci, as well as Gram-negative organisms such as *Salmonella* spp., *Shigella* spp., *E. coli*, *Haemophilus influenzae*, and *Proteus* spp. The extended-spectrum penicillin ***ticarcillin*** is active against Pseudomonas infections, as is the acylureidopenicillin ***piperacillin*** in combination with tazobactam (Tazocin).

Pivmecillinam is used for the treatment of UTI and has activity against Gram-negative bacteria, including ESBL-producing *E. coli*, *Klebsiella*, *Enterobacter*, and *Salmonella* spp. but not *Pseudomonas aeruginosa*.

Temocillin is active against Gram-negative bacteria only, including many ESBL producers. It appears to be less relatively likely to trigger *C. difficile* infection and is increasingly used because of resistance but also as a carbapenem-sparing agent. It is not active against *Pseudomonas* or *Acinetobacter* spp.

Generally, the penicillins are safe. Hypersensitivity (skin rash, urticaria, and anaphylaxis), encephalopathy, and tubulointerstitial nephritis can occur. Amoxicillin and ampicillin produce a hypersensitivity rash in approximately 90% of patients with infectious mononucleosis (glandular fever) who receive this drug. Amoxicillin/clavulanate (co-amoxiclav) (see below) causes cholestatic jaundice six times more frequently than amoxicillin, as does flucloxacillin.

Cephalosporins

The cephalosporins (**Fig. 12.3**) have an advantage over penicillins in that they are not inactivated by staphylococcal penicillinases. Except for the newer fifth-generation cephalosporins ***ceftaroline*** and ***ceftobiprole***, they are not active against MRSA. Activity extends to many Gram-negative and positive organisms, except for enterococci and Gram-negative anaerobic bacteria. Only certain cephalosporins (e.g., ***ceftazidime*** and ***cefepime***) are active against *P. aeruginosa*.

Second- and third-generation cephalosporins are reserved for the treatment of specific serious infections, as empirical use against Gram-negative infections can be rendered ineffective due to ESBL production and has been associated with an increased risk of *C. difficile* infection.

The toxicity is similar to that of the penicillins but is less commonly encountered. Some 10% of patients allergic to one group are also allergic to the other. The early cephalosporins caused proximal tubule damage, although the newer derivatives have fewer nephrotoxic effects.

Monobactams

Aztreonam is the only member of this class available. It is a synthetic β-lactam and, unlike the penicillins and cephalosporins, has no ring other than the β-lactam one: hence its description as a monobactam. Aztreonam spectrum of activity is limited to aerobic Gram-negative bacilli. It is a useful alternative to aminoglycosides in combination therapy, largely for the treatment of intra-abdominal sepsis, as it has activity against some carbapenem-resistant strains

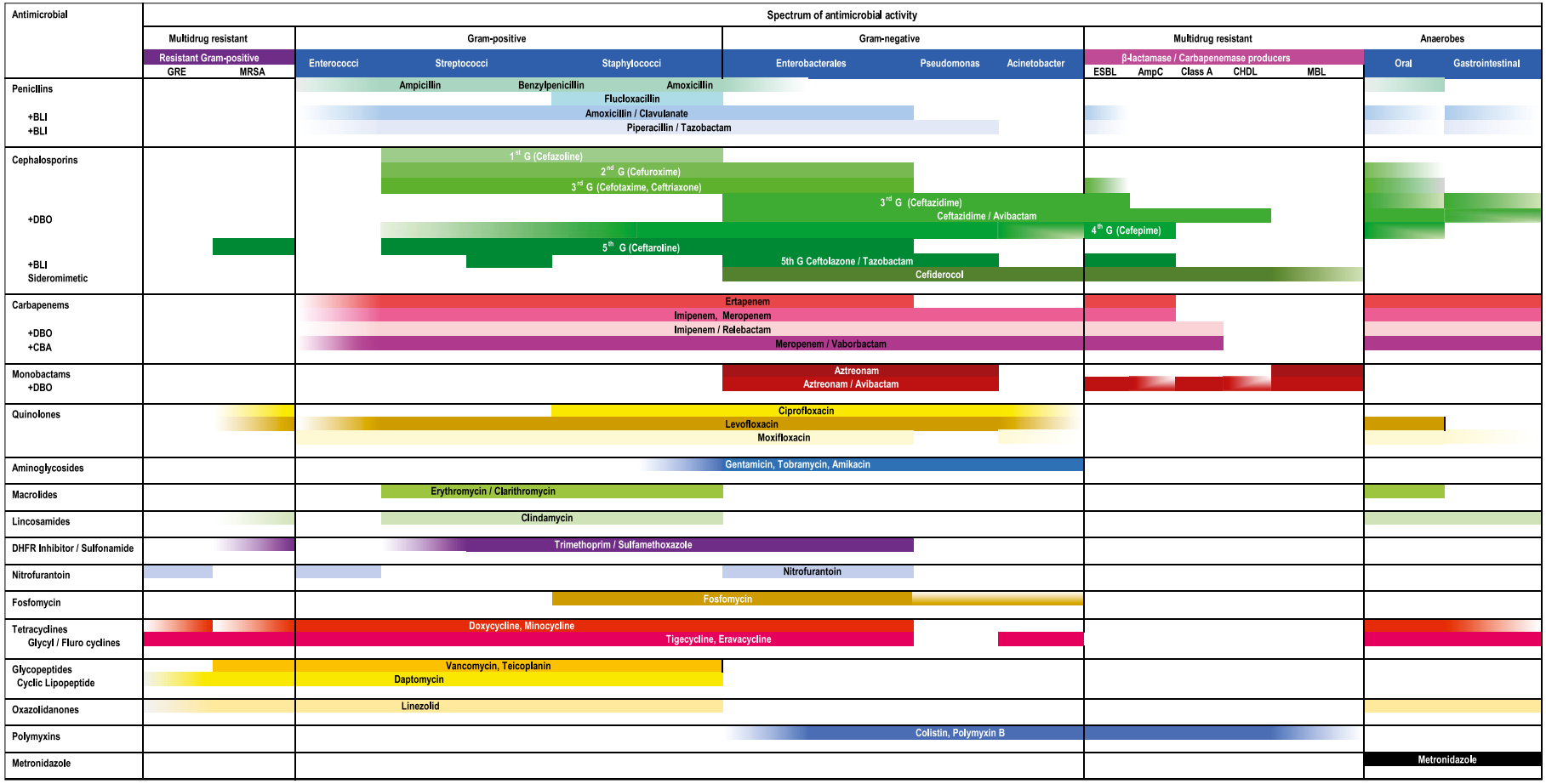


Fig. 12.2 The spectrum of antimicrobial activity. *1st 5th G,* Generation; *BLI,* β-lactamase inhibitor; *CBA,* cyclic boronic acid; *CHDL,* carbapenem-hydrolyzing class D β-lactmase; *DBO,* diazabicyclooctane (a non-β-lactam BLI); *DHFR,* dihydrofolate reductase; *ESBL,* extended-spectrum β-lactamase; *GRE,* glycopeptide-resistant *Enterococcus*; *MBL,* metallo-β-lactamase; *MRSA,* meticillin-resistant *Staphylococcus aureus*.

Box 12.9 Clinical uses of antibiotics

		SSTI	CAP	HAP	UTI	Meningitis	Other uses/notes
PENICILLINS	Benzylpenicillin	+	+			+	
	Amoxicillin		+		+		
	Flucloxacillin	+					Susceptible staphylococcal infections
	Amoxicillin/clavulanate	+	+	+	+		Pyelonephritis, abdominal infections
	Piperacillin/tazobactam	+	+	+	+		Pyelonephritis, abdominal infections, septicaemia
CEPHALOSPORINS	1st G cephalosporins (cefazoline)	+					Surgical prophylaxis
	2nd G cephalosporins (cefotaxime)	+	+	+	+		Pyelonephritis, abdominal infections, surgical prophylaxis
	3rd G cephalosporins (ceftazidime)				+	+	UTI, susceptible Gram-positive and Gram-negative infections
	Ceftazidime/avibactam			+	+		VAP, abdominal infections, susceptible Gram-negative infections
	4th G cephalosporins (cefepime)				+	+	Susceptible Gram-positive and Gram-negative infections
	5th G cephalosporins (ceftaroline)	+	+				
	Ceftolazone/tazobactam			+	+		VAP, abdominal infections
	Cefiderocol						MDR Gram-negative infections with limited treatment options
CARBAPENEMS	Ertapenem	+	+				Abdominal infections
	Meropenem						Susceptible Gram-positive and Gram-negative infections, hospital-acquired sepsis
	Imipenem/relebactam						MDR Gram-negative infections with limited treatment Options
	Meropenem/vaborbactam			+	+		VAP, Gram-negative infections with limited treatment options
MONOBACTAMS	Aztreonam				+		Severe Gram-negative infections
	Aztreonam/avibactam						MDR Gram-negative infections with limited treatment options
QUINOLONES	Ciprofloxacin				+		VAP, pseudomonas infections
	Levofloxacin	+	+	+	+		
	Moxifloxacin	+	+				
AMINOGLYCOSIDES	Aminoglycosides (e.g., gentamicin)				+		Septicaemia
MACROLIDES	Clarithromycin	+	+				
	Clindamycin	+					Abdominal infections
TRIMETHOPRIM	Trimethoprim/ sulfamethoxazole	+			+		
NITROFURANTOIN	Nitrofurantoin				+		
FOSFOMYCIN	Fosfomycin				+		Osteomyelitis
TETRACYCLINES	Doxycycline	+		+			
	Tigecycline	+					Abdominal infections
GLYCOPEPTIDES	Vancomycin	+	+	+			
	Daptomycin	+					Endocarditis
OXAZOLIDINONES	Linezolid	+	+				
POLYMYXINS	Colistin						MDR Gram-negative infections with limited treatment options
METRONIDAZOLE	Metronidazole						Anaerobic infections

The table should not be considered exhaustive guidance for use in clinical practice and does not replace specific local, national, or regional guidelines. Guidance on licensed indications and clinical dosing regimens can be found in the British National Formulary (BNF) | NICE and European Committee on Antimicrobial Susceptibility Testing (EUCAST): Clinical breakpoints and dosing of antibiotics.

1st 5th G, Generation; *CAP,* community-acquired pneumonia; *DHFR,* dihydrofolate reductase; *HAP,* hospital-acquired pneumonia; *MDR,* multidrug resistant; *SSTI,* skin and skin structure infection; *UTI,* urinary tract infection; *VAP,* ventilator-acquired pneumonia.

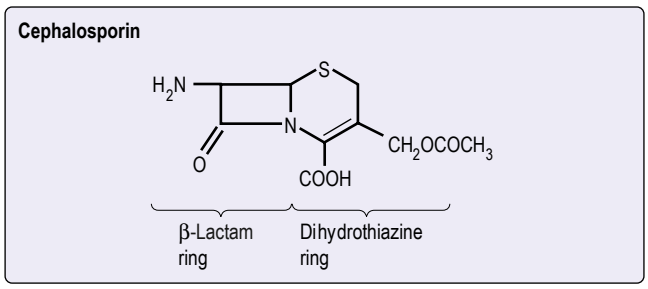


Fig. 12.3 **The structure of a cephalosporin.**

that produce metallo-β-lactamases. It is also used in *P. aeruginosa* infection (including lung infection in cystic fibrosis).

Carbapenems

The carbapenems are semi-synthetic β-lactams and include ***imipenem***, ***meropenem***, ***doripenem***, and ***ertapenem***. They currently have the broadest spectrum of the antibiotics, being active against the majority of Gram-positive, Gram-negative, and anaerobic bacterial pathogens (but not against MRSA). Ertapenem, unlike the others, is not active against *Pseudomonas* or *Acinetobacter* spp. They differ in their dosage and frequency of administration. Imipenem is partially inactivated in the kidney by enzymatic inactivation and is therefore administered in combination with cilastatin.

Carbapenems are the mainstay of treatment for severe ESBL-producing Gram-negative infections (e.g., urinary tract or biliary tract sepsis) and are also used for severe hospital-acquired infections when multidrug-resistant Gram-negative bacilli or mixed aerobic and anaerobic infections are suspected.

Their side-effect profile is similar to that of β-lactam antibiotics. Nausea, vomiting, and diarrhoea occur in less than 5% of cases. Imipenem and ertapenem may cause seizures and should not be used to treat meningitis. Meropenem is safe for this indication.

β-lactam/β-lactamase inhibitor combinations

One approach to the problem of β-lactamase-mediated resistance is to combine β-lactams with 'inhibitor' molecules (BLIs). Classical 'inhibitors' include clavulanic acid, sulbactam and tazobactam given in combination with amoxicillin (amoxicillin/clavulanate), ticarcillin (ticarcillin/clavulanate), ampicillin (ampicillin/sulbactam), piperacillin (piperacillin/tazobactam), and ceftolozane (ceftolozane/tazobactam). These BLIs are themselves β-lactams and act as inhibitors by protecting the partner molecule as suicide compounds that are more readily hydrolysed than the active compound, preserving its antimicrobial activity.

Avibactam is the first non-β-lactam inhibitor of β-lactamase enzymes; it has recently been developed and combined with ceftazidime (ceftazidime/avibactam). The combination appears to be stable to most ESBLs and some carbapenemase (KPC, OXA-48) enzymes in Enterobacterales.

Quinolones

The quinolone group of drugs inhibits bacterial DNA synthesis by inhibiting topoisomerase i.v. and DNA gyrase, the enzymes responsible for maintaining the super-helical twists in DNA.

Extended-spectrum quinolones such as ***ciprofloxacin*** have activity against Gram-negative bacteria, including *P. aeruginosa*, and some Gram-positive bacteria. They are useful in Gram-negative bloodstream infections, bone and joint infections, urinary and respiratory tract infections, meningococcal carriage, some sexually transmitted diseases such as gonorrhoea and non-specific urethritis due to *Chlamydia trachomatis*, as well as in severe cases of travellers' diarrhoea. The newer oral quinolones (e.g., ***levofloxacin***, ***moxifloxacin***) provide an alternative to β-lactams in the treatment of community-acquired lower respiratory tract infections and are effective against *S. pneumoniae*, *H. influenzae*, and the 'atypical' respiratory pathogens. They can also be used in the treatment of tuberculosis.

In many countries a high proportion of *E. coli* and *Klebsiella* spp. are now resistant (>80%). Resistance is also an emerging problem among *Salmonella*, *Vibrio cholerae*, *S. pneumoniae*, and *S. aureus*.

Gastrointestinal disturbances, photosensitive rashes, and occasional neurotoxicity can occur. Use should be avoided in pregnancy and childhood, and in patients taking corticosteroids (due to increased risk of tendinitis), unless the benefit outweighs the risk. Tendon damage, including rupture, can occur within 48 hours of use. MRSA and *C. difficile* infections in hospitals have been linked to high prescribing rates of quinolones, particularly when the O27 hypervirulent strain of *C. difficile* was endemic in most UK hospitals. Use is discouraged where an effective alternative is available. There is concern about QTc prolongation, and concomitant prescribing with other QTc prolongation drugs should be avoided whenever possible.

Aminoglycosides

Aminoglycosides (Fig. 12.4) interrupt bacterial protein synthesis by inhibiting ribosomal function (messenger RNA reading and transfer RNA binding). ***Gentamicin*** and ***tobramycin*** are given parenterally. They are highly effective against many Gram-negative organisms, including *Pseudomonas* spp. They are synergistic at low doses with penicillins against *Enterococcus* spp. and *Streptococcus viridans* and are therefore often used in endocarditis. ***Amikacin*** has a similar spectrum but is more resistant to the aminoglycoside-modifying enzymes (phosphorylating, adenylating, or acetylating) produced by some bacteria, in particular ESBL-producing Enterobacterales. Its use should be restricted to gentamicin-resistant organisms.

Dose-related nephrotoxicity and ototoxicity (vestibular and auditory) can occur, particularly in the elderly. Enhanced nephrotoxicity occurs with other nephrotoxic drugs, and enhanced ototoxicity with some diuretics. The m.1555A > G mutation has been associated with gentamicin ototoxicity but appears to occur in less than 1% of the general population. Monitoring is necessary to ensure therapeutic and non-toxic drug concentrations. Once-daily dosing is used for most indications, with a serum drug level taken at 6 to 14 hours post dose, followed by application of an appropriate nomogram to determine the subsequent frequency of dosing (e.g., every 24 or 48 hours). When aminoglycosides are used for endocarditis, low doses are prescribed every 12 hours with different target pre- and post-dose level ranges than for once-daily dosing. Alternatively, and more simply, a pre-dose level can be checked prior to drug administration and appropriate drug adjustments made with subsequent doses. Neuromuscular blockade can occur with curariform drugs and aminoglycosides should be avoided in patients with myasthenia gravis.

Glycopeptides

The glycopeptides are active against Gram-positive bacteria and act by inhibiting cell-wall synthesis.

Vancomycin is given intravenously for MRSA and other multiresistant, Gram-positive organisms. It is only slowly bactericidal (in contrast to β-lactams). It is also used for treatment of and prophylaxis against Gram-positive infections in penicillin-allergic patients. It is given in *S. pneumoniae* meningitis, in combination with other

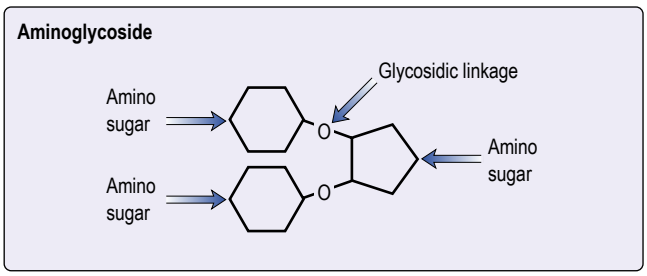


Fig. 12.4 The structure of an aminoglycoside.

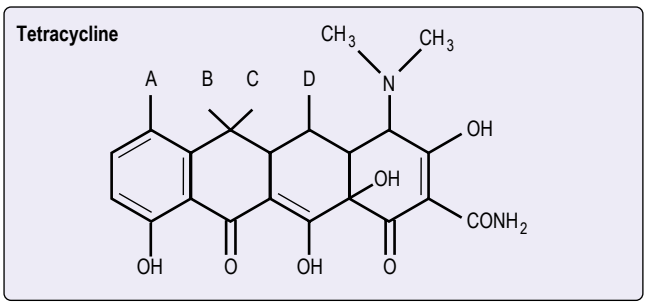


Fig. 12.5 The structure of a tetracycline. Substitution of CH_3, OH, or H at positions A–D produces variants of tetracycline.

effective antibiotics, when disease is caused by penicillin-resistant strains.

Vancomycin can cause ototoxicity and nephrotoxicity, and thus pre-dose (trough) serum levels should be monitored regularly. Serum levels at 1 hour post dose (peak) are also monitored to optimize drug efficacy. Care must be taken to avoid extravasation at the injection site, as this causes necrosis and thrombophlebitis. Too rapid infusion can produce symptomatic release of histamine (red man syndrome).

Teicoplanin is less nephrotoxic than vancomycin. It has more favourable pharmacokinetic properties, allowing once-daily or thrice-weekly dosage. It is given intravenously and pre-dose (trough) serum levels are monitored to optimize efficacy.

Lipopeptides

Daptomycin is a lipopeptide with a similar spectrum to that of vancomycin and is given by the intravenous route. It is used particularly for complicated skin and soft tissue infections, including those caused by MRSA, and is also a useful alternative agent for endocarditis, bone and joint infections, and Gram-positive bloodstream infections. Lipopeptides with very long elimination times, such as ***dalbavancin***, are used as single-dose therapies for skin and soft tissue infections.

Oxazolidinones

Linezolid was the first oxazolidinone antibacterial to be developed. ***Tedizolid*** has recently been approved for use in skin and skin structure infections. These drugs act by inhibiting protein synthesis, binding to the bacterial 23S ribosomal RNA of the 59S subunit, thereby preventing the formation of a functional 70S complex that is essential to bacterial translation.

Oxazolidinones are active against a variety of Gram-positive pathogens, including vancomycin-resistant *Enterococcus faecium* (although resistant organisms have been reported), MRSA, and penicillin-resistant *S. pneumoniae*. They are also active against group A and group B streptococci. Clinical experience with linezolid has demonstrated efficacy in a variety of hospitalized patients with severe to life-threatening infections, including bacteraemia, hospital-acquired pneumonia, skin and soft tissue infection, and bone and joint infection. These drugs can be given both intravenously and by mouth and are almost 100% bioavailable by the oral route in patients with normal gastrointestinal absorption.

Oxazolidinones interact reversibly as non-selective inhibitors of monoamine oxidase and have the potential for interacting with serotoninergic and adrenergic agents. Side-effects include gastrointestinal disturbances, headache, rash, hypertension, reversible but potentially severe cytopenias, and occasional reports of optic and peripheral neuropathy in patients receiving linezolid for longer than 28 days. Tedizolid is less myelotoxic and can also be given once daily (in contrast to twice-daily linezolid). Weekly monitoring of the full blood count for cytopenias, and for other serious adverse effects, is mandatory. Safety has not yet been shown in pregnancy but linezolid has been used successfully for serious infections in children.

Tetracyclines

These are bacteriostatic drugs possessing a four-ring hydronaphthacene nucleus (Fig. 12.5). Included among the tetracyclines are ***tetracycline***, ***oxytetracycline***, ***doxycycline***, and ***minocycline***. ***Tigecycline*** is an injectable glycylcycline that is structurally related to the tetracyclines.

Tetracyclines inhibit bacterial protein synthesis by interrupting ribosomal function (transfer RNA binding) and are active against Gram-positive and Gram-negative bacteria. Tigecycline is active against many organisms resistant to tetracycline, as it avoids common bacterial efflux pump systems. This includes vancomycin-resistant enterococci (VRE), MRSA, and Gram-negative bacilli, such as *Acinetobacter baumannii*, but not *Pseudomonas* or *Proteus* spp. Tigecycline is increasingly used in combination with other antibiotics (e.g., polymyxins) to treat infections caused by highly resistant, carbapenemase-producing, Gram-negative bacteria. The licensed indications are complicated skin and soft tissue infections and intra-abdominal sepsis. Tigecycline is also used to treat spirochaetal and rickettsial infections and also has a role as malaria prophylaxis. However, a 2010 US Food and Drug Administration (FDA) alert raised concerns about the efficacy of tigecycline in some serious infections (notably ventilator-associated pneumonia) and it should be used only on expert advice. The efficacy of tetracyclines is reduced by antacids and oral iron-replacement therapy.

Tetracyclines are generally safe drugs but they may enhance established or incipient renal failure, although doxycycline is safer than others in this group. They cause brown discoloration of growing teeth and thus are not given to children or pregnant women. Photosensitivity occurs in approximately 1 in 20 patients. Nausea and vomiting are the most frequent adverse effects of tigecycline.

Macrolides

Macrolides inhibit protein synthesis by interrupting ribosomal function. ***Erythromycin*** has a similar (but not identical) antibacterial spectrum to that of penicillin and may be useful in individuals with penicillin allergy, especially in the management of bacterial respiratory infections. It can be given orally or parenterally, but oral intake is associated with significant gastrointestinal side-effects, while the intravenous formulation is very irritant and causes phlebitis. For these reasons, ***clarithromycin*** (which has similar antimicrobial properties but fewer side-effects) is often preferred. These drugs are useful in the treatment of pneumonias caused by *Legionella* and *Mycoplasma* spp. They are also effective in the treatment of

infections due to *Bordetella pertussis* (whooping cough), *Campylobacter*, and *Chlamydia* spp. Macrolides are otherwise not usually used for life-threatening or serious infections, such as endocarditis and meningitis.

Other macrolides include ***azithromycin*** and ***telithromycin***. They have a broad spectrum of activity that covers selective Gram-negative organisms *(Salmonella, Shigella)*. Compared with erythromycin, they have superior pharmacokinetic properties with enhanced tissue and intracellular penetration and a longer half-life that allows once-daily dosage. Concern has been raised about the use of azithromycin in bloodstream infections (bacteraemia) because of low serum bioavailability. Azithromycin is also employed for trachoma, cholera, and some sexually transmitted infections.

Erythromycin and other macrolides interact with theophyllines, carbamazepine, digoxin, and calcineurin inhibitors (e.g., ciclosporin), occasionally necessitating dose adjustment of these agents. Diarrhoea, vomiting, and abdominal pain are the main side-effects of erythromycin (less so with clarithromycin and azithromycin) as a consequence of the intestinal prokinetic properties of the macrolides. QTc prolongation is a recognized cardiac effect of the macrolides and may lead to the potentially life-threatening syndrome of 'torsades de pointes' (see Chapter 17). Concomitant use of other drugs that cause QTc prolongation should be avoided unless absolutely essential.

Polymyxins (polymyxin B, colistimethate sodium [polymyxin E])

This is an old antibiotic class that, until recently, was rarely used in clinical practice because of concerns about neuro- and nephrotoxicity. The emergence of multidrug-resistant Gram-negative bacteria, in particular carbapenemase-producing Enterobacterales, has resulted in the increased use of drugs in this class as an agent of last resort, often in combination with other antibiotics. Colistimethate sodium (CMS) is an inactive prodrug, metabolized *in vivo* to the active component ***colistin***. This complicates dosing strategies and increases the potential for toxicity. The mechanism of bacterial killing of polymyxins is thought to be disruption of the bacterial cell membrane following binding to the lipopolysaccharide (LPS) component. They are active against most Gram-negative bacilli (except *Proteus* and *Providencia* spp.) and are administered intravenously for serious infections. Plasmid-mediated resistance, due to the MCR-1 phosphoethanolamine enzyme that modifies bacterial LPS, has recently been described, although mostly in veterinary strains in South and South-east Asia.

Sodium fusidate

Sodium fusidate has a structure resembling that of bile salts and is a potent inhibitor of bacterial protein synthesis. Its entry into cells is facilitated by the detergent properties inherent in its structure. It is mainly used for penicillinase-producing *S. aureus* infections, such as osteomyelitis (it is well concentrated in bone) or endocarditis, and as an adjunctive agent for other staphylococcal infections accompanied by septicaemia. The drug is well absorbed orally but must be given in combination with another staphylococcal agent to prevent resistance, which may occur rapidly.

Sodium fusidate commonly causes gastrointestinal adverse effects and may occasionally be hepatotoxic; however, it is generally a safe drug and can be given during pregnancy if necessary. Concomitant statin use is avoided. Sodium fusidate is also available in topical preparations for use in minor skin conditions (impetigo) but these should be avoided to limit the risk of emerging resistance.

Sulphonamides and trimethoprim

The sulphonamides are all derivatives of the prototype sulfanilamide and act by blocking thymidine and purine synthesis by inhibiting microbial folic acid synthesis. Trimethoprim is a 2,4-diaminopyrimidine, which prevents the bacterial reduction of dihydrofolate to tetrahydrofolate.

Sulfamethoxazole is mainly used in combination with trimethoprim (as ***co-trimoxazole***). Because of its adverse effect profile, use in high-income countries was largely restricted to the treatment and prevention of *Pneumocystis jirovecii* infection and listeriosis; however, it is increasingly being prescribed in hospitals again for other infections, such as acute exacerbations of chronic bronchitis and UTI, as it appears to carry a relatively lower risk for triggering *C. difficile* infection and remains useful for some resistant Gram-negative infections. It may also be given for toxoplasmosis and nocardiosis. ***Trimethoprim*** alone is often used for empirical treatment of UTIs.

Resistance to sulphonamides is often plasmid-mediated and results from the production of a sulphonamide-resistant dihydropteroate synthase. Sulphonamides potentiate oral anticoagulants and some hypoglycaemic agents.

The adverse effects of co-trimoxazole are most commonly due to the sulphonamide component. Sulphonamides cause a variety of skin eruptions, including toxic epidermal necrolysis, the Stevens–Johnson syndrome, thrombocytopenia, folate deficiency, and megaloblastic anaemia, with prolonged usage. They can provoke haemolysis in individuals with glucose-6-phosphate dehydrogenase deficiency and therefore should not be used in such people. Trimethoprim is similar in molecular structure to the potassium-sparing diuretic amiloride; monitoring of electrolytes and renal function is required when using trimethoprim or co-trimoxazole, particularly when the patient is prescribed other potassium-sparing drugs (e.g., angiotensin-converting enzyme [ACE] inhibitors) and during prolonged therapy to avoid hyperkalaemia.

Nitroimidazoles

These agents are active against anaerobic bacteria and some pathogenic protozoa. The most widely used drug is ***metronidazole*** (Fig. 12.6). Others include ***tinidazole*** and ***nimorazole***. After reduction of their 'nitro' group to a nitrosohydroxyl amino group by microbial enzymes, nitroimidazoles cause strand breaks in microbial DNA.

Metronidazole plays a major role in the treatment of anaerobic bacterial infections, particularly those due to *Bacteroides* spp. It is also used prophylactically in colonic surgery. It may be given orally, by suppository (well absorbed and cheap), or intravenously (more expensive). It is also the treatment of choice for mild *C. difficile* infection, amoebiasis, giardiasis, and infection with *Trichomonas vaginalis*.

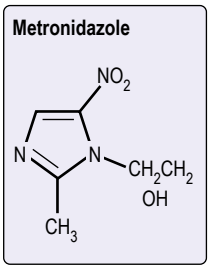


Fig. 12.6 The structure of metronidazole, a nitroimidazole.

Nitroimidazoles can produce a disulfiram-like reaction with ethanol and enhance the anticoagulant effect of warfarin; patients should be warned not to drink alcohol while taking them. They are tumorigenic in animals and mutagenic for bacteria, although carcinogenicity has not been described in humans. They cause a metallic taste and polyneuropathy with prolonged use. High-dose regimens should be avoided in pregnancy and during breast-feeding, unless the benefit is deemed to outweigh the risk.

Other antibiotics

Clindamycin

This is not widely used because of its strong association with *C. difficile* infection. It is active against Gram-positive cocci, including some penicillin-resistant staphylococci, and is a useful agent for severe streptococcal or staphylococcal cellulitis. It has the added effect of inhibiting staphylococcal toxic shock syndrome toxin 1 (TSST-1) and alpha toxin production and has a role in infections caused by *S. aureus*-secreting Panton–Valentine leukocidin (PVL). It is also active against anaerobes, especially *Bacteroides*. It is well concentrated in bone and used for osteomyelitis.

Nitrofurantoin

This is an old agent that is still widely used as an oral treatment for uncomplicated UTI. It is active against almost all the common urinary pathogens, with the exception of *Proteus* spp., and resistance remains rare. It is extensively metabolized, with only about 20% of the active component excreted in the urine, which makes it unsuitable for complicated UTI or pyelonephritis. Nausea is the main side-effect and there are rare reports of peripheral neuropathy, usually in patients with pre-existing renal impairment. It can cause haemolysis in patients with glucose-6-phosphate dehydrogenase deficiency and should be avoided in this group.

Mupirocin

Mupirocin prevents bacterial RNA and protein synthesis. It is only used topically, mainly for the nasal eradication of *S. aureus*, including MRSA, but can also be employed for minor skin infections and treatment of peritoneal dialysis catheter site infection. High-level resistance to mupirocin in MRSA due to plasmid-encoded MupA may lead to failure of topical therapies.

Fosfomycin

This is a relatively old antibiotic that has been used in some European countries for many years. It inhibits bacterial peptidoglycan biosynthesis. It requires a functional sugar transport system (glucose-6-phosphate) for uptake by Enterobacterales, thus requiring modifications to standard susceptibility testing methods. It is active against many Gram-positive organisms and *E. coli*, but many other Gram-negative bacteria have been slow to produce an enzyme (FosA) that is able to destroy the drug. The clinical significance of FosA-mediated resistance is unclear. Fosfomycin retains activity against many ESBL-producing *E. coli* and is therefore the subject of renewed interest. It is increasingly used in the UK, particularly for resistant UTI and is available in oral and intravenous formulations.

Rifaximin

This is a rifamycin with poor gastrointestinal absorption. It is used in portosystemic encephalopathy (see Chapter 23) and the prevention of travellers' diarrhoea; it may be of short-term benefit in irritable bowel syndrome.

Treatment of infection caused by multidrug-resistant organisms

Meticillin-resistant *Staphylococcus aureus*

Meticillin resistance is mediated through changes in the cell-wall penicillin-binding protein 2a, an enzyme that catalyzes cell-wall cross-linking, which meticillin or flucloxacillin is unable to inhibit. In the UK, most MRSA infections are healthcare associated. Uncomplicated infections are typically associated with intravascular catheters or devices; in these cases, line removal (source control) should form part of management. In approximately 5% of cases, infection may disseminate, with haematogenous spread to bones, joints, and heart valves.

For decades the treatment drug of choice has been intravenous ***vancomycin***; pre-levels of 15 to 20 mg/L are required for maximum therapeutic efficacy. Renal function should be closely monitored, as vancomycin is nephrotoxic. Treatment is for a minimum of 2 weeks, and for complicated infection, the addition of a second agent should be considered (e.g., fusidic acid, rifampicin, gentamicin, or ciprofloxacin). There is, however, no good evidence that this improves clinical outcomes: recently, a large UK trial of adjunctive rifampicin demonstrated no clinical benefit in staphylococcal bacteraemic infection.

An alternative to intravenous vancomycin is the oxazolidinone ***linezolid***. The major advantage is that it can be taken orally and is not nephrotoxic. However, it can cause reversible myelosuppression, in particular thrombocytopenia, and so weekly full blood count monitoring is required. In the longer term, there is a smaller risk of peripheral neuropathy or optic neuritis.

While MRSA is primarily seen in the UK in association with intravascular devices, in the USA it has been described as frequently causing severe skin and soft tissue infections. Empirical treatment therefore includes antibiotics such as linezolid, clindamycin, or co-trimoxazole.

Vancomycin-resistant enterococci

Vancomycin resistance in enterococci is mediated through changes in the peptidoglycan precursor from D-Ala-D-Ala to D-Ala-D-Lac, preventing glycopeptides (vancomycin and teicoplanin) binding to these cell-wall precursors and inhibiting cell-wall synthesis. Most infections caused by VRE are healthcare associated. This organism is of low virulence but is difficult to treat, and with the exception of infective endocarditis, infection with VRE is generally of low consequence. More common sites of infection include the urinary and biliary tracts. The only oral treatment is ***linezolid***; intravenous options include ***daptomycin*** and ***tigecycline***.

Carbapenem-resistant *Pseudomonas*

The most common cause of carbapenem resistance in *Pseudomonas* is increased expression of efflux systems or increased impermeability due to decreased porin expression. Most commonly, carbapenem-resistant *Pseudomonas* causes device-related infections or infection of the lower respiratory tract. If disease is mono-resistant, treatment options include ***ciprofloxacin***, ***gentamicin***, ***ceftazidime***, and ***piperacillin/tazobactam*** (Tazocin). Often, however, isolates are multidrug-resistant, with the exception of intravenous ***colistin***. This needs to be dosed appropriately and pre-drug levels should be taken to minimize the risk of both nephrotoxicity and neurotoxicity. More recently, ceftolazone/tazobactam has become available and is effective against multidrug-resistant *Pseudomonas*, as well as less toxic than colistin.

Extended-spectrum β-lactamase-producing Enterobacteriaceae

ESBLs are inactivating enzymes that confer resistance to most β-lactam antibiotics, including penicillins, cephalosporins, and aztreonam. In the UK, ESBL-producing *E. coli* of the CTX-M-15 type suddenly emerged in 2004 and approximately 15% of all *E. coli* bacteraemic isolates are now ESBL producers. Other Enterobacteriaceae, most commonly *K. pneumoniae*, can also be ESBL producers. The most common sites of infection for ESBL-producing Enterobacteriaceae are the urinary tract and, less frequently, the gastrointestinal or hepatobiliary tract. For both organisms, these isolates are often multidrug resistant and may express resistance to antibiotics such as ciprofloxacin, co-trimoxazole, and gentamicin. The mainstay of treatment is a carbapenem, either intravenous ***meropenem*** three times daily or, if OPAT is considered, once-daily intravenous ***ertapenem***. Potential carbapenem-sparing options include intravenous ***amikacin*** or intravenous ***temocillin*** but the suitability of these choices depends on sites of infection. Even when the infecting organism is shown to be susceptible to penicillins by minimum inhibitory concentration (MIC), a carbapenem should be used, as piperacillin/tazobactam has been found to be inferior to carbapenems in a recent trial.

Carbapenem-resistant Enterobacteriaceae

Carbapenem resistance is mediated through the production of enzymes that inactivate carbapenems, together with most β-lactams, and are therefore called carbapenemases. Like ESBL-producing Enterobacteriaceae, carbapenem-resistant Enterobacteriaceae (CRE) most frequently cause UTIs. Worldwide, the five most common types of carbapenemase are *K. pneumoniae* carbapenemases (KPC), New Delhi metallo-β-lactamases (NDM), Verona integron-encoded metallo-β-lactamase (VIM), imipenemase (IMP) metallo-β-lactamase, and the oxacillin carbapenemase (OXA). In the UK, the most common CRE types are OXA-48, NDM, and KPC. Carbapenem-resistant isolates with OXA-48 or KPC production may be susceptible to quinolones and aminoglycosides, and so other commonly used therapeutic options are available for treatment. Resistance to ceftazidime/avibactam in KPC and OXA-48 producers has been reported but remains rare. Isolates that produce NDMs tend to be pan resistant, with the exception of ***colistin***, ***fosfomycin***, and ***tigecycline***. Susceptibility testing for each of these drugs against CRE isolates is not well standardized. Genetic analysis suggests that *Klebsiella* is intrinsically resistant to fosfomycin and that heteroresistance to colistin and tigecycline can emerge during therapy. In these circumstances, colistin is the mainstay of treatment, with tigecycline and sometimes fosfomycin used as adjuvant therapy, depending on susceptibility data. Meropenem, given as a bolus or an infusion, is also of benefit as targeted therapy when organisms demonstrate intermediate sensitivity, as evidenced by a MIC of 8 to 16 mg/L.

Until recently an understanding of the mechanisms of CRE production was helpful when choosing different treatment options. Many of these treatments, described above, have been superseded by cefiderocol, a 5th-generation cephalosporin, which is also effective against NDM producers. The advantages of cefiderocol are monotherapy, simplicity, better tolerability, and less toxicity. While currently most isolates are susceptible, isolates need to be referred to a reference laboratory for testing.

Further reading

Tängdén T, Ramos Martín V, Felton TW et al. The role of infection models and PK/PD modelling for optimising care of critically ill patients with severe infections. *Intensive Care Med* 2017; 43:1021–1032.
http://www.eucast.org *European Committee on Antimicrobial Susceptibility Testing*.

Questions for this chapter can be found on Elsevier eBooks+.

13 Critical care medicine

Stephen J. Shepherd, Colette Coyle, Andrew W. Leitch, and Parjam Zolfaghari

Background 297
Clinical approach to the critically ill patient 298
Recognition and diagnosis of critical illness 298
General aspects of managing the critically ill 298
Discharge from the ICU/HDU 300
Applied cardiorespiratory physiology 300
Oxygen delivery and consumption 300
Cardiac output 303
Cardiovascular assessment and monitoring of critically ill patients 304
Less invasive techniques for assessing cardiac function and guiding volume replacement 306
Disturbances of acid–base balance 307
Shock, sepsis, and acute disturbances of haemodynamic function 308
Respiratory failure 318
Acute respiratory distress syndrome 325
Acute kidney injury 327
Neurocritical care 328
Outcomes 330
Withholding and withdrawing treatment 330
Death by neurological criteria 330

CORE SKILLS AND KNOWLEDGE

Critical care medicine is concerned predominantly with the management of patients with acute life-threatening conditions in specialized units. In addition to emergency cases, intensive care units (ICUs) electively admit high-risk patients after major surgery. The proportion of a hospital's resources dedicated to intensive care varies widely internationally, with the UK providing 6.6 ICU and high-dependency unit (HDU) beds per 100 000 population and Germany providing 29.2.

ICU staff also provide care throughout the hospital in the form of medical emergency and outreach teams. These teams recognize critically ill patients and provide resuscitation and safe transport to the ICU. They also identify patients who are deteriorating on the ward and intervene early, potentially preventing the need for intensive care admission.

Key skills provided by specialists in critical care medicine include:

- Recognizing and managing acutely ill and deteriorating patients and implementing basic resuscitation with advanced medical care and organ support
- Working as part of a multidisciplinary team that includes medical, nursing, and allied health professionals
- Developing an ability to communicate effectively, taking into consideration the risks and benefits of critical care management and advanced life support, especially when these treatments may be futile and not in the best interest of the patient.

Opportunities for learning critical care medicine include attending ward rounds on ICUs and HDUs, accompanying critical care outreach teams who are reviewing deteriorating patients on the wards, observing key procedures such as endotracheal intubation or arterial cannulation, and observing the titration of organ support systems as patients' clinical parameters change.

BACKGROUND

Critical care medicine describes the care given to some of the sickest patients (**Box 13.1**). This might be provided initially by outreach teams in clinical areas such as the emergency department or general medical or surgical wards, but appropriate patients are transferred to either an ICU or an HDU.

ICUs are usually reserved for patients with established or impending organ failure and provide facilities for the diagnosis, prevention, and treatment of multiple organ dysfunction. They are fully equipped with monitoring and technical facilities, including an adjacent laboratory and 'near-patient testing' devices for the rapid determination of blood gases and simple biochemical data such as serum potassium, blood glucose, and serum lactate levels. Technological advances have introduced more compact and complex devices to manage ventilation, cardiac and renal support, and extracorporeal therapies. Portable ultrasound and echocardiography equipment is commonly available. Most importantly, patients receive continuous expert nursing care and the constant attention of appropriately trained medical staff.

HDUs offer a level of care between that available in the general ward and that provided in an ICU. They ensure monitoring of and support for patients with acute (or acute-on-chronic) single-organ failure and for those who are at risk of developing organ failure. They can also provide a 'step-down' facility for patients being discharged from intensive care.

The different levels of hospital critical care bed are discussed further in Chapter 11.

Box 13.1 Some common indications for admission to intensive care

Surgical emergencies
- Acute intra-abdominal catastrophe:
  - Perforated viscus, especially with faecal soiling of peritoneum (often complicated by sepsis/septic shock)
  - Ruptured/leaking abdominal aortic aneurysm
- Trauma (often complicated by hypovolaemic shock and later sepsis/septic shock):
  - Multiple injuries
  - Massive blood loss
  - Severe head injury

Medical emergencies
- Respiratory failure:
  - Exacerbation of chronic obstructive pulmonary disease (COPD)
  - Acute severe asthma
  - Severe pneumonia (often complicated by sepsis/septic shock)
- Cardiac failure (cardiogenic shock) often following myocardial infarction, arrhythmias, and cardiac arrest
- Meningococcal infection
- Status epilepticus
- Severe diabetic ketoacidosis
- Coma

Elective surgical admissions
- Extensive/prolonged procedure (e.g., oesophagogastrectomy)
- Cardiothoracic surgery
- Major head and neck surgery
- Coexisting cardiovascular or respiratory disease

Obstetric emergencies
- Severe pre-eclampsia/eclampsia
- Haemorrhage
- Amniotic fluid embolism

CLINICAL APPROACH TO THE CRITICALLY ILL PATIENT

Recognition and diagnosis of critical illness

Early recognition, targeted resuscitation, and stabilization are fundamental to the successful management of the critically ill. In order to facilitate the identification of 'at-risk' patients on the ward and early referral to the critical care team, a number of early warning systems have been devised (e.g., the Modified Early Warning Score, MEWS). These are based primarily on bedside observations, assigning a single score reflecting the degree of physiological abnormality, which can be used to flag patients who require further assessment. The MEWS score has been superseded by the National Early Warning Score (NEWS2) in the UK, which has been optimized to recognize sepsis and new-onset confusion (**Fig. 13.1**). On average, 5% of hospital inpatients in the UK will score 5 (key threshold for urgent response) at some point during their admission; of these, 5% will suffer a cardiac arrest, die, or require ICU admission in the following 24 hours.

These early warning systems are not infallible and have not been universally implemented. It is therefore imperative that clinicians are trained to recognize critically ill patients at the bedside. The initial assessment may elicit obvious signs. The patient may be unduly agitated or, perhaps more worryingly, unresponsive. Of particular concern is the obtunded patient who is unable to protect their own airway from aspiration of gastric or oral contents. Snoring, grunting, or other respiratory sounds may indicate an obstructed airway, which can be caused by posterior displacement of the tongue due to lax oropharyngeal musculature in a comatose patient or perhaps by secretions pooling in the oropharynx in a patient with a depressed cough reflex. Obvious use of the accessory respiratory muscles and a tracheal tug are sensitive signs of impending respiratory decompensation. Patients in severe respiratory distress frequently sit forwards, grip the sides of the bed, and cannot complete sentences in a single breath. Review of the nursing observations may reveal a sudden deterioration in recorded variables, such as a sharp rise in temperature, increasing heart rate, a fall in blood pressure, or decreased urine output.

On examination, cool peripheries in conjunction with diaphoresis indicate increased sympathetic drive and may be a sign of cardiogenic shock, hypovolaemia, or hypoglycaemia. In contrast, flushed, warm peripheries may be a sign of a hyperdynamic circulation consistent with sepsis (see Chapter 12). ***Abdominal catastrophes*** are a common cause of an acute deterioration; an abdominal examination should always be performed and may reveal a distended, tender abdomen and often absent or altered bowel sounds consistent with a perforated viscus or ischaemic bowel. ***Blood gas analysis*** should be performed as soon as possible. Acid–base status, haemoglobin concentration, blood glucose, and electrolyte levels obtained from an arterial blood gas sample are helpful when assessing the cause and severity of acute illness. Increased lactate levels usually indicate severe illness. An ***electrocardiogram (ECG)*** can allow the rapid diagnosis of treatable conditions, as can a portable ***chest X-ray***.

Frequently, the precise underlying diagnosis is initially unclear but, in all cases, the immediate objective is to preserve life and prevent, reverse, or minimize damage to vital organs such as the lungs, brain, kidneys, and liver. A systematic approach to the recognition and initial treatment of acute illness should be adopted, as well as the performance of investigations to search for the underlying cause. A rapid assessment following the ABC approach (*A*irway, *B*reathing, *C*irculation; see **Fig. 13.22**) should be undertaken, with simultaneous institution of treatment. In practice, resuscitation, assessment, and diagnosis usually proceed in parallel.

General aspects of managing the critically ill

Critically ill patients require multidisciplinary care that should include the following elements.

Intensive skilled nursing care

In the UK, this usually entails a 1:1 nurse-to-patient ratio in an ICU (level 3 care) or 1:2 in an HDU (level 2 care). Frequent clinical observations and treatment are required.

Monitoring

Every patient in critical care is monitored to a far greater extent than generally practiced on wards. Most patients undergo continuous monitoring of ECG, oxygen saturation, end-tidal CO_2 (if intubated), invasive arterial blood pressure, and central venous

Physiological parameter	3	2	1	Score 0	1	2	3
Respiration rate (per minute)	≤8		9–11	12–20		21–24	25
SpO_2 Scale 1 (%)	≤91	92–93	94–95	≥96			
SpO_2 Scale 2 (%)	≤83	84–85	86–87	88–92 ≥93 on air	93–94 on oxygen	95–96 on oxygen	≥97 on oxygen
Air or oxygen?		Oxygen		Air			
Systolic blood pressure (mmHg)	≤90	91–100	101–110	111–219			≥220
Pulse (per minute)	≤40		41–50	51-90	91–110	111–130	≥131
Consclousness				Alert			CVPU
Temperature (°C)	≤35.0		35.1–36.0	36.1–38.0	38.1–39.0	≥39.1	

Fig. 13.1 The National Early Warning Score (NEWS2) system. S_pO_2 Scale 1 should be used in all patients except those where there is previously confirmed and documented hypercapnoeic respiratory failure demonstrated on blood gas analysis. Aggregate scores of 0 to 4 (low clinical risk) require a ward-based response, and those of 5 to 6 (medium risk) require an urgent response from the clinical team. Aggregate scores of 7 or more (high clinical risk) should receive an urgent emergency response from critical care staff. *CVPU,* Confusion, voice, painful stimulus, unresponsive. *(Reproduced from Royal College of Physicians. National Early Warning Score (NEWS) 2: Standardising the assessment of acute-illness severity in the NHS. Updated report of a working party. London: RCP, 2017.)*

pressure (CVP), with periodic recording of values. Settings and measurements from devices are also recorded (e.g., ventilator settings and measurements such as airway pressures, tidal and minute volume). In addition, strict fluid balance monitoring is essential, with a full record kept of fluids entering the patient (i.v. fluids and blood products, medications, liquid feed, and even fluid orally consumed by awake patients) and fluids leaving (urine output, drain fluid, CSF, with some units making an estimate of insensible losses). Furthermore, systematic regular assessments (e.g., GCS, sedation level, pupillary responses, limb perfusion) are also recorded. Increasingly, records are kept electronically, with monitor and device data downloaded directly into the record after ratification by the bedside nurse.

Specialized physiotherapy

This should include chest physiotherapy, mobilization, and rehabilitation.

Management of pain and distress

There should be judicious administration of analgesics and sedatives (see 'Sedation, analgesia, and muscle relaxation' section).

Constant reassurance and support

Critically ill patients easily become disorientated; delirium (a transient alteration in consciousness, attention, orientation, perception, or behaviour) is common. Delirium may be hypoactive, agitated (hyperactive), or a combination of the two. Pain, advanced age, sleep/sensory deprivation, sedative administration (especially benzodiazepines), alcohol/drug withdrawal, neurological injury, severe illness, and medical co-morbidities all play a role. Patients with delirium are more difficult to wean from ventilation, require larger doses of sedatives, and have an associated increased mortality and length of hospital stay. Treatment focuses on effective pain control, minimizing sensory deprivation, early mobilization, avoidance of benzodiazepines, and limiting sedative administration. Newer sedative agents, such as the α_2-agonist dexmedetomidine, may reduce the incidence of delirium and time on the ventilator, as do processes of care that mandate daily interruptions to sedative infusions and careful titration to the minimum required.

Gastric protection

H_2-receptor antagonists or proton pump inhibitors reduce gastric acidity and prevent stress-induced ulceration. They may, however, encourage bacterial overgrowth in the upper gastrointestinal tract and predispose patients to ventilator-associated pneumonia; they are also associated with *Clostridium difficile* colitis.

Deep vein thrombosis prevention

All critically ill patients are at increased risk of deep venous thrombosis and pulmonary embolism. Prophylaxis should be instigated with the use of compression stockings, pneumatic compression devices and regular subcutaneous heparin.

Mouth care, tooth-brushing, and oropharyngeal suction

These measures help to reduce hospital-acquired infections and ventilator-associated pneumonia by reducing the burden of pathogenic oral flora that may be aspirated into the respiratory tree. Chlorhexidine body washes may reduce the total carriage of resistant microorganisms.

Prevention of constipation and pressure ulcers

Steps should be taken to prevent constipation and pressure ulcers from developing.

Organ support

For example, inotropes and vasopressors may be required for cardiovascular support, invasive and non-invasive ventilation for respiratory failure, and dialysis for renal failure. Specialized units may provide extracorporeal membrane oxygenation (ECMO) for severe respiratory failure, mechanical support of the circulation for cardiac failure, and advanced liver support in the form of liver dialysis (e.g., the molecular adsorbent recirculation system, MARS).

Nutritional support

(See Chapter 25). Protein–energy malnutrition is common in critically ill patients and is associated with muscle wasting, weakness, delayed mobilization, difficulty weaning from ventilation, immune compromise, and impaired wound healing. ***Enteral nutrition***, usually delivered via a fine-bore nasogastric tube, is preferred because it is less expensive, preserves gut mucosal integrity, is more physiological, and is associated with fewer complications. However, the value of enteral nutrition early in the course of an acute illness remains uncertain. If enteral feeding is not possible due to gut dysmotility and malabsorption associated with critical illness, the alternative is ***intravenous (parenteral) nutrition***. This is more invasive and expensive and can be complicated by deranged liver function tests, hypertriglyceridaemia, hyperglycaemia, and increased susceptibility to hospital-acquired infections. Hypocaloric enteral nutrition is usually continued for up to a week in previously well-nourished patients before parenteral nutrition is considered. Administering parenteral nutrition early in order to maintain caloric input and prevent an energy deficit may be detrimental.

Although all nutrition should contain carbohydrates, proteins, lipids, and some micronutrients, the precise optimal formulation is unclear. Supplementation with the amino acid glutamine has theoretical advantages in some conditions (burns and trauma) but not in multiorgan failure. The omega-3 fatty acids derived from fish oils also have potentially beneficial antioxidant activity but have not been convincingly shown to improve outcome, as has been the case for micronutrient supplementation with selenium, copper, manganese, zinc, iron, and vitamins.

Many critically ill patients are at risk of developing a ***'re-feeding syndrome'*** (see Chapter 25) when nutritional support is first initiated and should be monitored closely.

Insulin treatment

Treatment should be given for critical illness-induced hyperglycaemia (which is associated with hospital-acquired infections, renal impairment, and poor wound healing, see Chapter 35). Insulin is often required in high doses due to insulin resistance. Although the use of intensive insulin therapy to achieve 'tight glycaemic control' (blood glucose level between 4.4 and 6.1 mmol/L) was initially shown to improve outcome, subsequent studies found that this approach is associated with an unacceptably high incidence of hypoglycaemia, and possibly increased mortality. Current recommendations suggest that blood glucose levels should be maintained below 10 mmol/L.

Discharge from the ICU/HDU

Discharge of patients from intensive care should normally be planned in advance and ideally take place during normal working hours. Frequently, when the condition of critically ill patients improves, they are initially 'stepped down' to HDU (level 2) care. Premature or unplanned discharge from the ICU or HDU, especially during the night, has been associated with higher hospital mortality and is best avoided. A discharge summary, including 'points to review', should be included in the clinical notes, with a detailed handover to the receiving team (medical, nursing, and therapies). The intensive care team should continue to review the patient (who might deteriorate following discharge) on the ward and be available for advice on further management (e.g., tracheostomy care, nutritional support). In this way, deterioration and re-admission to intensive care (which is associated with a particularly poor outcome), or even cardiorespiratory arrest, are hopefully avoided.

This chapter concentrates on cardiovascular, respiratory, renal, and neurological problems. Many patients also have failure of other organs, such as the liver; treatment of these is dealt with in more detail in the relevant chapters.

Further reading

Adhikari NKJ, Fowler RA, Bhagwanjee S et al. Critical care and the global burden of critical illness. *Lancet* 2010; 376:1339–1341.
Casaer MP, Van den Berghe G. Nutrition in the acute phase of critical illness. *N Engl J Med* 2014; 370:1227–1236.
Coursin DB, Skrobik Y. What is safe sedation in the ICU? *N Engl J Med* 2019; 380:2577–2578.
Frost PJ, Wise MP. Early management of acutely ill ward patients. *BMJ* 2012; 345:e5677.
Fullerton JN, Perkins GD. Who to admit to intensive care? *Clin Med* 2011; 11:601–604.
Gerry S, Bonnici T, Birks J et al. Early warning scores for detecting deterioration in adult hospital patients: systematic review and critical appraisal of methodology. *BMJ* 2020; 369:m1501.
Gunst J, Debaveye Y, Güiza F, et al. Tight blood-glucose control without early parenteral nutrition in the ICU. *N Engl J Med* 2023; 389:1180–1190.
Herridge, MS and Azoulay E. Outcomes after critical illness. *N Engl J Med* 2023; 388:913–992.
NICE–SUGAR Study Investigators. Intensive versus conventional glucose control in critically ill patients. *N Engl J Med* 2009; 360:1293–1297.
Pierre Singer P, Blaser AR, Berger M et al. ESPEN practical and partially revised guideline: clinical nutrition in the intensive care unit. *Clin Nutr* 2023; 42:1671–1689.
Royal College of Physicians. *National Early Warning Score (NEWS) 2: Standardising the Assessment of Acute-Illness Severity in the NHS* Updated report of a working party. London: RCP, 2017.
Vincent JL. Critical care—where have we been and where are we going? *Crit Care* 2013; 17(suppl 1):S2.

APPLIED CARDIORESPIRATORY PHYSIOLOGY

Oxygen delivery and consumption

Normal cellular, tissue, and organ function depend on adequate delivery of oxygen and fuel to mitochondria to produce energy, a process integrating the heart, lungs, and vascular and haematological systems. Adequate blood flow requires proper functioning of the three components of the circulation: the muscle pump, circulating volume, and arterial tone or resistance to flow to perfuse distal tissues. Each area has its own regulatory system and may develop pathology either independently or as a part of multisystem response to an insult such as sepsis.

The term oxygen delivery (DO_2) describes the total amount of oxygen supplied to the body per unit time (**Fig. 13.2**). Its main determinants are the amount of blood pushed through the microcirculation by the cardiac output (Q_t) and the amount of oxygen carried in that blood (i.e., arterial oxygen content, C_aO_2). Oxygen is transported primarily in combination with haemoglobin:

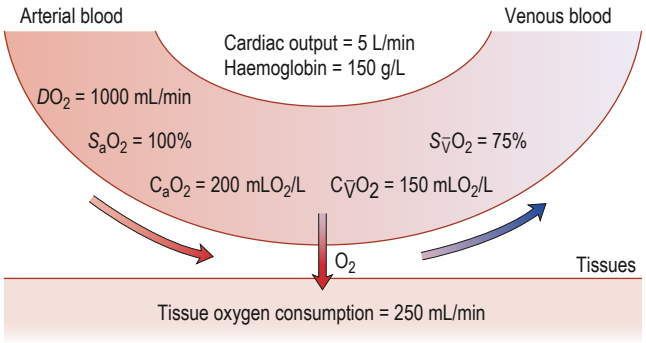


Fig. 13.2 Tissue oxygen delivery and consumption in a normal 70 kg person breathing air. Oxygen delivery (DO_2) = cardiac output × (haemoglobin concentration × oxygen saturation [S_aO_2] × 1.34). In normal adults, oxygen delivery is roughly 1000 mL/min, of which 250 mL is taken up by tissues. Mixed venous blood is thus 75% saturated with oxygen. C_aO_2, Arterial oxygen content; $C_{\bar{v}}O_2$, mixed venous oxygen content; $S_{\bar{v}}O_2$, mixed venous oxygen saturation.

the amount combined is determined by its oxygen capacity (usually 1.34 mL of oxygen per gram) and percentage saturation (SO_2). Oxygen is also dissolved in plasma at a proportion dependent on the partial pressure of oxygen (PO_2), but this is usually insignificant.

'Global' oxygen delivery as a concept is of limited clinical value, as differing tissues have differing metabolic requirements and blood flow. Generally, delivery is matched to individual metabolic requirements; some organs (such as the heart) have high oxygen requirements relative to their blood flow and may receive insufficient oxygen, whereas others such as the kidneys extract relatively little oxygen compared to the 20% cardiac output that flows through them. Microcirculatory flow is also reduced by increased blood viscosity and needs to be considered when thinking about oxygen delivery to the tissues, even if overall measured delivery appears adequate.

Oxygenation of the blood

Oxyhaemoglobin dissociation curve

The saturation of haemoglobin with oxygen is determined by the partial pressure of oxygen (PO_2), with this relationship being described by the oxyhaemoglobin dissociation curve (Fig. 13.3). The sigmoid shape of this curve is significant for a number of reasons:

- Modest falls in the partial pressure of oxygen in arterial blood (P_aO_2) have little effect on oxygen content, provided percentage saturations remain above about 92%.
- Increasing the P_aO_2 to above normal has only a minimal effect on oxygen content, unless hyperbaric oxygen is administered (when the amount of oxygen in solution in plasma becomes significant).
- On the slope of the curve (percentage saturation below 90%), a small decrease in P_aO_2 can cause large falls in oxygen content, whereas increasing P_aO_2 only slightly, e.g., by administering 28% oxygen to a patient with chronic obstructive pulmonary disease (COPD), can lead to a useful increase in oxygen saturation and content.

If the PCO_2 increases, the oxyhaemoglobin curve moves to the right, facilitating oxygen unloading to the tissues (Bohr effect).

The P_aO_2 is influenced, in turn, by the alveolar oxygen tension (P_AO_2), the efficiency of pulmonary gas exchange, and the partial pressure of oxygen in mixed venous blood ($P_{\bar{v}}O_2$).

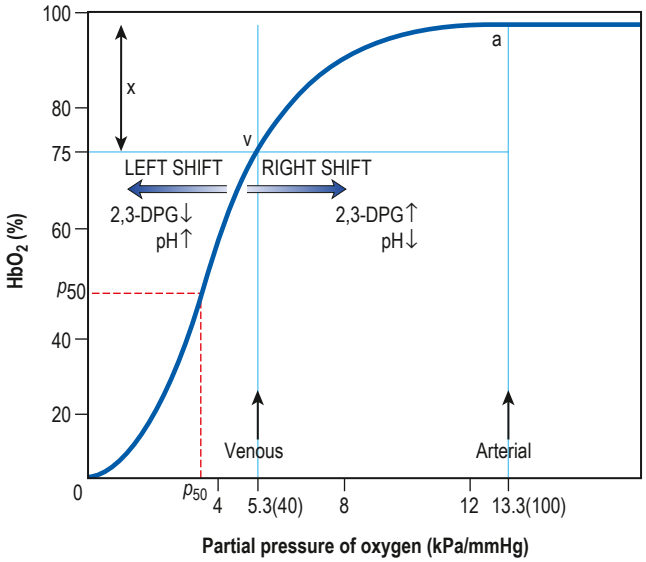


Fig. 13.3 The oxyhaemoglobin dissociation curve. HbO_2 (%) is the percentage saturation of haemoglobin with oxygen. The curve will move to the right in the presence of acidosis (metabolic or respiratory), pyrexia, or an increased red cell 2,3-diphosphoglycerate (2,3-DPG) concentration. For a given arteriovenous oxygen content difference, the mixed venous PO_2 will then be higher. Furthermore, if the mixed venous PO_2 is unchanged, the arteriovenous oxygen content difference increases, and more oxygen is off-loaded to the tissues (see Chapter 38). P_{50} (the PO_2 at which haemoglobin is half-saturated with O_2) is a useful index of these shifts—the higher the P_{50} (i.e., shift to the right), the lower the affinity of haemoglobin for O_2. *a,* Arterial point; *v,* venous point; *x,* arteriovenous oxygen content difference.

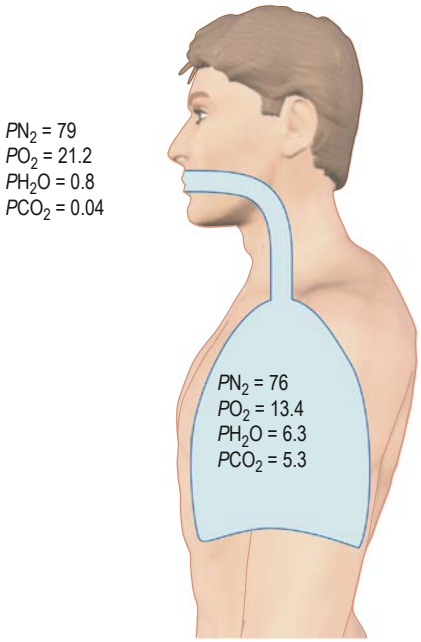


Fig. 13.4 The composition of inspired and alveolar gas. (Partial pressures in kPa.)

Alveolar oxygen tension (P_AO_2)

The partial pressures of inspired gases are shown in Fig. 13.4. By the time the inspired gases reach the alveoli, they are fully saturated with water vapour at body temperature (37°C), which has a

partial pressure of 6.3 kPa (47 mmHg), and contain CO_2 at a partial pressure of approximately 5.3 kPa (40 mmHg); the P_AO_2 is thereby reduced to approximately 13.4 kPa (100 mmHg).

The clinician can influence P_AO_2 by administering oxygen or by increasing the barometric pressure (hyperbaric therapy).

Pulmonary gas exchange

In *normal* subjects, a small alveolar–arterial oxygen difference ($P_{A–a}O_2$) exists due to:

- a small (0.133 kPa [1 mmHg]) pressure gradient across the alveolar membrane;
- a small amount of blood (2% of total cardiac output) bypassing the lungs via the bronchial and cardiac Thebesian veins; and
- a small degree of ventilation/perfusion mismatch.

There are three pathological causes of an increased $P_{A–a}O_2$ difference:

- ***Diffusion defect.*** This is *not* a major cause of hypoxaemia, even in conditions such as lung fibrosis, in which the alveolar–capillary membrane is considerably thickened. Carbon dioxide is also not affected, as it is more soluble than oxygen.
- ***Right-to-left shunts.*** In cyanotic congenital heart disease or following collapse of a lung segment, a proportion of venous blood passes to the left side of the heart without taking part in gas exchange, causing arterial hypoxaemia that cannot be corrected by administering oxygen to increase the P_AO_2 because blood leaving normal alveoli is already fully saturated. Further increases in PO_2 will not, therefore, significantly affect its oxygen content. On the other hand, because of the shape of the carbon dioxide dissociation curve (**Fig. 13.5**), the high PCO_2 of the shunted blood is compensated for by over-ventilating patent alveoli; hence many patients with acute right-to-left shunts hyperventilate, producing a low or normal P_aCO_2.
- ***Ventilation/perfusion mismatch*** (see Chapter 20). This happens when parts of the lung ventilate without much blood flow (e.g., pulmonary embolism), or when there is blood flow but no ventilation, such as in parenchymal lung disease (pneumonia, pulmonary oedema). Both of these states result in hypoxaemia. The increased dead space can be compensated for by increasing overall ventilation. In contrast to the hypoxia resulting from a true right-to-left shunt, that due to areas of low $\dot{V}/\dot{Q}$ can be partially corrected by administering oxygen and thereby increasing the P_AO_2.

Oxygen cascade

Oxygen levels fall further as it is unloaded into the tissues and diffuses to the mitochondria. Tissue oxygen content varies depending on the distance travelled from the local capillary network. Some mitochondria continue to function at a PO_2 as low as 0.07 kPa (0.5 mmHg) (**Fig. 13.6**).

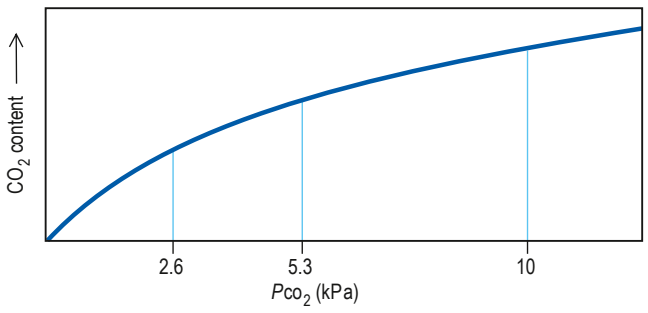


Fig. 13.5 The carbon dioxide dissociation curve. Note that, in the physiological range, the curve is essentially linear.

Mixed venous oxygen tension and saturation

Mixed venous oxygen tension ($P_{\bar{v}}O_2$) is the partial pressure of oxygen in pulmonary arterial blood that has been thoroughly mixed in the right heart. Assuming P_aO_2 remains constant, both $P_{\bar{v}}O_2$ and mixed venous oxygen saturation ($S_{\bar{v}}O_2$) fall as more oxygen is extracted by tissues. Low $S_{\bar{v}}O_2$ therefore indicates either inadequate oxygen delivery or a rise in tissue requirements beyond compensations in cardiac output. This may exacerbate pulmonary shunting; hence worsening hypoxaemia does not necessarily indicate declining pulmonary function. Conversely, rises in venous oxygen content may reflect impaired tissue extraction due to microcirculatory abnormalities and/or reduced utilization due to mitochondrial dysfunction, as in severe sepsis or certain poisons (see later). Monitoring central venous oxygen saturation from the superior vena cava ($S_{cv}O_2$) rather than the pulmonary artery is less invasive and approximates $S_{\bar{v}}O_2$ to within around 5% in most situations, although this relationship becomes less predictable at the extremes of physiology.

Adaptation to hypoxia

Acute exposure to severe hypoxia may lead to sudden death if the immediate adaptive responses fail to maintain mitochondrial oxygen delivery. Chronic exposure to low oxygen tension, on the other hand, allows time for compensatory mechanisms to develop. Among the first is an increase in cardiac output, achieved primarily by increasing heart rate. Given the reciprocal relationship between the partial pressures of oxygen and carbon dioxide in the alveoli, as defined by the alveolar gas equation ($P_AO_2 = P_IO_2 - P_ACO_2/R$, where R is the respiratory quotient [usually 0.8]), an increase in respiratory rate serves to decrease alveolar PCO_2 and thereby increase alveolar PO_2 (see also **Fig. 13.4**). Over time, increased erythropoietin production stimulates haemoglobin synthesis, leading to a marked increase in haematocrit. Over the longer term, capillary bed density in specific tissues adjusts to the physiological demand. In those

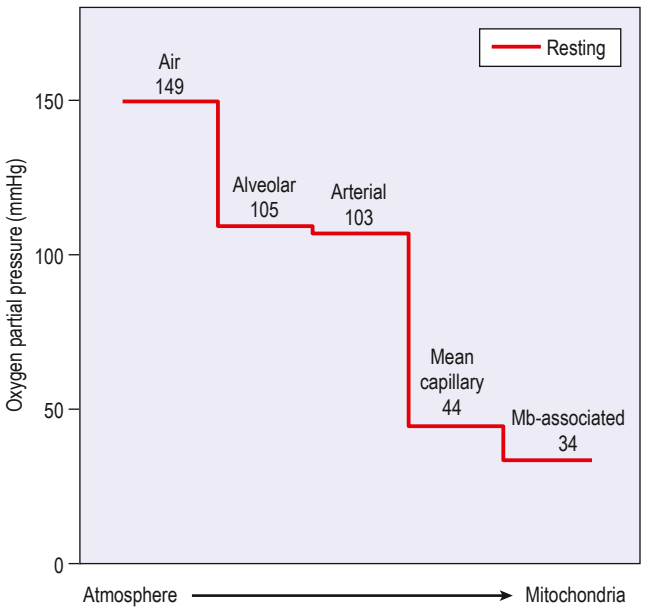


Fig. 13.6 The oxygen cascade. Oxygen levels decrease (shown in graph) as oxygen is unloaded into the tissues and diffuses to the mitochondria. *Mb,* Myoglobin.

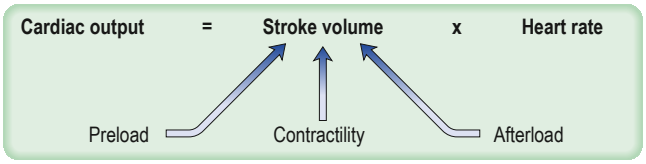


Fig. 13.7 The determinants of cardiac output.

residing at altitude over generations, the hypoxic ventilatory drive and pulmonary vasoconstrictor response evolve to maximize oxygen uptake. In 2007 a series of arterial blood samples were obtained by a group of critical care doctors on the summit of Mount Everest (see 'Further reading'). The project included a 4-month acclimatization period at altitude. The samples were obtained at 8400 m altitude while breathing air. The average P_aO_2 was 3.2 kPa (24.6 mmHg). These figures neatly demonstrate features of both acute (hyperventilation) and chronic (increased haematocrit) acclimatization to hypobaric hypoxia.

Further reading

Dunn JO-C, Mythen MG, Grocott MP. Physiology of oxygen transport. *BJA education* 2016; 16(10):341–348.
Grocott MPW, Martin DS, Levett DZH et al. Arterial blood gases and oxygen content in climbers on Mount Everest. *N Engl J Med* 2009; 360:140–149.

Cardiac output

Cardiac output is the product of heart rate and stroke volume (Fig. 13.7). When the heart rate increases, the duration of systole is largely unchanged but diastole, and thus time for ventricular filling progressively shortens and stroke volume eventually falls. In normal hearts this occurs at rates of more than about 160 beats per minute, but in those with cardiac pathology, particularly restricted ventricular filling (e.g., mitral stenosis), it may be seen at much lower rates. Tachycardia increases myocardial oxygen consumption and may precipitate ischemia in areas with restricted coronary perfusion. Conversely, when heart rate falls, the compensatory increase in stroke volume eventually becomes insufficient and cardiac output falls. Furthermore, rhythm disturbances (e.g., atrial fibrillation, complete heart block) may prevent atrial augmentation of ventricular filling, exacerbating the fall in stroke volume.

Stroke volume

The volume of blood ejected in a single contraction is the difference between the ventricular end-diastolic volume (VEDV) and ventricular end-systolic volume (VESV) (i.e., stroke volume = VEDV – VESV). The ejection fraction describes the stroke volume as a percentage of VEDV (i.e., ejection fraction = (VEDV – VESV)/VEDV × 100%) and is an indicator of myocardial performance.

Three interdependent factors determine the stroke volume:

Preload

This is defined as the tension of the myocardial fibres at the end of diastole and is related to the degree of stretch of the fibres by intracardiac volume. As the end-diastolic volume of the ventricle increases, tension in the myocardial fibres is increased and stroke volume rises (Fig. 13.8). Myocardial oxygen consumption ($\dot{V}_mO_2$) increases only slightly with an increase in preload (produced, e.g., by a 'fluid challenge', see later) and this is therefore the most efficient way of improving cardiac output.

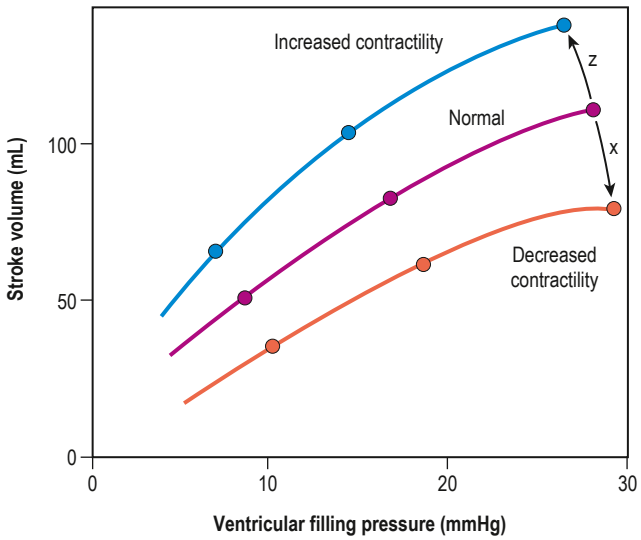


Fig. 13.8 The Frank–Starling relationship: as preload is increased, stroke volume rises. If the ventricle is overstretched, stroke volume will fall (x). In myocardial failure, the curve is depressed and flattened. Increasing contractility, such as that due to sympathetic stimulation, shifts the curve upwards and to the left (z).

Myocardial contractility

This refers to the ability of the heart to perform work, independent of changes in preload and afterload. The state of myocardial contractility determines the response of the ventricles to changes in preload and afterload. Contractility is often reduced in critically ill patients, because of either pre-existing myocardial dysfunction (e.g., ischaemic heart disease) or the acute disease process itself (e.g., sepsis). Changes in myocardial contractility alter the slope and position of the Starling curve; worsening ventricular performance is manifested as a depressed, flattened curve (see Fig. 13.8). Inotropic drugs are used to increase myocardial contractility (see later).

Afterload

This is defined as the myocardial wall tension developed during systolic ejection. In the case of the left ventricle, the resistance imposed by the aortic valve, the peripheral vascular resistance, and the elasticity of the major blood vessels are the major determinants of afterload. Ventricular wall tension will also be increased by ventricular dilation, an increase in intraventricular pressure, or a reduction in ventricular wall thickness.

Decreasing the afterload (through vasodilation due to exercise, sepsis, or vasodilating agents) can increase the stroke volume achieved at a given preload (Fig. 13.9) while reducing $\dot{V}_mO_2$. The reduction in wall tension also leads to an increase in coronary blood flow, improving the myocardial oxygen supply/demand ratio. Excessive reductions in afterload will cause systemic arterial hypotension and potentially an increase in cardiac output.

Increasing the afterload (vasoconstriction due to increased sympathetic activity, vasoconstrictor agents), on the other hand, can cause a fall in stroke volume and an increase in $\dot{V}_mO_2$.

Right ventricular afterload is normally negligible because the resistance of the pulmonary circulation is extremely low but is increased in pulmonary hypertension, the causes of which are multiple and include left-sided heart disease, lung disease, pulmonary emboli, and significant pulmonary emboli. Unlike the left ventricle, the right is poorly tolerant of acute changes in this parameter, and acutely elevated pulmonary vascular resistance may cause shock due to inadequate delivery of blood to the left ventricle, causing falls in cardiac output.

Cardiovascular assessment and monitoring of critically ill patients

Monitoring helps establish or confirm a diagnosis, gauge the severity of a condition, follow its evolution, guide interventions, and assess response to treatment. Invasive monitoring is usually indicated in critically ill patients and those unresponsive to initial resuscitation. These techniques carry a risk of significant complications, are costly, and may lead to patient discomfort. They should be used only when potential benefits outweigh their dangers and for the minimum time necessary.

Assessment of tissue perfusion

- ***Pale, cold skin***, delayed capillary refill, and the absence of visible veins in the hands and feet indicate poor perfusion. Although peripheral skin temperature measurements can help clinical evaluation, the earliest compensatory response to hypovolaemia or low cardiac output, and the last to resolve after resuscitation, is splanchnic vasoconstriction.
- ***Metabolic acidosis with raised lactate concentration*** suggests that tissue perfusion is sufficiently compromised to cause cellular hypoxia and anaerobic glycolysis. Persistent, severe lactic acidosis is associated with an extremely poor prognosis. In addition to serving as a screen for cardiovascular insufficiency and poor tissue perfusion, lactate levels are frequently used to guide the adequacy of resuscitation. High lactate levels that do not respond to resuscitation suggest reduced splanchnic blood flow and bowel ischemia or mitochondria cytopathy. In sepsis, lactic acidosis can also be caused by metabolic disorders unrelated to tissue hypoxia and exacerbated by reduced clearance owing to hepatic or renal dysfunction and the administration of adrenaline (epinephrine) which impairs lactate metabolism.
- ***Urinary flow*** is a sensitive indicator of renal perfusion and haemodynamic performance in the absence of intrinsic renal disease or obstruction.

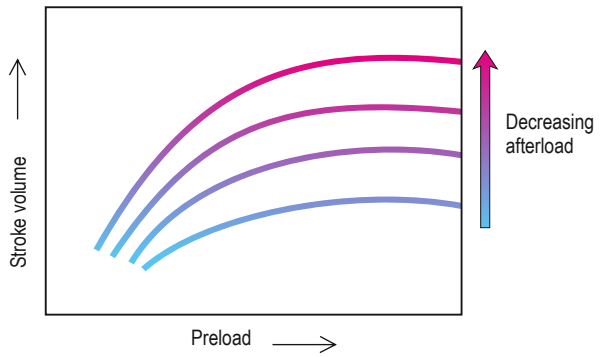


Fig. 13.9 The effect of changes in afterload on the ventricular function curve. At any given preload, decreasing afterload increases the stroke volume.

Blood pressure

Alterations in blood pressure are often interpreted as reflecting changes in cardiac output. However, if there is vasoconstriction with a high peripheral resistance, the blood pressure may be normal, even when the cardiac output is low. Conversely, the vasodilated patient may be hypotensive, despite a very high cardiac output.

Hypotension jeopardizes perfusion of vital organs. The adequacy of blood pressure in an individual patient should always be assessed in relation to the pre-morbid value. Blood pressure is traditionally measured using a sphygmomanometer, but if rapid alterations are anticipated or administration of vasoactive agents is required, continuous monitoring using an intra-arterial cannula is indicated (Box 13.2 and Fig. 13.10).

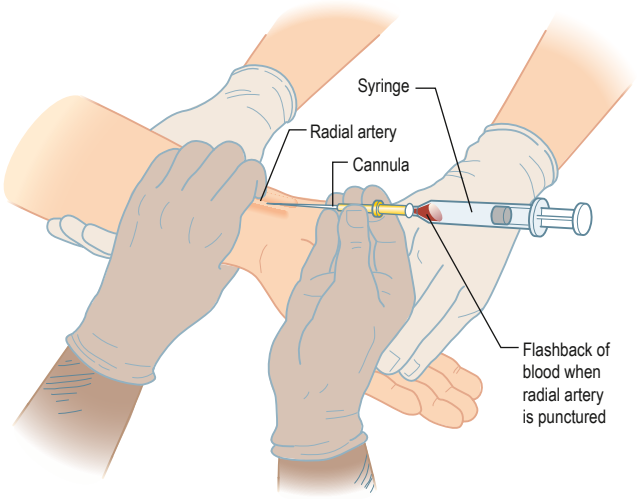


Fig. 13.10 Percutaneous cannulation of the radial artery.

Box 13.2 Radial artery cannulation

Technique

1. Explain the procedure to the patient and, if possible, obtain consent.
2. Ask an assistant to support the patient's arm, with the wrist extended. (Gloves should be worn.)
3. Clean the skin with chlorhexidine. Take sterile precautions throughout the procedure.
4. Palpate the radial artery where it arches over the head of the radius.
5. In conscious patients, inject local anaesthetic to raise a weal over the artery, taking care not to puncture the vessel or obscure its pulsation.
6. Make a small skin incision over the proposed puncture site.
7. Use a small, parallel-sided cannula (20 gauge for adults, 22 gauge for children) in order to allow blood flow to continue past the cannula.
8. Insert the cannula over the point of maximal pulsation and advance it in line with the direction of the vessel at an angle of approximately 30 degrees.
9. Look for 'flashback' of blood into the cannula, which indicates that the radial artery has been punctured.
10. To ensure that the shoulder of the cannula enters the vessel, lower the needle and cannula and advance them a few millimetres into the vessel.
11. Thread the cannula off the needle into the vessel and withdraw the needle.
12. Connect the cannula to a non-compliant manometer line filled with saline. Then connect this via a transducer and continuous flush device to a monitor, which records the arterial pressure.

Possible complications

- Thrombosis
- Loss of arterial pulsation
- Distal ischaemia, e.g., digital necrosis (rare)
- Infection
- Accidental injection of drugs—can produce vascular occlusion
- Disconnection—rapid blood loss

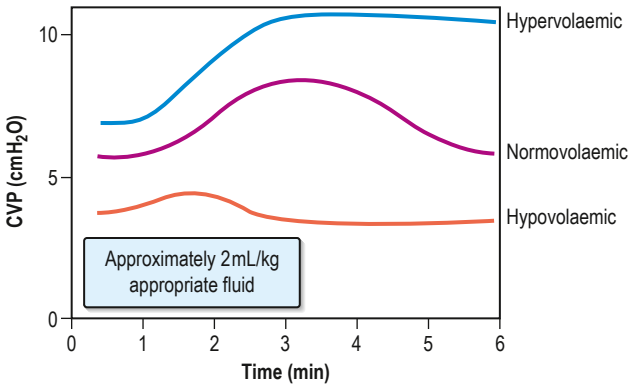


Fig. 13.11 The effects on the central venous pressure (CVP) of rapid administration of a 'fluid challenge' to patients with a CVP within the normal range.

Central venous pressure

Assessment of CVP provides a simple but approximate method of gauging the adequacy of a patient's circulating volume and preload. The absolute value of the CVP is not as informative as its response to a fluid challenge (the infusion of 100–200 mL of fluid over a few minutes; **Fig. 13.11**). The hypovolaemic patient may initially respond to fluid infusion with little or no change in CVP but will exhibit some improvement in cardiovascular function (falling heart rate, rising blood pressure, increased peripheral temperature, and urine output). As the normovolaemic state is approached, the CVP may rise slightly and plateau, while other cardiovascular values stabilize. At this stage, volume replacement should be slowed or even stopped. In cardiac failure, the venous pressure is generally high; the patient will not improve in response to volume replacement, which will cause a further, sometimes dramatic, rise in CVP. However, these numbers cannot be taken in isolation and must be placed against the clinical status. The use of CVP to assess cardiovascular function is controversial, and newer methods, particularly pulse contour analysis and oesophageal Doppler techniques, have largely replaced it (see later).

Central venous catheter placement is described in Chapter 11. Usually, these catheters consist of more than one lumen allowing use for drug or fluid administration in addition to monitoring CVP. The CVP is usually displayed continuously using a transducer and bedside monitor but can also be recorded intermittently using a manometer system. It is essential for the recorded pressure always to be related to the level of the right atrium as a zero point. Various landmarks are advocated (e.g., sternal notch with the patient supine; sternal angle or mid-axilla when the patient is at 45 degrees), but the choice is immaterial, provided the landmark is used consistently in an individual patient.

Left atrial pressure

In uncomplicated cases, careful interpretation of the CVP may provide a reasonable guide to the filling pressures of both sides of the heart. However, in most critically ill patients, there is a disparity in function between the two ventricles related to variation in systemic and pulmonary vascular resistance and differing function. Most commonly, left ventricular performance is worse so that the left ventricular function curve is displaced downwards and to the right (**Fig. 13.12**). High right ventricular filling pressures, with normal or low left atrial pressures, are less common but occur with right ventricular dysfunction and with raised pulmonary vascular resistance (i.e., right ventricular afterload), such as in acute respiratory failure and pulmonary embolism.

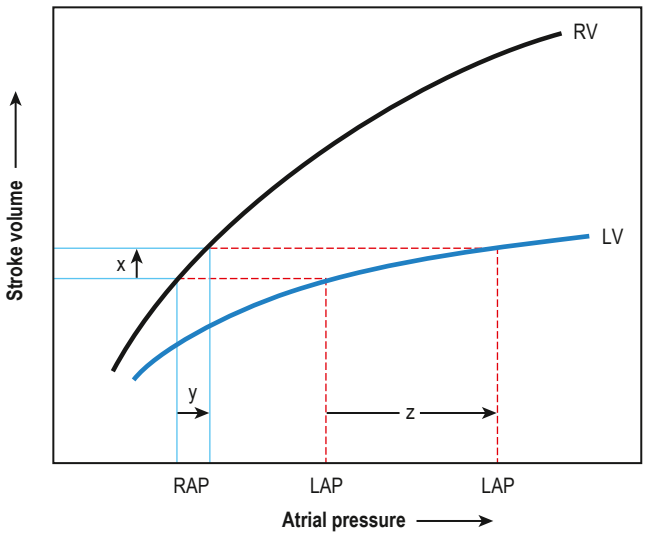


Fig. 13.12 Left ventricular (LV) and right ventricular (RV) function curves in a patient with left ventricular dysfunction. Since the stroke volume of the two ventricles must be the same (except, perhaps, for a few beats during a period of circulatory adjustment), left atrial pressure (LAP) must be higher than right atrial pressure (RAP). Moreover, an increase in stroke volume (x) produced by expanding the circulatory volume may be associated with a small rise in RAP (y) but a marked increase in LAP (z).

Box 13.3 Passage of a pulmonary artery balloon flotation catheter

The catheter is passed through the chambers of the heart into the 'wedged' position.

1. Explain the procedure to the patient and, if possible, obtain consent. Place the patient in a 'head down' position.
2. Insert a balloon flotation catheter through a large vein (see text).
3. Look for respiratory oscillations once the catheter is in the thorax. Advance it further towards the lower superior vena cava/right atrium (see **Fig. 13.13A**), where pressure oscillations become more pronounced. Then inflate the balloon and advance the catheter.
4. When the catheter is in the right ventricle (see **Fig. 13.13B**), there is no dicrotic notch and the diastolic pressure is close to 0. Return the patient to the horizontal or slightly head-up position before advancing the catheter further.
5. When the catheter reaches the pulmonary artery (see **Fig. 13.13C**), a dicrotic notch appears and there is elevation of the diastolic pressure. Advance the catheter further with the balloon inflated.
6. Reappearance of a venous waveform indicates that the catheter is 'wedged'. Deflate the balloon to obtain the pulmonary artery pressure. Inflate the balloon intermittently to obtain the pulmonary artery occlusion pressure (also known as pulmonary artery or capillary 'wedge' pressure; see **Fig. 13.13D**).

Pulmonary artery pressures

A 'balloon flotation catheter' enables reliable catheterization of the pulmonary artery. These are often called 'Swan–Ganz' catheters after their inventors and are inserted through the central veins. Passage of the catheter from the major veins, through the chambers of the heart into the pulmonary artery and into the wedged position, is monitored and guided by the pressure waveforms recorded from the distal lumen (**Box 13.3** and **Fig. 13.13**). A chest X-ray should always be obtained to check the final position of the catheter (**Fig. 13.14**).

Once in position, the catheter measures pulmonary artery mean, systolic, and end-diastolic pressures (PAEDP). The pulmonary artery occlusion pressure (PAOP, or capillary 'wedge' pressure) is

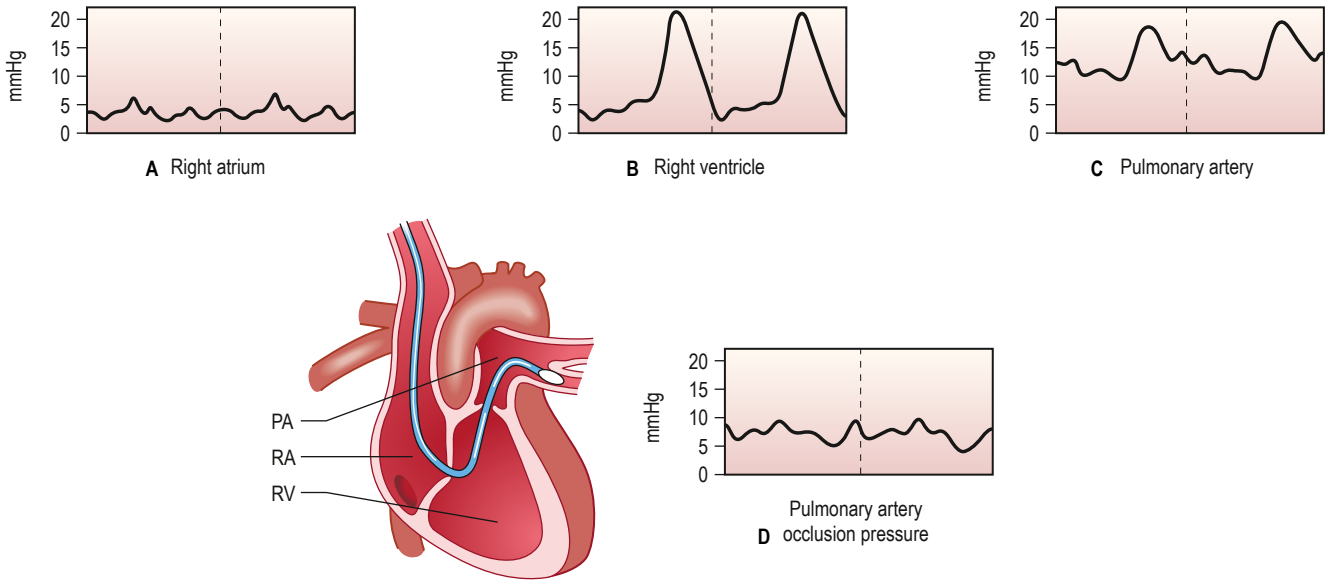


Fig. 13.13 Passage of pulmonary artery balloon flotation catheter through the chambers of the heart. The catheter is passed into the 'wedged' position to measure the pulmonary artery occlusion pressure (see **Box 13.4**). *PA,* Pulmonary artery; *RA,* right artery; *RV,* right ventricle.

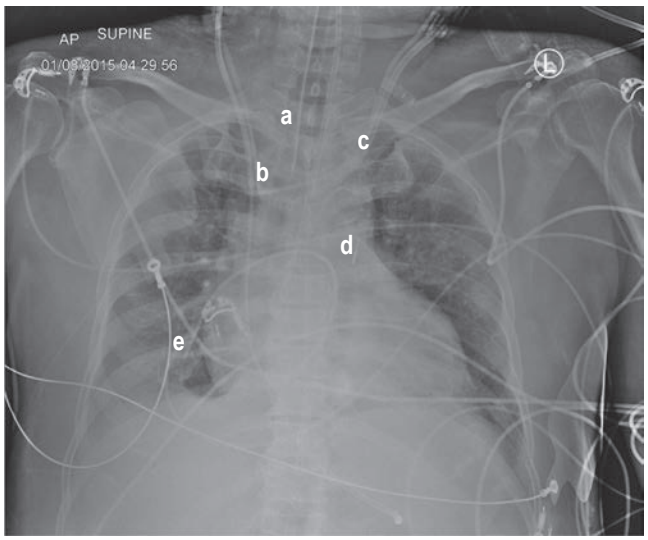


Fig. 13.14 Chest X-ray of a critically ill, mechanically ventilated patient. The X-ray shows **(a)** endotracheal tube; **(b)** end of central venous line seen here medial to the proximal end of the pulmonary artery catheter; **(c)** double-lumen catheter for continuous renal replacement therapy placed via the left internal jugular vein; **(d)** tip of the intra-aortic balloon pump travelling superiorly from a femoral artery; and **(e)** pulmonary artery catheter. Note that the tip of the catheter is too distal; it was therefore subsequently withdrawn by a few centimetres.

measured by inflating the balloon at the tip of the catheter. This creates a continuous column of fluid between the distal lumen of the catheter and the left atrium so that the PAOP is usually a reasonable reflection of left atrial pressure.

The technique is generally safe; the majority of complications, such as 'knotting', valve trauma, and pulmonary artery rupture (which can be fatal), are related to user inexperience. However, they have largely been superseded by less invasive technologies.

Cardiac output

The gold standard cardiac output monitor uses a modified pulmonary artery catheter that transmits low-energy heat from a heating element in the catheter into the surrounding blood. A 'thermodilution curve' is constructed by measuring dissipation of the heat using a thermistor located distal to the heating coil. The dissipation of heat is directly proportional to the cardiac output. These catheters also optically measure and continuously display S_VO_2.

In general, pulmonary artery catheters may help the clinician optimize cardiac output and oxygen delivery while minimizing the risk of volume overload. They guide the rational use of inotropes and vasoactive agents, particularly in patients with pulmonary hypertension. The unselective use of this monitoring device in the absence of evidence-based haemodynamic goals does not, however, lead to improved outcomes, and less invasive techniques such as oesophageal Doppler and pulse contour analysis devices are preferred, except in the most complex cases (e.g., high-risk cardiac surgery) or where there are significant differences in ventricular performance such as an isolated right ventricular infarct.

Less invasive techniques for assessing cardiac function and guiding volume replacement

Arterial pressure variation as a guide to hypovolaemia

Systolic arterial pressure decreases during the inspiratory phase of intermittent positive pressure ventilation (IPPV) (see 'Respiratory support' section). The magnitude of this cyclical variability has been shown to correlate more closely with hypovolaemia than other monitored variables. Systolic pressure (or pulse pressure) variation during mechanical ventilation can be used as a simple and reliable guide to the circulatory volume's adequacy. The response to fluid loading can

also be predicted simply by observing the changes in pulse pressure during passive leg-raising.

Oesophageal Doppler

Stroke volume, cardiac output, and myocardial function can be assessed non-invasively using Doppler ultrasonography. A probe is passed into the oesophagus to monitor velocity waveforms from the descending aorta continuously (**Fig. 13.15**). Although reasonable estimates of stroke volume, and hence cardiac output, can be obtained, the technique is best reserved for trend analysis rather than for making absolute measurements. It is particularly valuable for perioperative optimization of the circulating volume and cardiac performance in the unconscious patient. It is contraindicated in patients with oropharyngeal/oesophageal pathology.

Arterial waveform analysis

Lithium dilution/pulse contour analysis does not require pulmonary artery catheterization or instrumentation of the oesophagus and is suitable for use in conscious patients. A bolus of lithium chloride is administered via a central venous catheter and the change in arterial plasma lithium concentration is detected by a lithium-sensitive electrode. This sensor can be connected to an existing arterial cannula via a three-way tap. The cardiac output determined in this way can be used to calibrate an arterial pressure waveform ('pulse contour') analysis programme that will continuously monitor changes in cardiac output. Devices that employ uncalibrated pulse contour analysis to estimate cardiac output are also available.

Other devices utilize a thermodilution technique, where a small volume of cold fluid is injected into a large central vein and a temperature washout curve is detected via a sensor in a modified arterial line. Following this calibration, beat-to-beat cardiac output is generated by computer-based analysis of the arterial pressure waveform.

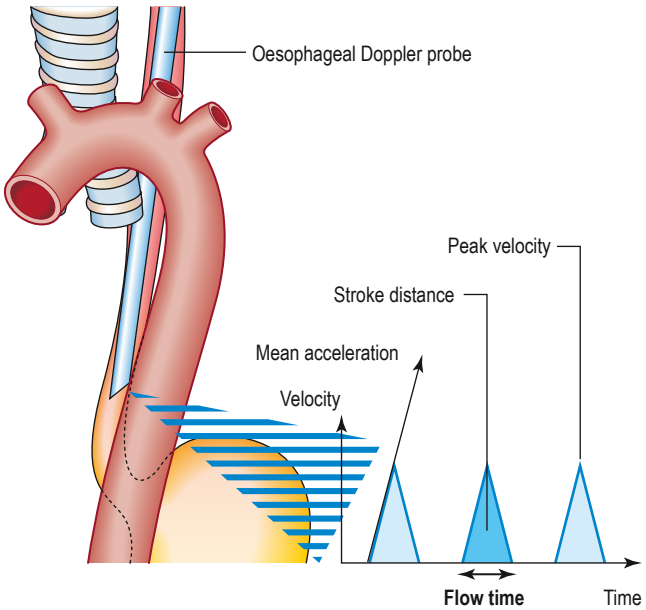


Fig. 13.15 Doppler ultrasonography. Velocity waveform traces are obtained using oesophageal Doppler ultrasonography. *(Reproduced from Singer M, Bennett, ED. Noninvasive optimization of left ventricular filling using esophageal Doppler*. Crit Care Med *1991;19:1132–1137, with permission. © 1991 The Williams & Wilkins Co., Baltimore.)*

As with pulse pressure variation, stroke volume variation, determined by oesophageal Doppler or arterial waveform analysis, can be used to guide fluid replacement.

Echocardiography

Focused transthoracic echocardiography (TTE) is increasingly used to provide immediate diagnostic information about cardiac structure and function (myocardial contractility, ventricular filling) in the critically ill patient. Although transoesophageal echocardiography (TOE) may offer superior image clarity (**Fig. 13.16**), TTE is non-invasive, more readily available, and can easily guide haemodynamic queries, with TOE reserved for complex structural assessment or where there are poor acoustic echo windows.

Key points in monitoring cardiac function

These are shown in **Box 13.4**.

Further reading

Arya VK, Al-Moustadi W, Dutta V. Cardiac output monitoring—invasive and noninvasive. *Curr Opin Crit Care* 2022; 28(3):340–347.
Lamperti M, Bodenham AR, Pittiruti M et al. International evidence-based recommendations on ultrasound-guided vascular access. *Intensive Care Med* 2012; 38:1105–1117.
Marik PE, Monnet X, Teboul JL. Hemodynamic parameters to guide fluid therapy. *Ann Intensive Care* 2011; 1:1.
Zampierei F, Bagshaw G, Semler M. Fluid therapy for critically ill adults with sepsis. *JAMA* 2023; 329(22):1967–1980.

DISTURBANCES OF ACID–BASE BALANCE

The physiology of acid–base control is discussed in Chapter 27. Acid–base disturbances can be described in relation to **Fig. 27.17** (which shows P_aCO_2 plotted against arterial $[H^+]$).

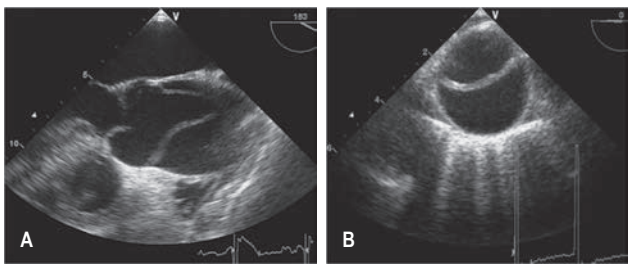


Fig. 13.16 Aortic dissection (transoesophageal echocardiography, TOE). **(A)** Mid-oesophageal, long-axis view showing type A aortic dissection. **(B)** Short-axis view of descending aorta showing intimal flap with false and true lumen. *(From Hinds CJ, Watson JD*. Intensive Care: A Concise Textbook, *3rd edn. Edinburgh: Saunders; 2008. Courtesy of Dr C. Rathwell.)*

i Box 13.4 Key points in monitoring cardiac function

- If there is disagreement between clinical signs and a monitored variable, assume that the monitor is incorrect until all sources of potential error have been checked and eliminated.
- Remember that changes and trends in monitored variables are more informative than a single reading.
- Use non-invasive monitoring where possible.
- Remove invasive devices as soon as possible

Box 13.5 Arterial blood gas and acid–base values

Analyte	Normal reference range
H^+	35–45 nmol/L (pH 7.35–7.45)
PO_2 (breathing room air)	10.6–13.3 kPa (80–100 mmHg)
PCO_2	4.8–6.1 kPa (36–46 mmHg)
Base deficit	±2.5
Plasma HCO_3^-	22–26 mmol/L
O_2 saturation	95–100%

Both acidosis and alkalosis can occur, each of which is either metabolic (primarily affecting the bicarbonate component of the system) or respiratory (primarily affecting P_aCO_2). Compensatory changes may also be apparent. In clinical practice, pH values outside of 6.9–7.7 (arterial $[H^+]$ values 20 to 126 nmol/L) are rarely encountered.

Blood gas and acid–base values (normal ranges) are shown in Box 13.5. (For blood gas analysis, see 'Respiratory failure' section.)

Respiratory acidosis

This is caused by retention of carbon dioxide. The P_aCO_2 and $[H^+]$ rise. A chronically raised P_aCO_2 is compensated by renal retention of bicarbonate and the $[H^+]$ returns towards normal. A constant arterial bicarbonate concentration is then usually established within 2 to 5 days. This represents a primary respiratory acidosis with a compensatory metabolic alkalosis. Common causes of respiratory acidosis include ventilatory failure (e.g., patients in coma or those with neuromuscular disorders) and COPD (type II respiratory failure, where there is a high P_aCO_2 and a low P_aO_2, see Chapter 20).

Respiratory alkalosis

This results from an increase in minute ventilation (higher respiratory rate and/or tidal volumes) in which there is a fall in P_aCO_2 and $[H^+]$, often with a small reduction in bicarbonate concentration. If hypocarbia persists, some degree of renal compensation may occur, producing a metabolic acidosis, although in practice, this is unusual. A respiratory alkalosis may be produced, intentionally or unintentionally, when patients are mechanically ventilated; it may also be seen in patients with hypoxaemic (type I) respiratory failure (see Chapter 20), those with spontaneous hyperventilation, and those living at high altitudes.

Metabolic acidosis

Metabolic acidosis (see Chapter 27) may be due to excessive acid production, often lactic acid (lactic acidosis), as a consequence of anaerobic metabolism during an episode of shock or following cardiac arrest or after a prolonged seizure. A metabolic acidosis may develop as a consequence of chronic renal failure or in diabetic ketoacidosis. It can also follow the loss of bicarbonate from the gut or from the kidney in renal tubular acidosis. Respiratory compensation for a metabolic acidosis occurs when the patient hyperventilates and 'blows off' carbon dioxide to produce a compensatory respiratory alkalosis. However, there is a limit to this respiratory compensation since, in practice, values for P_aCO_2 of less than 1.4 kPa (11 mmHg) are rarely achieved. Spontaneous respiratory compensation cannot occur if the patient's ventilation is controlled or if the respiratory centre is depressed, for example, by drugs or head injury.

Metabolic alkalosis

This can be caused by loss of acid—for example, from the stomach with excessive vomiting or with nasogastric suction, or in high intestinal obstruction, or excessive administration of absorbable or intravenous alkali. Respiratory compensation for a metabolic alkalosis is often slight, and it is rare to encounter a P_aCO_2 above 6.5 kPa (50 mmHg), even with severe alkalosis.

Box 13.6 Causes of shock

Hypovolaemic
- Exogenous losses (e.g., haemorrhage, burns)

Cardiogenic
- 'Myocardial failure' (e.g., ischaemic myocardial injury)

Obstructive
- Obstruction to cardiac outflow (e.g., pulmonary embolus)
- Restricted cardiac filling (e.g., cardiac tamponade, tension pneumothorax)

Distributive
- Vascular dilation (e.g., sepsis, anaphylaxis)
- Sequestration
- Arteriovenous shunting
- Maldistribution of flow

SHOCK, SEPSIS, AND ACUTE DISTURBANCES OF HAEMODYNAMIC FUNCTION

Shock is the term used to describe acute circulatory failure with inadequate or inappropriately distributed tissue perfusion, resulting in generalized cellular hypoxia and/or an inability of the cells to utilize oxygen.

Aetiology of shock

Abnormalities of tissue perfusion can result from:
- failure of the heart to act as an effective pump
- mechanical impediments to forward flow
- loss of circulatory volume
- abnormalities of the peripheral circulation.

The causes of shock are shown in Box 13.6; see also Chapter 12 for definitions of sepsis. Often, shock can result from a combination of these factors (e.g., in sepsis, distributive shock is frequently complicated by hypovolaemia and myocardial depression).

Pathophysiology

The sympatho-adrenal response to shock

Hypotension stimulates baroreceptors and, to a lesser extent, chemoreceptors, causing increased sympathetic nervous activity with 'spill-over' of noradrenaline (norepinephrine) into the circulation. Later, this is augmented by the release of catecholamines (predominantly adrenaline [epinephrine]) from the adrenal medulla. The resulting vasoconstriction, together with increased myocardial contractility and heart rate, helps to restore blood pressure and cardiac output (Fig. 13.17).

Reduced perfusion of the renal cortex stimulates the juxtaglomerular apparatus to release renin. This converts angiotensinogen to angiotensin I, which, in turn, is converted in the lungs and by the vascular endothelium to the potent vasoconstrictor angiotensin II. Angiotensin II also stimulates secretion of aldosterone by the adrenal cortex, causing sodium and water retention (see Chapter 26). This helps to restore the circulating volume.

The neuroendocrine response

- ***Release of pituitary hormones*** includes adrenocorticotrophic hormone (ACTH), vasopressin (arginine vasopressin, AVP), and

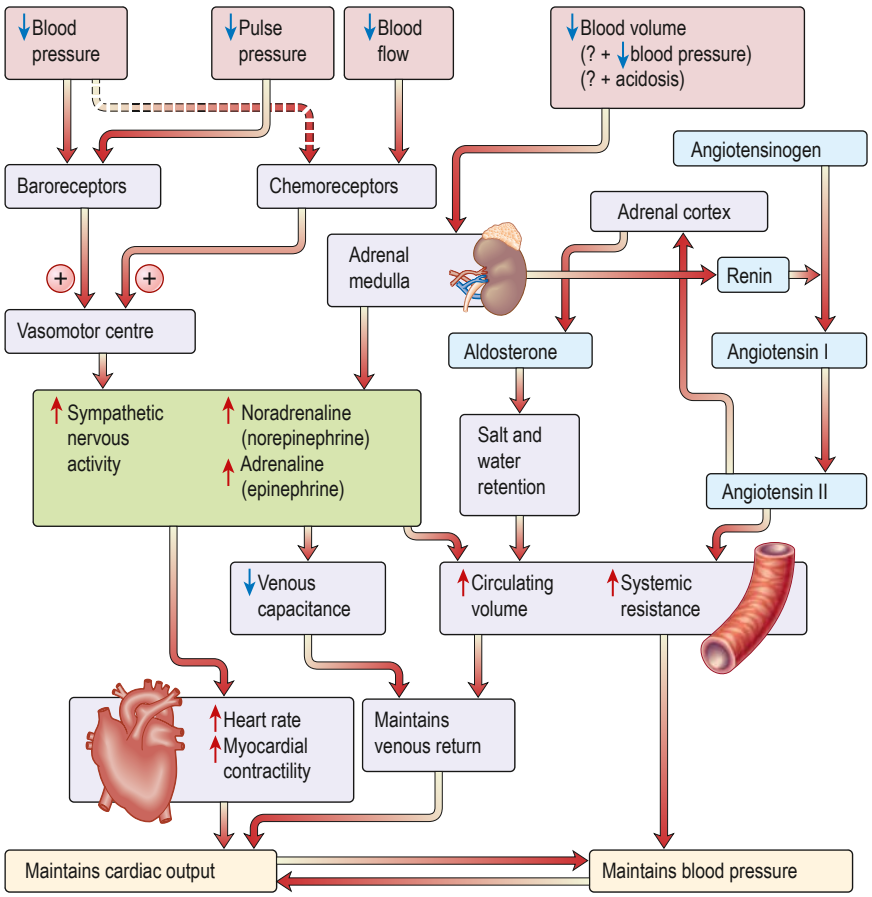


Fig. 13.17 The sympatho-adrenal response to shock. The effect of increased catecholamines is shown on the left and the release of angiotensin and aldosterone on the right. Both mechanisms help to maintain cardiac output and blood pressure in shock.

endogenous opioid peptides. (In septic shock, there may be a relative deficiency of vasopressin.)

- ***Release of cortisol*** causes fluid retention and antagonizes insulin.
- ***Release of glucagon*** raises the blood sugar level.

Although absolute adrenocortical insufficiency (e.g., due to bilateral adrenal haemorrhage or necrosis in meningococcal infection) is rare, a blunted response to exogenous ACTH (so-called ***'relative'*** or ***'occult' adrenocortical insufficiency***) has been observed in septic shock patients with impaired vasoconstrictor response to noradrenaline (norepinephrine). Steroid supplementation improves vasomotor response with a rise in blood pressure but without mortality benefit.

Release of immune mediators

Severe infection (often with bacteraemia or endotoxaemia), the presence of large areas of damaged tissue (e.g., following trauma or extensive surgery), hypoxia, or prolonged/repeated episodes of hypoperfusion can all trigger a dysfunctional immune response with alterations in leucocyte activation and release of a variety of potentially damaging 'mediators' (see also Chapter 6). Although an appropriate immune response is clearly beneficial, the disseminated, dysregulated response observed in some patients can lead to shock and organ failure. This immune response is complex, with upregulation and downregulation of both the innate and the adaptive immune pathways at different stages of the illness. Characteristically, the later phase is typified by a period of immune suppression, during which the patient is at increased risk of developing secondary infections.

Microorganisms and their toxic products

In sepsis/septic shock, the innate immune response (Fig. 13.18) and inflammatory cascade are triggered by the recognition of pathogen-associated molecular patterns (PAMPs), including bacterial and fungal DNA, cell-wall components (e.g., endotoxin), and/or exotoxins (antigenic proteins produced by bacteria such as staphylococci, streptococci, and *Pseudomonas*) (see also Chapter 12).

Endotoxin is a lipopolysaccharide (LPS) derived from the cell wall of Gram-negative bacteria and a potent trigger of the immune response. The lipid A portion of LPS is bound by a serum protein known as lipopolysaccharide-binding protein (LBP). The LBP–LPS complex attaches to the cell surface marker CD14 and, combined with a secreted protein (MD2), then binds to a member of the Toll-like receptor family (TLR4), which transduces the activation signal into the cell. Activation of these receptors promotes the synthesis of a wide variety of immune mediators. Gram-positive bacteria have cell-wall components that are similar in structure to LPS (e.g., lipoteichoic acid) and can also trigger an inflammatory response through similar pathways (see Fig. 13.18).

Toxic products of tissue injury (surgery and trauma)

Following traumatic or surgical tissue injury, inflammatory pathways may be triggered by damage-associated molecular

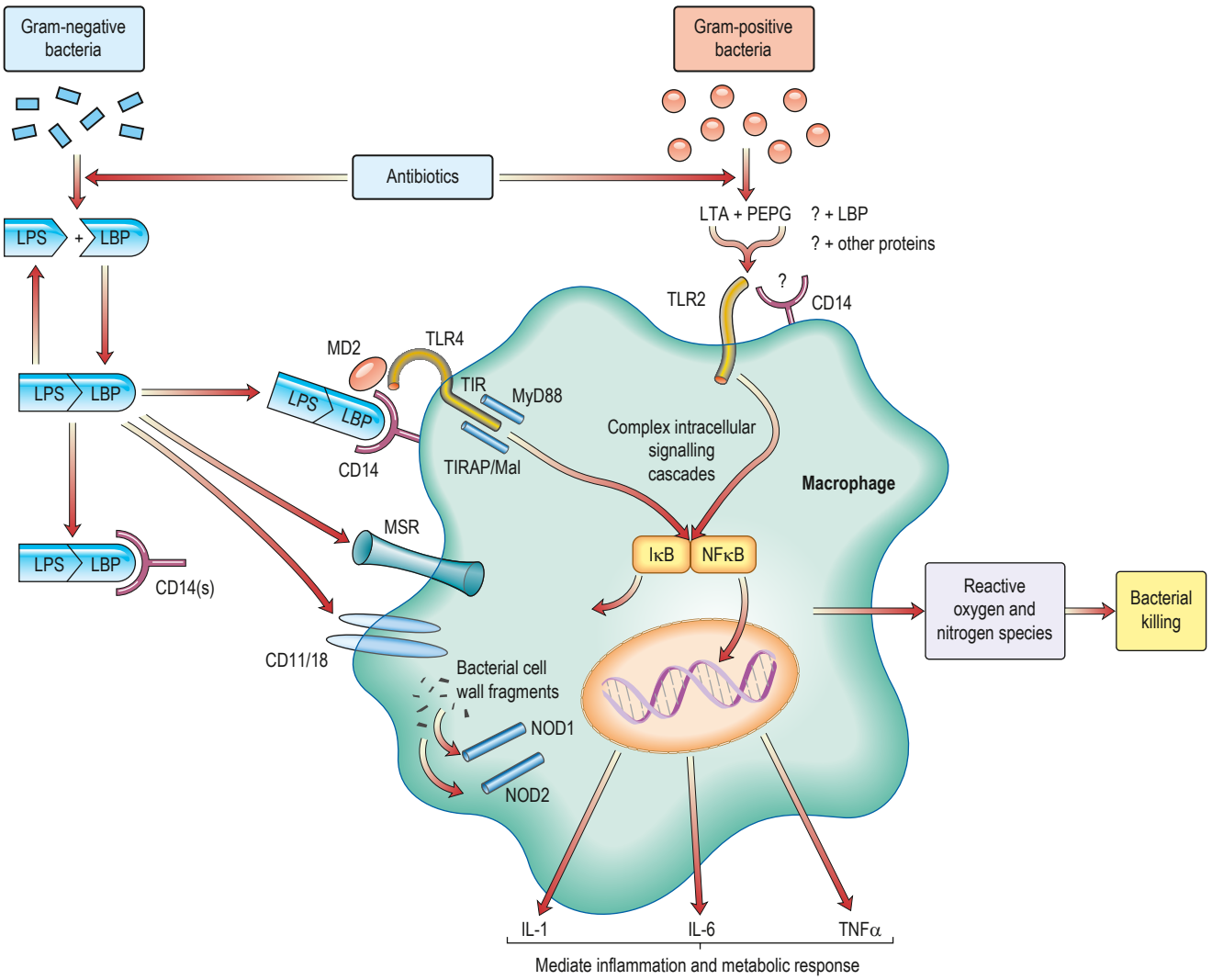


Fig. 13.18 Induction of the innate immune response by the lipopolysaccharide–lipopolysaccharide-binding protein (LPS–LBP) complex. The intracellular events (simplified) initiated by Gram-negative and Gram-positive bacteria, which eventually lead to bacterial killing. *IκB,* Inhibitory factor kappa B; *IL,* interleukin; *LTA,* lipoteichoic acid; *MD2,* a secreted protein involved in binding liposaccharide with TLR4; *MSR,* macrophage scavenger receptor; *MyD88,* myeloid differentiation factor 88; *NFκB,* nuclear factor kappa B; *NOD,* nucleotide-binding oligomerization domain; *PEPG,* peptidoglycan-N; *TIR,* Toll-interleukin receptor; *TIRAP,* Toll-interleukin 1 receptor adaptor protein; *TIRAP/Mal,* an adaptor protein for TLR2 and TLR4; *TLR,* Toll-like receptors; *TNF-α,* tumour necrosis factor-alpha.

patterns (DAMPs), such as DNA fragments. Of particular importance is circulating mitochondrial DNA (mtDNA), which is similar to bacterial DNA and can be 'mistaken' by the immune system for bacteria. This initiates an immune response identical to that observed following bacterial invasion. This is why distributive shock secondary to infection can be difficult to distinguish clinically from that caused by a 'sterile' insult such as severe trauma or surgery.

Cytokines and other immune mediators

An array of cytokines and chemokines (small proteins that orchestrate immune cell behaviour), prostaglandins, leukotrienes, heat shock proteins, adhesion molecules, and endothelial-derived vasoactive factors are induced, particularly during distributive shock. These include pro-inflammatory cytokines (tumour necrosis factor-alpha [TNF-α] and interferon-gamma [IFN-γ]), anti-inflammatory cytokines (interleukin 10 [IL-10]), mediators promoting extracellular extravasation (intracellular adhesion molecule 1 [ICAM-1]), and vasodilation (inducible nitric oxide synthase [iNOS], Fig. 13.19). Although there was initial enthusiasm for treatments that dampened the immune response in septic shock (e.g., high-dose steroids and anti-TNF-α compounds), these strategies failed to improve survival and some worsened it.

Activation of the complement cascade

Fragments of C3 act as opsonins and co-stimulatory molecules that assist lymphocytes with the adaptive immune response, while small peptides derived from C3, C4, and C5 cause leucocyte chemotaxis, release of cytokines, and increased vascular permeability (see Chapter 6).

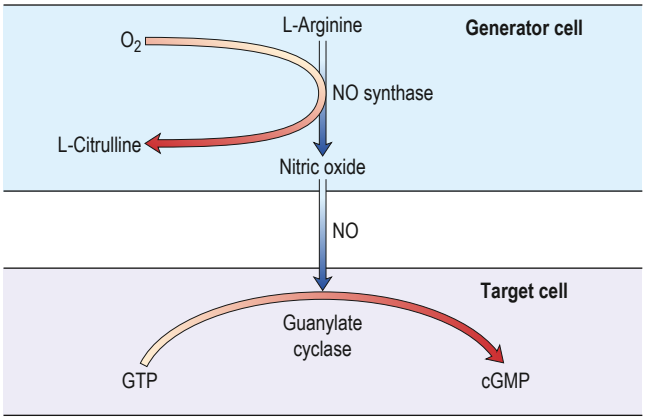


Fig. 13.19 Synthesis and biochemical action of nitric oxide (NO). *cGMP,* Cyclic guanosine monophosphate; *GTP,* guanosine triphosphate.

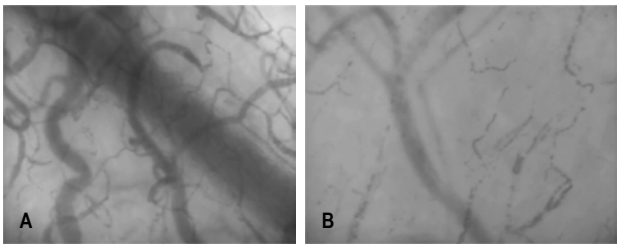


Fig. 13.20 Still side-stream dark field (SDF) images of the microcirculation. (A) Normal subject. **(B)** Septic shock.

Influence of genetic variation

Individuals vary considerably in their susceptibility to infection, as well as their ability to recover from apparently similar infections, illnesses, or traumatic insults; this is explained in part by genetic variations.

Haemodynamic and microcirculatory changes

The dominant haemodynamic feature of severe sepsis/septic shock is peripheral vascular failure with:

- vasodilation;
- maldistribution of regional blood flow; and
- abnormalities in the microcirculation (Fig. 13.20):
  1. 'stop-flow' capillaries (flow is intermittent)
  2. 'no-flow' capillaries (capillaries are obstructed)
  3. failure of capillary recruitment
  4. increased capillary permeability with interstitial oedema.

Although these ***vascular and microvascular abnormalities*** may partly account for the reduced oxygen extraction seen in septic shock, there is also a ***primary defect of cellular oxygen utilization*** caused by mitochondrial dysfunction (see earlier). These changes may be associated with impaired oxygen consumption, a reduced arteriovenous oxygen content difference, an increased $S_{\bar{v}}O_2$, and a lactic acidosis (so-called ***'tissue dysoxia'***). The rise in lactate levels is partly due to tissue hypoxia as well as the stress response leading to increased glucose metabolism and lactate formation. Similar vasodilation and increased vascular permeability also occur in anaphylactic shock.

The glycocalyx consists of a tight, negatively charged, extremely thin meshwork of proteoglycans on the luminal surface of the vascular endothelium. Oxidative stress, free radicals, and endotoxaemia all have specific detrimental effects on the glycocalyx that promote fluid movement into the tissues. Hence, 'pitting' peripheral oedema is often seen in the most seriously ill patients.

Activation of the coagulation system

The immune response to shock, tissue injury, and infection is frequently associated with systemic activation of the clotting cascade, leading to platelet aggregation and widespread microvascular thrombosis, resulting in turn in reduced tissue perfusion.

Initially, the production of prostaglandin I_2 (PGI_2) by the capillary endothelium is impaired. Vascular endothelial injury leads to exposure to tissue factor (see Chapter 38), which triggers coagulation. In severe cases, these changes are compounded by impairment of fibrinolysis, as well as by deficiencies in physiological inhibitors of coagulation (including antithrombin, proteins C and S, and tissue factor-pathway inhibitor). This procoagulant state can lead to end-artery thrombosis and digital and mesenteric ischaemia. Following clot formation, plasminogen is converted to plasmin, which breaks down thrombus, liberating fibrin/fibrinogen degradation products (FDPs). These processes continue in parallel, leading to a 'consumptive' coagulopathy known as ***disseminated intravascular coagulation (DIC)***. Circulating levels of FDPs and D-dimers are therefore increased; the thrombin time, partial thromboplastin time (PTT), and prothrombin time (PT) are prolonged; and platelet and fibrinogen levels fall. Clinically, the patient will bleed from trivial venepuncture sites but, paradoxically, may still be at risk of thrombosis. The development of DIC is often associated with multiorgan failure. In some cases, a microangiopathic haemolytic anaemia develops. DIC is relatively uncommon but is associated particularly with septic shock, especially when due to meningococcal infection. Management targets the ***underlying cause***, as well as supportive treatment with fresh frozen plasma, platelets, and cryoprecipitate when fibrinogen levels are low and the patient is bleeding.

Clinical features of shock and sepsis

Although many clinical features are common to all types of shock, the causes may be distinguished by the associated haemodynamic changes (Box 13.7) and by examination.

Hypovolaemic shock

- ***Inadequate tissue perfusion:***
  1. Skin: cold, pale, slate-grey, slow capillary refill, 'clammy'
  2. Kidneys: oliguria, anuria
  3. Brain: drowsiness, confusion, and irritability
- ***Increased sympathetic tone:***
  1. Tachycardia, narrowed pulse pressure, 'weak' or 'thready' pulse
  2. Sweating
  3. Blood pressure: may be maintained initially (despite up to a 25% reduction in circulating volume if the patient is young and fit) but hypotension later supervenes
- ***Lactic acidosis:*** compensatory tachypnoea.

Extreme hypovolaemia may be associated with bradycardia.

Cardiogenic shock

The signs in cardiogenic shock are the same as those in hypovolaemic shock, with the addition of myocardial failure: for example, raised jugular venous pressure (JVP), pulsus alternans, 'gallop' rhythm, basal crackles, and pulmonary oedema.

Obstructive shock

Again, the signs are as in hypovolaemic shock with the following additions:

- Elevated JVP
- Pulsus paradoxus and muffled heart sounds in cardiac tamponade
- Signs of pulmonary embolism (see Chapter 11).

Box 13.7 Haemodynamic changes in shock

Hypovolaemic shock

Low central venous pressure (CVP) and pulmonary artery occlusion pressure (PAOP)

- Low cardiac output
- Increased systemic vascular resistance

Cardiogenic shock

Signs of myocardial failure

- Increased systemic vascular resistance
- CVP and PAOP high (except when hypovolaemic)

Cardiac tamponade

Parallel increases in CVP and PAOP

- Low cardiac output
- Increased systemic vascular resistance

Pulmonary embolism

Low cardiac output

- High CVP, high pulmonary artery pressure but low PAOP
- Increased systemic vascular resistance

Anaphylaxis

Low systemic vascular resistance

- Low CVP and PAOP
- High cardiac output

Septic shock

Low systemic vascular resistance

- Low CVP and PAOP
- Cardiac output usually high
- Myocardial depression—low ejection fraction
- Stroke volume maintained by ventricular dilation
- Cardiac output maintained or increased by tachycardia

Distributive shock: anaphylactic shock

(See Chapter 11).

- Signs of profound vasodilation:
  1. Warm peripheries
  2. Low blood pressure
  3. Tachycardia.
- Erythema, urticaria, angio-oedema, pallor, cyanosis.
- Bronchospasm, rhinitis.
- Oedema of the face, pharynx, and larynx.
- Pulmonary oedema.
- Hypovolaemia due to vascular leak.
- Nausea, vomiting, abdominal cramps, diarrhoea.

Ideally, 10 mL of clotted blood should be taken within 45 to 60 minutes of the reaction for confirmation of the diagnosis by measurement of mast cell degranulation (tryptase levels). Immunology follow-up of these patients is essential.

Distributive shock: sepsis and septic shock

Sepsis and ***septic shock*** are defined in Chapter 12. The ***Sequential Organ Failure Assessment (SOFA) score***, used to identify organ dysfunction, is shown in **Box 12.2**.

Symptoms and signs of sepsis include:

- Pyrexia and rigors, or hypothermia (unusual, but more common in the elderly and associated with worse prognosis)
- Nausea, vomiting
- Vasodilation, warm peripheries
- Bounding pulse
- Rapid capillary refill
- Hypotension, low diastolic pressure, widened pulse pressure

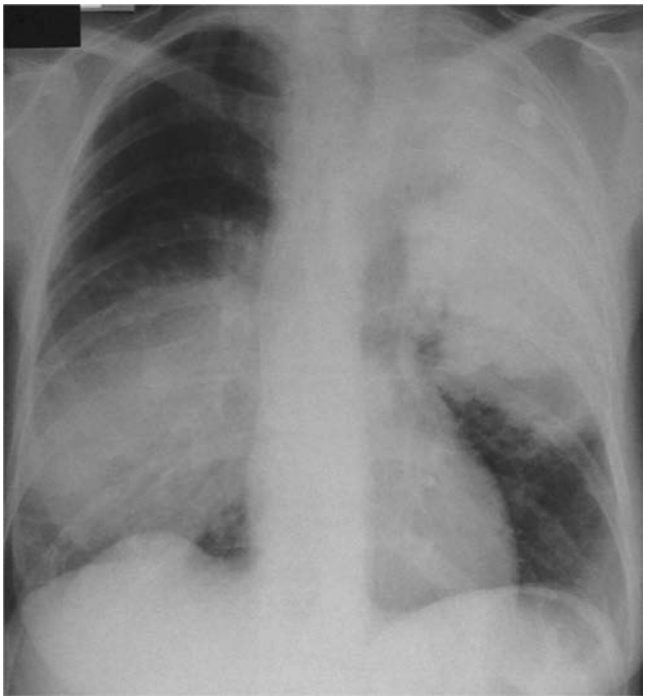


Fig. 13.21 Bilateral pneumococcal pneumonia. Community-acquired pneumonia is the most common cause of sepsis requiring admission to intensive care. *(From Hinds CJ, Watson JD. Intensive Care: A Concise Textbook, 3rd edn. Edinburgh: Saunders; 2008. Courtesy of Dr SPG Padley.)*

- Occasionally, signs of cutaneous vasoconstriction
- Other signs:
  1. Coma, stupor
  2. Bleeding due to coagulopathy (e.g., from vascular puncture sites, gastrointestinal tract, and surgical wounds)
  3. Rash and meningism
  4. Hyperglycaemia; in more severe cases, hypoglycaemia
  5. Jaundice

The ***diagnosis of sepsis*** is easily missed, particularly in the elderly. Clues include mild confusion, tachycardia, tachypnoea, unexplained hypotension, a reduction in urine output, a rising plasma creatinine, and glucose intolerance.

Sepsis and multiple organ failure (multiple organ dysfunction syndrome)

Sepsis is being diagnosed with increasing frequency and is now the most common cause of death in non-coronary adult ICUs. The estimated incidence of severe sepsis is 77 to 300 cases per 100 000 of the population. The most common cause is community-acquired pneumonia (**Fig. 13.21**). Mortality rates are high (between 20% and 60%) and are closely related to the severity of illness and the number of organs that fail. Although some early deaths are caused by cardiovascular collapse, advances in ICU care and supportive therapy have reduced early deaths, and most of those who die are overwhelmed by persistent or recurrent sepsis, with fever, intractable hypotension, and failure of several organs.

Sequential failure of vital organs occurs progressively over days or weeks, although the pattern of organ dysfunction is variable. In many cases, the lungs are the first to be affected (acute respiratory distress syndrome, ARDS; see later) in association with cardiovascular instability and deteriorating renal function. Central

Box 13.8 The metabolic response to trauma, major surgery, and severe infection

- ↑ Energy expenditure
- ↑ Gluconeogenesis
- ↑ Mobilization of glucose from glycogen
- Insulin resistance and hyperglycaemia
- ↑ Free fatty acid synthesis
- ↑ Protein breakdown
- ↑ Liver synthesis of acute phase reactants

nervous system (CNS) dysfunction is also an early feature, including impaired consciousness, disorientation, and coma, and is included in the qSOFA score (see Box 12.3) commonly used to identify septic patients outside of the ICU. Liver dysfunction may develop later. Gastrointestinal failure, with an inability to tolerate enteral feeding and paralytic ileus, is common. Ischaemic colitis, acalculous cholecystitis, pancreatitis, and gastrointestinal haemorrhage may also occur. Characteristically, these patients initially have a hyperdynamic circulation with vasodilation and a high cardiac output, associated with an increased metabolic rate. Eventually, however, cardiovascular collapse supervenes. It is often possible to support such patients for weeks or months; most die following a decision to withdraw or not to escalate treatment (see 'Withholding and withdrawing treatment' section).

Metabolic response to trauma, major surgery, and severe infection

This is initiated and controlled by the neuroendocrine system and various cytokines (e.g., IL-6) acting in concert and is characterized initially by an increase in energy expenditure ('hypermetabolism'; **Box 13.8**). Gluconeogenesis is stimulated by increased glucagon and catecholamine levels, while hepatic mobilization of glucose from glycogen is increased. Catecholamines inhibit insulin release and reduce peripheral glucose uptake. Combined with elevated circulating levels of other insulin antagonists such as cortisol and downregulation of insulin receptors, these changes mean that the majority of patients are hyperglycaemic ('insulin resistance'). Later, hypoglycaemia may be precipitated by depletion of hepatic glycogen stores and inhibition of gluconeogenesis. Free fatty acid synthesis is also increased, leading to hypertriglyceridaemia.

Protein breakdown is initiated to provide energy from amino acids, and hepatic protein synthesis is preferentially augmented to produce the 'acute phase reactants'. The amino acid glutamine (which is indispensable in this situation) is mobilized from muscle for use as a metabolic fuel in rapidly dividing cells such as leucocytes and enterocytes. Glutamine is also required for hepatic production of the free radical scavenger glutathione. When severe and prolonged, this catabolic response can lead to considerable weight loss. Protein breakdown is associated with wasting and weakness of skeletal and respiratory muscle, prolonging the need for mechanical ventilation and delaying mobilization. Tissue repair, wound healing, and immune function are also compromised.

Management of shock and sepsis

Delays in recognizing sepsis and initiating resuscitation and treatments (**Fig. 13.22**) (particularly antibiotics when infection is the underlying cause) are associated with substantially increased morbidity and mortality.

A patent airway must be maintained and oxygen given, as is standard for any acutely ill patient. Some may require endotracheal intubation and those with impending respiratory arrest will require immediate mechanical ventilation.

The underlying cause of shock should be sought and corrected where possible; for example, haemorrhage should be controlled or infection treated. In patients with septic shock, every effort must be made to identify the source of infection and isolate the causative organism. A thorough history, clinical examination, and imaging (X-ray, ultrasonography, or computed tomography [CT] scanning) should be undertaken. Appropriate samples (blood, urine, sputum, cerebrospinal fluid [CSF], pus drained from abscesses) should be taken for microscopy, culture, and sensitivities. Empirical, broad-spectrum antibiotic therapy (see Chapter 12) should be commenced *within the first hour* of recognition of sepsis. If an organism is isolated later, therapy can be adjusted accordingly. The choice of antibiotic depends on the likely source of infection, previous antibiotic therapy, and known local resistance patterns, as well as if the infection was acquired in hospital or in the community. Sources of sepsis such as abscesses and collections must be drained, and infected indwelling catheters removed.

Whatever the aetiology of the haemodynamic abnormality:

- ***Tissue perfusion*** must be restored promptly by resuscitation and re-establishment of the circulation and cardiac output, as well as haemoglobin levels (oxygen carrying capacity).
- ***Arterial blood pressure*** should be restored to maintain perfusion of vital organs.
- ***Published guidelines*** for adult patients suffering septic shock should be followed. These advocate targeting a mean arterial pressure (MAP) of >65 mmHg, a CVP of 8–12 mmHg (>12 mmHg if mechanically ventilated), a urine output of >0.5 mL/kg per hour, and an $S_{CV}O_2$ of ≥70% as the initial goals of resuscitation.
- The ***aim of current sepsis guidelines*** (see 'Further reading') is to combine evidence-based interventions with early effective resuscitation (aimed especially at achieving an adequate circulating volume, combined with the rational use of inotropes and/or vasoactive agents to maintain blood pressure, cardiac output, and oxygen transport), in order to create 'bundles of care' delivered within specific time limits.

Preload and volume replacement

Optimizing preload is the most efficient way of increasing cardiac output. Volume replacement is obviously essential in hypovolaemic shock but is also required in anaphylactic and septic shock because of vasodilation, sequestration of blood, and loss of circulating volume because of vascular leak.

In obstructive shock, high filling pressures may be required to maintain an adequate stroke volume. Even in cardiogenic shock, careful volume expansion may, on occasion, lead to a useful increase in cardiac output. On the other hand, patients with severe cardiac failure, in whom ventricular filling pressures are markedly elevated, often benefit from measures to reduce preload (and afterload), such as the administration of vasodilators and diuretics (see later). Adequate perioperative volume replacement also reduces morbidity and mortality in high-risk surgical patients.

The circulating volume must be replaced quickly in order to reduce tissue damage and prevent acute kidney injury (AKI). Continuous observation of clinical signs, biochemical variables (acid–base and lactate levels), and stroke volume is vital for management and avoidance of volume overload. Pulmonary and peripheral oedema is more likely in seriously ill patients receiving large-volume fluid resuscitation because of a low colloid osmotic pressure (usually due to a low serum albumin), disruption of the alveolar–capillary

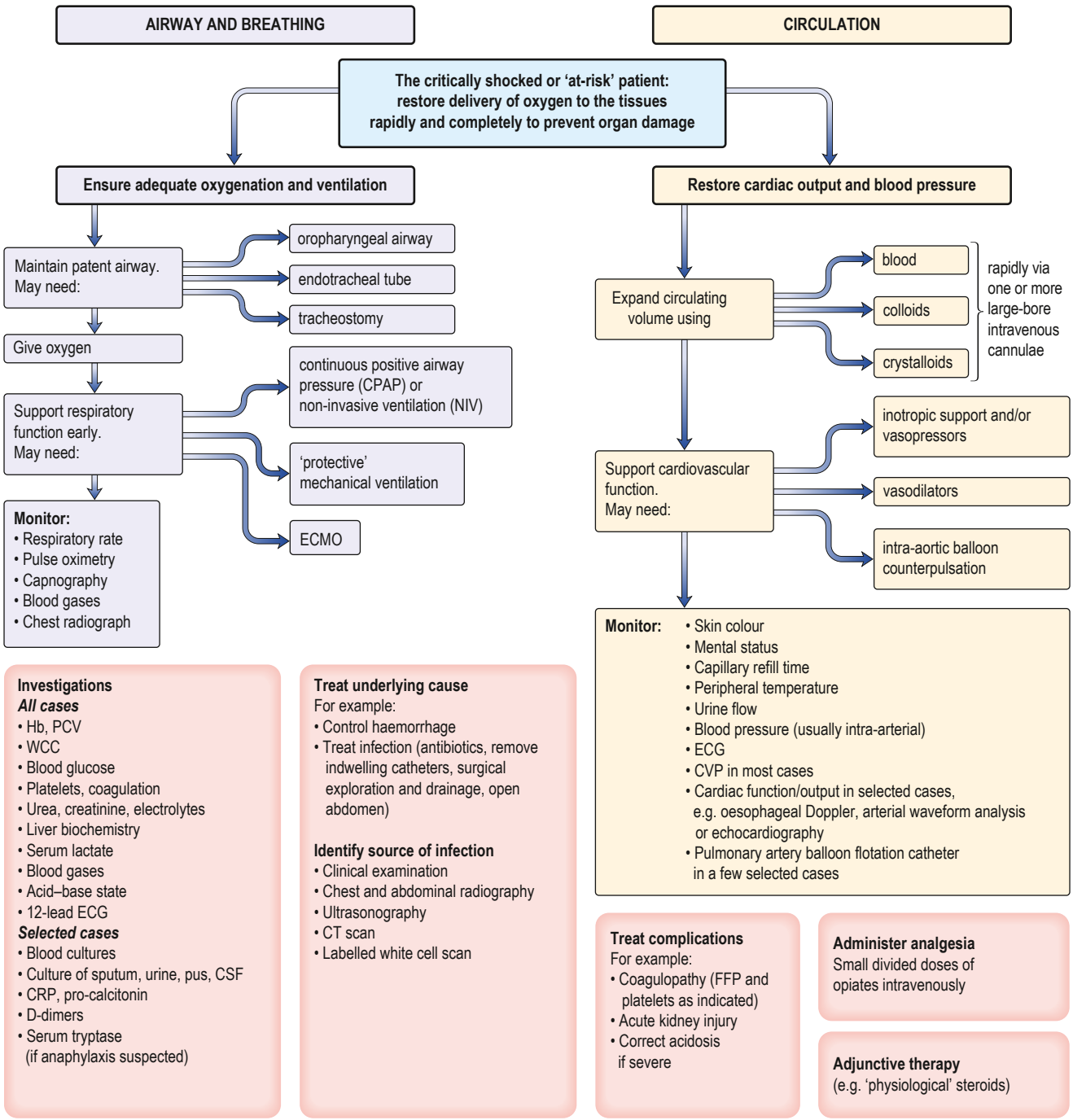


Fig. 13.22 Management of the critically ill, shocked, or 'at-risk' patient. *CRP,* C-reactive protein; *CSF,* cerebrospinal fluid; *CT,* computed tomography; *CVP,* central venous pressure; *ECG,* electrocardiogram; *ECMO,* extracorporeal membrane oxygenation; *FFP,* fresh frozen plasma; *Hb,* haemoglobin; *PCV,* packed cell volume; *WCC,* white cell count.

membrane, microcirculatory changes, and destruction of the endothelial glycocalyx.

Choice of fluid for volume replacement

Blood

This is conventionally given for haemorrhagic shock as soon as it is available. In extreme bleeding emergencies, blood group O Rhesus D negative (universal donor) red cells can be used, although group-specific crossmatch can be performed in minutes (see Chapter 38). When available and not contraindicated, blood salvage may be employed for those with severe ongoing bleeding.

Transfusion of whole blood has largely been replaced by that of red cell concentrates. The use of leucodepleted blood is considered to be safer in terms of immune-mediated transfusion reactions, disease transmission, and immune suppression. Although red cell transfusion can augment oxygen-carrying capacity, and hence global oxygen delivery, transfusion of old stored red blood cells, which become spherical rather than biconcave and poorly

deformable with increased adhesiveness, can compromise microvascular flow. In the case of traumatic or intraoperative haemorrhagic shock, many now favour 'hypotensive resuscitation' with blood and blood products until surgical control of the bleeding has been achieved. This partial resuscitation has been shown to reduce the rate of bleeding and allows the surgeon a greater chance of achieving definitive haemostasis. Once the patient is stable, higher haemoglobin and blood pressure targets may be pursued.

The absolute haemoglobin target in critically ill patients has changed with increasing awareness of the subtle detrimental immune effects associated with transfusing blood. In patients not actively bleeding and without significant cardiovascular risk factors, it is usual to transfuse in order to achieve a target haemoglobin level of 70 g/L or more ('restrictive transfusion strategy'). In patients with significant coronary artery disease or limited respiratory reserve, a target level of 90 g/L ('liberal transfusion strategy') may be more appropriate but remains controversial. Healthy individuals can tolerate haemoglobin levels as low as 40 g/L before displaying signs of cardiac ischaemia.

Massive blood transfusion can be defined as a volume of more than 8 to 10 units of red cells transfused within a 24-hour period, and ***massive haemorrhage*** as a loss of 50% or more of blood volume within 3 hours or a rate of blood loss exceeding 150 mL/min.

Complications of blood transfusion are discussed in Chapter 38. Special problems arise as a result of massive transfusion:

- ***Temperature changes.*** Bank blood is stored at 4°C; transfusion may result in hypothermia, peripheral venoconstriction, and arrhythmias. Blood should be warmed during massive transfusion and in those at risk of hypothermia (e.g., prolonged major surgery with open body cavity).
- ***Coagulopathy.*** Stored blood has virtually no effective platelets or clotting factors. Massive blood transfusions are now accompanied by administration of fresh frozen plasma and platelet concentrates in a balanced ratio of 1:1:1, especially in the treatment of severe traumatic haemorrhage. Cryoprecipitate also plays a key role in securing haemostasis. Recombinant factor VIIa is rarely indicated in those with uncontrollable bleeding, and prothrombin complex concentrates have the advantage that they do not need to be crossmatched or thawed but are expensive. There is significant evidence to support the use of tranexamic acid in patients following major haemorrhage as a result of trauma, postpartum, and major surgery and should be administered early.
- ***Hypocalcaemia.*** Citrate is used to stop coagulation in stored blood by binding calcium ions. As a result, during large transfusion, total body ionized calcium levels often drop, causing myocardial depression and exacerbating coagulation defects. This is corrected by administering intravenous calcium chloride.
- ***Increased oxygen affinity.*** In stored blood, the red cell 2,3-diphophoglycerate (2,3-DPG) content is reduced so that the oxyhaemoglobin dissociation curve is shifted to the left. The oxygen affinity of haemoglobin is therefore increased and oxygen unloading is impaired. Red cell levels of 2,3-DPG are substantially restored within 12 hours of transfusion.
- ***Hyperkalaemia.*** Plasma potassium levels rise progressively as blood is stored. However, hyperkalaemia is rarely a problem, as rewarming of the blood increases red cell metabolism and reactivates the sodium/potassium membrane pumps resulting in uptake of potassium by the red cells. However, there is the occasional need for managing hyperkalaemia with insulin dextrose boluses.
- ***Microembolism.*** Microaggregates in stored blood may obstruct the pulmonary capillaries. This process is thought by some to contribute to acute lung injury.
- ***Immunity.*** Blood transfusion has subtle detrimental effects on the host immune system and has been implicated in a greater susceptibility to nosocomial infections and an increase in the risk of cancer recurrence.

Crystalloids and colloids

The choice of intravenous fluid for resuscitation and the relative merits of crystalloids or colloids have long been controversial. Crystalloid solutions are cheap and convenient to use. However, they are not completely free of side-effects, as large volumes of solutions with high chloride content, such as 0.9% saline, can cause a hyperchloraemic acidosis and are associated with a greater risk of renal impairment. Balanced solutions, such as compound sodium lactate or Plasma-Lyte 148, are preferred.

Crystalloid solutions are made up of small molecules, and due to the increased capillary leak seen in conditions such as sepsis, larger volumes are required to achieve the end points of resuscitation. As a result, colloid solutions (gelatin or hydroxyethyl starch based), with larger molecules, were previously more extensively used as they were more efficient at stabilizing resuscitation end points with lower volumes required to do so. However, recent evidence has suggested higher mortality rates, as well as more AKI and clotting disorders. Furthermore, colloid solutions are more expensive to make and can cause anaphylactic reactions. As a result, their use is no longer recommended.

Human albumin solution (HAS) is a natural colloid that has been used for volume replacement in shock and burns and for the treatment of hypoproteinaemia. In critically ill patients, resuscitation with equivalent volumes of either 0.9% saline or the colloid 4% albumin has been shown to result in similar outcomes. As a result, HAS is not generally recommended for routine volume replacement because supplies are limited, and other cheaper solutions are equally effective. However, it may have a role in the treatment of hepatorenal syndrome in patients with liver cirrhosis or low albumin.

Myocardial contractility and inotropic agents

Myocardial contractility can be impaired by many factors, such as hypoxaemia and hypocalcaemia, as well as by some drugs (e.g., beta-blockers, antiarrhythmics, and sedatives).

Severe metabolic acidosis is conventionally said to depress myocardial contractility and limit the response to vasopressor agents. Attempted correction of acidosis with intravenous sodium bicarbonate, however, generates additional carbon dioxide, which diffuses across cell membranes, producing or exacerbating intracellular acidosis. Other disadvantages of bicarbonate therapy include sodium overload and a left shift of the oxyhaemoglobin dissociation curve. Ionized calcium levels may be reduced and, combined with the fall in intracellular pH, this may impair myocardial performance. Treatment of lactic acidosis should therefore concentrate on correcting the cause. Bicarbonate should be administered only to correct ***extreme persistent metabolic acidosis*** (see Chapter 27).

If the signs of shock persist, despite adequate volume replacement, and perfusion of vital organs is jeopardized, inotropic/vasopressor agents should be administered to improve cardiac output and blood pressure. Vasopressor therapy may also be required to maintain perfusion in those with life-threatening hypotension, even when volume replacement is incomplete. All inotropes increase myocardial oxygen consumption, particularly if a tachycardia develops, and this can lead to an imbalance between myocardial

oxygen supply and demand, with the development or extension of ischaemic areas. Inotropes should therefore be used with special caution, particularly in cardiogenic shock following myocardial infarction and in known ischaemic heart disease.

Many of the most seriously ill patients become increasingly resistant to the effects of pressor agents, an observation attributed to 'downregulation' of adrenergic receptors and nitric oxide-induced 'vasoplegia' (see later).

All inotropic agents should be administered via a large central vein and their effects continually monitored (**Box 13.9**).

Adrenaline (epinephrine)

Adrenaline stimulates both α- and β-adrenergic receptors, but β effects predominate at low doses. Heart rate and cardiac index increase while peripheral resistance is reduced. If there is an associated increase in perfusion pressure, urine output may improve. Adrenaline at higher doses can cause excessive (α-mediated) vasoconstriction with reductions in splanchnic flow, and cardiac output may fall. Prolonged high-dose administration can cause peripheral gangrene and lactic acidosis. The minimum effective dose of adrenaline should therefore be used for as short a time as possible.

Noradrenaline (norepinephrine)

This is predominantly an α-adrenergic agonist. It is particularly useful in patients with hypotension and a low systemic vascular resistance, such as is seen in septic shock. There is a risk of producing excessive vasoconstriction with impaired organ perfusion and increased afterload, particularly if noradrenaline is administered when fluid resuscitation is inadequate, and the circulating volume is reduced.

Dopamine

The haemodynamic effects of dopamine are dose dependent. At low doses, dopamine is a positive inotrope with vasodilator actions on the renal and splanchnic circulation. Higher doses may be complicated by tachycardia, arrhythmias, and vasoconstriction. Compared to noradrenaline, the use of dopamine is associated with increased risk of death and arrhythmic events in patients with septic shock and is no longer recommended.

Dopexamine

Dopexamine is an analogue of dopamine that activates β_2 receptors, as well as DA_1 and DA_2 receptors. Dopexamine is a weak positive inotrope but a powerful splanchnic vasodilator, reducing afterload and improving blood flow to vital organs, including the kidneys. It has been used as an adjunct to the perioperative management of high-risk surgical patients (see below).

Dobutamine

Dobutamine is closely related to dopamine and has predominantly β_1 activity. Dobutamine has no specific effect on the renal vasculature, but urine output often increases as cardiac output and blood pressure improve. It reduces systemic vascular resistance, as well as improves cardiac performance, thereby decreasing afterload and ventricular filling pressures. Dobutamine is therefore useful in patients with cardiogenic shock and cardiac failure.

Box 13.9 Receptor actions of sympathomimetic and dopaminergic agents

Agent	Receptor						Dose dependence
	β_1	β_2	α_1	α_2	DA_1	DA_2	
Adrenaline (epinephrine)							++++
Low-dose	++	+	+	±	–	–	
Moderate-dose	++	+	++	+	–	–	
High-dose	++(+)	+(+)	++++	+++	–	–	
Noradrenaline (norepinephrine)	++	0	+++	+++	–	–	+++
Isoprenaline	+++	+++	0	0	–	–	
Dopamine							+++++
Low-dose	±	0	±	+	++	+	
Moderate-dose	++	+	++	+	++(+)	+	
High-dose	+++	++	+++	+	++(+)	+	
Dopexamine	+	+++	0	0	++	+	++
Dobutamine	++	+	±	?	0	0	++
Action	*Postsynaptic:* positive inotropism and chronotropism; renin release	*Presynaptic:* stimulation of noradrenaline release *Postsynaptic:* positive inotropism and chronotropism; vascular dilation; relaxation of bronchial smooth muscle	*Postsynaptic:* constriction of peripheral, renal and coronary vascular smooth muscle; positive inotropism; antidiuresis	*Presynaptic:* inhibition of noradrenaline release; vasodilation *Postsynaptic:* constriction of coronary arteries; promotion of salt and water excretion	*Postsynaptic:* dilation of renal, mesenteric, and coronary vessels; renal tubular effect (natriuresis, diuresis)	*Presynaptic:* inhibition of adrenaline release	

Phosphodiesterase inhibitors (e.g., milrinone, enoximone)

These agents have both inotropic and vasodilator properties. Because the phosphodiesterase type III inhibitors bypass the β-adrenergic receptor, they may cause less tachycardia and fewer arrhythmias than β agonists. They are useful in patients with receptor 'downregulation', those receiving beta-blockers, those being weaned from cardiopulmonary bypass, and those with cardiac failure.

Vasopressin

Low-dose vasopressin can increase blood pressure and systemic vascular resistance in patients with vasodilatory septic shock and a high cardiac output unresponsive to other vasopressors ('vasoplegia'). Low-dose vasopressin is often added to treatment of patients with septic shock with escalating doses of conventional vasopressors such as noradrenaline.

Levosimendan

Levosimendan is a myofilament calcium sensitizer. Unlike other inotropes, levosimendan does not exert its action through increases in intracellular Ca^{2+} and, as a result, does not impair diastolic relaxation of the heart. Levosimendan has phosphodiesterase inhibitor actions, but these are not thought to be clinically significant. The dose is usually infused over 24 hours but, significantly, a long-acting metabolite of levosimendan has similar calcium-sensitizing actions, maintaining the inotropic effect of levosimendan once an infusion is stopped. Adverse cardiovascular effects of levosimendan include tachycardia and hypotension; as a consequence, the addition of a vasopressor may be required.

Summary for use of inotropic and vasopressor agents

- In ***vasodilated septic patients*** with an adequate cardiac output, noradrenaline (norepinephrine) is often used alone to augment the blood pressure during and following fluid resuscitation. A combination of dobutamine and noradrenaline (norepinephrine) can be used for those showing signs of low cardiac output: Dobutamine is given to achieve optimal cardiac output.
- Vasopressin is used to supplement escalating doses of noradrenaline in the patient with vasodilatory shock.

There is evidence to suggest that adrenaline (epinephrine) may be equally safe and effective as a dobutamine/noradrenaline combination. Because of its potency, adrenaline is particularly useful in patients with ***refractory hypotension***. Phosphodiesterase inhibitors can be used in the management of cardiac failure, especially when associated with pulmonary hypertension, and perioperatively in those undergoing cardiac surgery. Dobutamine is an alternative that is also used in septic patients with fluid overload or myocardial failure. The role of levosimendan in the management of shock is less clear due to high cost and lack of mortality benefit, but it may be useful in resistant cardiogenic shock.

Targeting haemodynamics and oxygen transport

Critically ill patients who survive often have raised values for cardiac output and oxygen delivery, DO_2 and VO_2. However, elevation of DO_2 and VO_2 to these 'supranormal' levels following admission to intensive care may in fact be harmful. Despite its initial promise, early goal-directed therapy to resuscitate patients in the emergency room aimed at maintaining a central venous oxygen saturation of more than 70% with augmentation of cardiac output and haemoglobin concentration does not appear to improve outcome in patients with severe sepsis or septic shock. However, the principles of optimization of cardiac output through fluid resuscitation and targeting mean arterial blood pressures of 65 mmHg or more are the basis of early management of patients with septic shock.

Box 13.10 Patients at risk of developing perioperative multiorgan failure

- The elderly
- Patients with co-morbidity, especially limited cardiorespiratory reserve
- Patients with trauma to two body cavities requiring multiple blood transfusions
- Patients undergoing surgery involving extensive tissue dissection, e.g., oesophagectomy, pancreatectomy, aortic aneurysm surgery
- Patients undergoing emergency surgery for intra-abdominal or intrathoracic catastrophic states, e.g., faecal peritonitis, oesophageal perforation

High-risk surgical patients

These patients benefit from intensive perioperative monitoring and circulatory support in a critical care area (**Box 13.10**), in particular, maintenance of oxygenation, adequate circulating volume, and blood pressure. Circulating volume replacement and administration of inotropes or vasopressors should be guided by monitoring of stroke volume/cardiac output. The value of the routine use of inodilators such as dopexamine remains unclear.

Vasodilator therapy

In selected cases, afterload reduction is used to increase stroke volume and decrease myocardial oxygen requirements by reducing the systolic ventricular wall tension. Vasodilation also decreases heart size and the diastolic ventricular wall tension so that coronary blood flow is improved. Vasodilators also improve microcirculatory flow.

Vasodilator therapy can be particularly helpful in patients with cardiac failure in whom the ventricular function curve is flat (see **Fig. 13.8**) so that falls in preload have only a limited effect on stroke volume. This form of treatment, combined in selected cases with inotropic support, is therefore useful in the management of patients with cardiogenic shock and in those with cardiogenic pulmonary oedema or mitral regurgitation.

Nitroglycerine (NTG), at low doses, is predominantly a venodilator, but as the dose is increased, it also causes arterial dilation, thereby decreasing both preload and afterload. Nitrates are particularly useful in the treatment of cardiac failure with pulmonary oedema and are usually given in combination with intravenous furosemide. NTG reduces pulmonary vascular resistance, an effect that can be exploited in patients with a low cardiac output secondary to pulmonary hypertension.

Sodium nitroprusside (SNP) dilates arterioles and venous capacitance vessels, as well as the pulmonary vasculature, by donating nitric oxide. SNP therefore reduces the afterload and preload of both ventricles and can improve cardiac output and the myocardial oxygen supply/demand ratio. The effects of SNP are rapid in onset and spontaneously reversible within a few minutes of discontinuing the infusion. Cyanide poisoning is a risk with high-dose, prolonged infusions.

Mechanical support of the myocardium

Intra-aortic balloon pump counterpulsation (IABP) is the technique used most widely for mechanical support of the failing myocardium. It is discussed in Chapter 17. In specialized centres, ventricular

assist devices and veno-arterial ECMO (see below) may be used in the treatment of cardiac failure.

Further reading

Angus DC, Barnato AE, Bell D et al. A systematic review and meta-analysis of early goal-directed therapy for septic shock: the ARISE, ProCESS and ProMISe investigators. *Intensive Care Med* 2015; 41:1549–1560.
DeBacker D, Aldecoa C, Nijimi H, et al. Dopamine versus norepinephrine in the treatment of septic shock: a meta-analysis. *Crit Care Med* 2012; 40(3):725–730.
Evans L, Rhodes A, Alhazzani W et al. Surviving sepsis campaign: International guidelines for management of sepsis and septic shock 2021. *Crit Care Med* 2021; 49(11):e1063–e1143.
Fajgenbaum DC, June CH. Cytokine storm. *N Engl J Med* 2020; 383:2255–2273.
Finfer S, Myburgh J, Bellomo R. Intravenous fluid therapy in critically ill adults. *Nat Rev Nephrol* 2018; 14:541–557.
Hunt BS. Bleeding and coagulopathies in critical care. *N Engl J Med* 2014; 370:847–859.
Jha A, Zilahi G, Rhodes A. Vasoactive therapy in shock. *BJA Education* 2021; 21:270–277.
Meyhoff TS, Hjortrup, PB, Wetterslev J. Restriction of intravenous fluid in ICU patients with septic shock. *N Engl J Med* 2022; 386:2459–2470.
Pearse RM, Harrison DA, MacDonald A et al. Effect of a perioperative, cardiac output-guided hemodynamic therapy algorithm on outcomes following major gastrointestinal surgery: a randomized clinical trial and systematic review. *JAMA* 2014; 311:2181–2190.
Pirracchio R, Annane D, Waschka AK et al. Patient-level meta-analysis of low-dose hydrocortisone in adults with septic shock. *NEJM Evid* 2023; 2(6).
Rautanen A, Mills TC, Gordon AC et al. Genome-wide association study of survival from sepsis due to pneumonia: an observational cohort study. *Lancet Respir Med* 2015; 3:53–60.
Russell JA. Is there a good MAP for septic shock? *N Engl J Med* 2014; 370:1649–1651.
Venkatesh B, Finfer S, Cohen J et al. Adjunctive glucocorticoid therapy in patients with septic shock (ADRENAL study). *N Engl J Med* 2018; 378:797–808.
Zampieri FG, Bagshaw SM, Semler MW. Fluid therapy for critically ill adults with sepsis: a review. *JAMA* 2023; 329(22):1967–1980.
http://www.survivingsepsis.org *Surviving Sepsis campaign*.

RESPIRATORY FAILURE

(See also Chapter 20).

Classification and aetiology

The respiratory system consists of a gas-exchanging organ (the lungs) and a ventilatory pump (respiratory muscles/thorax), either or both of which can fail and precipitate respiratory failure. Respiratory failure occurs when pulmonary gas exchange is sufficiently impaired to cause hypoxaemia with or without hypercarbia. In practical terms, respiratory failure is present when the P_aO_2 is below 8 kPa (60 mmHg) or the P_aCO_2 is above 6.7 kPa (50 mmHg). It can be divided into:

- type I respiratory failure, in which the P_aO_2 is low and the P_aCO_2 is normal or low
- type II respiratory failure, in which the P_aO_2 is low and the P_aCO_2 is high.

Type I or 'acute hypoxaemic' respiratory failure often occurs as a result of right-to-left shunts or $\dot{V}/\dot{Q}$ mismatch. Common causes include pneumonia, ARDS, cardiogenic pulmonary oedema, pulmonary embolism, and lung fibrosis.

Type II or 'ventilatory failure' occurs when alveolar ventilation is insufficient to remove the volume of carbon dioxide being produced by tissue metabolism. Inadequate alveolar ventilation may be due to reduced ventilatory effort, inability to overcome an increased resistance to ventilation, failure to compensate for an increase in dead space and/or carbon dioxide production, or a combination of these factors. The most common cause is COPD. Other causes include chest-wall deformities, respiratory muscle weakness (e.g., Guillain–Barré syndrome), and depression of the respiratory centre (e.g., drug overdose).

Deterioration in the mechanical properties of the lungs and/or chest wall increases the work of breathing and the oxygen consumption/carbon dioxide production of the respiratory muscles. Respiratory muscle fatigue is a factor in the pathogenesis of respiratory failure.

Clinical features

A clinical assessment of respiratory distress should be made on the following criteria:

- The use of accessory muscles for respiration
- Intercostal recession
- Tachypnoea
- Tachycardia
- Sweating
- Pulsus paradoxus (rarely present)
- Inability to speak, unwillingness to lie flat
- Agitation, restlessness, diminished conscious level
- Asynchronous respiration (a discrepancy in the timing of movement of the abdominal and thoracic compartments)
- Paradoxical respiration (abdominal and thoracic compartments move in opposite directions)
- Respiratory alternans (breath-to-breath alteration in the relative contribution of the intercostal/accessory muscles and the diaphragm).

Investigations

The ***most sensitive clinical indicator*** of increasing respiratory difficulty is a rising respiratory rate. ***Vital capacity*** is a better guide to deterioration, particularly in patients with respiratory inadequacy due to neuromuscular problems such as the Guillain–Barré syndrome or myasthenia gravis, where the vital capacity decreases as weakness worsens. Measurement of ***forced expiratory volume in 1 second (FEV_1)*** is useful in the assessment of patients suffering from acute asthma or COPD. ***Blood gas analysis*** should be performed to guide oxygen therapy and to provide an objective assessment of the severity of the respiratory failure.

Monitoring

Pulse oximetry

Pulse oximeters applied to a finger or earlobe measure the changing amount of light transmitted through the pulsating arterial blood and provide a continuous, non-invasive measurement of arterial oxygen saturation (S_pO_2). These devices are reliable and easy to use and do not require calibration.

Pulse oximetry is not a sensitive guide to *changes* in oxygenation. An S_pO_2 within normal limits in a patient receiving supplemental oxygen does not exclude the possibility of hypoventilation with carbon dioxide retention. Readings can be inaccurate in those with poor peripheral perfusion.

Blood gas analysis

Normal values of blood gas analysis are shown in **Box 13.5**. Disposable pre-heparinized syringes are available for blood gas analysis.

- The sample should be analyzed immediately. Alternatively, the syringe should be put on ice to prevent the continuing metabolism of white cells from causing a reduction in PO_2 and a rise in PCO_2.

- Air almost inevitably enters the sample and air bubbles should be ejected immediately. The gas tensions within these air bubbles will equilibrate with those in the blood, thereby lowering the PCO_2 and usually raising the PO_2 of the sample.

Interpretation of the results of blood gas analysis can be considered in two separate parts:

- Disturbances of acid–base balance (see Chapter 27)
- Alterations in oxygenation.

Correct interpretation requires a knowledge of the patient's clinical history and age, the inspired oxygen concentration, patient temperature, and any other relevant treatment (e.g., the ventilator settings for those on mechanical ventilation or the administration of sodium bicarbonate). The oxygen content of the arterial blood is determined by the percentage saturation of haemoglobin with oxygen. The relationship between the latter and the P_aO_2 is determined by the oxyhaemoglobin dissociation curve (see **Fig. 13.3**).

P_aO_2/F_iO_2 ratio

This calculated variable, the ratio of partial pressure of oxygen in arterial blood to the fractional inspired oxygen, is a simple approximation to the alveolar-arterial oxygen difference ($P_{A-a}O_2$) and can be used to assess the severity of respiratory failure, particularly in patients with ARDS (see 'Acute respiratory distress syndrome' section). It is calculated by taking the arterial PO_2 in mmHg and dividing by the F_iO_2 expressed as a fraction of 1.

Capnography

Continuous breath-by-breath analysis of the expired carbon dioxide concentration can be used to:

- confirm tracheal intubation (standard of care for in-hospital practice);
- continuously monitor end-tidal PCO_2, which approximates to P_aCO_2 in normal subjects and is used to assess effectiveness of ventilation;
- detect acute airway problems, e.g., blocked or dislodged tracheal tube/tracheostomy, in all mechanically ventilated patients (essential when transporting the critically ill); and
- detect acute alterations in cardiorespiratory function (e.g., a sudden fall in cardiac output).

Management

Standard management of patients with respiratory failure includes:

- Administration of supplemental oxygen through a patent airway
- Treatment for distal airway obstruction
- Measures to limit pulmonary oedema
- Control of secretions
- Treatment of pulmonary infection.

The load on the respiratory muscles should be reduced by improving lung mechanics. Correction of abnormalities that may lead to respiratory muscle weakness, such as hypophosphataemia and malnutrition, is also necessary.

Oxygen therapy

Methods of oxygen administration

Oxygen is initially given via a face mask. In the majority of patients (except those with COPD with chronically elevated P_aCO_2), the precise concentration of oxygen given is not vital and oxygen can therefore be delivered by a 'variable performance' device such as a simple face mask or nasal cannula (**Fig. 13.23**).

With these devices, the inspired oxygen concentration varies from about 35% to 55%, with oxygen flow rates of 6 to 10 L/min. Nasal cannulae are often preferred because they are less claustrophobic and do not interfere with feeding or speaking, but they can cause ulceration of the nasal mucosa and the inspired oxygen concentration is diluted by mouth breathing. Higher concentrations of oxygen can be administered by using a mask with a reservoir bag attached (**Fig. 13.23C**). **Fig. 13.23** should be compared with the fixed-performance mask shown in **Fig. 13.24**. The latter allows the oxygen concentration to be controlled and is used in patients with COPD and chronic type II failure who rely on their hypoxic drive to stimulate respiration. The hazards of reducing hypoxic drive can at times be overemphasized and are less dangerous when the patient is in a critical care unit. *Remember*, severe hypoxaemia is more dangerous than hypercapnia.

Oxygen toxicity

Hyperoxia can cause pulmonary damage due to increased oxygen free radicals and oxidative damage to the lung tissue. Although dangerous hypoxia through fear of pulmonary oxygen toxicity should never be tolerated, oxygen saturations between 90% and 92% are probably adequate for most patients and there is some suggestion that targeting higher oxygen saturations using greater concentrations of inspired oxygen may be detrimental (e.g., post-myocardial infarction). Recent evidence suggests that even a brief period of hyperoxia may increase mortality compared with normoxia, even hyperoxia following emergency intubations for any cause in the emergency department.

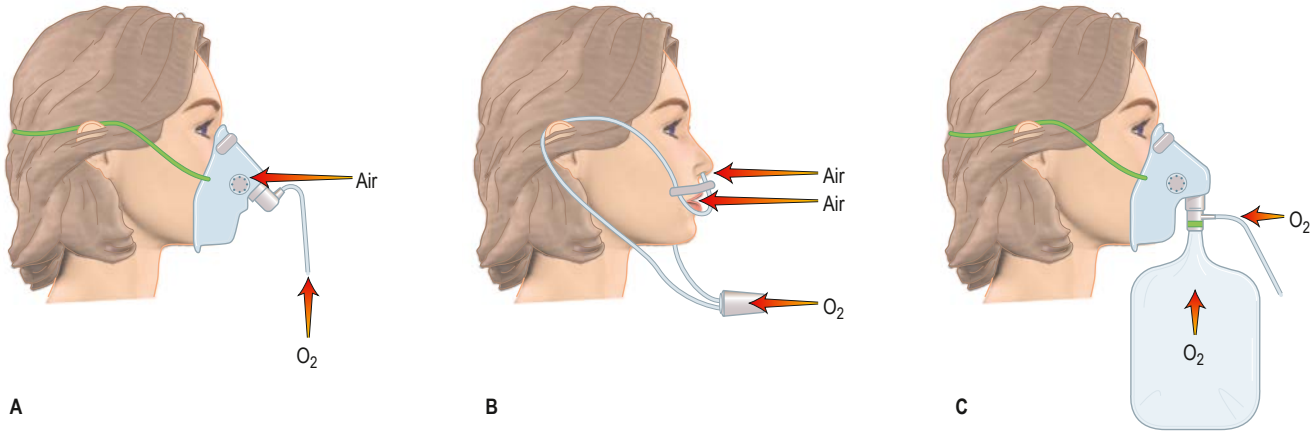


Fig. 13.23 Methods of administering supplemental oxygen to the unintubated patient. (A) Simple face mask. **(B)** Nasal cannulae. **(C)** Mask with reservoir bag and non-rebreathing valve.

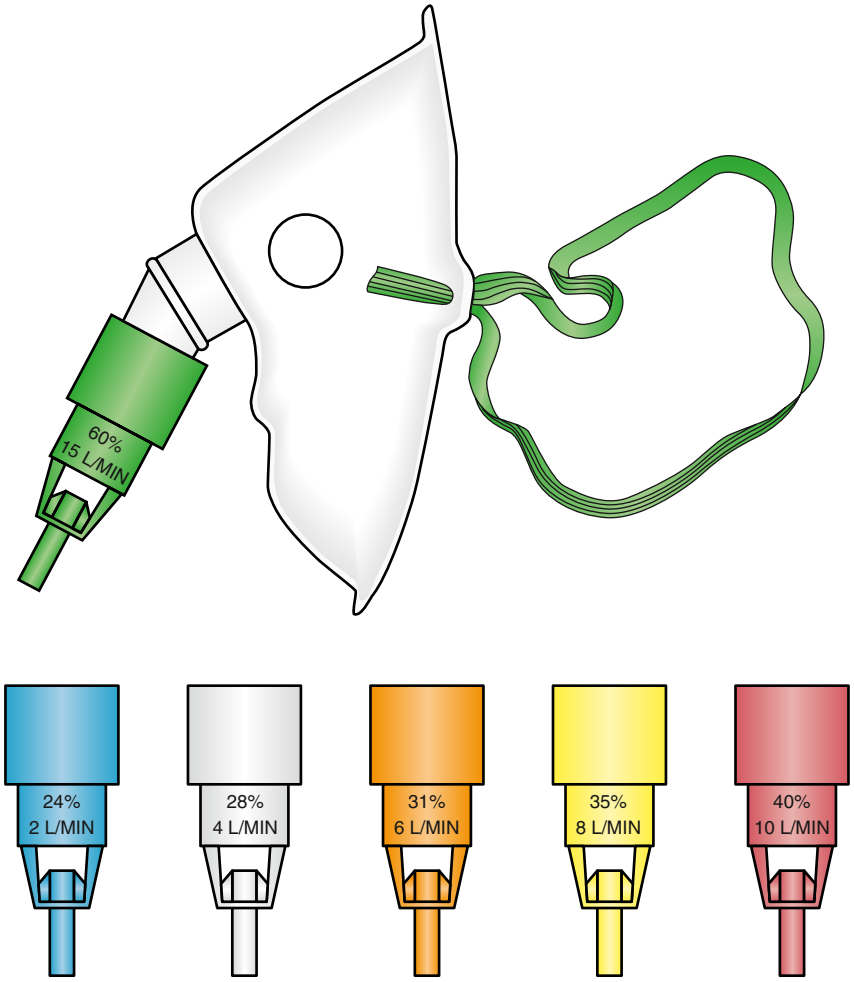


Fig. 13.24 **Fixed-performance Venturi valves attached to a face mask to deliver a fixed percentage of oxygen.**

Respiratory support

If, despite the above measures, the patient continues to deteriorate or fails to improve, the institution of some form of respiratory support is necessary (**Box 13.11**). ***Non-invasive ventilation (NIV)*** via a mask or hood (see 'Non-invasive ventilation' section and also Chapter 11) can be used, but in many critically ill patients ***invasive ventilation*** through an endotracheal tube or tracheostomy is required.

Invasive mechanical ventilation is achieved by intermittently inflating the lungs with positive pressure delivered by a mechanical ventilator through an endotracheal or tracheostomy tube, commonly carried out in the absence of spontaneous breaths. This is called ***IPPV*** or controlled mechanical ventilation (CMV). However, technological development has meant that it is now more commonly carried out by augmenting the patient's own spontaneous breaths (see 'Techniques for respiratory support' section). This has the advantage of maintaining the patient's own respiratory muscle activity and function, thereby avoiding muscle atrophy.

Indications for mechanical ventilation

Acute respiratory failure, with signs of severe respiratory distress (e.g., respiratory rate >40 breaths/min, inability to speak, patient exhaustion) persisting despite maximal therapy, requires mechanical ventilation. Confusion, restlessness, agitation, a decreased conscious level, a rising P_aCO_2 (>8 kPa, >60 mmHg), and extreme hypoxaemia (<8 kPa, <60 mmHg), despite oxygen therapy, are further indications.

In ***acute ventilatory failure*** due, for example, to myasthenia gravis, Guillain–Barré syndrome, or high spinal cord injuries, mechanical ventilation should usually be instituted when the vital capacity has fallen to 10 mL/kg or less. This will avoid complications such as atelectasis and infection, as well as preventing respiratory arrest. The tidal volume and respiratory rate are relatively insensitive indicators of respiratory failure in these conditions and change late in the course of the disease. A high P_aCO_2 (particularly if rising) is an indication of urgent mechanical ventilation.

Not all patients with respiratory failure and/or a reduced vital capacity require ventilation; clinical assessment of each individual case is essential. The patient's general condition, degree of exhaustion, level of consciousness, and ability to protect the airway are often more useful than blood gas values. However, such patients require very close observation of signs of deterioration.

Other indications include:

- Postoperative ventilation in high-risk patients
- Head injury: to avoid hypoxia and hypercarbia, which increase cerebral blood flow and intracranial pressure (ICP, see later)

Box 13.11 Techniques for respiratory support

Technique	Delivery	
Intermittent positive-pressure ventilation (IPPV)		
Controlled mechanical ventilation (CMV) (volume-controlled or pressure-controlled) Synchronized intermittent mandatory ventilation (SIMV)	Usually with	Positive end-expiratory pressure (PEEP) *Lung protective ventilation* Pressure-limited or volume-controlled pressure-limited (peak airway pressure <35–40 cmH_2O) Low tidal volume 6–8 mL/kg (±'permissive hypercarbia') maintain alveolar volume with PEEP and, in some cases, prolonged inspiratory phase
Pressure support ventilation (PSV)		
Biphasic positive airway pressure (BiPAP)		
Non-invasive ventilation	Nasal mask Face mask Hood	
Continuous positive airway pressure (CPAP)	Endotracheal tube, tracheostomy, mask or hood, high-flow nasal prongs	
Extracorporeal techniques	Extracorporeal membrane oxygenation (ECMO), extracorporeal CO_2 removal ($ECCO_2$-R)	

- Trauma: chest injury with lung contusions and high spinal cord injuries
- Severe left ventricular failure with pulmonary oedema
- Coma with airway compromise or breathing difficulties, e.g., following drug overdose.

Institution of invasive respiratory support

This requires tracheal intubation, which should be performed under anaesthesia only by experienced staff. If the patient is conscious, the procedure should be fully explained and consent obtained. The complications of tracheal intubation are given in **Box 13.12**.

Intubating a critically ill patient is very different from intubating a patient in the operating theatre for elective surgery. The patient is usually hypoxic and hypercarbic, with increased sympathetic activity; the stimulus of laryngoscopy and intubation can precipitate dangerous arrhythmias, bradycardia, and even cardiac arrest. In extreme emergencies, it may be necessary to ventilate the patient by hand using a bag-valve-mask (supplemented with high-flow oxygen) and an oropharyngeal airway or a laryngeal mask airway, until experienced help arrives.

The ECG and oxygen saturation should be monitored, and the patient pre-oxygenated with 100% oxygen before intubation. Resuscitation drugs should be immediately available. If time allows, the circulating volume should be optimized and, if necessary, inotropes commenced before intubation is attempted. In some cases, it is appropriate to establish intra-arterial and CVP monitoring before instituting mechanical ventilation, although many patients will not tolerate the supine or head-down position. In some deeply comatose patients, no sedation may be required, but in the majority, a short-acting intravenous anaesthetic agent, usually with an opiate followed by muscle relaxation, will be necessary. Capnography must be used to confirm tracheal intubation.

Box 13.12 Complications of tracheal intubation

Complication	Comments
Immediate	
Trauma to the upper airway Tube in oesophagus	Affects lips, teeth, gums, trachea Gives rise to hypoxia and abdominal distension Detected by absence of capnography trace Requires immediate removal, bag-mask ventilation with oxygen, and re-insertion of tracheal tube
Tube in one or other (usually the right) main bronchus	Avoid by checking both lungs are being inflated, i.e., both sides of chest move, and air entry is heard bilaterally on auscultation Obtain chest X-ray to check position of tube and to exclude lung collapse
Early	
Migration of tube out of trachea Leaks around tube Obstruction of tube because of kinking or secretions	These are dangerous complications Patient becomes distressed and cyanosed and has poor chest expansion The following should be performed immediately: Manual inflation with 100% oxygen Tracheal suction Check position of tube Deflation of cuff Check tube for 'kinks' or blockage with secretions or blood (common) If no improvement, remove tube, ventilate with face mask, and then insert new endotracheal tube
Late	
Sinusitis Mucosal oedema and ulceration Laryngeal injury Tracheal narrowing and fibrosis Tracheomalacia	

Sedation, analgesia, and muscle relaxation

Most critically ill patients require analgesia and many will receive sedatives. The combination of an opiate with propofol or a benzodiazepine is often used to facilitate mechanical ventilation. Heavy sedation is indicated in those with severe respiratory failure, especially since 'lung-protective' ventilatory strategies (see later) are inherently uncomfortable. A few may require neuromuscular blockade. It is recognized, however, that minimizing sedation levels using 'sedation scores' and 'daily wakening', or even the avoidance of sedatives altogether, often in combination with spontaneous breathing modes of respiratory support (see later), is associated with reductions in the duration of mechanical ventilation and more rapid discharge from the ICU and hospital.

Tracheostomy

Tracheostomy may be required for long-term control of the airway in some patients with persistent reduced conscious state, to manage excessive bronchial secretions, facilitate respiratory weaning, reduce sedation requirements, and improve patient comfort.

Tracheostomy is now performed most commonly on the ICU by the intensivist, adopting a percutaneous dilational approach. This

approach is quicker, less resource intensive, safe, and associated with a greatly reduced wound infection rate than surgical tracheostomy performed in the operating theatre. Surgical tracheostomy is reserved for when anatomical factors preclude a percutaneous approach or when there are concerns regarding abnormal coagulation that may require surgical haemostasis.

A life-threatening obstruction of the upper respiratory tract that cannot be bypassed with an endotracheal tube can be relieved by a ***cricothyroidotomy***, which is safer, quicker, and easier to perform than a formal tracheostomy.

Tracheostomy has a small but significant mortality rate. Complications associated with the technique are shown in **Box 13.13**.

Complications associated with mechanical ventilation

Airway complications

See **Boxes 13.12 and 13.13**.

Disconnection, failure of gas or power supply, and mechanical faults

These are unusual but dangerous. A method of manual ventilation, a face mask, and oxygen must always be available by the bedside.

Cardiovascular complications

The application of positive pressure to the lungs impedes venous return and distends alveoli, thereby 'stretching' the pulmonary capillaries and causing a rise in pulmonary vascular resistance. Both of these mechanisms can produce a fall in cardiac output.

Respiratory complications

Mechanical ventilation can be complicated by a deterioration in gas exchange because of $\dot{V}/\dot{Q}$ mismatch, fluid retention, and collapse of peripheral alveoli. Traditionally, the latter was prevented by using high tidal volumes (10–12 mL/kg) but high inflation pressures, with over-distension of compliant alveoli, perhaps exacerbated by the repeated opening and closure of distal airways, can disrupt the alveolar–capillary membrane. There is an increase in microvascular permeability and release of inflammatory mediators, leading to ***'ventilator-associated lung injury'***. This ***'barotrauma'*** and ***'volutrauma'*** can result in rupture of the alveoli and may be complicated by pneumomediastinum, subcutaneous emphysema, pneumothorax, and intra-abdominal air. The risk of pneumothorax is increased in those with destructive lung disease (e.g., necrotizing pneumonia, emphysema), asthma, or fractured ribs.

A ***tension pneumothorax*** can be rapidly fatal in ventilated patients. Signs include the development or worsening of hypoxia, hypercarbia, respiratory distress, and an unexplained increase in airway pressure, as well as hypotension and tachycardia, sometimes accompanied by a rising CVP. Examination may reveal unequal chest expansion, mediastinal shift away from the side of the pneumothorax (deviated trachea, displaced apex beat), and a hyper-resonant hemithorax. Although breath sounds are often diminished over the pneumothorax, this sign can be misleading in ventilated patients. If there is time, the diagnosis can be confirmed by chest X-ray. In a rapidly deteriorating patient with a high index of suspicion for a tension pneumothorax, a needle decompression into the second intercostal space in the mid-clavicular line may be indicated, followed by insertion of a definitive chest drain.

Ventilator-associated pneumonia

Hospital-acquired pneumonia occurs in as many as one-third of patients receiving mechanical ventilation and is associated with a significant increase in length of stay and mortality. The diagnosis can be controversial, as pyrexia and infiltrates on the chest X-ray that may not be infective in origin are common. Organisms commonly isolated are aerobic Gram-negative bacilli (e.g., *Pseudomonas aeruginosa*, *Klebsiella pneumoniae*, *Escherichia coli*, *Acinetobacter* spp.) and *Staphylococcus aureus*, including methicillin-resistant *S. aureus* (MRSA). Deciding whether an organism that has been isolated is causing ventilator-associated pneumonia or is simply colonizing the respiratory tract can also be difficult. Leakage of infected oropharyngeal secretions past the tracheal tube cuff is thought to be largely responsible for ventilator-associated pneumonia, and many ICUs now utilize endotracheal tubes with modern tube cuffs and subglottic aspiration ports to minimize the risk. Bacterial colonization of the oropharynx may be promoted by regurgitation of colonized gastric fluid, so the risk of ventilator-associated pneumonia can be reduced by nursing patients in the semi-recumbent position and by oropharyngeal decontamination. Treatment is with an appropriate broad-spectrum antibiotic, which can be modified if a causative organism is isolated.

Techniques for respiratory support

(See **Box 13.11**).

Controlled mechanical ventilation

CMV is used when respiratory efforts are absent or have been abolished. It involves one of two types:

- ***Volume-controlled ventilation.*** The tidal volume and respiratory rate are preset on the ventilator. The airway pressure varies according to both the ventilator setting and the patient's lung mechanics (airway resistance and compliance).
- ***Pressure-controlled ventilation.*** Both the inspiratory pressure and the respiratory rate are preset but the tidal volume varies according to the patient's lung mechanics.

Positive end-expiratory pressure

A positive airway pressure can be maintained at a chosen level throughout expiration using a threshold resistor valve in the

Box 13.13 Complications of tracheostomy

The complications for tracheal intubation shown in **Box 13.12** *plus:*

Early

- Death
- Pneumothorax
- Haemorrhage
- Hypoxia
- Hypotension
- Cardiac arrhythmias
- Tube misplaced in pretracheal subcutaneous tissues
- Subcutaneous emphysema

Intermediate

- Mucosal ulceration
- Erosion of tracheal cartilages (can cause tracheo-oesophageal fistula)
- Erosion of innominate artery (occasionally leads to fatal haemorrhage)
- Stomal infection
- Pneumonia

Late

- Failure of stoma to heal
- Tracheal granuloma
- Tracheal stenosis at level of stoma, cuff, or tube tip
- Collapse of tracheal rings at level of stoma
- Cosmetic factors

expiratory limb of the circuit. Positive end-expiratory pressure (PEEP) re-expands under-ventilated lung units and redistributes lung water from the alveoli to the perivascular interstitial space, thereby reducing shunt and increasing P_aO_2. The inevitable rise in mean intrathoracic pressure associated with the application of PEEP may, however, further impede venous return, increase pulmonary vascular resistance, and reduce cardiac output. The fall in cardiac output can be ameliorated by expanding the circulating volume, although inotropic or vasopressor support is required in some cases. Thus, although arterial oxygenation is often improved by the application of PEEP, a simultaneous fall in cardiac output can lead to a reduction in total oxygen delivery.

Low levels of PEEP (5–8 cmH_2O) are used in the majority of mechanically ventilated patients in order to maintain lung volume, as well as in those with basal atelectasis and in selected cases with airways obstruction.

Continuous positive airway pressure

The application of continuous positive airway pressure (CPAP) achieves for the spontaneously breathing patient what PEEP does for the ventilated patient. Oxygen and air are delivered under pressure via an endotracheal tube, a tracheostomy, a tightly fitting face mask, or a hood (Fig. 13.25). Not only can CPAP improve oxygenation, but also the lungs can become more compliant and the work of breathing is reduced.

An alternative is to use a very high flow of oxygen and air delivered via large-diameter nasal prongs to create a modest amount of CPAP. This high-flow nasal oxygen (HFNO) system is often better tolerated, as the patient is able to cough, expectorate, communicate, and eat. Commencing therapy with HFNO in patients with type I respiratory failure may be equivalent to, and in some circumstances more effective than, NIV in preventing patient deterioration and avoiding subsequent invasive ventilation.

Pressure support ventilation

In pressure support ventilation (PSV), spontaneous breaths are augmented by a preset level of positive pressure (usually between 5 and 20 cmH_2O), triggered by the patient's spontaneous respiratory effort and applied for a given fraction of inspiratory time or until inspiratory flow falls below a certain level. Tidal volume is determined by the set pressure, the patient's effort, and pulmonary mechanics. The level of pressure support can be reduced progressively as the patient improves.

Fig. 13.25 Use of a continuous positive airway pressure (CPAP) hood.

Intermittent mandatory ventilation

The intermittent mandatory ventilation (IMV) technique allows the patient to breathe spontaneously between the 'mandatory' tidal volumes delivered by the ventilator. These mandatory breaths are timed to coincide with the patient's own inspiratory effort (synchronized IMV, or SIMV). The spontaneous breaths are usually assisted with PSV.

'Lung-protective' ventilation

This is an ideal method of mechanical ventilation that avoids overdistension of alveoli, as well as repeated energetic opening and closing of distal airway and alveoli. Alveolar volume is maintained with PEEP, and sometimes by prolonging the inspiratory phase, while tidal volumes are limited to 4 to 8 mL/kg ideal body weight (ideally, 6 mL/kg). Plateau pressures (P_{Plat}) should not exceed 30 cmH_2O. Respiratory rate can be increased to improve CO_2 removal and avoid severe acidosis (pH >7.2), but permissive hypercarbia is common and acceptable. Ventilation with low tidal volumes has been shown to improve outcome in patients with ARDS (see later). Lung-protective ventilation should be used in almost all patients undergoing mechanical ventilation.

High-frequency oscillation

High-frequency oscillation (HFO) is administered using a purpose-designed ventilator. With HFO, there is no bulk flow of gas; rather, gas oscillates to and fro at rates of 4 to 15 Hz in a CPAP circuit set at a high mean airway pressure (often 20–30 cmH_2O). The mechanism of gas exchange is not fully understood, but lung volume is well maintained and oxygenation may be improved. While it is still used in paediatric ICUs, recent evidence in adult patients has not shown any mortality benefit, and one trial reported increased mortality, and consequently HFO is no longer advocated.

Extracorporeal gas exchange

In patients with severe refractory respiratory failure, pumped, high-flow, veno-venous blood flow through an artificial lung is used in specialized centres to achieve adequate oxygenation and CO_2 removal (ECMO; Fig. 13.26). ECMO provides respiratory support while the underlying lung disease improves and allows avoidance of excessive injury by mechanical ventilation. Extracorporeal carbon dioxide removal ($ECCO_2R$) is a simpler device with lower blood flows to allow CO_2 removal (and hence lower tidal volumes) with only marginal oxygenation, but recent trials have not shown outcome benefits.

Adjuncts and newer ventilation modes

Lung recruitment manoeuvres are transient, sustained increased airway pressure to open up collapsed airless alveoli. They can improve oxygenation in patients with refractory hypoxaemia but can be associated with significant morbidity and need to be carried out by clinical experts. Newer modes of ventilation such as airway pressure release ventilation (APRV), an inverse ratio, and pressure-controlled, intermittent mandatory mode of ventilation with unrestricted spontaneous breathing have shown promise in clinical trials to improve oxygenation.

Non-invasive ventilation

Non-invasive ventilation (NIV) is suitable for patients who are conscious, cooperative, and able to protect their airway; they must also

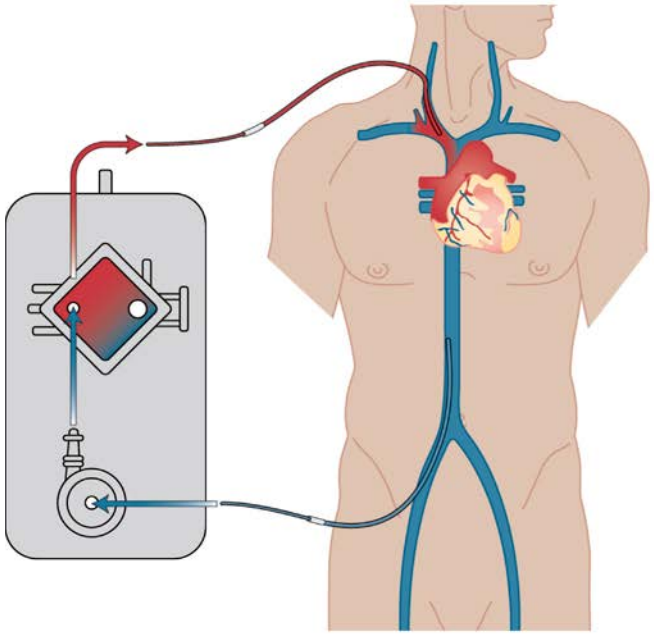


Fig. 13.26 Veno-venous extracorporeal membrane oxygenation (ECMO). A pump pulls blood off the patient through a large cannula with tip in the inferior vena cava. The blood then passes through a membrane oxygenator and is returned to the superior vena cava fully oxygenated and with CO_2 removed.

be able to expectorate effectively. Positive pressure is applied to the airways using a tight-fitting full-face/nasal mask or a hood. Modern ventilators are cheap, easy to use, and flexible. They deliver bi-level positive airway pressure (BiPAP) with inspiratory (IPAP) and expiratory (EPAP) pressure levels, and times are set independently. Unrestricted spontaneous respiration is possible throughout the respiratory cycle. BiPAP can also be patient triggered. There is a reduced risk of ventilator-associated pneumonia and improved patient comfort, with preservation of airway defence mechanisms, speech, and swallowing (which allows better nutrition). Spontaneous coughing and expectoration are not hampered, permitting effective physiotherapy, and sedation is usually unnecessary. Institution of non-invasive respiratory support can reduce respiratory muscle fatigue, reduce respiratory acidosis and breathlessness, improve clearance of secretions, and re-expand collapsed lung segments. The intubation rate, length of ICU and hospital stay, and, in some categories of patient, mortality may all be reduced. NIV is particularly useful in acute hypercapnic respiratory failure associated with COPD, provided the patient is not profoundly hypoxic or obtunded. NIV is also valuable in immunocompromised patients with acute respiratory failure and following extubation of hypercapnic patients with respiratory disorders to reduce the risk of subsequent respiratory failure and mortality. **Box 13.14** lists some indications and contraindications for the use of NIV when standard medical treatment has failed.

Box 13.14 Indications for/contraindications to use of non-invasive ventilation

Some indications

- Acute exacerbation of chronic obstructive pulmonary disease (H^+ >44 nmol/L; pH <7.35)
- Cardiogenic pulmonary oedema
- Chest wall deformity/neuromuscular disease (hypercapnic respiratory failure)
- Obstructive sleep apnoea
- Severe pneumonia (with caution)
- Asthma (occasionally)
- Weaning patients from invasive ventilation

Some contraindications

- Facial or upper airway surgery
- Reduced conscious level
- Inability to protect the airway

Weaning

Weakness and wasting of respiratory muscles are inevitable consequences of the catabolic response to critical illness and the disuse atrophy that will occur as a result of mechanical ventilation. Often, abnormalities of gas exchange and lung mechanics persist. Not surprisingly, therefore, many patients experience difficulty in resuming unsupported spontaneous ventilation.

Neuromuscular weakness complicating critical illness

Polyneuromyopathies have most often been described in association with persistent sepsis and multiple organ failure.

Critical illness polyneuropathy is characterized by a primary axonal neuropathy involving both motor and, to a lesser extent, sensory nerves. Clinically, the initial manifestation is often difficulty in weaning the patient from respiratory support. There is muscle wasting, the limbs are weak and flaccid, and deep tendon reflexes are reduced or absent. Cranial nerves are relatively spared. Nerve conduction studies confirm axonal damage. The CSF protein concentration is normal or minimally elevated. These findings differentiate critical illness neuropathy from Guillain–Barré syndrome, in which nerve conduction studies nearly always provide evidence of demyelination and CSF protein is usually high.

The cause of critical illness polyneuropathy is not known and there is no specific treatment. Weaning from respiratory support and rehabilitation are likely to be prolonged. With resolution of the underlying critical illness, recovery can be expected after 1 to 6 months but weakness and fatigue frequently persist.

Critical illness myopathies can also occur, often in association with a neuropathy. A severe quadriplegic myopathy has been associated in particular with the administration of steroids and muscle relaxants to mechanically ventilated patients who have acute, severe asthma.

Criteria for weaning patients from mechanical ventilation

Clinical assessment is the best way of deciding whether a patient can be weaned from the ventilator. A key part of the assessment is deciding whether the precipitating need for mechanical ventilation is no longer present or has vastly improved, to allow a trial without ventilation. The patient's conscious level and cardiovascular performance, as well as the effects of drugs, must all be taken into account. A subjective evaluation by an experienced clinician of the patient's response to a short period of spontaneous breathing (spontaneous breathing trial) is the most reliable predictor of weaning success or failure. Objective criteria are based on an assessment of pulmonary gas exchange (blood gas analysis), lung mechanics, and muscular strength.

Techniques for weaning

Patients who have received mechanical ventilation for less than 24 to 48 hours—after elective major surgery, for example—can usually resume spontaneous respiration immediately and no weaning process is required. This procedure can also be adopted for those who have been ventilated for longer periods but who tolerate a spontaneous breathing trial and clearly fulfill objective criteria for weaning.

Techniques of weaning include the following:

- The traditional method is to allow the patient to breathe entirely spontaneously for a short time, following which respiratory support is resumed. These periods of spontaneous breathing are gradually increased and the periods of respiratory support are progressively reduced. Initially, it is usually advisable to ventilate the patient throughout the night. This method can be stressful and tiring for both patients and staff, although it is sometimes successful when other methods have failed.
- SIMV (see earlier) involves a progressive reduction in the frequency of mandatory breaths. Spontaneous breaths are usually pressure supported.
- Gradual reduction of the level of pressure support is thought to be the preferred technique.
- CPAP (see above) can prevent alveolar collapse, hypoxaemia, and fall in compliance that might otherwise occur when patients start to breathe spontaneously. It is therefore used during weaning with SIMV or pressure support, during spontaneous breathing trials, and in spontaneously breathing patients prior to extubation.
- Tracheostomy can be used to facilitate weaning from mechanical ventilation and is better tolerated by patients than endotracheal tubes allowing sedation weaning and comfort.
- Non-invasive respiratory support (BiPAP, CPAP) can be used following extubation to prevent respiratory failure and re-intubation.

Extubation and tracheostomy decannulation

These should not be performed until patients can cough, swallow, and protect their own airway and are sufficiently alert to be co-operative. Together with spontaneous breathing trials to show adequate respiratory function, these parameters are used to make the decision to extubate patients. Similar considerations guide the elective removal of tracheostomy tubes.

The 'rapid shallow breathing index'—the ratio of spontaneous respiratory rate (in breaths/minute) to tidal volume (in litres)—of less than 100 is another relatively good predictor of extubation success and represents a patient who breathes slowly and comfortably with adequate tidal volumes.

Further reading

Combes A, Hajage D, Capellier G et al. Extracorporeal membrane oxygenation for severe acute respiratory distress syndrome. *N Engl J Med* 2018; 378:1965–1975.
Davidson AC, Banham S, Elliott M et al. BTS/ICS guideline for the ventilatory management of acute hypercapnic respiratory failure in adults. *Thorax* 2016; 71(suppl. 2):1–35.
Huang DT, Yealy, Filbin MR et al. Procalcitonin-guided use of antibiotics for lower respiratory tract infection. *N Engl J Med* 2018; 379:236–249.
Jonkman AH, Ranieri VM, Brochard L. Lung recruitment. *Intensive Care Med* 2022; 48:936–938.
Malhotra A, Drazen JM. High-frequency oscillatory ventilation on shaky ground. *N Engl J Med* 2013; 368:863–865.
McNamee J, Gilles M, Barrat N, et al. Effect of lower tidal volume ventilation facilitated by extracorporeal carbon dioxide removal vs standard care ventilation on 90-day mortality in patients with acute hypoxemic respiratory failure. The REST randomized clinical trial. *JAMA* 2021; 326(11):1013–1023.
Olsen HT, Nedergaard HK, Strøm T et al. Nonsedation or light sedation in critically ill, mechanically ventilated patients. *N Engl J Med* 2020; 382: 103–1111.
Schmidt GA, Girard TD, Kress JP, et al. Official executive summary of an American Thoracic Society/American College of Chest Physicians clinical practice guideline: liberation from mechanical ventilation in critically ill adults. *Am J Respir Crit Care Med* 2017; 195:115–119.
Semler MW, Casey JD, Lloyd BD. Oxygen-saturation targets for critically ill adults receiving mechanical ventilation. *N Engl J Med* 2022; 387:1759–1769.
Stollings JL, Balas MC, Chanques G. Evolution of sedation management in the intensive care unit (ICU). *Intensive Care Med* 2022; 48:1625–1628.
Swindin J, Sampson C, Howatson A. Airway pressure release ventilation. *BJA Ed* 2020; 20(3):80–88.

ACUTE RESPIRATORY DISTRESS SYNDROME

Definition and aetiology

The acute respiratory distress syndrome (ARDS, **Box 13.15**) can be defined as follows:

- Respiratory distress with hypoxaemia
- Chest X-ray, CT, or lung ultrasound findings of new bilateral, diffuse, patchy, or homogeneous pulmonary infiltrates
- Pulmonary oedema not due to cardiogenic oedema or fluid overload
- Gas exchange abnormalities, all with a PEEP ≥ 5 cmH_2O **(Berlin definition):**
  1. Mild: P_aO_2/F_iO_2 ratio 300–200 mmHg (40–26.6 kPa)
  2. Moderate: P_aO_2/F_iO_2 200–100 mmHg (26.6–13.3 kPa)
  3. Severe: P_aO_2/F_iO_2 ratio <100 mmHg (<13.3 kPa).

Stiff lungs with reduced pulmonary compliance resulting in high inflation pressures are often a feature of ARDS but not a defining element. ARDS can occur as a non-specific reaction of the lungs to a wide variety of direct pulmonary and indirect non-pulmonary insults. Approximately 10% of admissions to critical care fit the Berlin definition for ARDS, and the classifications above predict mortality which ranges from 27% in mild, 32% in moderate, and 45% in severe ARDS. By far, the most common predisposing factor is

Box 13.15 Disorders associated with acute respiratory distress syndrome

Direct lung injury	Indirect lung injury
Common causes	
Pneumonia COVID-19 Aspiration of gastric contents	Sepsis Severe trauma with shock and multiple transfusions
Less common causes	
Pulmonary contusion Blast injury Fat embolism Near-drowning Inhalational injury (smoke or corrosive gases) Reperfusion lung injury after lung transplantation or pulmonary embolectomy Amniotic fluid embolism	Cardiopulmonary bypass Drug overdose (heroin, barbiturates) Acute pancreatitis Transfusion-associated lung injury (TRALI) Eclampsia High altitude Severe burns Pulmonary vasculitis

sepsis, and 20% to 40% of patients with severe sepsis will develop ARDS.

Pathogenesis and pathophysiology

ARDS can be viewed as an early manifestation of a generalized inflammatory response with endothelial dysfunction and therefore is frequently associated with the development of multiple organ dysfunction syndrome (MODS, see 'Sepsis and multiple organ failure [multiple organ dysfunction syndrome]' section).

Non-cardiogenic pulmonary oedema

This is the cardinal feature of ARDS and is the first and clinically most evident sign of a generalized increase in vascular permeability, caused by the microcirculatory changes and release of immune mediators described earlier, with activated neutrophils playing a key role. The pulmonary epithelium is also damaged in the early stages, reducing surfactant production and predisposing to alveolar collapse.

Pulmonary hypertension

Pulmonary hypertension, sometimes complicated by right ventricular failure (see Chapter 17), is a common feature of ARDS. Initially, mechanical obstruction of the pulmonary circulation may occur as a result of vascular compression by interstitial oedema, while local activation of the coagulation cascade leads to thrombosis and obstruction in the pulmonary microvasculature. Later, pulmonary vasoconstriction may develop in response to increased autonomic nervous activity and circulating substances such as catecholamines and thromboxane. Those vessels supplying alveoli with low oxygen tensions constrict (the 'pulmonary hypoxic vasoconstrictor response'), diverting pulmonary blood flow to better-oxygenated areas of lung and thus limiting the degree of shunt.

Haemorrhagic intra-alveolar exudate

This exudate is rich in platelets, fibrin, fibrinogen, and clotting factors and may inactivate surfactant and stimulate inflammation, as well as promote hyaline membrane formation and the migration of fibroblasts into the air spaces.

Resolution, fibrosis, and repair

Within days of the onset of lung injury, formation of a new epithelial lining is underway and activated fibroblasts accumulate in the interstitial spaces. However, in some cases, there is progressive interstitial fibrosis. In those who recover the lungs are substantially remodelled.

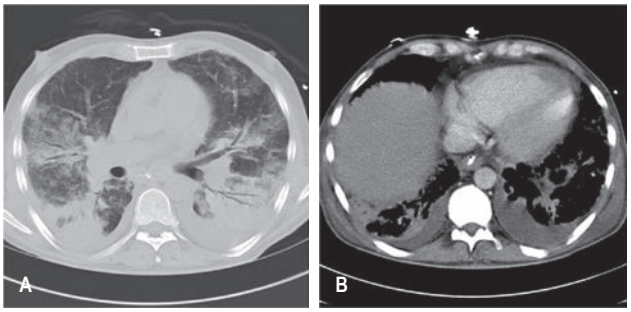


Fig. 13.27 Acute respiratory distress syndrome. **(A)** Computed tomography scan of the lung showing ground-glass opacification in non-dependent regions, with atelectasis and consolidation in dependent regions. There are small pleural effusions. **(B)** The same patient as shown in (A) using soft tissue window settings to demonstrate small bilateral effusions layering in the dependent region of both haemithoraces. *(From Hinds CJ, Watson JD. Intensive Care: A Concise Textbook, 3rd edn. Edinburgh: Saunders; 2008, with permission. Courtesy of Dr SPG Padley.)*

Physiological changes

Shunt and dead space increase, compliance falls, and there is evidence of airflow limitation. Although the lungs in ARDS are diffusely injured, the pulmonary lesions, when identified as densities on a CT scan, are predominantly located in dependent regions (Fig. 13.27). This is partly explained by the effects of gravity on the distribution of extravascular lung water and areas of lung collapse. Pleural effusions are common.

Clinical features

The first sign of the development of ARDS is often an unexplained tachypnoea, followed by increasing hypoxaemia and breathlessness. Fine crackles are heard throughout both lung fields. Later, the chest X-ray shows bilateral diffuse shadowing, interstitial at first, but subsequently with an alveolar pattern and air bronchograms (Fig. 13.28). The differential diagnosis includes cardiac failure and lung fibrosis.

Management

This is based on treatment of the underlying condition (e.g., eradication of sepsis), supportive measures, and avoidance of complications such as ventilator-induced lung injury and ventilator-associated pneumonia.

Mechanical ventilation

This is the cornerstone of management; strategies designed to minimize ventilator-induced lung injury and encourage lung healing should be used (see earlier).

Pulmonary oedema limitation

While most patients require an initial period of resuscitation, it is important to limit pulmonary oedema formation. Fluid restriction strategies and the use of diuretics may be required to prevent fluid overload with the aim of achieving a consistently neutral to negative fluid balance.

Prone position

When the patient is changed from the supine to the prone position, collapsed alveoli in the dependent region are redistributed and

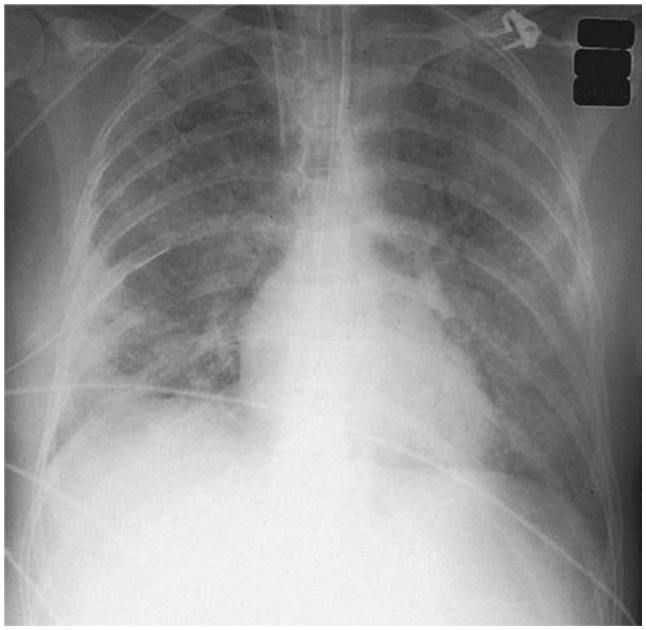


Fig. 13.28 Chest X-ray appearances in acute respiratory distress syndrome. Bilateral diffuse alveolar shadowing with air bronchograms and no cardiac enlargement.

shunt fraction is reduced. More uniform alveolar ventilation, caudal movement of the diaphragm, redistribution of perfusion, and recruitment of collapsed alveoli all contribute to the improvement in gas exchange. Recent findings suggest that proning early in the course of the disease process and spending longer periods prone than supine are associated with substantial mortality benefits.

Inhaled nitric oxide

This vasodilator, when inhaled, may improve ventilation-perfusion matching by increasing perfusion of ventilated lung units, as well as reducing pulmonary hypertension. It has been shown to improve oxygenation in so-called 'responders' with ARDS but has not been shown to enhance survival and may lead to higher rates of AKI.

Aerosolized prostacyclin

This appears to have similar effects to inhaled nitric oxide and is easier to deliver. As with inhaled nitric oxide, the response to aerosolized prostacyclin is variable: although it improves oxygenation, its effect on outcome is unclear.

Steroids

Administration of steroids to patients with persistent ARDS may improve lung function but does not appear to improve outcome. Their use remains controversial.

Prognosis

Mortality from ARDS has fallen over the last two decades, from around 60% to 20% to 40%, perhaps as a consequence of improved general care, lung-protective ventilation strategies, the increasing use of management protocols, and attention to infection control and nutrition. Prognosis is, however, still very dependent on aetiology. When ARDS occurs in association with intra-abdominal sepsis, mortality rates remain very high, whereas much lower mortality is to be expected in those with 'primary' ARDS (pneumonia, aspiration, lung contusion). Mortality rises with increasing age and failure of other organs. Most of those dying with ARDS do so as a result of MODS and haemodynamic instability rather than impaired gas exchange.

COVID-19

ARDS secondary to viral pneumonitis can be particularly challenging to treat. Critical care units around the world were required to grapple with this problem during the COVID-19 pandemic. Pneumonitis due to SARS-CoV-2 (see Chapter 20) proved slow to heal, with patients requiring aggressive ARDS management (paralysis, prone ventilation, or extra-corporeal support) for far longer than had generally been the case with 'classical' ARDS. These clinical features probably reflect differences in the inflammatory response to bacterial sepsis and SARS-CoV-2; initial inflammation is less severe in the latter but more prolonged, with accelerating immunosuppression leaving patients particularly vulnerable to secondary bacterial infection. Early multicentre RCT evidence did demonstrate a mortality benefit for corticosteroid treatment in these patients. The greatest impact of COVID-19 on critical care, however, was the organizational response required to care for far greater than usual patient numbers, as well as the psychological effect of a protracted surge on staff in the specialty.

Further reading

ARDS Definition Task Force. Acute respiratory distress syndrome: the Berlin definition. *JAMA* 2012; 307:2526–2533.
Bellani G, Laffey J, Pham T et al. Epidemiology, patterns of care and mortality for patients with acute respiratory distress syndrome in intensive care units in 50 countries. *JAMA* 2016; 315:788–800.
Faculty of Intensive Care Medicine/Intensive Care Society. *Guidelines on the Management of Acute Respiratory Distress Syndrome, version 1* 2018; https://www.ficm.ac.uk/sites/default/files/ficm_ics_ards_guideline_-_july_2018.pdf.
Grasselli G, Calfee CS, Camporota L, et al. ESICM guidelines on acute respiratory distress syndrome: definition, phenotyping and respiratory support strategies. *Intensive Care Med* 2023; 49:727–759.
Guérin C, Reignier J, Richard JC et al. Prone positioning in severe acute respiratory distress syndrome. *N Engl J Med* 2013; 368:2159–2168.
Hall JP, Kress JP. The burden of functional recovery from ARDS. *N Engl J Med* 2011; 364:1360–1366.
Hao D, Low S, Di Fenza R et al. Prone positioning of intubated patients with an elevated body-mass index. *N Engl J Med* 2022; 386:e34.
Herridge MS, Tansey CM, Matté A et al. Functional disability 5 years after acute respiratory distress syndrome. *N Engl J Med* 2011; 364:1293–1304.
Kress JP, Hall JB. ICU-acquired weakness and recovery from critical illness. *N Engl J Med* 2014; 370:1626–1635.
Loftus TJ, Ungaro R, Dirain M, et al. Overlapping but disparate inflammatory and immunosuppressive responses to SARS-CoV-2 and bacterial sepsis: An immunological time course analysis. *Front Immunol* 2021; 12:792448.
Madahar P, Beitler J. Emerging concepts in ventilation-induced lung injury. *F1000Res* 2020; 9:F1000 Faculty Rev-222.
Meyer M, Gattinoni L, Calfee C. Acute respiratory distress syndrome. *Lancet* 2021; 398:622–637.
The RECOVERY Collaborative Group. Dexamethasone in hospitalized patients with Covid-19. *NEJM* 2021; 384:693–704.

ACUTE KIDNEY INJURY

Acute kidney injury (AKI) is a common and serious complication of critical illness and is strongly associated with increased morbidity and mortality. The presence and severity of AKI are now formally classified by relative increases in serum creatinine and/or a persistent reduction in urine output (see Chapter 11).

The importance of preventing renal injury with rapid and effective resuscitation, the avoidance of nephrotoxic drugs (especially non-steroidal anti-inflammatory drugs [NSAIDs]), and control of infection cannot be overemphasized. While shock and sepsis are the most common causes of AKI in the critically ill, causes requiring specific treatment should be excluded, especially urinary tract obstruction

and acute intrinsic renal disease such as rapidly progressive glomerulonephritis (see Chapter 11).

Acutely, oliguria often accompanies shock and renal hypoperfusion and should prompt attempts to optimize cardiovascular function by expanding the circulating volume and restoring blood pressure; restoration of urine output is a good indicator of successful resuscitation. However, when AKI is established, urine output is often normal or high due to impaired concentrating capacity, so the development of oliguria then implies almost complete loss of renal function; in this case, continued fluid resuscitation will be ineffective and potentially harmful.

There is no evidence to support specific pharmacological interventions to prevent or treat AKI; in particular, low-dose dopamine is ineffective for preventing or reversing renal impairment in sepsis. Thus, treatment of those developing or at risk of AKI should focus on prompt resuscitation with fluid and/or vasopressor therapy, as well as treatment of the underlying causes of shock and sepsis. If these measures fail to reverse oliguria, loop diuretics such as furosemide may be used to treat or prevent fluid overload. However, there is no evidence that diuretics alter the clinical course of AKI, and renal replacement therapy (RRT) should not be delayed if indicated.

In AKI, RRT is indicated for refractory fluid overload, major electrolyte disturbances (especially hyperkalaemia), severe acidosis, and, less often, uraemia. There is no evidence to support commencement of RRT for less severe AKI in the absence of these indications. Continuous RRT (e.g., haemofiltration) is associated with greater haemodynamic stability and better control of fluid balance than intermittent haemodialysis, while the use of peritoneal dialysis is unsatisfactory in critically ill patients (although this may be an option in resource-limited environments). Thus, continuous RRT is now the preferred method of renal support in the critically ill.

In critical illness complicated by AKI, if the underlying problems resolve, renal function usually recovers over a period of a few days to several weeks. However, there is good evidence to suggest that, after recovery, subtle defects in renal function persist, so patients requiring RRT in the ICU have a significantly increased risk of developing chronic kidney disease in the months and years after critical illness and should ideally be followed up.

Further reading

Pickkers P, Darmon M, Hoste E, et al. Acute kidney injury in the critically ill: an updated review on pathophysiology and management. *Intensive Care Med* 2021; 47:835–850.

RENAL Replacement Therapy Study Investigators. Intensity of continuous renal-replacement therapy in critically ill patients. *N Engl J Med* 2009; 361:1627–1638.

STARRT-AKI Investigators. Timing of initiation of renal-replacement therapy in acute kidney injury. Timing of initiation of renal-replacement therapy in acute kidney injury. *N Engl J Med* 2020; 383:240–251.

NEUROCRITICAL CARE

Care of patients with a critically injured CNS can be provided in a general ICU but is also sometimes delivered in dedicated neurocritical care units. Examples of lesions found in such patients include traumatic brain injury, acute intracranial haemorrhage (usually subarachnoid bleeding from a berry aneurysm or a hypertensive intraparenchymal haemorrhage), severe ischaemic stroke, and spinal cord injury, as well as medical lesions such as Guillain–Barré syndrome.

Physiology

Neurones are particularly susceptible to acute ischaemic/hypoxic insults, as they have very limited capacity for anaerobic metabolism and are irreversibly lost when blood flow is restricted for as little as 3 to 8 minutes.

The intracranial contents—brain parenchyma, blood, and CSF—are contained in the rigid skull vault. ICP (pressure within the skull) is approximately 7 to 15 mmHg in a resting, supine adult. Changes in intra-abdominal and intrathoracic pressures, such as occurring during coughing or a Valsalva manoeuvre, may result in transient increases in ICP that are of no consequence. The ***Monro–Kellie doctrine*** states that because the intracranial volume is fixed, any increase in the volume of one of the intracranial constituents must be compensated by a decrease in the volume of the others. As space-occupying lesions expand or brain tissue swells, the rise in ICP is limited by displacement of CSF and venous blood from the intracranial compartment. Once these compensatory mechanisms are exhausted, further increases in intracranial volume cause exponentially greater increases in ICP (**Fig. 13.29**).

In health, ***autoregulation*** maintains constant cerebral blood flow over a range of blood pressures. However, these mechanisms are lost following significant intracranial injury, and cerebral blood flow then becomes directly dependent on cerebral perfusion pressure (CPP). CPP is calculated as the MAP minus the ICP. Occasionally, a very high CVP, greater than the ICP, will also reduce CPP (**Fig. 13.30**).

Clinical features and investigations

The causes of coma are listed in 'The unconscious patient' section, Chapter 11. Conscious level should be documented regularly. The Glasgow Coma Scale (GCS) score (see also Chapter 11) is widely used as a semi-quantitative, yet crude, means of gauging the level of consciousness. Assessment of the airway is of particular importance. Patients with airway obstruction due to lax oropharyngeal musculature should be managed as discussed in Fig. 13.22. Patients without a cough or gag reflex may need to be intubated in order to protect the airway. Pupillary size and reactivity is a crucial part of the assessment. A unilaterally dilated pupil that reacts sluggishly or not at all is frequently a sign of increasing ICP. This is caused by compression of the third cranial nerve by herniating brain; if untreated, it can be rapidly fatal. Immediate measures to decrease ICP (see **Box 13.16**) are required while a CT scan of the brain is obtained. In some situations, patients may be taken directly to the operating theatre in order to alleviate the raised ICP and to

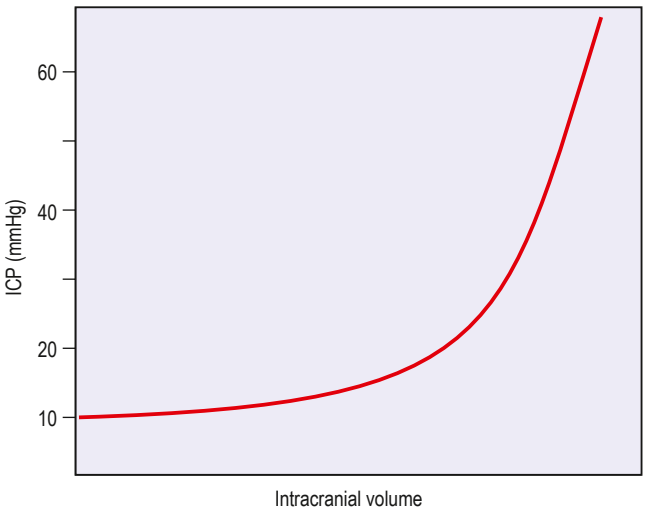


Fig. 13.29 Intracranial compliance curve. *ICP,* Intracranial pressure. *(From Hinds CJ, Watson JD.* Intensive Care: A Concise Handbook, *3rd edn. Edinburgh: Saunders; 2008, with permission.)*

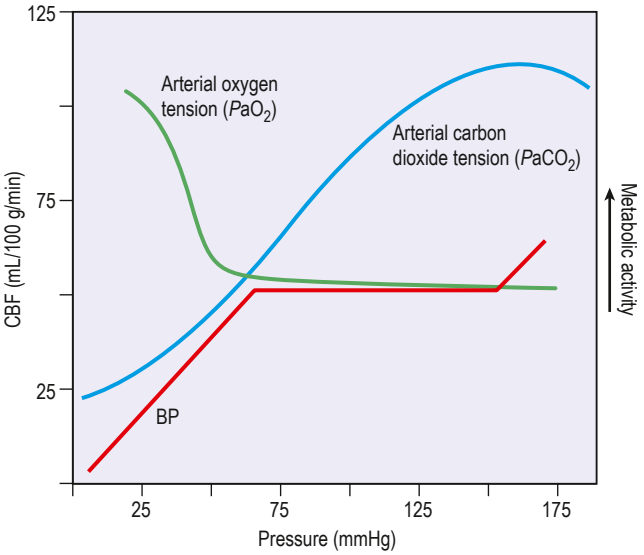


Fig. 13.30 Factors influencing cerebral blood flow (CBF). *BP,* Mean arterial blood pressure.

Box 13.16 Measures to control raised intracranial pressure and prevent secondary brain injury

- Control/restoration of blood pressure
- Control of intracranial pressure:
  - Reduction of intracranial blood volume
  - Mechanical ventilation
  - Sedation
  - Unobstructed venous drainage
  - Control of cerebral oedema
  - Osmotic diuretics (e.g., mannitol)
  - Loop diuretic (e.g., furosemide)
- Other measures:
  - Removal of cerebrospinal fluid via intraventricular catheter
  - Decompressive craniectomy
  - Excision of brain tissue
- Control of seizure activity

remove expanding space-occupying lesions such as a haematoma. Small or pinpoint pupils may be a sign of mid-brain injury or opioid overdose. New focal neurological signs should be sought, such as cranial nerve palsies, limb weakness, or hemiparesis. Seizures should be recognized and treated appropriately (see Chapter 11).

Monitoring

Regular clinical assessments by the bedside nurse are essential, including GCS score and pupillary size and reaction to light at least every hour. Assessment for focal neurological signs and evolving weakness (e.g., regular forced vital capacity measurement) may also be indicated. Neurological assessments are inevitably limited in patients requiring sedation and mechanical ventilation. In these individuals, invasive monitoring of ICP may be necessary.

Invasive ICP monitoring devices may be extradural, subdural, subarachnoid, intraparenchymal, or intraventricular (**Fig. 13.31**). The external ventricular drain (EVD) is the most difficult to insert, particularly when increased ICP causes the ventricles to collapse. EVDs are normally placed through the non-dominant (usually right) hemisphere into the lateral ventricle in the operating theatre. Their main advantage over the other monitoring devices is that they can also be used to treat raised ICP by draining CSF from the ventricular system. They are more prone to infection, however.

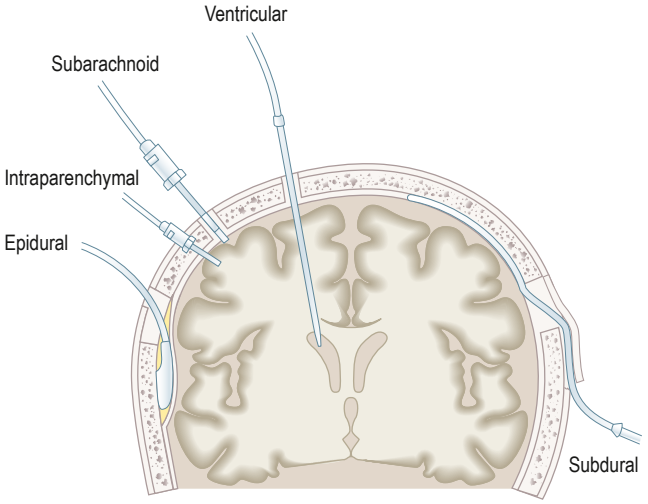


Fig. 13.31 Methods for measuring intracranial pressure.

Multi-electrode electroencephalography (EEG) monitors give both a waveform and an analogue display that can help in the diagnosis of non-convulsive seizures and are also useful in monitoring the depth of sedation.

Oximeter-tipped catheters can be placed via the internal jugular vein and advanced in a cranial direction to lie in the jugular bulb just external to the skull. These 'jugular bulb oximetry' catheters monitor venous saturation in blood leaving the brain and give an indication of the balance between cerebral oxygen supply and demand. Measurements can be interpreted similarly to the mixed venous oxygen content (see 'Mixed venous oxygen tension and saturation' section).

Cerebral oximetry uses near-infrared spectroscopy (NIRS) technology to estimate the oxygenation of a small portion of the cerebral cortex. Adhesive pads that both emit and capture near-infrared light waves are placed over the frontal cortex. Cerebral oximetry can provide an early warning of decreased cerebral oxygen delivery.

Management

Little can be done to reverse the immediate primary brain injury and irreversible loss of neurones that occur as a direct consequence of intracerebral haemorrhage, infarct, or traumatic brain injury. ICU care is focused on minimizing secondary brain injury (**Box 13.16**). This commonly occurs as a result of decreases in oxygen delivery secondary to hypoxaemia or reduced CPP due to systemic hypotension and/or raised ICP. Cerebral oedema, expanding mass lesions (e.g., haemorrhage), prolonged seizure activity, and hypercarbia (which causes cerebral vasodilation) are all examples of mechanisms leading to secondary injury. Neurocritical care has therefore focused on monitoring ICP and maintaining CPP. The Brain Trauma Foundation (BTF) guidelines state that an ICP of 22 mmHg or less and a CPP of between 60 and 70 mmHg should be the aim, with vasopressors and fluid given as required in conjunction with supportive care; the latter should include airway protection, maintenance of oxygenation and normocarbia, reduction of the risk of nosocomial infections, and prevention of thromboembolic complications.

Further reading

Hawryluk GWJ, Rubiano AM, Totten AM, O'Reilly C, et al. Guidelines for the management of severe traumatic brain injury: 2020 update of the decompressive craniectomy recommendations. *Neurosurgery* 2020; 87(3):427–434.
Rajesh S, Wonderling D, Bernstein I et al. Head injury: assessment and early management—summary of updated NICE guidance. *BMJ* 2023; 381.
http://www.braintrauma.org *Brain Trauma Foundation*.

OUTCOMES

Withholding and withdrawing treatment

(See also Chapter 50).

For many critically ill patients, intensive care is undoubtedly life-saving, and resumption of a normal lifestyle is to be expected. It is also widely accepted that the elective admission of high-risk patients into an ICU or HDU, particularly in the immediate post-operative period, can minimize morbidity and mortality and lower costs, as well as reduce the demands on medical and nursing personnel on general wards. In the most seriously ill patients, however, mortality rates are high, and a significant number die soon after discharge from the ICU. Mortality rates are particularly high in those who require re-admission to intensive care. Moreover, patients surviving a prolonged ICU admission often do not regain their premorbid functional status, and longer-term mortality rates (for at least 10 years post discharge) remain higher than in a general population matched for age and illness severity.

Inappropriate use of intensive care facilities has other implications. The patient may experience unnecessary suffering and loss of dignity, while relatives may also have to endure considerable emotional pressures. In some cases treatment may simply prolong the process of dying, and in others, the risks of interventions outweigh the potential benefits. Lastly, intensive care is expensive, particularly for those with the worst prognosis, and resources are limited.

To ensure both a humane approach to the management of critically ill patients and the appropriate use of limited resources, it is necessary to:

- avoid admitting patients who cannot benefit from intensive care; and
- limit further aggressive therapy when the prognosis is clearly hopeless.

Such decisions are extremely difficult; each case is assessed individually, taking into account previous health and quality of life, the patient's expressed wishes, the primary diagnosis, medium- and long-term prognosis of the underlying condition, and survivability of the acute illness. Age alone should not be the sole consideration. When in doubt, active measures should continue but with regular review in light of response to treatment and any other changes.

Decisions to limit therapy, not to resuscitate or to withdraw treatment should be made jointly by the ICU medical staff, the primary physician or surgeon, nursing staff, and, if possible, the patient, normally in consultation with the patient's family. Limitation of active treatment is *not* the cessation of medical or nursing care; rather, a caring approach must be adopted to ensure a dignified death, free of pain and distress, with support for family and friends.

Scoring systems

A variety of scoring systems have been developed to evaluate the severity of critical illness. Some include an assessment of the patient's previous state of health and the severity of the acute disturbance of physiological function (***acute physiology, age, chronic health evaluation***, or ***APACHE II***; the ***simplified acute physiology score***, or ***SAPS II and the SOFA score***). Other systems have been designed for particular categories of patient (e.g., the Injury Severity Score [ISS] for trauma victims).

The APACHE II (**Box 13.17**), SAPS II (**Box 13.18**), and SOFA (**Box 12.2**) scores are widely applicable and have been extensively validated. They can accurately quantify the severity of illness and predict the overall mortality for large groups of critically ill patients and are therefore useful for defining the 'case mix' of patients when auditing a unit's clinical activity, for comparing results nationally or internationally, and for characterizing groups of patients in clinical studies.

APACHE II is measured in the first 24 hours of admission. It has a maximum score of 71. A score of 25 represents a predicted mortality of 50% and a score of over 35 to 80%.

SAPS II is also measured in the first 24 hours of critical care admission and gives a mortality prediction.

SOFA score should also be undertaken in the first 24 hours of admission but can also be recalculated at every 24-hour period of admission. Although the APACHE II, SAPS II, and SOFA methodologies can also estimate mortality risk, no scoring system has yet been devised that can predict with certainty the outcome in an individual patient. They must not, therefore, be used in isolation as a basis for limiting or discontinuing treatment.

Death by neurological criteria

Traditionally, death is thought of as cessation of breathing and the heartbeat. Death by neurological criteria, sometimes referred to as brain death, means 'the irreversible loss of the capacity for consciousness combined with the irreversible loss of the capacity to breathe'. Both of these are essentially functions of the brainstem. With the advent of mechanical ventilation, it became possible to support such a dead patient temporarily, although, in all cases, cardiovascular failure eventually supervenes and progresses to asystole.

Before deciding on a diagnosis of death by neurological criteria, it is essential for certain preconditions and exclusions to be fulfilled. Two doctors need to agree on these preconditions—one consultant and one doctor who has been fully registered with the GMC for more than 5 years. These same two doctors will then perform the tests once the preconditions have been met.

Preconditions

- The patient must be in apnoeic coma (i.e., unresponsive and on a ventilator, with no spontaneous respiratory efforts).
- Irremediable structural brain damage due to a disorder that can cause brainstem death must have been diagnosed with certainty (e.g., head injury, intracranial haemorrhage).

Exclusions

- The possibility that unresponsive apnoea is the result of poisoning, sedative drugs, or neuromuscular blocking agents must be excluded.
- Hypothermia must be excluded as a cause of coma (core temperature >34°C).
- There must be no significant circulatory, metabolic, or endocrine disturbance that could produce or contribute to coma or cause it to persist (including acid–base, electrolyte, and glucose abnormalities).

Diagnostic tests for the confirmation of death by neurological criteria

All brainstem reflexes are absent to confirm death by neurological criteria.

The following tests should not be performed in the presence of seizures or abnormal postures.

- The pupils are fixed and unresponsive to bright light. Both direct and consensual light reflexes are absent. The size of the pupils is irrelevant, although most often, they will be dilated.

Box 13.17 The APACHE II scoring system

Physiologic variable	+4	+3	+2	+1	0	+1	+2	+3	+4
Rectal temperature °C	≥41	39–40.9		38.5–38.9	36–38.4	34–35.9	32–33.9	30–31.9	≤29.9
Mean arterial pressure mmHg	≥160	130–159	110–129		70–109		50–69		≤49
Heart rate	≥180	140–179	110–139		70–109		55–69	40–54	≤39
Respiratory rate	≥50	35–49		25–34	12–24	10–11	6–9		≤5
Oxygenation (mmHg) a. FIO_2 >0.5 record A-aDO_2 b. FIO_2 <0.5 record P_aO_2	500	350–499	200–349	<200	 >70	 61–70		 55–60	 <55
Acid–base balance a. Arterial pH b. Serum HCO_3^- (mEq/L)	 ≥7.7 ≥52	 7.6–7.69 41–51.9		 7.5–7.59 32–40.9	 7.33–7.49 22–31.9		 7.25–7.32 18–21.9	 7.15–7.24 15–17.9	 <7.15 <15
Sodium mM/L	≥180	160–179	155–159	150–154	130–149		120–129	111–119	≤110
Potassium mM/L	≥7	6–6.9		5–5.9	3.5–4.9	3–3.4	2.5–2.9		<2.5
Creatinine mg/dL (Double point score if acute renal failure)	≥3.5	2–3.4	1.5–1.9		0.6–1.4		<0.6		
Hct %	≥60		50–59.9	46–49.9	30–45.9		20–29.9		<20
WBC (in 1000's)	≥40		20–39.9	15–19.9	3–14.9		1–2.9		<1

GCS	Score = 15 minus actual GCS
Age	**Points**
<44	0
45–54	2
55–64	3
65–74	5
>75	6
Chronic health condition	**Points**
None	0
Yes, and admitted after elective surgery	2
Yes, and admitted after emergency surgery or any reason other than elective surgery	5

A-aDO_2, Alveolar-arterial oxygen gradient, calculated as A-aDO_2 = (FIO_2 × 710) − ($PaCO_2$ × 1.25) − PaO_2

- Corneal reflexes are absent.
- There are no vestibulo-ocular reflexes on caloric testing.
- There is no motor response within the cranial nerve territory to painful stimuli applied centrally or peripherally. Spinal reflex movements may be present.
- There is no gag or cough reflex in response to pharyngeal, laryngeal, or tracheal stimulation.
- Spontaneous respiration is absent. The patient should be ventilated with 100% O_2 for 10 minutes and then temporarily disconnected from the ventilator for 5 minutes. Oxygenation is maintained by insufflation with 100% oxygen via the endotracheal tube. The patient is observed for any signs of spontaneous respiratory efforts. A blood gas sample should be obtained during this period to ensure that the P_aCO_2 is sufficiently high to stimulate spontaneous respiration (starting P_aCO_2 ≥ 6.0 kPa and pH < 7.4, with P_aCO_2 rise of >0.5 kPa after 5 minutes).

The examination should be performed twice; the completion of the first set of tests is the legal time of death, and the second set of tests is confirmatory.

In the UK, it is not considered necessary to perform routine confirmatory tests, such as EEG or cerebral angiography; however, there are some 'red flags' listed on the faculty of intensive care guidance for brainstem testing that suggest performing CT angiography to confirm loss of blood flow is advisable. These include decompressive craniectomy, inability to perform all tests due to severe facial injuries, ongoing metabolic derangement that cannot be remedied, and situations where families are finding diagnosis difficult to accept and understand by neurological testing alone. The primary purpose of establishing a diagnosis of death by neurological criteria is to demonstrate beyond doubt that ongoing treatment is futile.

Organ donation

Organ donation may be possible in situations where ongoing life-sustaining treatment is deemed futile and a decision is reached in conjunction with the family (and, in rare circumstances, the patient) to withdraw treatments such as invasive ventilation and inotropes that are simply delaying an inevitable death. Organ donation can occur after death by neurological criteria—known as donation after brain death (DBD) or after withdrawal of life-sustaining therapies such as ventilation and inotropes and subsequent cardiac death, known as donation after cardiac death (DCD). In the circumstance of DCD, the withdrawal of life-sustaining therapies usually occurs in a planned fashion and

Box 13.18 The SAPS II score

SAPS II score	Mortality
29	10%
40	25%
52	50%
64	75%
77	90%

SAPS II criteria	0 points	Abnormal points				
Age	<40	50–59 7 points	60–69 12 points	70–74 15 points	75–80 16 points	≥80 18 points
Heart rate	70–119	40–69 2 points	120–159 4 points	≥160 7 points	<40 11 points	
SBP mmHg	100–199	≥200 2 points	70–99 5 points	<70 13 points		
Temp °C	<39	≥39 3 points				
P_aO_2 mmHg/F_iO_2 (if mechanically ventilated)		≥200 6 points	100–199 9 points	<100 11 points		
Urine output L/day	≥1	0.5–0.9 4 points	<0.5 11 points			
Urea mmol/L	>10	10–29.9 6 points	≥30 10 points			
WBC /mm^3	1–19.9	≥20 3 points	<1 12 points			
K^+ mmol/L	3–4.9	<3 or >5 3 points				
Sodium mmol/L	125–144	≥145 1 point	<125 5 points			
Bicarbonate mmol/L	≥20	15–19 3 points	<15 6 points			
Bilirubin μmol/L	<68.4	68.4–102.5 4 points	≥102.6 9 points			
GCS	14–15	11–13 5 points	9–10 7 points	6–8 13 points	<6 26 points	
Chronic disease		Metastatic cancer 9 points	Haematological malignancy 10 points	AIDS 17 points		
Type of admission	Scheduled surgical	Medical 6 points	Unscheduled surgical 8 points			

often in the operating department suite. If death occurs relatively quickly, then organ retrieval may proceed immediately following cardiac death. In DBD mechanical ventilation continues during organ retrieval.

Both of these situations require rapport and trust to be established with the family. In the UK, there are separate teams of nurse specialists to help the families through these very difficult situations and coordinate organ retrieval. A local and regional transplant coordinator can help with the process, as well as provide information, training, and advice about organ donation. The medical teams involved in the ICU care of the patient and those involved in the organ retrieval process must remain completely independent. The role of the intensive care physician once a patient has been identified as a donor and consented for organ donation is to optimize the donor physiology in order that the donor's organs are in the best possible condition at the point of donation. This may require escalation of some therapies and includes correcting hypovolaemia, treating AVP-deficiency (AVP-D, formerly cranial diabetes insipidus), administration of steroids, and use of vasopressors to maintain organ perfusion. NHS Blood and Transplant has a donor optimization care bundle to guide physicians through the management of this phase.

The UK now operates an opt-out policy for the presumed consent to organ donation. This means that a person is presumed to consent to donation unless they have actively opted out of the organ donor register. The final consent still lies with the next of kin.

Follow-up and care after discharge

Critical care survivors often have ongoing health problems that can last for many years, and a full recovery to their premorbid state is not always possible. Many patients will require months of inpatient rehabilitation before being able to return home. Both the patients

and their families/carers can experience something called post-intensive care syndrome (PICS). PICS is a combination of physical, psychological, cognitive, and socioeconomic problems that can be seen in critical care survivors and their loved ones. Approximately 30% of critical care patients that survive to discharge exhibit some signs of PICS.

Physical symptoms may include chronic pain, weakness, fatigue, breathlessness, and a reduction in mobility and exercise tolerance. Psychological problems such as post-traumatic stress, depression, and anxiety are common. Cognition deficiencies can manifest as memory loss, concentration difficulties, and slow mental processing. There are often employment issues and a lot of patients do not return to work and family members take on the role of caregiver which can have huge financial implications for families.

In order to identify PICS and help these patients and their families, it is recommended that critical care survivors are offered a multidisciplinary assessment approximately 3 months after hospital discharge. This usually occurs in a critical care follow-up outpatient clinic. A detailed assessment of the patient's pre-morbid and current state is undertaken including screening for physical, psychological, cognitive, and socioeconomic problems that the patient may be experiencing since their illness. Following this, patients are often referred to other services which may include therapies and rehabilitation, psychological support in the form of counselling, peer support groups, one-to-one psychological support, and other medical or surgical specialties for ongoing management of physical issues.

Further reading

Academy of Royal Medical Colleges. *A Code of Practice for the Diagnosis and Confirmation of Death* London: Academy of Royal Medical Colleges. 2008.
Curtis JR, Vincent JL. Ethics and end of life care for adults in the intensive care unit. *Lancet* 2010; 376:1347–1353.
Desai N, Goss J. Scoring systems in the critically ill: uses, cautions, and future directions. *BJA Educ* 2019; 19(7):212–18.
Finfer S, Vincent JL. Dying with dignity in the intensive care unit. *N Engl J Med* 2014; 370:2506–2514.
General Medical Council's. *Treatment and Care Towards the End of Life: Good Practice in Decision Making* https://www.gmc-uk.org/ethical-guidance/ethical-guidance-for-doctors/treatment-and-care-towards-the-end-of-life.
Herridge, MS and Azoulay E. Outcomes after critical illness. *N Engl J Med* 2023; 388:913–924.
NICE Quality Standard 158—*Rehabilitation After Critical Illness in Adults* 2017; https://www.nice.org.uk/guidance/qs158.
Oram J, Murphy P. Diagnosis of death, *BJA Education* 2011; 11(3):77–81.

Bibliography

Bersten AD, Soni N. *Oh's Intensive Care Manual* 7th edn. Oxford: Butterworth–Heinemann. 2014.

Significant websites

http://www.esicm.org *European Society of Intensive Care Medicine*.
http://www.ficm.ac.uk *Faculty of Intensive Care Medicine*.
http://www.icnarc.org *Intensive Care National Audit & Research Centre*.
http://www.ics.ac.uk *UK Intensive Care Society*.
http://www.sccm.org/Research *Society of Critical Care Medicine guidelines*.
http://odt.nhs.uk NHS Blood and Transplant Organ Donation and Transplantation.

Questions for this chapter can be found on Elsevier eBooks+.

14 Anaesthesia and perioperative medicine

Sohail Bampoe, Peter M. Odor, Stuart Connal, and Elliott Ridgeon

Clinical skills 334
Background 334
Pathophysiology 337
Preoperative care 337
Emergency versus elective surgery 337
Consent 337
Shared decision-making 338
Risk assessment 338
Anaesthetic assessment and perioperative clinics 340
Specific common diseases in the perioperative patient 340
Optimization and prehabilitation 342
Fasting and nutrition 342
Intraoperative care 343
Types of anaesthesia 343
Anaesthetic drugs 343
Additional drugs 345
Emergency drugs 345
Intraoperative fluids 345
Airway management 345
The anaesthetic machine 346
Theatre safety and human factors 346
Postoperative care 346
Location of postoperative care 346
Important and common complications after surgery 347
Enhanced recovery after surgery and day-case surgery 349
Orthogeriatrics 349
Perioperative fluid balance 349
Special populations 350
Emergency laparotomy 350
Hip fracture 350
Cancer 350
Obstetrics 351
Trauma 351

CORE SKILLS AND KNOWLEDGE

Every year, more than 300 million operations are performed around the world. As global life expectancy increases, so too does the complexity of procedures that surgeons are able to perform. As a result, as time passes, patients presenting for surgery will be older, with multiple co-morbidities and having increasingly complex surgery. Therefore, the risk of poor outcomes after surgery is elevated and perioperative medicine has emerged as a multidisciplinary subspecialty focused on improving surgical outcomes. Anaesthetists form a central part of perioperative teams and often have oversight and influence over the entire surgical pathway, working closely with surgeons, physicians, nurses, and other allied health professionals.

Key skills to learn in anaesthesia and perioperative medicine include:

- Understanding the process of risk assessment, shared decision-making, and consent in the preparation of a patient for surgery
- Knowledge of the spectrum of anaesthetic agents and principles of intraoperative monitoring
- Recognizing common postoperative complications and understanding how to prevent and treat them.

Opportunities to learn skills in anaesthesia and perioperative medicine include attending surgical and anaesthetic pre-assessment clinics and multidisciplinary teams (MDTs), 'shadowing' an anaesthetist during an operating list, and joining specialist postoperative ward rounds (e.g., orthogeriatrics).

CLINICAL SKILLS

The preoperative anaesthetic assessment:

The preoperative assessment is a key step in patient preparation for surgery, providing opportunities to pre-empt or avoid predictable complications, evaluate and optimize chronic health conditions ahead of surgery, and provide a bespoke estimate of 'risk', to better inform and counsel both the patient and other health professionals involved in their care (**Box 14.1**).

The perioperative checklist:

All hospitals should have a localized perioperative checklist based on the World Health Organization Surgical Safety template to minimize risks due to human error and oversight and avoid 'never events' such as 'wrong site surgery' (**Fig. 14.1**).

BACKGROUND

Perioperative medicine is a medical subspeciality dedicated to comprehensive care of patients undergoing surgery, encompassing the assessment, optimization, and management of patients in the period spanning the decision to undergo surgery, through the surgical procedure itself and into the postoperative recovery phase (**Fig. 14.2**). This discipline is essential for ensuring the safety, well-being, and optimal outcomes of patients having surgery.

Perioperative medicine specialists include doctors trained in anaesthesia, surgery, or medicine working collaboratively alongside allied health professionals, such as psychologists, physiotherapists, and dietitians to provide patients contemplating surgery with integrated and personalized care throughout their surgical pathway.

Box 14.1 The preoperative anaesthetic assessment

Preoperative assessment		Implications/further actions
Anaesthetic history	Previous anaesthetics? Any problems? Any family problems with anaesthesia?	Explore events further. Review previous anaesthetic charts. Consider rare inherited problems with volatile anaesthetics and muscle relaxants.
	Fasting status	Allowed to drink clear fluids until 2 hours preop. Allowed food until 6 hours preop.
	History of acid reflux Controlled or symptomatic?	Implications for risk of pulmonary aspiration of gastric contents and airway management.
Medical history	Hypertension Ischaemic heart disease Heart failure Cerebrovascular disease Diabetes Asthma/COPD Snoring/obstructive sleep apnoea Obesity Liver disease Renal impairment Venous thromboembolism Anaemia	Contribute to risk assessment and optimization. Influence delivery and choice of anaesthetic care.
Drug history	**Allergies** Regular medications Any issues taking NSAIDs	Influence delivery and choice of anaesthetic care.
	Specific medicines—steroids, anticoagulation, insulin, SGLT-2 inhibitors	Perioperative management of specific medicines.
Social history	Alcohol intake	Influences delivery and choice of anaesthetic care.
	Smoking including vaping or cannabis use Other recreational drug use	Increased risk of laryngospasm, bronchospasm, postoperative pulmonary complications.
	Home environment	Planning for rehabilitation and postoperative care.
Functional assessment	Subjective assessments	Contribute to risk assessment and optimization.
	Objective tests, for example, CPET	
Physical examination	Airway assessment	Implications for airway management
	Cardiorespiratory examination	
Investigations	ECG Echocardiogram Blood tests Respiratory function tests CPET	Depends on age, ASA status, and surgical complexity. Guided by hospital and NICE guidelines (Routine preoperative tests for elective surgery).
Summary	ASA status Risk assessment tools Frailty assessment	Contribute to risk assessment and optimization, shared decision-making, and resource planning.
Communication with patient, families, or carers	Explain conduct of anaesthetic options Explain the expected pre/intra/postoperative journey Allow time for questions, concerns, or clarifications	Describe the advantages/benefits and disadvantages/risks of options for anaesthesia. Facilitate shared decision-making.
Multidisciplinary communication	Request additional investigations Refer for optimization of co-morbidities, nutrition, physical fitness, lifestyle interventions (prehabilitation) Arrange additional perioperative resources including critical care admission	Optimize progress through perioperative journey. Reduce or modify risk of perioperative complications or mortality. Reduce cancellations at short notice or on the day of surgery.

Over the past two decades, the landscape of medical care for surgical patients has undergone a marked transformation. Evidence that pre-existing medical conditions and development of complications following surgery significantly impact long-term outcomes has driven development of perioperative medicine teams, breaking traditional silos that separated surgical, anaesthetic, and medical care. For example, while preoperative assessments have traditionally focused on surgical fitness and procedural planning, the perioperative medicine approach places a stronger emphasis on *optimizing patients' overall health and individual needs before surgery*. This involves meticulous evaluation of medical co-morbidities, nutritional status, and psychosocial factors to identify and address potential risks, ensuring patients are in the best possible condition before surgery.

Advances in technologies including real-time data on vital signs, tissue oxygenation, and anaesthesia depth allow for immediate intervention and optimization of patient physiology. Adoption of minimally invasive surgical techniques has revolutionized many procedures, enabling smaller incisions, reduced tissue trauma, and quicker recovery times minimizing surgical stress on the patient. Postoperative management has evolved beyond immediate recovery to focus on long-term outcomes, with enhanced

Surgical Safety Checklist

World Health Organization | Patient Safety
A World Alliance for Safer Health Care

Before induction of anaesthesia (with at least nurse and anaesthetist)	Before skin incision (with nurse, anaesthetist and surgeon)	Before patient leaves operating room (with nurse, anaesthetist and surgeon)
Has the patient confirmed his/her identity, site, procedure, and consent? ☐ Yes	☐ **Confirm all team members have introduced themselves by name and role.**	**Nurse Verbally Confirms:** ☐ The name of the procedure
Is the site marked? ☐ Yes ☐ Not applicable	☐ **Confirm the patient's name, procedure, and where the incision will be made.**	☐ Completion of instrument, sponge and needle counts
Is the anaesthesia machine and medication check complete? ☐ Yes	**Has antibiotic prophylaxis been given within the last 60 minutes?** ☐ Yes ☐ Not applicable	☐ Specimen labelling (read specimen labels aloud, including patient name)
Is the pulse oximeter on the patient and functioning? ☐ Yes	**Anticipated Critical Events** **To Surgeon:** ☐ What are the critical or non-routine steps? ☐ How long will the case take? ☐ What is the anticipated blood loss?	☐ Whether there are any equipment problems to be addressed
Does the patient have a: **Known allergy?** ☐ No ☐ Yes **Difficult airway or aspiration risk?** ☐ No ☐ Yes, and equipment/assistance available **Risk of >500ml blood loss (7ml/kg in children)?** ☐ No ☐ Yes, and two IVs/central access and fluids planned	**To Anaesthetist:** ☐ Are there any patient-specific concerns? **To Nursing Team:** ☐ Has sterility (including indicator results) been confirmed? ☐ Are there equipment issues or any concerns? **Is essential imaging displayed?** ☐ Yes ☐ Not applicable	**To Surgeon, Anaesthetist and Nurse:** ☐ What are the key concerns for recovery and management of this patient?

This checklist is not intended to be comprehensive. Additions and modifications to fit local practice are encouraged.

Revised 1 / 2009

© WHO, 2009

Fig. 14.1 **World Health Organization (WHO) surgical safety checklist.** *(Reproduced with permission from www.who.int, Tool and Resources.)*

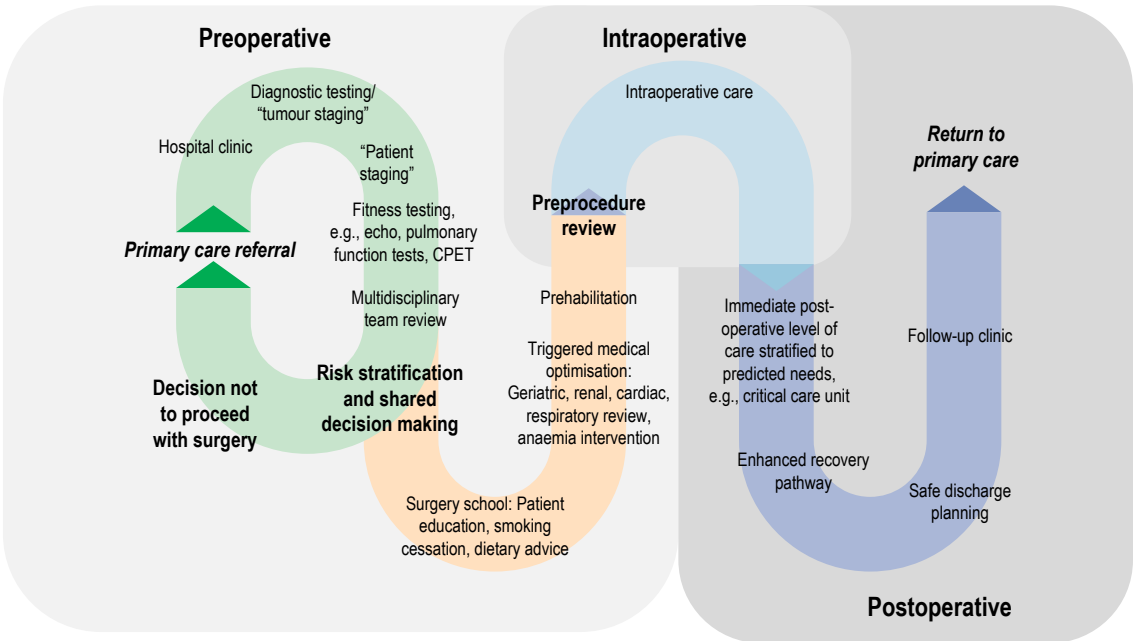


Fig. 14.2 **Modern perioperative medicine: the contemporary 'journey' of a patient undergoing elective surgery.**

pain management, early mobilization, and proactive rehabilitation integral components.

This more holistic and proactive approach to the surgical patient enables greater efficiency and targeted allocation of expensive resources, such as postoperative critical care best enhances patient safety, minimizes the impact of complications, and improves surgical outcomes.

PATHOPHYSIOLOGY

Major surgery constitutes a significant injury to the body. Surgery causes acute inflammation both local to and distant from the site of surgery. This is known as the 'surgical stress response' and is mediated by:

- ***Inflammatory cells***, including macrophages, monocytes and neutrophils
- ***Cytokines and other inflammatory mediators***, such as tumour necrosis factor-alpha (TNF-α) and various interleukins
- ***Activation of the endothelium*** with production of kinins, prostaglandins, and nitric oxide, leading to vasodilation and increased vessel permeability
- ***Release of stress hormones***, mediated primarily via:
  1. the sympathetic nervous system, through noradrenaline (norepinephrine) from sympathetic nerve endings and adrenaline (epinephrine) from the adrenal glands
  2. activation of the hypothalamic–pituitary axis with release of arginine vasopressin (AVP) and adrenocorticotrophic hormone (ACTH), leading to direct effects, as well as the subsequent release of other hormones such as glucagon and aldosterone.

This inflammation presents a significant challenge to organ systems in terms of metabolic supply and demand. **In particular, oxygen supply/demand mismatch may occur:** delivery of oxygen to metabolically active tissue can fail to meet requirements, which have dramatically increased, leading to disrupted cell function. This can affect any tissue or organ, but incision/wound sites, the heart, lungs, kidneys, and brain are often affected. For this reason, the perioperative physician focuses on assessing and optimizing a patient's ability to deliver oxygen to tissues when presented with a surgical stress. Cardiac, lung, and circulatory performance are the main determinants of oxygen delivery. **Patients presenting for surgery with existing organ dysfunction or co-morbid disease are therefore at significant risk of morbidity and mortality.**

The need for energy to support inflammation and tissue repair triggers a catabolic state in the surgical patient. This is characterized by:

- ***Sodium and water retention*** mediated by AVP and the renin–angiotensin–aldosterone pathway
- ***An increase in carbohydrate, fat, and protein catabolism***, mediated by catecholamines and cytokines, with increased glucose and fat turnover
- ***The breakdown of adipose tissue and catabolism of skeletal muscle proteins*** to provide amino acids for both gluconeogenesis and acute-phase proteins by the liver
- ***Hypercoagulability***, typically in the first 24 to 48 hours after surgery and mediated via endothelial injury and activation, which increases the risk of thromboembolism.

This may be on top of existing malnutrition and metabolic deficit, for example, in the elderly or patients with cancer.

Postoperative complications (such as sepsis) can increase inflammation, energy demand, and catabolism yet further.

Our understanding of the pathophysiology of surgical stress, oxygen delivery, and metabolic supply/demand is developing. Newer work includes focus on dysfunction at the level of the mitochondria and genetic drivers of an individual's response to surgery.

PREOPERATIVE CARE

Preoperative care of the surgical patient can be thought of as occurring in three phases:

First, the assessment and diagnosis of the presenting surgical disease, usually by a surgeon. This involves relevant examination (e.g., of the abdomen, or limb), imaging (e.g., CT scanning), and further investigation (e.g., blood tests).

Second, medical (i.e., non-operative) treatments, administered by a multidisciplinary team of surgeons, physicians, anaesthetists, and allied health professionals (e.g., physiotherapists). Measures can include analgesia if the patient is in pain, resuscitation, and organ support if the patient is significantly unwell and is hospitalized in an emergency, and chemotherapy or radiotherapy if appropriate for a patient's cancer.

Third, preparation for surgery, by surgeons, anaesthetists, and other perioperative teams. This requires careful assessment of the patient's pre-existing disease (whether chronic co-morbidity, or acute related to the active surgical pathology) and functional state. Optimization of a patient's physiology, timing of surgery, planning the surgery itself (e.g., open vs. laparoscopic, anatomical approach, availability of equipment and resources), and careful consent and shared decision-making are all fundamental.

These phases can be interrupted at any time to delay or even cancel surgery if the patient's condition changes, or if they withdraw their consent.

Emergency versus elective surgery

Surgery can be elective (i.e., planned in advance) or emergency. Certain conditions (e.g., rapidly advancing cancer) require surgery that is both elective and urgent.

The phases of preoperative care detailed above are similar for elective and emergency surgery but may be delivered in a different setting and in a different time frame (**Box 14.2**). Time to prepare for surgery may be limited if the patient needs an emergency procedure, and resuscitation may occur simultaneous to the operation itself. Optimization of existing disease may not be possible at all—hours before an emergency surgery, a patient's exercise capacity or lung function, for example, are unlikely to be modifiable. For these reasons, emergency surgery is understood to be higher risk in terms of morbidity and mortality.

Consent

All surgery requires informed consent from the person undergoing the operation and in the UK, this must be written. The principles of obtaining informed consent, including assessment of mental capacity, are discussed in Chapter 3. If a patient requires emergency surgery but is unable to provide consent (e.g., if they are unconscious as a result of trauma), clinicians and patient representatives (e.g., family) can consider options based on the best interests of that person.

Information provided during the consent process must be detailed and include risks, benefits, and alternatives. Risks must be those that are common, those that are significant or major (e.g.,

Box 14.2 Emergency versus elective surgery: differences in preoperative care

	Elective	Emergency
Example	Knee replacement	Laparotomy and Hartmann's procedure for perforated bowel
Diagnosis and investigation	Outpatient *Orthopaedic clinic, X-ray/CT/MRI scans, repeat appointments for reassessment*	Emergency department, inpatient *Urgent examination and CT scan, vital sign monitoring*
Non-surgical treatment	Usually outpatient, often multiple steps/stages *Weight loss, physiotherapy, analgesia*	Resuscitation, simultaneous bundles of care alongside operation *i.v. antibiotics, i.v. fluid resuscitation, vasopressors*
Preparation for surgery	Outpatient, over weeks *'Prehabilitation' programmes, nutrition and fitness, 'surgery school', blood tests, functional tests, referral to other clinics for management of existing chronic diseases*	Inpatient where possible, rapid *Resuscitation as above, blood tests and blood gas analysis, minimal delay of surgery for optimization of chronic diseases*

death or serious disability, even if rare), and those that are relevant to the particular patient (e.g., a risk that, while rare, might impact the person significantly because of their particular job). Individualized descriptions of risk are essential after the key legal case of Montgomery versus Lanarkshire Health Board (2015).

A description of the planned surgery must also be given, along with relevant additional steps that may become necessary as the operation progresses. A common example is consent for a laparoscopic procedure, with simultaneous consent for conversion to an open approach if required.

Best practice is to provide information much in advance of the proposed date of surgery, so that a patient can discuss it with trusted friends and family and come to a measured decision. Clearly, in the setting of emergency surgery, this may not be possible, but care should be taken not to rush or pressure a person, even if time to surgery is short.

Site of surgery is an important detail not to be missed during consent. Laterality (i.e., left- or right-sided surgery) or level (e.g., in spinal surgery) in particular should be noted. On the day of surgery, patients should have a mark (usually an arrow) drawn on their skin by a surgeon to indicate where on the body the operation will take place (Fig. 14.3). A patient needing a right nephrectomy should have an arrow on the right side of their abdomen, even if a midline incision is planned. Wrong-site surgery is catastrophic for the patient.

Multiperson checks, performed separately at multiple times during the journey from admission to operation, are essential. These confirm consent and ensure the correct surgery for the correct patient at the correct time.

A consent form is not a binding contract. A person can withdraw their consent for a surgery or procedure at any time.

Shared decision-making

Shared decision-making describes a collaborative process of actively involving people in their care, moving beyond informed consent to reach more patient-centred decisions. In the UK, *recently updated General Medical Council (GMC) and National Institute for Health and Care Excellence (NICE) guidance highlight the legal, ethical, and regulatory frameworks underpinning this approach to decision-making*. Every test or intervention has advantages and disadvantages, and how these are balanced will be different for each patient. Choosing Wisely, an international initiative aims to improve communication around decision-making and advocates for use of the BRAN model:

Fig. 14.3 The left leg has been marked with an upward arrow pointing to the left knee in this patient prepared for arthroscopic meniscectomy.

- **B**enefits—what are the intended benefits of the proposed investigation or treatment?
- **R**isks—what are the risks of each option, individualized to a person's specific circumstances?
- **A**lternatives—are there any alternatives to the proposed intervention? What would be the relative benefits and risks?
- The option to do **N**othing—what would happen if we chose not to proceed? What might the consequences involve over the short term and longer term?

Risk assessment

'Risk' is a concept that describes the likelihood of developing a specified outcome in either an individual or population. The term is often used and interpreted differently by clinicians and patients, making communication about risk related to surgery challenging. Risk is also subjective: what is considered 'low' risk to one person may not be for another.

Historically, short-term mortality has been considered an important outcome in both surgery and perioperative medicine, due to its significance and ease of measurement. However, short-term mortality is highly limited as a quality measure and of little relevance to patients wanting to understand how surgery will impact their quality of life and longer-term health. As a consequence, greater weight is now being given to *predicting postoperative morbidity and effects on long-term survival after surgery*.

From the perspective of a perioperative physician, accurate risk assessment is a key step in preparing and counselling patients for surgery and planning individualized perioperative care.

In general terms, perioperative risk can be assessed using several methods:

1. **Scoring systems and prediction models:**
   a. Population-based risk scores, for example, American Society of Anesthesiologists Physical Status score (ASA Score)
   b. Organ-specific risk scores, for example, Lees Cardiac Risk Index
   c. Individualized risk prediction models, for example, Surgical Outcome Risk Tool (SORT) and American College of Surgeons National Surgical Quality Improvement tool
2. **Functional capacity assessments:**
   a. Pulmonary function testing (PFT)
   b. Echocardiography
   c. Cardiopulmonary exercise testing (CPET)
3. **Serum biomarkers:**
   a. High-sensitivity Troponin-T (hsTnT)
   b. Brain natriuretic peptide (BNP)

Scoring systems and prediction models

Appraisal and categorization of underlying fitness is the most fundamental component of predicting outcomes following surgery. The American Society of Anesthesiologists Physical Status score (ASA Score, **Box 14.3**) is a simple and easily replicable system for achieving this, and despite limitations, correlates surprisingly well with outcomes in multiple surgical patient settings.

Urgency of surgery also impacts rates of postoperative morbidity with **emergency surgery being consistently higher risk for both anaesthetic and surgical-related complications**. Non-elective surgical cohorts differ markedly from populations undergoing planned procedures in key aspects: the presence of acute physiological and organ system derangements, differences in innate compensatory capacity, and the limited potential for preoperative modification of patient risk. The effects of variations in the delivery of healthcare in non-elective settings also play a role.

More sophisticated risk prediction tools involve inputting preoperative parameters about patients, including planned surgical procedure, age, co-morbidity status, and basic investigation results. Computer models used to generate risk estimates are based on comprehensive data from thousands of patients undergoing surgery. This data powers tools, such as the American College of Surgeons, National Surgical Quality Improvement Program, which include freely available online calculators that allow clinicians to estimate risk of various postoperative complications for individual patients.

All risk prediction tools have limitations, most importantly that they are technically only valid for the specific patient population whose data was used to design and validate them.

Functional capacity assessments

Functional capacity assessments help gauge a patient's ability to withstand the physiological stress of surgery. These can be subjective, including structured self-reports on exercise tolerance or graded ability to carry out a variety of daily activities, or objective assessments based on clinical investigations and measurements.

Common tests of cardiorespiratory function include echocardiography and pulmonary function tests. Such assessments are non-invasive, relatively cheap, and provide information about cardiorespiratory health, but are best considered diagnostic tools to determine the presence of abnormal function or disease, with opportunities for targeted optimization prior to surgery. Echocardiography, in particular, provides limited information about dynamic performance when the heart is under stress and can therefore be misleading when used to estimate perioperative risk.

In contrast, **cardiopulmonary exercise testing (CPET)** is an objective measure of a patient's aerobic capacity and cardiopulmonary function (**Fig. 14.4**). CPET quantifies variables such as oxygen consumption (VO_2), carbon dioxide production (VCO_2), and the anaerobic threshold during exercise. Patients with reduced aerobic

Box 14.3 The American Society of Anaesthesiologists Physical Status Score

ASA grade[a]	Assessment	Description
I	A normal healthy patient	Healthy, non-smoking, no or minimal alcohol use
II	A patient with mild systemic disease	Mild diseases only without substantive functional limitations. Examples include current cigarette smoking, social alcohol drinking, pregnancy, obesity (BMI 30–40), well-controlled diabetes or hypertension, mild lung disease
III	A patient with severe systemic disease	Substantive functional limitations; one or more moderate to severe diseases. Examples include poorly controlled diabetes or hypertension, COPD, morbid obesity (BMI ≥40), active hepatitis, alcohol dependence or abuse, moderate reduction of ejection fraction, end-stage renal failure undergoing regular dialysis, recent (but >3 months) myocardial infarction or stroke
IV	A patient with severe systemic disease that is a constant threat to life	Examples include recent (<3 months) myocardial infarction, stroke or coronary intervention (such as coronary angioplasty), ongoing cardiac ischaemia or severe valve dysfunction, severe reduction of ejection fraction, sepsis, advanced renal impairment not on dialysis
V	A moribund patient who is not expected to survive without the operation	Examples include ruptured abdominal/thoracic aortic aneurysm, massive trauma, intracranial bleed with mass effect, ischaemic bowel, multiple organ/system dysfunction
VI	A declared brain-dead patient whose organs are being removed for donor purposes	

[a]The addition of 'E' denotes emergency surgery (when delay in treatment would lead to a significant increase in the threat to life or body part).
BMI, Body mass index; *COPD*, chronic obstructive pulmonary disease.

Fig. 14.4 A patient undergoing cardiopulmonary exercise testing (CPET).

capacity are at higher risk for postoperative complications, including adverse cardiac events, respiratory failure, and longer hospital stays.

Identifying patients as high-risk for surgery helps guide decisions about surgical candidacy and procedural planning and may prompt further interventions, such as cardiac rehabilitation, pulmonary rehabilitation, or nutritional support. CPET data can also be used to facilitate informed consent discussions with patients by providing a quantitative and objective way of describing fitness.

Serum biomarkers

A relatively new and still controversial approach to help predict risk after surgery is by monitoring for serum biomarkers indicating early end-organ damage to vital organs, such as the heart (via serial troponin or BNP concentration changes) or kidneys. An increase in troponin levels indicates myocardial injury or strain, which is most frequently due to imbalance between myocardial oxygen demand and supply during the perioperative period (rather than coronary artery plaque rupture). Identifying troponin rises early and optimizing cardiovascular status to minimize further injury to the heart may improve patient outcomes.

Frailty

Frailty is a state of vulnerability leading to poor restoration of homeostasis after a stressor event. It is a result of cumulative functional decline of the body's organ systems with age, leading to reduced physiological reserve. Frailty assessment is a comprehensive evaluation that considers multiple domains, including physical function (e.g., grip strength, gait speed), nutritional status, cognitive function, and psychosocial factors. A detailed description of frailty, including tools for diagnosis and grading of severity, is provided in Chapter 49.

Frailty assessments are highly predictive of adverse perioperative outcomes, including postoperative delirium (POD), prolonged hospital stays, falls, disability, and mortality.

Anaesthetic assessment and perioperative clinics

The preoperative anaesthetic assessment is a comprehensive evaluation of the patient conducted by the anaesthetist before surgery to ensure the safety and well-being of a patient before they have their operation and incorporates anaesthetic, surgical, and medical history, focused clinical examination and outputs from relevant clinical investigations, functional assessments, and risk calculations (Box 14.1).

Preoperative fasting status should be determined. Fasting is required to minimize the risk of regurgitation and pulmonary aspiration of gastric contents under anaesthesia. In general, patients should be fasted for 6 hours for solid foods and 2 hours for clear liquids. Non-clear liquids, such as milk, should be treated as solids.

Airway assessment

The anaesthetic airway assessment ensures safe and effective management of a patient's airway during surgery. It involves a thorough examination of anatomical factors, such as neck mobility, mouth opening, and the presence of any abnormalities that may impede safe airway access. Assessing factors like Mallampati score (Fig. 14.5), thyromental distance, and the ability to protrude the mandible helps predict ease of intubation. This assessment guides the choice of airway management techniques and equipment, minimizing the risk of complications and ensuring adequate oxygenation and ventilation throughout the surgical procedure.

Specific common diseases in the perioperative patient

Any patient with any pre-existing disease can present for surgery at any time. The student or early career doctor learning about a condition should ask themselves: 'What if this patient needed an operation?'.

A number of conditions merit special attention because they are common and associated with morbidity and mortality across surgical populations.

Diabetes mellitus

Diabetes is common and predisposes to other conditions requiring surgery (e.g., abscesses, vascular pathology). Perioperative priorities are to optimize chronic glucose control where time allows or acute control if surgery is emergent. Any patient normally using insulin (whether type 1 or type 2 diabetic) should have their blood ketones monitored perioperatively—surgical stress is a trigger of ketoacidosis—and may require variable-rate insulin infusion. Preoperative fasting can cause problems for diabetic patients, and common practice is to schedule their surgery at the start of an operating list, to minimize fasting duration.

Targets: HbA1c < 69 mmol/L to proceed for elective surgery. Blood glucose 6 to 12 mmol/L throughout surgical journey. Minimize fasting time preoperatively, restore diet as quickly as possible postoperatively.

Drugs: Most can be continued as normal (including metformin), although any agents that predispose to hypoglycaemia should be stopped if the patient is not eating at the time of their normal dose. This includes sulphonylureas such as gliclazide. Insulin doses may need to be adjusted down or doses omitted if not eating (although basal insulins should continue in those with type 1 diabetes). Sodium/glucose cotransporter 2 (SGLT2) inhibitors (e.g., dapagliflozin) can cause euglycaemic ketoacidosis in patients undergoing starvation and surgical stress and should be omitted on the day before and day of surgery. Ketones should be checked regularly.

Risks and complications: Hypoglycaemia, hyperglycaemia. Increased risk of acute kidney injury (AKI). Predisposition to local (wound) and systemic infection postoperatively. Diabetic ketoacidosis (more likely in type 1 diabetes), hyperosmolar hyperglycaemic state, euglycaemic ketoacidosis (if taking SGLT-2 inhibitors).

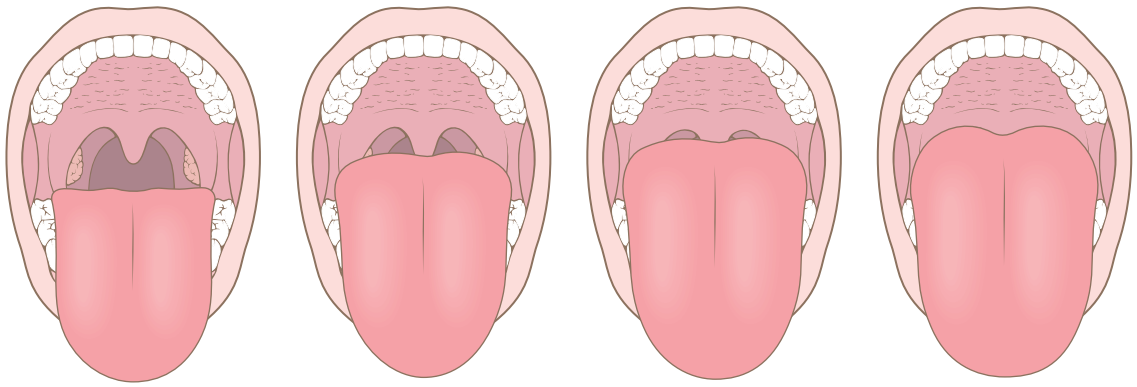


Fig. 14.5 The Mallampati score. Scores of ≥3 are associated with more difficult endotracheal intubation.

Key investigations: HbA1c, capillary blood glucose, blood ketones, renal function, fluid balance assessment.

Hypertension

High blood pressure (BP) affects approximately one in four adults. It is associated with other medical and surgical diseases, in particular vascular pathology. Cardiovascular risk related to hypertension is the major perioperative concern. Hypertension is a chronic condition, and management occurs over months to years, usually in the primary care setting. Acute perioperative BP treatment is therefore only advised in the setting of severe hypertension, that is, BPs high enough to present significant risk of end-organ damage (e.g., stroke). The most common cause of hypertension postoperatively is pain, including pain caused by urinary retention. Exclude triggers for high BP before treating the hypertension directly.

Targets: Aim BP < 160/100 mmHg in primary care, prior to referral for surgery. BP < 180/110 mmHg in hospital pre-assessment clinic or on day of elective surgery. BP reduction in the setting of emergency surgery is unlikely to be beneficial unless a hypertensive crisis is suspected. Maintaining intraoperative mean arterial pressure > 60 mmHg and systolic BP > 100 mmHg in a hypertensive patient may reduce risk—the usually hypertensive patient can be *relatively* hypotensive with a BP that would not normally be considered low.

Drugs: ACE inhibitors (e.g., lisinopril) and angiotensin receptor blockers (e.g., losartan) should be stopped 24 hours prior to surgery to reduce risk of AKI and refractory hypotension under anaesthesia and restarted within 48 hours of operation. Thiazides, calcium channel blockers, and beta-blockers should be continued. Beta-blockers should not be newly started preoperatively.

Risks and complications: Stroke, myocardial infarction, heart failure, increased risk of AKI, hypotension secondary to treatments.

Key investigations: Serial BP measurement, electrocardiogram (ECG), renal function

Anaemia

Anaemia (low haemoglobin) has a variety of causes and is always associated with increased perioperative risk. Low haemoglobin equates to a reduced capacity for oxygen carriage and delivery to tissues, which can lead to organ dysfunction, impaired healing, and other complications. The preoperatively anaemic patient has less capacity to accommodate a fall in haemoglobin caused by intra- or postoperative bleeding. Modern approaches aim to avoid excessive blood transfusion—the traditional multi-unit transfusion to target haemoglobin > 100 g/L is incorrect. Patients who refuse blood products (commonly on religious grounds) present a perioperative challenge, and additional advice should be sought.

Targets: Anaemia in the non-pregnant adult patient is defined by Hb < 130 g/L. Blood transfusion is indicated if Hb is below 70 g/L, or 80 g/L in patients with ischaemic heart disease. Post-transfusion targets are 70 to 90 g/L and 80 to 100 g/L, respectively.

Drugs: Oral iron—can take up to four weeks to be effective. i.v. iron can be used if surgery is less than four weeks away or oral iron fails. B12 and folate replacement based on blood tests. Oral therapies should be continued throughout the perioperative period. Erythropoietin is often given in patients with renal disease and should be continued. Tranexamic acid should be given intraoperatively during major surgery.

Risks and complications: Untreated anaemia increases risk of local (wound) and systemic infection, cardiovascular complications (stroke, myocardial infarction), and death. Anaemic patients will tolerate bleeding poorly.

Key investigations: Full blood count, iron studies (iron level, ferritin, transferrin saturation), B12 and folate levels, C-reactive protein, renal function, urinalysis, referral for endoscopy if GI bleed suspected as cause of anaemia.

Ischaemic heart disease

Ischaemic heart disease is defined by acute or chronic failure of oxygen delivery to meet myocardial demand. This is normally because of partial or complete occlusion of one or more coronary arteries. Surgery leads to increased myocardial work and can unmask latent disease or worsen existing disease. Minimizing surgical stress is therefore essential, and approaches to anaesthesia and surgery must be planned accordingly. Patients with ischaemic heart disease are at high risk of perioperative cardiovascular complications, principally myocardial infarction, dysrhythmia, and death. In addition, many patients (with or without existing ischaemic heart disease) experience **m**yocardial **i**njury after **n**on-cardiac **s**urgery (MINS), an ischaemic episode not meeting criteria for myocardial infarction. Investigations and treatments for MINS are the subject of ongoing debate. The Lee Revised Cardiac Risk Index is a scoring system that can be used to predict cardiac complications after non-cardiac surgery.

Targets: Reduce risk of perioperative cardiovascular event by avoiding sustained hypo/hypertension, uncontrolled pain, anaemia, tachycardia and using goal-directed haemodynamic therapy. Optimize associated risk factors preoperatively (BP, haemoglobin, cholesterol). Delay elective non-cardiac surgery for 6 months after elective percutaneous coronary intervention (PCI) and 12 months after an acute coronary syndrome event.

Drugs: After elective PCI, delay urgent non-cardiac surgery until a minimum of one month of dual antiplatelet therapy (DAPT) has been given. Continue aspirin throughout the perioperative period unless surgery is neurosurgical (including spine) or vitreal ophthalmology. Clopidogrel and ticagrelor should be stopped for 5 days prior to surgery; prasugrel should be stopped for 7 days. Direct oral anticoagulants (DOACs, e.g., dabigatran, rivaroxaban) should be stopped for at least 2 days (longer if a patient has impaired renal function). In patients at very high risk of thrombosis (e.g., mechanical heart valves), a 'bridging' plan may be commenced, using low molecular weight heparin (e.g., enoxaparin) to provide anticoagulation while a patient is not taking a direct oral anticoagulant. Secondary prevention drugs for patients with ischaemic heart disease (e.g., statins, beta blockers) should not be discontinued.

Risks and complications: Myocardial infarction, MINS, dysrhythmia, heart failure. Thrombosis due to stopping anticoagulants, or bleeding when continuing.

Key investigations: Based on patient and surgical risk. ECG, troponin if myocardial injury or ACS suspected, echocardiography, stress imaging (e.g., myocardial perfusion scintigraphy), BNP.

Chronic obstructive pulmonary disease (COPD)

Patients with chronic respiratory diseases such as COPD are at high risk of perioperative complications. They are more likely to develop postoperative pneumonia, particularly if surgery is prolonged, involves the abdomen (since abdominal pain due to surgery impairs deep breathing), or involves the thoracic cavity. COPD itself is almost always caused by smoking, and hence, patients are at risk of additional complications (myocardial infarction [MI], stroke) or surgical disease (cancers) related to this. Severe COPD makes intraoperative and postoperative ventilation more difficult and may be an indication to avoid general anaesthesia.

Targets: Smoking cessation is vital, and patients should be referred to relevant support services. Patients should also have up-to-date pneumococcal and influenza vaccinations. Although spirometry is used for diagnosis of COPD, preoperative pulmonary rehabilitation is better guided by symptom assessment and functional tests (see below).

Drugs: Inhalers should be continued and should be used on the morning of surgery as normal. Most patients will be using inhaled steroids and may be at risk of adrenal suppression—additional i.v. or oral steroids should be given perioperatively if so.

Risks and complications: Pneumonia, pneumothorax, prolonged and difficult ventilation, bronchospasm, respiratory failure, death.

Key investigations: Spirometry (forced expiratory volume in 1 second [FEV1], forced vital capacity [FVC]), functional assessment (e.g., Chronic Respiratory Questionnaire, 6-minute walk test, CPET), arterial blood gases in major surgery, chest X-ray.

Optimization and prehabilitation

Screening and risk assessment to identify patients at increased risk of perioperative complications presents an opportunity to optimize modifiable risk factors and plan care.

Preoperative optimization describes a holistic approach to preparing patients physically and psychologically for surgery. In addition to optimizing common co-morbidities described above, preoperative preparation includes identifying and addressing additional modifiable risk factors such as:

- Smoking and vaping
- Alcohol use
- Recreational drug use
- Nutritional state and obesity
- Physical fitness

'Prehabilitation' is an umbrella term describing multidisciplinary interventions to help patients prepare for surgery, reduce complications, and improve outcomes. The most common interventions include:

- **Psychological interventions:** signposting patient information resources, support groups, counselling, and other therapeutic techniques such as motivational interviewing and coaching.
- **Exercise and fitness interventions:** delivered virtually or in-person, most often led by physiotherapists, sports and exercise therapists, or personal trainers. Improving cardiorespiratory fitness through group or individual exercise sessions, particularly high-intensity interval training, may be associated with reduced perioperative complications.
- **Nutritional screening and interventions:** led by dietitians to optimize energy reserves and metabolic status before surgery, particularly during neoadjuvant cancer therapy.

High-quality evidence supporting the use of prehabilitation has yet to be established. Interventions and outcome measures in the available evidence are heterogenous, leading to low-certainty conclusions in recent meta-analyses.

Preoperative optimization and prehabilitation offer the opportunity for 'teachable moments' regarding a patient's long-term health. Improvements to chronic health problems and risk factors such as smoking or cardiorespiratory fitness may have long-term benefits for patients beyond the perioperative period.

Fasting and nutrition

Surgical pathology and systemic illness are associated with malnutrition. Patients have inadequate energy reserves in the context of increased energy demands and/or impaired fuel utilization associated with cancer, cancer treatments, infection, and inflammation. Nutritional deficits may be further exacerbated in hospital by restricted intake, intestinal malabsorption, and systemic catabolism.

Daily requirements after surgery range from 30 to 40 kCal/kg and 1 to 2 g/kg of protein per day. Assessment of nutritional status and screening for malnutrition is discussed in Chapter 25.

Patients at risk of malnutrition warrant early specialist input from a dietitian, with nutritional supplementation.

Preoperatively, patients undergoing both elective and emergency surgery are instructed to avoid eating and consuming dairy products for 6 hours before surgery and to stop drinking fluids 2 hours before, to reduce the risk of aspirating gastric contents into the airway and lungs during anaesthesia. Increasingly, hospitals are introducing 'sip until send' policies which allow patients to sip water until surgery to avoid excessive periods of time 'nil by mouth' (NBM). Many elective surgical pathways utilize carbohydrate drinks before surgery to minimize periods of nutritional deficit as part of **e**nhanced **r**ecovery **a**fter **s**urgery (ERAS): multicomponent bundles of care which seek to reduce complications and length of hospital stay.

Postoperatively, patients may be unable to meet the increased energy requirements associated with healing and postoperative inflammation through oral intake alone. Malnutrition adversely affects patients' recovery after surgery through reduced mobility and wound healing, in addition to increased risk of complications such as wound infection. Options for supplementation include energy-dense drinks or enteral nutrition via nasogastric, nasojejunal, or percutaneous feeding tubes. In patients with significant gut dysfunction, total parenteral nutrition (TPN) may be delivered intravenously via central venous lines or peripherally inserted central catheters (PICC). TPN requires intensive monitoring of fluids and electrolytes and patients are at increased risk of infection.

INTRAOPERATIVE CARE

Types of anaesthesia

Anaesthesia exists to make safe and painless surgery possible and can be administered in three main ways, which can also be combined: general anaesthesia, neuraxial anaesthesia, and regional or local anaesthesia.

General anaesthesia

General anaesthesia, or 'GA', is the administration of drugs to a patient to render them unconscious. This is often described to patients as 'going to sleep' for an operation, although brain activity in sleep states is markedly different from that under anaesthetic. GA can involve inhalation of volatile gases (e.g., sevoflurane) or giving intravenous medications (e.g., propofol). Patients can be given a 'push' of i.v. drugs to render them unconscious, prior to maintenance of anaesthesia with gas, have inhalational induction and maintenance (e.g., in small children), or have i.v. induction and maintenance using infusions (known as total intravenous anaesthesia). A variety of agents are available for each technique. All will suppress respiratory drive and airway reflexes, meaning that patients must have airway support (e.g., intubation or supraglottic airway insertion) and ventilation provided. Most will cause a fall in blood pressure and myocardial contractility to some degree. GA can also involve muscle relaxation (with, e.g., atracurium or rocuronium), but this is not a requirement.

Neuraxial anaesthesia

Neuraxial anaesthesia is the blockade of the central nervous system below the brain. This involves passing a needle intrathecally (i.e., into the cerebrospinal fluid) for a spinal anaesthetic, or into the epidural space for an epidural anaesthetic, and administering local anaesthetic drugs (e.g., lidocaine, bupivacaine) to bathe the spinal cord and/or nerve roots. These techniques can be truly anaesthetic (i.e., surgery can be completed without any additional anaesthesia) or analgesic (used in combination with a general anaesthetic for postoperative pain relief, e.g., in laparotomy). As sole techniques, they are used for surgery below the umbilicus of short to medium duration (e.g., caesarean section, knee replacement). Spread of local anaesthetic to high thoracic or cervical spinal levels can cause bradycardia, respiratory depression, and coma. Spinal anaesthetic, in particular, causes a loss of vascular tone through blockade of sympathetic fibres and falls in blood pressure are likely.

Regional anaesthesia

Regional anaesthesia is the blockade of specific nerves (or their roots) by needle injection of local anaesthetic into the anatomical area around those nerves. The aim is to interrupt sensory signals from relevant areas distal to the blockade. For example, an interscalene brachial plexus block involves injection of local anaesthetic around the C5, C6, and C7 roots of the brachial plexus in the neck between the scalene muscles, such that the shoulder and arm become numb, and a procedure (e.g., shoulder surgery) can be performed. Again, this could be used as a sole technique (i.e., with an awake patient) or in combination with general anaesthesia, as pain relief. Historically, nerve blocks were done using landmark anatomy only, but modern practice is to use ultrasound to guide the path of the needle towards neurologic structures of interest. Some nerve blocks do not target specific structures, but rather fill a space or potential space with local anaesthetic, with action on fibres passing through, such as the rectus sheath block for midline abdominal incisions. These are analgesic rather than anaesthetic. Field blocks, or simple wound infiltration, involve injection of local anaesthetic without targeting a specific nerve or plane. This provides analgesia to a small area of tissue and allows minor surgery (e.g., lipoma excision) or postoperative analgesia at the surgical site.

Anaesthetic drugs

The anaesthetic pharmacy is large and varied, but some agents in regular use merit brief introduction.

Volatiles—anaesthetic gases

Volatile gases are used to induce and maintain anaesthesia by inhalation. In modern practice they include sevoflurane, desflurane, and isoflurane, but historically have covered a multitude of compounds, including ether, cyclopropane, and halothane. A vaporizer mounted on the anaesthetic machine is used to produce a precisely controlled mixture of anaesthetic and life support gases (i.e., oxygen and air) to be delivered to the patient. The percentage of anaesthetic gas within this mixture can be altered by the anaesthetist to control depth of anaesthesia. The partial pressure of anaesthetic gas in the alveoli dictates the amount in the patient's bloodstream and hence the amount reaching the brain to exert the anaesthetic effect. Mechanism of action is likely via lipophilic protein binding, but a variety of theories remain under review—no individual target channel or receptor has been identified conclusively. Volatiles cause dose-dependent hypotension, myocardial depression, and postoperative nausea and vomiting. They form a significant part of healthcare-related greenhouse gas emissions.

Propofol

Propofol is an intravenous anaesthetic, formulated as a milky-white lipid emulsion, at 10 mg/mL or 20 mg/mL. It was first described in 1980 and has rapidly become the most commonly used anaesthetic drug in the UK. Propofol can be used as a bolus dose to induce anaesthesia, an infusion to maintain anaesthesia, or for sedation delivered in either way at lower doses. It has a rapid onset and offset, with relatively little accumulation when infused for a long time (known as a short context-sensitive half time). This makes it excellent for use in anaesthesia and intensive care for longer procedures or sedation of the critically ill intubated patient. In these settings, it is almost always used alongside an infusion of opioid, for example, remifentanil. Propofol also reliably and rapidly blunts airway reflexes, making insertion of supraglottic airway devices (e.g., a laryngeal mask or iGel) straightforward and allowing swift progress between cases on a surgical list. It does not cause nausea and vomiting but does cause a fall in blood pressure

due to vasodilation, which can be severe and dangerous in the elderly or shocked patient.

Opioids

Derived from *Papaver somniferum*, the opium poppy, opioids are effective analgesics. The class includes morphine (the reference compound to which all other opioids are compared), fentanyl, remifentanil, alfentanil, oxycodone, codeine, and diamorphine, among others. All act on the mu opioid receptor, found throughout the CNS. Naloxone is a competitive antagonist at this receptor, used to reverse the effects of opioids. Most agents can be used i.v. (as a bolus or infusion) and orally, while some are also used intrathecally or epidurally. Speed of onset and offset and potency vary significantly between compounds. Remifentanil is used as a component of total intravenous anaesthesia because of its rapid onset and offset, non-hepato-renal metabolism, and because it does not accumulate when infused for long durations. Morphine, fentanyl, and oxycodone are commonly used to treat acute postoperative pain but should be avoided in the chronic pain setting due to risk of addiction. Opioids can cause sedation, respiratory depression, nausea, itching, and constipation.

Muscle relaxants

Muscle relaxants, or neuromuscular blocking drugs (NMBDs), are used to temporarily paralyse a patient. This can be for anaesthetic reasons (to improve conditions when instrumenting the airway), surgical reasons (to improve operating conditions, e.g., in laparoscopy or when closing an abdominal incision), or in critical care to improve ventilation in severe respiratory failure. They are always given alongside general anaesthetic drugs. Agents include atracurium, rocuronium, and suxamethonium. All are chemical derivatives of curare, found in several plant species native to South America and used on the tips of arrows for hunting. NMBDs act by competitive antagonism at the neuromuscular acetylcholine receptor, except suxamethonium which acts as an agonist, depolarizing the neuromuscular end-plate and rendering it refractory to further stimulus. Speed of onset and offset, alongside duration of action, usually drive the choice of a specific agent for a given task. Level of paralysis is monitored by delivering small currents transcutaneously, to stimulate a nerve and observe the extent of muscle contraction ('twitches') that follows (e.g., the ulnar nerve, causing movement of the hand). Non-depolarizing NMBDs (those other than suxamethonium) are reversed by increasing the concentration of acetylcholine at the neuromuscular junction, such that they are out-competed. This is achieved by administering neostigmine, which blocks the metabolism of acetylcholine by acetylcholinesterase. Rocuronium and vecuronium can be reversed rapidly by sugammadex, a drug which chemically envelops the muscle relaxant. NMBDs cause voluntary respiration to cease and obtund airway reflexes. They are important triggers for anaphylaxis.

Local anaesthetics

Local anaesthetics are used to block transmission of action potentials along a nerve, such that sensation (pain) is not felt in the area supplied by that nerve. They cross neuronal cell membranes and block sodium channels. Examples include lidocaine, bupivacaine, and prilocaine. Choice of agent is driven by duration of action, speed of onset and offset, and route of administration. Local anaesthetics can be given via ultrasound-guided needle injection around a nerve or nerve root (e.g., popliteal sciatic nerve block, Fig. 14.6), into a plane or potential space (e.g., erector spinae plane block), into a wound or tissue area (including the eye), intrathecally or

Fig. 14.6 **Popliteal sciatic nerve block performed by the supine position lateral approach under ultrasound guidance.**

epidurally. Toxicity is seen at high doses when significant absorption into the bloodstream occurs. This is characterized initially by excitatory phenomena such as tingling lips, visual disturbance, tinnitus, and tremor, followed by seizure, then coma. Cardiac dysrhythmia may also occur, particularly with bupivacaine.

Additional drugs

Antiemetics

Antiemetic drugs are designed to prevent and treat nausea and vomiting, common symptoms associated with anaesthesia. These drugs exert their effects by targeting neurotransmitter pathways involved in the emetic (vomiting) reflex.

Commonly used anti-emetics, their mechanisms of action, and potential adverse effects are described in Box 22.7.

Non-opioid analgesia

Non-opioid analgesic drugs play a crucial role in managing pain during and after surgery, offering effective pain relief with fewer opioid-related side-effects. Common non-opioid analgesics include:

Non-Steroidal Anti-Inflammatory Drugs (NSAIDs): Drugs like ibuprofen, parecoxib, and ketorolac are commonly used to reduce inflammation and alleviate pain. They work by inhibiting enzymes (COX-1 and COX-2) involved in prostaglandin synthesis, thereby reducing pain and inflammation.

Paracetamol: Widely used for its analgesic and antipyretic properties, paracetamol works centrally to modulate pain perception. It is often combined with NSAIDs for synergistic pain relief.

Local Anaesthetics: Administered through injections or topical applications, local anaesthetics like lidocaine provide targeted pain relief by blocking nerve signals in specific areas.

N-methyl-D-aspartate Receptor Antagonists: Drugs like ketamine act on the N-methyl-D-aspartate receptors in the central nervous system, modulating pain perception. They are used as adjuncts to other analgesics to enhance pain control.

Emergency drugs

Anaesthetic emergency drugs are essential components of any medical facility, ready to address unforeseen complications during surgery. Key drugs include:

Adrenaline: Acts rapidly to counteract severe allergic reactions (anaphylaxis) and cardiac arrest by increasing heart rate and blood pressure.

Atropine: Reverses bradycardia and counters excessive vagal stimulation.

Vasopressors: Raise blood pressure by constricting blood vessels, counteracting anaesthesia-induced hypotension.

Suxamethonium (Succinylcholine): A neuromuscular blocking agent used for rapid sequence intubation to secure the airway.

Intraoperative fluids

Intravenous (i.v.) fluids are vital during surgery to maintain hydration and electrolyte balance and support physiological haemodynamics, such as blood pressure and heart rate. The choice of i.v. fluids depends on the patient's needs and the surgical context.

Crystalloids: These are solutions containing water, electrolytes, and, sometimes, sugar. Normal saline (0.9% sodium chloride) is a common crystalloid used to replace extracellular fluid losses and restore electrolyte balance. Balanced crystalloids, like Ringer's lactate or Plasma-Lyte, mimic the body's electrolyte composition and are often preferred.

Colloids: These solutions contain larger molecules, such as starches or albumin, which tend to stay in the intravascular space longer, helping to maintain blood volume. However, the use of colloids has been debated due to concerns about adverse effects, including allergic reactions and acute kidney injury.

Blood and Blood Products: In certain scenarios, there may be a need for transfusions of whole blood, packed red blood cells, platelets, or fresh frozen plasma to address significant blood loss and maintain adequate oxygen-carrying capacity.

Dextrose Solutions: Solutions with varying concentrations of dextrose are used to prevent hypoglycaemia, especially during prolonged surgeries.

The choice of i.v. fluids is carefully considered based on the patient's co-morbidities, the extent of surgery, and specific fluid requirements. Maintaining fluid balance is crucial to prevent complications such as electrolyte imbalances, fluid overload, or inadequate tissue perfusion during and after surgery.

Airway management

Almost all hypnotic anaesthetic drugs can suppress ventilatory drive and airway reflexes and relax muscles supporting the mouth and tongue. This process is carefully controlled and monitored by anaesthetists, but without intervention can cause airway obstruction. Airway management involves procedures and techniques aimed at establishing and maintaining a clear and secure airway. In turn, this ensures adequate oxygenation and ventilation while patients receive GA.

Preoperative airway assessment is discussed earlier. Proper positioning of the patient is crucial to maintain airway patency. The head-tilt, chin-lift manoeuvre, or the use of specialized devices (e.g., Guedel airways) may be employed to open and align the airway during induction of GA.

During GA, airway patency is secured using specialist devices categorized according to their positioning relative to patient anatomy. **Endotracheal tubes** are inserted beyond a patient's vocal cords into the trachea, where a balloon is inflated to protect the lungs from potential gastric content regurgitation. **Supraglottic airway devices (SGAs)**, which include laryngeal masks and iGels, are simpler to insert and less invasive, being situated above the vocal cords, but do not provide definitive protection against aspiration. Other airway devices are used for certain surgeries, including double-lumen endotracheal tubes, which can isolate lungs for thoracic surgery, or small or armoured tubes that can be used when surgeons need to operate around a patient's airway. Airway devices are most commonly inserted using a technique called laryngoscopy, in which an instrument is used to visualize the larynx and vocal cords. Laryngoscopy can offer direct views or involve cameras, mirrors, or fibreoptic scopes to provide indirect inspection of airway structures.

After securing the airway, anaesthetists ensure proper mechanical ventilation to maintain oxygenation and prevent hypercarbia. Ventilator settings are adjusted to individual patient needs. Continuous monitoring of oxygen saturation (pulse oximetry), end-tidal carbon dioxide (capnography), and other vital signs helps detect changes in the patient's respiratory status promptly. Anaesthetists are trained to respond quickly to unexpected airway emergencies, such as laryngospasm, bronchospasm, or tube dislodgement and have a range of techniques and equipment at their disposal to manage these situations.

At the end of surgery, anaesthetists oversee the safe extubation of the patient. This includes assessing the patient's readiness to breathe independently and ensuring a smooth transition to spontaneous breathing and wakefulness.

The anaesthetic machine

The anaesthetic machine is a piece of equipment found in all areas where anaesthesia is delivered including operating theatres, diagnostic and therapeutic radiology suites, and some emergency departments. The machine is connected to piped supplies or large cylinders of oxygen and air. The machine consists of a combination of essential equipment for delivering anaesthesia including:

- **Physiological monitoring** including pulse oximetry (SpO_2), non-invasive blood pressure monitoring (NIBP), three-lead ECG. Additional monitoring may include invasive blood pressure monitoring via an arterial cannula, cardiac output monitoring, temperature, processed electroencephalogram (pEEG) to monitor depth of anaesthesia and neuromuscular monitoring to assess the effect of NMBDs.
- **Ventilator monitoring** including pressures and flow of gases. Gas analysers measure inspired and expired gases including oxygen, carbon dioxide, and volatile anaesthetic gases. End-tidal carbon dioxide monitoring ($etCO_2$) is essential for any patient undergoing GA or sedation.
- **A mechanical ventilator** similar to those found in critical care settings but with fewer advanced modes of ventilation. Positive pressure ventilation maintains delivery of oxygen ± anaesthetic gases because anaesthetic agents cause hypoventilation and apnoea. Depending on the mode selected, the anaesthetist may select the pressure, volume, number of breaths per minute, and concentrations of different gases delivered to the patient.
- **A breathing circuit** which comprises tubing inside and outside the machine to deliver oxygen, air, and volatile anaesthetic gases to the patient. The most used in modern practice is the circle system which facilitates efficient recycling of anaesthetic gases and helps to maintain moisture and humidity of inspired gases. Expired carbon dioxide, the product of aerobic respiration, is absorbed by granules of soda lime to remove it from the circuit.
- **Vaporisers** which deliver volatile anaesthetic gases into the breathing circuit and subsequently to the patient. The anaesthetic gases may be carried along with oxygen and air or directly injected into the breathing circuit. The concentration added to the circuit is controlled by the anaesthetist.
- **Other supplemental equipment** including suction, additional oxygen ports, and emergency equipment including a self-inflating bag for use in case of a power or gas supply failure.

Ensuring that the anaesthetic machine is functioning normally is an essential component of routine safety checks before delivering any form of anaesthesia to a patient.

Theatre safety and human factors

Human factors play a crucial role in ensuring safety within the operating theatre, significantly influencing the overall quality of patient care. Communication, teamwork, and situational awareness are pivotal aspects of these factors.

Clear and effective communication among members of the surgical team is paramount. Miscommunication can lead to errors, compromising patient safety. Open lines of communication, including briefings before surgery and timely updates during the procedure, are vital. The collaborative efforts of the perioperative team are essential for safety. Well-coordinated teamwork helps prevent errors and promotes a culture of safety. Each team member's role and responsibilities should be clearly defined to avoid misunderstandings.

Maintaining situational awareness of the surgical environment and anticipating potential challenges is critical. This involves constant vigilance and attention to detail, ensuring that the team is prepared to respond effectively to unexpected developments, for example, awareness of ongoing blood losses.

Healthcare professionals working in the operating theatre often experience high levels of stress and fatigue. Acknowledging these factors is essential for maintaining concentration and decision-making abilities. During long operations, surgeons may take short breaks and anaesthetists may handover patient care to a colleague to take a short break. Ongoing training and education contribute to the continuous improvement of skills and knowledge among the perioperative team. Regular drills and simulations can enhance preparedness for emergency situations in the operating theatre environment.

Encouraging a culture where all team members feel empowered to voice concerns or question decisions, regardless of hierarchy, is crucial. This fosters a climate where patient safety is prioritized over traditional hierarchies. By recognizing and addressing these human factors, perioperative teams can create a safer and more efficient operating theatre environment, ultimately improving patient outcomes and reducing the likelihood of adverse events.

POSTOPERATIVE CARE

Location of postoperative care

The most appropriate location for postoperative care is based on several factors, including the time point after surgery, type and complexity of the operation, the patient's medical history and comorbidities, and any procedure-specific monitoring and treatment requirements.

Immediate recovery:

- Patients are transferred to the recovery unit directly from the operating theatre, where they are closely monitored as they emerge from anaesthesia.
- Key responsibilities in recovery include assessing vital signs, oxygen saturation, consciousness level, surgical site condition, and pain management.
- Nurses and anaesthetists oversee this phase and ensure patients are stable and comfortable before discharge to a higher level of care or the surgical ward.

Surgical ward care:

- Patients who are stable and have recovered sufficiently may be transferred to a surgical ward.
- Nurses on the surgical ward continue to monitor vital signs, assess wound healing, administer pain medications, manage drains and tubes, and provide patient education on postoperative care.
- Physiotherapists may assist with early ambulation and mobility.

High dependency unit (HDU) or Level 2 care:

- Patients with more complex surgical procedures or higher acuity may be admitted to an HDU or enhanced recovery unit.
- In these units, patients receive more intensive monitoring, including continuous ECG, more frequent vital sign assessments, and specialized care for complex medical conditions.
- The nursing staff-to-patient ratio is often lower in HDUs to provide a higher level of care.

Intensive care unit (ICU) Care or Level 3 care:

- Patients with the most complex surgeries, major co-morbidity, or those who experience severe postoperative complications may require admission to an ICU.
- ICU care involves continuous monitoring of vital signs, invasive hemodynamic monitoring, mechanical ventilation if needed, and management of organ support.
- A multidisciplinary team, including intensivists, critical care nurses, and physiotherapists, manages care in the ICU. Nursing support is often provided at a 1:1 ratio.

Home Recovery Care:

- After being discharged from the hospital, many patients continue their recovery at home.
- Home recovery care involves following instructions for wound care, pain management, medication administration, and any physical therapy or exercises as prescribed by the surgical team.
- Patients and caregivers play an active role in managing home recovery care, with regular follow-up appointments scheduled with the surgeon.

Rehabilitation and Follow-Up:

- Some patients, especially those who have undergone orthopaedic, neurological, or major reconstructive surgeries, may require rehabilitation services.
- Rehabilitation can include physical therapy, occupational therapy, and speech therapy, depending on the specific surgical procedure and patient needs.
- Long-term follow-up appointments with the surgeon may be necessary to monitor progress and address any complications or concerns.

It is vital that a continuum of care is maintained across each different area as patients progress through the postoperative recovery process. Surgical teams provide overriding continuity, with effective communication and comprehensive handover processes by other parts of the perioperative team, ensuring that transition from one care setting to another (e.g., from operating theatre to recovery or ICU) involves sharing of all relevant patient information and ongoing care needs. Checklists are frequently used to minimize human factor errors.

Important and common complications after surgery

Bleeding

Bleeding usually occurs in the acute phase, in the first 12 to 24 hours after surgery. Effects depend on the volume and location of the bleed. Rapid large volume loss leads to shock, with hypotension and tachycardia. This is more likely when blood loss is into a major body cavity, for example, the abdomen, as a haematoma is unlikely to form and tamponade the damaged vessel. A normal haemoglobin in this setting is not necessarily reassuring—it is a poor marker of acute haemorrhage. In confined spaces, for example the brain, spine, or fascial compartments in the limbs, compressive effects may predominate, revealed as pain and derangement of function of the compressed tissues (e.g., altered neurology).

Treatment is with transfusion of red cells and other blood products to stabilize and resuscitate, as a bridge to definitive surgical (or radiological) control of the bleed.

Venous thromboembolism

Venous thromboembolism (VTE) comprises deep vein thrombosis (DVT) and pulmonary embolism (PE). Surgical patients are at higher risk of VTE because of their procoagulant state (due to the inflammatory response to surgery) and immobility (whether during or after surgery). Specific risk factors include prolonged surgery (>90 minutes), abdomino-pelvic procedures, lower limb orthopaedic surgery, and cancer. DVT presents as limb swelling and pain and is usually unilateral. PE presents as pleuritic chest pain, breathlessness, tachycardia, and, if massive, hypoxia and shock. These normally occur more than 72 hours after surgery and can present after more than a week. Prevention is mechanical (compression stockings or intermittent pneumatic compression devices) or pharmacological (commonly low molecular weight heparin). Both prevention and treatment can be complex because using anticoagulants after surgery can cause bleeding. Risk/benefit discussions are essential and should include anaesthetists, surgeons, haematologists, and patients/family. Diagnosis and treatment of VTE is discussed in Chapter 11.

Pneumonia/postoperative pulmonary complications

Postoperative pulmonary complications (PPCs), including pneumonia, are among the most common adverse events after surgery. Prevalence of postoperative pneumonia can vary widely depending on several factors, typically ranging from 1% to 5% of surgical cases. Factors predicting risk include the type of surgery, the patient's baseline respiratory function, and preventive measures employed. Proximity of surgical incision to the thorax, where pain disrupts the performance of respiratory muscles, is a strong predictor of PPCs, as is age, with even healthy older patients being at higher risk of PPCs. PPCs are also predictors of short-term and long-term health outcomes after surgery and are associated with increased risk of admission to critical care and prolonged length of hospital stay.

Atelectasis describes partial collapse or compression of lungs and is the most common PPC, occurring due to depressant effects of anaesthesia and mechanical ventilation causing small airway collapse within lungs and impairment to normal mucociliary clearance. Failure to reverse atelectasis can lead to a worsening cycle of ventilatory impairment, leading to trapped sputum, increased oxygen requirements, and respiratory infection.

Interventions aimed at preventing PPCs include pre-emptive strategies to optimize respiratory physiology (such as respiratory muscle training before surgery) and intraoperative and postoperative interventions to minimize the adverse effects of surgery and anaesthesia. Early mobilization, supervised respiratory physiotherapy, and ambulation help prevent atelectasis and promote normal lung function. Effective pain management to encourage coughing and deep breathing are also vital.

Pain

Acute postoperative pain is common: up to 20% of patients experience severe pain in the first 24 hours after surgery. Pain should be pre-empted and prevented, with appropriate analgesia given before and during surgery. The WHO pain ladder (Fig. 50.3) is an appropriate starting point to manage pain. Additional options include intravenous strong opioids (perhaps in the form of patient-controlled analgesia, or PCA), gabapentinoids, alpha-2 agonists (clonidine), ketamine, magnesium, and regional anaesthetic techniques. Uncontrolled pain leads to further morbidity, both physical and psychological. Increased sympathetic drive adds to surgical stress and challenges the myocardium and circulation through hypertension and tachycardia, with MI and stroke possible. Surgery to the abdomen or thorax can cause pain, which prevents

adequate deep breathing or coughing, leading to atelectasis and pneumonia.

Chronic post-surgical pain is defined as pain persisting beyond the healing process, that is, beyond 3 months postoperatively. Knee replacement, thoracotomy, amputation, and mastectomy all have notably high rates of chronic post-surgical pain. Neuropathic type features may develop, and treatment should be multidisciplinary, along the bio-psycho-social model of care. Strong opioids should be avoided where possible in chronic pain settings.

Delirium and perioperative neurocognitive disorders

Perioperative delirium (POD) is a common and potentially serious condition that affects individuals after surgery. It is characterized by acute confusion, disorientation, and altered cognitive function, often occurring within the first few days following an operation. While the exact causes are multifactorial, factors such as age, pre-existing cognitive impairment, type of surgery, and anaesthesia can contribute to its onset. POD poses risks, including prolonged hospital stays, increased morbidity, and higher healthcare costs. Preventative measures include careful medication management, minimizing sedatives, and promoting early mobilization. Timely recognition and management are crucial. The approach to a delirious patient is discussed further in Chapter 49.

Postoperative cognitive dysfunction (POCD) is a condition marked by a longer-term decline in cognitive function following surgery, with symptoms such as memory problems, reduced attention, and impaired information processing. It is particularly prevalent in older adults and can persist for weeks to months, affecting the patient's daily life and quality of life. The exact cause of POCD is not fully understood but is thought to involve factors like anaesthesia, surgery-related stress, and inflammation. Prevention and management strategies include careful anaesthesia selection, minimizing surgical trauma, and promoting early postoperative mobilization.

Myocardial infarction and myocardial injury

Postoperative myocardial infarction (POMI) is a critical, albeit rare, complication that occurs following surgery. It refers to the occurrence of a myocardial infarction in the immediate postoperative period or within 30 days of surgery. POMI is a multifactorial condition, often associated with underlying heart disease, atherosclerosis, and the stress of surgery. Patients with risk factors like hypertension, diabetes, or a history of heart disease are more susceptible. Early diagnosis (ECG, cardiac biomarkers) and intervention are essential, as POMI can be life-threatening. Preventative measures include preoperative cardiac evaluation, careful monitoring during and after surgery, and optimizing patient management to reduce the risk of cardiac events in the postoperative setting.

Myocardial injury after non-cardiac surgery (MINS) is a condition characterized by damage to the heart muscle during surgery unrelated to heart procedures. MINS often goes unnoticed but can lead to complications. It is most commonly seen in individuals with existing heart disease or risk factors but is thought to occur primarily as a result of a supply and demand mismatch in myocardial perfusion rather than as a direct result of underlying cardiac disease.

Acute kidney injury

Postoperative AKI is a common (up to 30% of surgical patients) and often avoidable complication. Like many other postoperative complications, even lesser grades of AKI are associated with longer-term impact, including increased hospital stay after surgery, mortality, and risk of developing chronic kidney disease. AKI is recognized through either reduced urine output or an elevation in serum creatinine on postoperative blood testing. Reduced urine volumes are common immediately after surgery due to adaptive neurohumoral responses (including AVP secretion and activation of the renal-angiotensin-aldosterone axis) and do not necessarily indicate AKI, but recognition should trigger patient assessment to include review of haemodynamic parameters and fluid balance and exclude post-renal factors (e.g., urinary retention or a blocked catheter). The STOP mnemonic (Sepsis and hypovolaemia, Toxins, Obstruction, Parenchymal) is a useful *aide memoire* to potential causes of AKI.

Attention to fluid balance is key to avoiding AKI after surgery. These measures include maintaining haemodynamic stability and avoiding intraoperative hypotension and extremes of fluid administration. Both too much and too little fluid may be deleterious, the former resulting in oedema or hyponatraemia, and the latter prerenal 'stress' to the kidneys, which are highly metabolically active organs and are particularly susceptible to impaired oxygen delivery and AKI is often hence an early sign of inadequate global tissue oxygen delivery. NSAIDs should be avoided or used sparingly in patients with pre-existing renal disease.

Patients are commonly monitored by measuring urine output and serial plasma creatinine after surgery. For high-risk patients, this should occur in a critical care environment. Specialized devices, including cardiac output monitors, may be employed to specifically guide fluid management during the perioperative period.

Direct complications of surgery

Anastomotic leak

Leaks from surgically created anastomoses, created using sutures or automated stapling devices during gastrointestinal or other luminal surgery, may result in contents from the gastrointestinal tract leaking into the peritoneal space. This may lead to sepsis, abscess, and fistulae. Investigation and management may involve CT imaging, treatment with antimicrobials, initiation of TPN and delayed restoration of enteral nutrition, drainage by interventional radiology, or unplanned return to theatre.

Ileus

Paralytic ileus occurs after intraperitoneal surgery associated with handling or instrumentation of the bowel. Clinical features are similar to mechanical bowel obstruction including vomiting, abdominal distension, and constipation. Management comprises gastric drainage and decompression via a wide bore nasogastric (NG) tube, intravenous fluid replacement and maintenance, optimization of electrolytes, and where possible, withdrawal or dose reduction of medications which may slow gastrointestinal motility such as opioids.

Wound infection

Localized infection may cause superficial or deep wound dehiscence, prompt return to theatre for exploration and washout, impair wound healing, and delay patients' recovery and hospital discharge.

Flap/graft failure

Surgery involving transfer of skin or tissue to another location relies on delivery of oxygen and venous drainage, often via surgically formed anastomoses of small blood vessels. Complications include infection and reduced perfusion which may require washout, unplanned return to theatre, or revision of the graft.

Other morbidity and mortality

Functional recovery/morbidity

Recovering from the physiological insult of surgery may be associated with a delay or failure to return to previous levels of function. Delayed recovery and rehabilitation are influenced by several patient factors including pre-existing functional capacity and co-morbidities, advanced age, and frailty. Surgical factors include unplanned or major surgery, critical care admission, or other postoperative complications.

Patients may experience new morbidity such as chronic pain, anaemia, or deterioration in renal function. Patients may take longer than expected to leave the hospital and may require additional support including carers at home or moving to a residential or nursing care institution.

Mortality

Overall mortality in the 30 days after surgery is estimated between 0.5% and 3.6%. High-risk patients represent 80% of postoperative deaths. Developing postoperative complications is also associated with increased mortality at 1 year after surgery. This highlights the importance of preoperative risk assessment (discussed earlier) to facilitate:

- Planning of treatment options
- Optimization of modifiable risk factors
- Shared decision-making with patients, families, and carers
- Resource planning including postoperative critical care admission

Enhanced recovery after surgery and day-case surgery

Patients recovering from surgery would traditionally expect long hospital stays and extended periods of bed rest. Modern approaches emphasize rapid return to normal function and avoid prolonged admission.

Enhanced recovery pathways (also known as fast-track pathways) map a patient's surgical journey. The aim is to achieve the earliest possible return to drinking, eating, and mobilizing ('DrEaMing') by focusing on streamlined, standardized, and protocolized care. An emphasis on patient-centred outcomes like these has translated into reduced surgical stress and hence complications, reduced hospital length of stay, and reduced cost. Enhanced recovery guidelines are designed for specific surgical procedures (e.g., cystectomy, hemicolectomy) but some core principles are shared:

- Minimize preoperative starvation time, avoid dehydration
- Laparoscopic rather than open surgery where possible
- Multimodal analgesia, with regional anaesthetic techniques where appropriate
- Minimize use of systemic (i.v.) opioids
- Use goal-directed fluid therapy, avoid excess i.v. fluid administration
- Avoid use of drains, NG tubes, and urinary catheters
- Early physiotherapy and mobilization
- Early restoration of diet
- Use appropriate antimicrobial prophylaxis, for the minimum required duration
- Use antiemetics and appropriate anaesthetic techniques to minimize risk of postoperative nausea and vomiting
- Use mechanical or chemical thromboprophylaxis where appropriate

Many surgical procedures can be completed as **day cases**, that is, without admitting the patient to hospital overnight. This approach allows patients to return rapidly to their own home and their normal function and routines. It implicitly includes the principles of enhanced recovery detailed above. Day surgery also avoids potential cancellations when inpatient beds are full due to high emergency caseload and, for this reason, day surgery units are often physically separate from inpatient areas within a hospital. Rapid turnover and elective work are prioritized, although emergency procedures are not excluded from the day case approach (e.g., superficial plastic surgery, drainage of some abscesses). Successful day surgery requires high-quality preoperative assessment to avoid unplanned conversion to an inpatient case.

There is no single list of procedures that are/are not suitable for day surgery, nor a list of particular patient groups. However, major surgery with expected bleeding, opening a major body cavity, prolonged surgery, and patients without suitable home support in the 24 hours following GA are usually excluded. Perhaps counter-intuitively, elderly or co-morbid patients may benefit from day surgery, through reduced risk of VTE and nosocomial infection linked to hospital admission, and can resume normal medication and care routines quickly. The best centres would consider day surgery the default approach and ask, 'what benefit is offered to this patient through inpatient admission?' rather than simply allocating an overnight bed automatically. Frontier models include remote home monitoring of postoperative patients and intensive recovery in non-hospital facilities.

Orthogeriatrics

An orthogeriatric service is a specialized healthcare approach that combines the expertise of orthopaedic and elderly care medicine to provide comprehensive care for older adults with fractures and complex medical needs. The principles of medicine for the older adult are discussed in Chapter 49. This model of care recognizes the unique needs of elderly patients who may have multiple co-morbidities and emphasizes a multidisciplinary and holistic approach.

Orthogeriatrics employs a multiprofessional approach to recognize, prevent, and treat common postoperative complications including delirium and pressure ulceration, navigate polypharmacy and drug-drug interactions, reduce falls risk, and facilitate smooth transitions of care from the hospital to rehabilitation centres or home. The orthogeriatric model has served to drive up quality and safety of care in this vulnerable population and improve patient satisfaction.

Perioperative fluid balance

Perioperative fluid management seeks to maintain physiological homeostasis of water, glucose, and electrolytes. The aims of fluid management are to maintain perfusion pressure and oxygen delivery to tissue capillaries in order to facilitate aerobic respiration (see also Chapter 27).

Fluid replacement is often described in terms of:

- Resuscitation: replacing acute fluid losses to restore physiological homeostasis
- Replacement: replacing fluid and electrolyte deficits
- Maintenance: providing weight-based daily requirements of water, glucose, and electrolytes.

In practice in the perioperative setting, these three indications for fluid are managed simultaneously.

Maintenance intravenous fluid therapy may be required in patients fasting for protracted periods or unable to consume fluid orally due to acute illness. The UK NICE Guidelines for intravenous fluid therapy in adults in hospital highlight core principles for safe prescription of maintenance fluids to address baseline requirements. Prescriptions of maintenance fluid should comprise:

- **25 to 30 mL/kg/day of water** (so a 70 kg patient would require approximately 2000 mL/24 hours).
- Approximately **1 mmol/kg/day of potassium, sodium, and chloride** (round potassium prescriptions to commonly available fluids which contain 20 mmol or 40 mmol of potassium; adding potassium to bags of i.v. fluid should be avoided to diminish the chance of dangerous errors).
- Approximately **50 to 100 g/day of glucose** to limit starvation ketosis.
- For patients living with obesity, adjust maintenance i.v. fluid prescriptions to their ideal body weight (patients rarely require more than 3000 mL/24 hours).

Exercise additional caution or consider lower volumes (e.g., 20–25 mL/kg/day) for patients who are older, frail, have renal impairment or cardiac failure, or are malnourished.

Regardless of initial prescription choice, monitoring of fluid balance and electrolytes is pivotal and the ongoing regimen should be tailored accordingly.

Resuscitation and replacement fluid requirements are determined by the clinical status of the patient (i.e., symptoms and signs of hypovolaemia) and fluid losses due to the underlying pathology or surgical procedure, respectively. Potential losses may include:

- Bleeding
- Loss of other body fluids via vomiting, diarrhoea, sweating, urine
- Shifts in extracellular fluid from the intravascular space to the interstitial space leading to functional reduction in circulating volume

Fluid balance is assessed through balancing measurable inputs and outputs and clinical assessment (heart rate, BP, urine output, etc.) ± blood gas sampling (serum lactate, point of care haemoglobin, creatinine, and electrolytes). Advanced perioperative haemodynamic monitoring includes cardiac output monitoring using medical devices which analyze arterial waveforms, Doppler ultrasound of blood flowing through the aorta, or temperature difference in blood samples from different parts of a patient's circulation. Devices and techniques include arterial waveform pulse contour analysis, oesophageal Doppler, transthoracic or transoesophageal echocardiography, and pulmonary artery flotation catheters. Each device has a different profile of invasiveness, accuracy, cost, and variables measured.

The most common intravenous fluids used for resuscitation and replacement are balanced crystalloid solutions. Bolus administration (e.g., 250–500 mL of Hartmann's solution) is preferred for resuscitation and response gauged by continuous reassessment of clinical and haemodynamic parameters. Repeated boluses may be required, although a requirement for volumes in excess of 30 mL/kg should generally trigger escalation to critical care services, as additional pharmacological treatments of shock may be required.

Administering too little or too much fluid are both associated with perioperative complications: hypovolaemia may contribute to hypoperfusion, acidosis, hypotension, and AKI while hypervolaemia may be associated with pulmonary complications, venous congestion, and anastomotic leak. Regular reassessment of fluid status and monitoring of electrolytes and renal function remains the best defence against fluid-related complications. The best method of assessing fluid requirements remains a debated topic; anaesthetists aim to deliver individualized fluid management to administer a 'Goldilocks' level of just enough but not too much fluid.

SPECIAL POPULATIONS

Emergency laparotomy

Emergency laparotomy is a common operation worldwide and is associated with a mortality risk of almost 10% in the UK. Indications for emergency laparotomy are broad, spanning acute abdominal trauma, bowel obstruction, perforations, and vascular emergencies. The complexity of surgery varies with indication and is one important determinant of outcomes. Surgery is often time-sensitive necessitating swift decision-making and leaving little time for detailed risk assessment or patient preparation, with perioperative resuscitation proceeding in parallel.

Mortality rates, reflecting the severity of underlying conditions, are higher than those observed in elective surgeries. Complications such as surgical site infections, wound dehiscence, respiratory issues, intra-abdominal collections, and sepsis pose additional challenges and prolonged hospital stays are common. Resource utilization is substantial, with emergency laparotomies demanding significant operating theatre time, intensive care unit resources, and postoperative care.

In the UK, the National Emergency Laparotomy Audit (NELA) allows institutions to report and benchmark outcomes at a national level and has led to reduced variations in the quality of care between institutions and a reduction in mortality and complications.

Hip fracture

Hip fractures are a serious and relatively common injury in the UK, with approximately 76 000 fractured neck-of-femurs occurring each year. Frail and older patients, particularly those over the age of 65 years, are most vulnerable. Approximately 75% of hip fracture patients are females, in part due to the higher prevalence of osteoporosis. Hip fractures are associated with increased mortality rates, particularly in the first year following the fracture. Approximately 10% to 30% of hip fracture patients in the UK do not survive the first year after the injury, with advanced age and underlying health conditions contributing to higher mortality rates.

In response to the high incidence of hip fractures, fall prevention programs and initiatives have been implemented in various regions of the UK (see Chapter 49 for further discussion). Treatment to reduce the risk of future fracture is discussed in Chapter 33.

Surgical recovery can be challenging for hip fracture patients, and orthogeriatric services help provide a holistic and 'joined-up' approach to care (see earlier). Analgesic nerve block injections (e.g., fascia iliaca blocks) are now widely used for pain control after hip fracture and help both ensure patient comfort and reduce the risk of complications such as delirium.

As surgical repair to hip fractures can offer pain benefits, surgery should not be delayed for more than 36 hours after hospital admission. Spinal anaesthesia is preferred over general anaesthesia and may have benefits for reducing delirium risk.

Early rehabilitation to restore mobility and function is crucial to patient outcomes.

Cancer

The incidence of cancer is increasing: one in three people in the UK will be diagnosed with cancer in their lifetime and one in four will die from it. More than 60% of patients with cancer require surgery

during their treatment and therefore high-quality perioperative care is a fundamental component of cancer care.

Patients with cancer are discussed in ***multidisciplinary meetings*** which combine the different expertise of radiologists, pathologists, oncologists, surgeons, palliative care physicians, anaesthetists, clinical nurse specialists (CNSs) and allied health professionals to consider the individualized risks and benefits of different treatment options (see Chapter 41).

The main ***role of surgery*** in cancer is to completely remove the primary tumour with clear margins of normal tissue around it. Depending on the tumour type and Tumour, Node, Metastasis (TNM) stage, surgical resection of a cancer presents a potentially curative intervention. Surgery may also play a role in cancer which has spread locally or metastasized to distant sites including removing large tumours causing symptoms due to mass effect or bleeding or resection of metastases.

Surgery is often supplemented by ***systemic anticancer therapy*** (including chemotherapy, immunotherapy, or targeted therapies) and/or radiotherapy to optimize conditions for surgery or reduce the risk of recurrence and metastases. Oncological therapies may be delivered before surgery (neoadjuvant) and/or after surgery (adjuvant). Multidisciplinary teams consider the optimal timing of systemic anticancer therapy in relation to surgical intervention: on one hand, neoadjuvant treatment aims to reduce the size or extent of the tumour to make surgery possible or more successful; on the other hand, systemic anticancer therapy is associated with reduced cardiorespiratory fitness. As discussed above, prehabilitation seeks to improve outcomes by optimizing or maintaining patients' cardiorespiratory fitness throughout the perioperative period including while undergoing neoadjuvant or adjuvant treatment.

The ***goals of the perioperative team*** are to minimize early morbidity and complications in order to improve survival and patient-reported outcome measures (PROMs) such as health-related quality of life and functional status. An additional priority is recovery from surgery to continue adjuvant oncological treatment, sometimes described as **r**eturn to **i**ntended **o**ncologic **t**reatment (RIOT).

Complications from cancer depend on the tumour type and location but may include anaemia, pleural and pericardial effusions, ascites, and VTEs. Cancer therapies are also associated with complications including bone marrow suppression, neuropathies, fatigue, and renal impairment. Patients from distant geographical areas may be referred to specialist centres for cancer surgery requiring a high standard of communication and clinical data-sharing to ensure adequate and timely preoperative assessment.

Obstetrics

Performing surgery on obstetric patients, whether it is for obstetric or non-obstetric reasons, involves care of not just one but two patients: mother and fetus. This situation presents unique perioperative considerations and implications to ensure the safety and well-being of all involved. In addition, maternal physiology undergoes adaptations across most organ systems from the first trimester, driven by a combination of progesterone release from the placenta and the increased metabolic demands associated with pregnancy (see Chapter 48). Such changes are vital adaptations to accommodate the growing fetus and prepare for delivery but can make perioperative care more challenging. Later in pregnancy, as the gravid uterus increases in size and weight, it exerts mechanical effects causing further potential challenges to safely performing surgery.

Cardiac output and blood volume both increase by approximately 30% during pregnancy. This can affect drug distribution and metabolism, potentially altering the effects of anaesthesia agents. The enlarging uterus pushes the diaphragm upward, restricting ventilatory capacity and leading to a higher risk of rapid desaturation after induction of GA. Delayed gastric emptying and increased intra-abdominal pressure during pregnancy can lead to a higher risk of aspiration of gastric contents during induction of anaesthesia. As the uterus enlarges, it can compress the inferior vena cava when the patient is in the supine position, potentially causing hypotension. Left lateral tilt or wedge positioning is often used to alleviate this. Finally, weight gain and increased breast size can make airway management more difficult, alongside affecting positioning of the patient for regional anaesthesia techniques like epidurals and spinals.

Surgery in obstetric patients involves close collaboration between surgical teams, obstetricians, anaesthetists, midwives, and neonatologists to ensure comprehensive care that addresses both the surgical and obstetric aspects. Timing of surgery is crucial. Whenever possible, non-emergency surgeries should be delayed until after delivery, especially during the first trimester when the fetus is most vulnerable. However, sometimes surgical interventions are necessary during pregnancy, such as for appendicitis or acute gynaecological conditions.

Anaesthesia choices should prioritize the safety of the mother and fetus. Regional anaesthesia (epidural or spinal) is often preferred over GA to avoid potential risks to the fetus.

Postoperatively, it is important to consider how medications used during surgery may affect breastfeeding. Consultation with a lactation specialist or obstetrician is often necessary to guide postoperative breastfeeding practices. Recognition of the potential emotional stress that pregnant patients may experience when undergoing surgery is also important. Providing psychological support and clear communication about the risks and benefits of the procedure is essential.

Trauma

Major trauma can be caused by accident, violence, or even a fall from standing in vulnerable patients. Perioperative care includes initial assessment and resuscitation (commonly in the emergency department), immediate surgery, secondary surgery, and rehabilitation.

The 'ABCDE' approach used in critically ill patients elsewhere in medicine is adapted to trauma, recognizing the physiological consequences of different types of traumatic injury:

- The ***airway*** is assessed while also ***protecting the cervical spine***, which may have been injured during the incident that led to the trauma.
- ***Breathing*** may be compromised by massive haemothorax or tension pneumothorax, requiring urgent chest drain insertion.
- ***Circulatory compromise*** from haemorrhagic shock or cardiac tamponade may occur, and ***haemorrhage control*** is undertaken before proceeding to further evaluation.
- ***Disability*** may occur from traumatic brain or spine injury.
- ***Exposure and environmental control*** is carried out to exclude other injuries not immediately obvious on initial assessment, and steps are taken to warm the patient and prevent hypothermia.

The metabolic response to injury is aggravated in major trauma by what has been termed the 'triad of death': ***hypothermia***, ***acidosis***, and ***coagulopathy***. Together, these can have devastating consequences for a rapidly exsanguinating, severely injured patient. While massive blood transfusion is often required, emergency surgery to

achieve control of a catastrophic bleed and '***turn off the tap***' is an immediate and life-saving intervention. A degree of hypotension can be permitted until surgical control can be achieved, in order to limit the loss of blood. Excessive crystalloid resuscitation must be avoided in the severely injured patient as this may exacerbate coagulopathy; instead, use of antifibrinolytics such as tranexamic acid, transfusion of packed red cells, fresh frozen plasma, and platelets in 1:1:1 ratios form part of major transfusion protocols in trauma.

Care has been revolutionized by a focus on early transfusion in the management of massive haemorrhage in trauma, in combination with damage control surgery. The availability of less invasive but equally effective interventional radiology angio-embolization techniques has facilitated a more conservative approach to blunt trauma in particular. The outcomes for severely injured patients continue to improve.

The key principles of major trauma are based around ***damage control***—operating early to avert life-threatening complications, leaving definitive surgical management until the patient is more stable.

Questions for this chapter can be found on Elsevier eBooks+.

15 Toxicology

Mark Pucci and Sally Marie Bradberry

Clinical skills 353
History 353
Examination 353
Risk assessment 354
Resources 354
Background 355
Approach to the poisoned patient 355
Investigations 355
Principles of management of poisoning 356
Care of the unconscious patient 356
Body 'packers' and body 'stuffers' 357
Specific management of the poisoned patient 358
Specific poisons 359
The top ten 359
Other drugs of importance in poisoning 363
Poisoning with chemicals 366
Poisons in the natural world 370

CORE SKILLS AND KNOWLEDGE

Poisoning is a significant global health problem most commonly seen in the context of deliberate self-harm but with a substantial proportion of inadvertent poisoning globally. Patients who have taken a drug overdose, a substance of abuse, or who have been accidentally poisoned will usually present to a hospital emergency department. Fortunately most can be managed with supportive care alone, although a few require critical care admission. Poisoning with rare agents may require the expertise of clinical toxicologists in a specialist poisons unit.

Key skills related to toxicology include:

- Understanding common feature clusters in acute poisoning
- Understanding treatment algorithms for patients poisoned by one of the 'top ten poisons' described in this chapter.
- Understanding features and management of poisoning with other important drugs, toxins, and chemicals.

The initial assessment and management of patients who have been poisoned, either intentionally or accidentally, is best observed in the emergency department. Subsequent management may be observed in the intensive care setting or on medical wards.

CLINICAL SKILLS

History

More than 80% of adults are conscious on arrival at hospital and the diagnosis of self-poisoning can usually be made from the history (Box 15.1). Acute poisoning should be considered in the differential diagnosis of any patient presenting with impaired consciousness.

Box 15.1 History-taking in poisoning (record in notes)

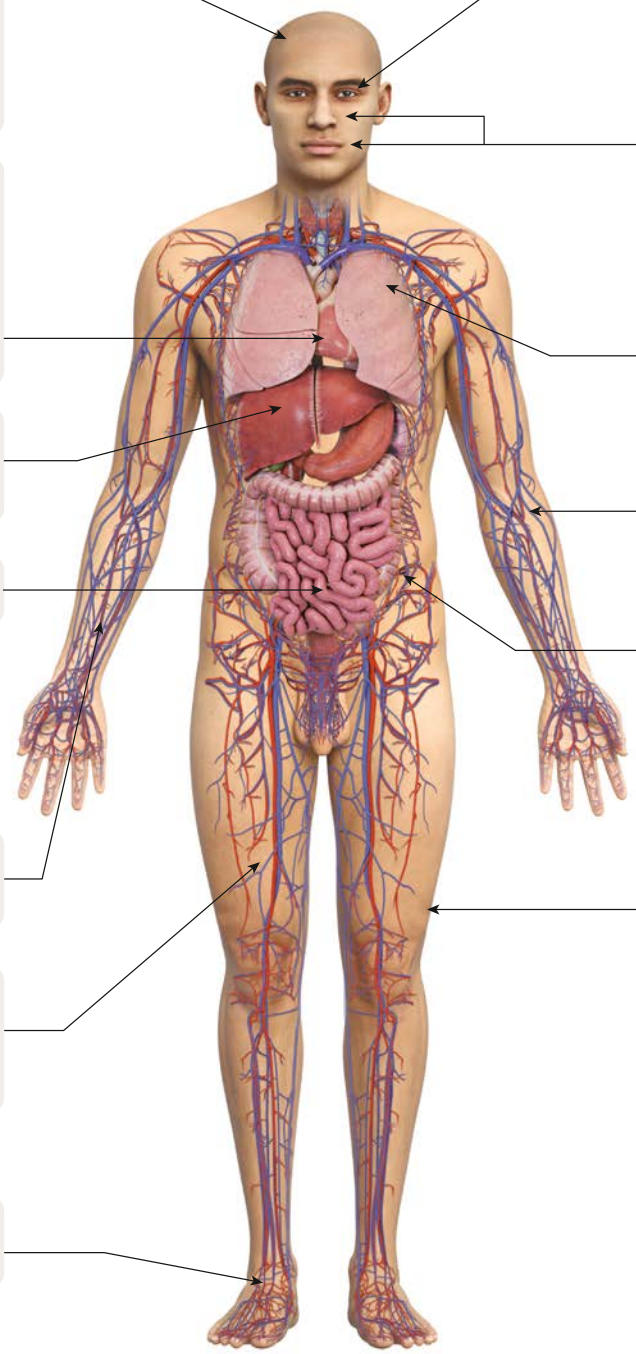
- Obtain history (if possible) from patient, relatives, friends, paramedics, witnesses
- Seek evidence of overdose
- Establish whether suicide note was left
- Ascertain what drugs/poisons were taken
- Establish time/route taken
- Ask about additional drugs, e.g., alcohol
- Seek details from general practitioner, e.g., prescribed medicines
- Assess suicide risk
- Assess capacity to make decisions
- Record a general history—past history, allergies, family history, social history

Examination

Box 15.2 provides a quick examination guide. The 'cluster of features' on presentation may be distinctive and diagnostic; for example, sinus tachycardia, fixed dilated pupils, exaggerated tendon reflexes, extensor plantar responses, and coma suggest tricyclic antidepressant poisoning (Box 15.3).

Box 15.2 Quick examination guide for the poisoned patient

- A B C D E (see Chapter 11)
- Level of consciousness (Glasgow Coma Scale; see Chapter 11)
- Ventilation—pulse oximetry
- Blood pressure and pulse rate
- Pupil size and reaction to light
- Temperature
- Head injury complicating poisonings
- If patient is unconscious, check cough and gag reflex



Brain
Coma – benzodiazepines, opioids, anticonvulsants, tricyclic antidepressants
Agitation – tricyclic antidepressants, synthetic cannabinoids, stimulants such as amfetamines and cathinones;
Seizures – tramadol, mefenamic acid, SSRIs, anticonvulsants, synthetic cannabinoids

Heart
Tachycardia – stimulants such as amfetamines, cathinones, cocaine, synthetic cannabinoids
Bradycardia – beta-blockers, calcium channel blockers, digoxin
QT prolongation – citalopram, amisulpride

Liver
Acute liver failure – paracetamol, mushroom poisoning (e.g. amanita phalloides)

Abdomen
Vomiting – ethanol, opioids

Forearm
Linear scars suggest previous deliberate self-harm

Thigh
Rhabdomyolysis – stimulants such as amfetamines, cathinones, cocaine, synthetic cannabinoids; long lie during period of coma

Ankle
Clonus – serotonin syndrome (e.g. SSRI + ecstasy)

Eyes
Miosis – opioids, organophosphorus and carbamate insecticides, nerve agents
Mydriasis – tricyclic antidepressants, ecstasy and other amfetamines, cathinones, cocaine

Mouth/Nose
Hypersalivation, rhinorrhea, bronchorrhoea – organophosphorus and carbamate insecticides, nerve agents

Lungs
Reduced respiratory rate – opioids
Hyperventilation – salicylates
Tachypnoea – may indicate aspiration pneumonitis

Antecubital fossa
'Track marks' suggest previous intravenous drug use (e.g. heroin, 'crack' cocaine)

Kidneys
Acute renal failure – ethylene glycol

Skin
Hyperthermia – stimulants such as ecstasy and other amfetamine derivatives, dinitrophenol (DNP); serotonin syndrome (e.g. SSRI + ecstasy)

Risk assessment

Risk assessment should be carried out following initial treatment and stabilization of the patient. The clinical course may be predictable based on the substance involved, the dose (e.g., in mg/kg), the timing, the clinical features observed, and the results of initial investigations. This can help guide decisions on the optimal location for on-going management of the patient.

Resources

If specialist advice is required for the management of a poisoned patient, regional or national poison centres should be the first port of call. In the UK, the National Poisons Information Service operates a 24-hour telephone advice service, Internet database, and app (Toxbase®). Toxbase® is continually updated and contains information on over 20000 products, together with generic advice on the management of poisoning.

Box 15.3 Common feature clusters in acute poisoning

Feature clusters	Poisons
Coma, hypertonia, hyper-reflexia, extensor plantar responses, myoclonus, strabismus, mydriasis, sinus tachycardia	Tricyclic antidepressants; less commonly, antihistamines, particularly diphenhydramine
Coma, hypotonia, hyporeflexia, plantar responses flexor or non-elicitable, hypotension	Benzodiazepines, with or without alcohol
Coma, miosis, reduced respiratory rate	Opioid analgesics
Nausea, vomiting, tinnitus, deafness, sweating, hyperventilation, vasodilation, tachycardia	Salicylates
Hyperthermia, tachycardia, delirium, agitation, mydriasis	Ecstasy (MDMA) or other amfetamines, cathinones, cocaine
Miosis, hypersalivation, rhinorrhoea, bronchorrhoea	Organophosphorus and carbamate insecticides, nerve agents

Box 15.4 Prevention of self-poisoning

Patients usually take what is readily available at home:

- Only small amounts of drugs should be sold in chemists/shops
- Foil-wrapped drugs are less likely to be taken in overdose
- Drugs should always be kept in a safe place
- Drugs and liquids should be kept in their original containers
- Child-resistant drug containers should be used
- Care should be taken in prescribing all drugs
- Prescriptions for any susceptible patient (e.g., the depressed) must be monitored
- Household products should be labelled and kept safely away from children

Box 15.5 Investigations that may be helpful in poisoned patients

- Serum creatinine and electrolytes, estimated glomerular filtration rate
- Creatine kinase activity
- Acid–base assessment (arterial gas, unless no concern regarding ventilation)
- Blood for specific poisons (see later)
- Electrocardiography (see Chapter 17)
- Urine saved (plain tube) for possible drug screen
- Radiology (see Chapter 8)

BACKGROUND

In high-income countries, poisoning is responsible for approximately 10% of acute hospital medical presentations and 1% of admissions. Most cases involve self-administration of prescribed or over-the-counter medicines and/or illicit drugs.

- ***Poisoning in children*** aged <6 months is usually iatrogenic. Children between 8 months and 5 years of age may ingest poisons accidentally (due to inappropriate storage of prescribed or illicit medicines intended for use by an older family member/carer). Drugs may also be administered deliberately to cause harm (abuse by proxy).
- ***Self-poisoning in adults*** is commonly a form of care-eliciting behaviour without fatal intent, particularly among females <35 years old who are otherwise in good physical health. The overdose is taken in circumstances where they are likely to be found, often in the presence of others. In the over-55s, there is a preponderance of males who are suffering from mental and/or physical ill health.
- ***Occupational poisoning*** due to dermal or inhalational exposure to chemicals is more common in the developing world.
- Poisoning may be ***iatrogenic*** and common in drugs with a narrow therapeutic window, for example, lithium, digoxin, theophylline, or phenytoin.
- The ***type of agent*** taken in overdose is influenced by availability and culture. In the UK, paracetamol poisoning remains the most common drug taken, whereas in Sri Lanka, for example, the agents ingested are more often pesticides (e.g., organophosphorus [OP] insecticides) or plants (e.g., oleander). In addition, ingestion of heating fuels, antimalarials, anti-tuberculous drugs, and traditional medicines is common in the developing world. Poisoning from snake venoms is also a problem in rural areas.

Most patients taking an overdose reach for a drug, or a combination of drugs, that is easily available. To this end, **Box 15.4** lists some significant aspects of poisoning prevention. Acute overdoses often involve more than one agent, particularly alcohol. There is often a poor correlation between the patient's history (which itself may change over time) and subsequent analytical findings.

The majority of patients do not require intensive medical management, but all require a sympathetic and caring approach, with a psychiatric and social assessment. Fatalities in the UK are due predominantly to antidepressants, paracetamol, or analgesic combinations containing paracetamol and an opioid, heroin, methadone, or cocaine. In North America there has been a substantial increase in recent years in fatal poisonings from prescription and illicit opioids (including fentanyl).

APPROACH TO THE POISONED PATIENT

The approach to history taking and examination of the poisoned patient are summarized in Boxes 15.1 and 15.2.

Investigations

Box 15.5 (and see also **Box 15.7**) sets out investigations that may be helpful in poisoned patients. On admission, or at an appropriate time post overdose, a timed blood sample should be taken if it is suspected that aspirin, digoxin, ethylene glycol, iron, lithium, methanol, paracetamol, paraquat, quinine, or theophylline has been ingested (**Box 15.6**). The determination of the concentrations of these drugs will be valuable in management. Drug screens on blood and urine are occasionally indicated in severely poisoned patients in whom the cause of coma is unknown. A poison information service will advise.

Some routine investigations (**Box 15.7**) are of value in the differential diagnosis of coma or the detection of poison-induced hypokalaemia, hyperkalaemia, hypoglycaemia, hyperglycaemia, hepatic or renal failure, or metabolic acidosis (**Box 15.8**). Measurement of carboxyhaemoglobin, methaemoglobin, and cholinesterase activities are of assistance in the diagnosis and management of cases

Box 15.6 Laboratory analysis for toxins

- Carbon monoxide (carboxyhaemoglobin) and methaemoglobin concentrations are available on most blood gas analysers
- Digoxin
- Ethanol (when monitoring treatment in glycol and methanol poisoning)
- Ethylene glycol
- Iron
- Lithium
- Methanol
- Paracetamol
- Salicylate
- Theophylline

Box 15.7 Relevant non-toxicological investigations with examples

Investigation	Situation
Serum sodium concentration	Hyponatraemia in ecstasy (MDMA) poisoning
Serum potassium concentration	Hypokalaemia in theophylline poisoning Hyperkalaemia in digoxin poisoning
Serum creatinine concentration	Acute kidney injury in ethylene and diethylene glycol poisoning
Acid–base disturbances	Including metabolic acidosis (see Box 15.8)
Blood glucose concentration	Hypoglycaemia in insulin poisoning Hyperglycaemia in salicylate poisoning
Serum calcium concentration	Hypocalcaemia in ethylene glycol poisoning
Liver function (prothrombin time)	In paracetamol poisoning
Carboxyhaemoglobin concentration	In carbon monoxide poisoning
Methaemoglobin concentration	In nitrite poisoning
Cholinesterase activity	Organophosphorus insecticide and nerve agent poisoning
ECG	Wide QRS in tricyclic antidepressant poisoning
X-ray	Detection of radio-opaque toxins, e.g., elemental mercury

Box 15.8 Some poisons that induce metabolic acidosis

- Carbon monoxide
- Cyanide
- Ethanol
- Ethylene glycol
- Iron
- Metformin
- Methanol
- Paracetamol
- Salicylates
- Tricyclic antidepressants

of poisoning due to carbon monoxide, methaemoglobin-inducing agents, such as nitrites, and organophosphorus insecticides, respectively.

ECG

Routine ECG is of limited diagnostic value, but continuous ECG monitoring should be undertaken in those ingesting potentially cardiotoxic drugs. Sinus tachycardia with prolongation of the PR and QRS intervals in an unconscious patient suggests tricyclic antidepressant overdose. QT interval prolongation is an adverse effect of several drugs (e.g., citalopram).

Radiology

Routine radiology is of little diagnostic value. It can confirm ingestion of metallic objects (e.g., coins, button batteries) or injection of globules of metallic mercury. Some enteric-coated or sustained-release drugs or ingested packets of illicit substances may be seen on plain abdominal X-rays. Radiology can confirm complications of poisoning, such as aspiration pneumonia, non-cardiogenic pulmonary oedema (salicylates), and acute respiratory distress syndrome (ARDS).

PRINCIPLES OF MANAGEMENT OF POISONING

Most people with self-poisoning require only general care and support of the vital systems (**Box 15.9**). However, for a few poisons, additional therapy is required.

Care of the unconscious patient

(See also Chapter 11.) In all cases, the initial priority is to ensure that basic resuscitative measures are instituted promptly, establishing a clear airway and supporting breathing as necessary. Remember that pulse oximetry alone will not detect hypercapnia. Loss of the cough or gag reflex is the prime indication for intubation. In many severely poisoned patients, the reflexes are depressed sufficiently to allow intubation without the use of sedatives or relaxants.

Immediate catheterization of the bladder is not usually required. Insertion of a venous cannula is usual, but administration of intravenous fluids is unnecessary unless the patient has been unconscious for more than 12 hours, is dehydrated, or is hypotensive.

Box 15.9 Management strategy in acute poisoning

Immediate decisions

- Supportive treatment
- Is an antidote appropriate? (see **Box 15.10**)
- Is it appropriate to try to reduce poison absorption?
- Is it appropriate to perform toxicological investigations?
- Is it appropriate to try to enhance elimination?

Practical points—contact with poison

Eyes

- Remove contact lenses
- Wash eyes with 0.9% saline or water for about 15 min
- Assess corneal damage (if necessary) with slit lamp and fluorescein stain

Skin

- Remove clothing
- Wash thoroughly with soap and water if, for example, chemical exposure

Cardiovascular support

Severe hypotension and shock are unusual in the acutely poisoned patient but may be caused by:

- ***a direct cardio-depressant action*** of the poison (e.g., beta-blockers, calcium-channel blockers, tricyclic antidepressants)
- ***vasodilation and venous pooling*** in the lower limbs (e.g., ACE inhibitors, phenothiazines)
- ***a decrease in circulating blood volume*** because of gastrointestinal losses (e.g., profuse vomiting), increased sweating, increased renal losses, and/or increased capillary permeability.

Hypotension may be exacerbated by coexisting hypoxia, acidosis, and dysrhythmias. Volume expansion with intravenous crystalloids is usually sufficient to restore blood pressure. Volume replacement and the use of inotropes are discussed in Chapter 13. All patients with cardiogenic shock should have ECG monitoring.

Systemic hypertension may occur in poisoning with drugs such as amfetamines and cocaine. If this is mild and associated with agitation, treatment with a benzodiazepine may suffice. In more severe cases there may be a risk of arterial rupture, particularly intracranial. To prevent this, intravenous nitrates such as glyceryl trinitrate can be given, starting at a dose of 1 to 2 mg/h, gradually increasing the dose to a maximum of 12 mg/h until blood pressure is controlled. Calcium antagonists, such as verapamil 240 to 480 mg daily in divided doses, are an alternative second-line therapy. Sodium nitroprusside 0.5 to 1.5 μg/kg/min (to a maximum of 8 μg/kg/min) is an option for patients with increased blood pressure when there is no evidence of cardiac ischaemia but caution is required as it may cause a rapid fall in blood pressure.

Arrhythmias can occur, such as tachyarrhythmias following ingestion of a tricyclic antidepressant or bradyarrhythmias with digoxin poisoning. Known arrhythmogenic factors, such as hypoxia, acidosis, and hypokalaemia, should be corrected.

Other problems

Hypothermia

A rectal temperature below 35°C is a recognized complication of poisoning, especially in older patients or in those who are comatose. The patient should be covered with a 'space blanket' and, if necessary, given intravenous fluids at normal body temperature. The administration of heated (37°C), humidified oxygen delivered by face mask is also useful.

Hyperthermia

Body temperature can, rarely, increase to being potentially fatal after poisoning with central nervous stimulants, such as cocaine, and amfetamines, including 'ecstasy' (MDMA) and the weight-loss/body-building agent dinitrophenol. Muscle tone is often increased and convulsions and rhabdomyolysis are common. Cooling measures, sedation with diazepam, and, in severe cases, intravenous dantrolene (a skeletal muscle relaxant) 2.5 mg/kg body weight should be given.

Skin blisters

Skin blisters may be found in poisoned patients who are, or have been, unconscious and do not move. Such lesions are not diagnostic of specific poisons but are sufficiently common in poisoned patients (and sufficiently uncommon in patients unconscious from other causes) to be of diagnostic value.

Rhabdomyolysis

Rhabdomyolysis can occur from pressure necrosis in drug-induced coma, or it may complicate poisoning in the absence of coma, such as in ecstasy (MDMA) abuse. People with rhabdomyolysis are at risk of developing acute kidney injury from myoglobinuria, particularly if they are hypovolaemic and have a metabolic acidosis. Severe rhabdomyolysis may lead to the development of a compartment syndrome (see Chapter 32), which can cause serious tissue damage if not treated promptly.

Convulsions

These may occur in poisoning from many drugs, including tricyclic antidepressants, mefenamic acid, and opioids such as tramadol. Usually, seizures are short-lived but, if they are prolonged, intravenous diazepam 10 to 20 mg or lorazepam 4 mg should be administered. Persistent fits must be controlled rapidly to prevent severe hypoxia, brain damage, and laryngeal trauma. If diazepam or lorazepam in repeated dose is ineffective, second-line treatments include intravenous levetiracetam (loading dose 60 mg/kg, maximum dose 4500 mg over 10 minutes) or intravenous phenobarbital sodium (10 mg/kg at not more than 100 mg/min). Phenytoin is contraindicated in cases of poisoning with sodium channel-blocking drugs (such as tricyclic antidepressants) and is relatively contraindicated in all cases of poisoning with cardiotoxic drugs.

Stress ulceration and bleeding

An intravenous proton pump inhibitor should be given to prevent stress ulceration of the stomach in the elderly and patients with a previous history of ulceration.

Body 'packers' and body 'stuffers'

Body 'packers' (sometimes called 'mules' or 'swallowers') are those who swallow a substantial number of packages containing illicit drugs as a means of smuggling. Heroin used to be the drug of choice but this has been superseded by cocaine. Although each package contains a potentially lethal amount of drug, packets are now usually machine manufactured using a material that does not leak. Body packers may ingest up to 100 to 200 packages.

Body 'stuffers' swallow a small number of packages containing an illicit drug, usually heroin, cocaine, cannabis, or an amfetamine, in an unplanned attempt to conceal evidence when on the verge of being arrested. These drugs are usually either unpackaged or poorly packaged and, as a consequence, leakage may occur over the ensuing 3 to 6 hours and cause significant symptoms. Some also hide illicit drug packages in their rectum or vagina with the same intent (these are sometimes known as ***body 'pushers'***).

The ***role of imaging*** is usually confined to body packers; imaging has little role in the care of body stuffers or pushers. Low-dose computed tomography (CT) is now considered the gold standard imaging investigation of choice for locating body packages. A urine screen for drugs of misuse should be performed. A screen that is positive for one or more drugs of misuse suggests that either the patient has used the drug in the previous few days or at least one packet is leaking. A negative screen strongly suggests that no packet is leaking. Screens should be repeated daily, or immediately if the patient develops features of intoxication, to confirm the diagnosis.

Management

Packages can be removed most expeditiously in body stuffers by employing whole-bowel irrigation (see 'Gut decontamination' section). Indications for urgent surgery in body packers include evidence of acute intestinal obstruction, radiological confirmation of many packets, and clinical or analytical evidence to suggest package leakage, particularly if the drug involved is cocaine.

Packets in the vagina can usually be removed manually.

Specific management of the poisoned patient

Antidotes

Specific antidotes are available for a small number of poisons only (Box 15.10).

Antidotes may exert a beneficial effect by:

- ***forming an inert complex*** with the poison (e.g., desferrioxamine, dicobalt edetate, digoxin-specific antibody fragments, dimercaprol, HI-6, hydroxocobalamin, obidoxime, pralidoxime, sodium calcium edetate, succimer (dimercaptosuccinic acid; [DMSA]), unithiol (dimercaptopropanesulphonate; [DMPS])
- ***accelerating detoxification*** of the poison (e.g., acetylcysteine, sodium thiosulphate)
- ***reducing the rate of conversion*** of the poison to a more toxic compound (e.g., ethanol, fomepizole)
- ***competing with the poison*** for essential receptor sites (e.g., oxygen, naloxone, phytomenadione)
- ***blocking essential receptors*** through which the toxic effects are mediated (e.g., atropine)
- ***bypassing the effect*** of the poison (e.g., oxygen, glucagon).

Gut decontamination

Although it seems logical to assume that removal of unabsorbed drug from the gastrointestinal tract will be beneficial (gut decontamination), the efficacy of gastric lavage and syrup of ipecacuanha (to induce vomiting) remains unproven and efforts to remove small amounts of non-toxic drugs are clinically not worthwhile or appropriate.

Gastric lavage should only be performed if a patient has ingested a potentially life-threatening amount of a poison, for example, iron, and the procedure can be undertaken within 60 minutes of ingestion. Ensure the airway is protected. Lavage is contraindicated if a hydrocarbon with high aspiration potential or a corrosive substance has been ingested.

Syrup of ipecacuanha to induce vomiting should not be used, as the amount of drug recovered is highly variable and diminishes with time; there is no evidence that it improves outcome.

Single-dose activated charcoal can adsorb a wide variety of compounds. Exceptions are strong acids and alkalis, ethanol, ethylene glycol, iron, lithium, mercury, and methanol. In studies in volunteers given 50 g activated charcoal, the mean reduction in absorption was 40%, 16%, and 21%, at 60, 120, and 180 min, respectively, after ingestion. Administration of oral activated charcoal (50 g for an adult, 1 g/kg for a child) should be considered following ingestion of a potentially toxic amount of a poison (except those listed above) if this is possible within 1 hour of ingestion. There are insufficient data to support or exclude its use after 1 hour. There is no evidence that administration of activated charcoal improves clinical outcome.

Cathartics have no role in the management of the poisoned patient.

Whole-bowel irrigation requires the insertion of a nasogastric tube into the stomach and the introduction of polyethylene glycol electrolyte solution 1500 to 2000 mL/h in an adult, which is continued until the rectal effluent is clear. Whole-bowel irrigation may be used for potentially toxic ingestions of sustained-release or enteric-coated drugs or to remove illicit drug packets.

Box 15.10 Antidotes of value in poisoning

Poison	Antidotes
Aluminium	Desferrioxamine
Arsenic	Succimer (DMSA), dimercaprol
Benzodiazepines	Flumazenil
β-Adrenoceptor-blocking drugs	Atropine, glucagon
Calcium-channel blockers	Atropine
Carbamate insecticides	Atropine
Carbon monoxide	Oxygen
Cyanide	Oxygen, hydroxocobalamin, sodium thiosulphate
Digoxin and digitoxin	Digoxin-specific antibody fragments
Ethylene glycol	Fomepizole, ethanol
Hydrogen sulphide	Oxygen
Iron salts	Desferrioxamine
Lead (inorganic)	Succimer (DMSA), sodium calcium edetate
Methaemoglobinaemia	Methylthioninium chloride (methylene blue)
Methanol	Fomepizole, ethanol
Mercury (inorganic)	Unithiol (DMPS)
Nerve agents	Atropine, HI-6, obidoxime, pralidoxime
Opioids	Naloxone
Organophosphorus insecticides	Atropine, HI-6, obidoxime, pralidoxime
Paracetamol	Acetylcysteine
Warfarin and similar anticoagulants	Phytomenadione (vitamin K)

Increasing poison elimination

Multiple-dose activated charcoal (MDAC) involves the repeated administration of oral activated charcoal to increase elimination of a drug that has already been absorbed. Most absorbed drugs re-enter the gut from the villi capillaries by passive diffusion (a few by an active process) if the concentration in the gut is lower than that in the blood (this is called the entero-enteric circulation). Some drugs are also secreted in the bile (enterohepatic circulation). Activated charcoal will bind any drug that is in the gut lumen, so keeping the concentration in the gut low and favouring further diffusion from blood to gut lumen.

Elimination of drugs with a small volume of distribution (<1 L/kg), low p*K*a (which maximizes transport across membranes), low binding affinity, and prolonged elimination half-life following overdose is particularly likely to be enhanced by MDAC. MDAC also improves total body clearance of the drug when endogenous processes are compromised by liver and/or renal failure.

Although MDAC has been shown to increase elimination significantly for a small number of drugs (e.g., carbamazepine, dapsone, phenobarbital, quinine, theophylline), it has not been shown to reduce morbidity and mortality in controlled studies.

In *adults*, charcoal should be administered in an initial dose of 50 to 100 g and then at a rate of not less than 12.5 g/h, preferably via a nasogastric tube; usually, 200 g is sufficient. Intravenous ondansetron 4 to 8 mg is an effective antiemetic and facilitates MDAC administration if vomiting is a problem.

Urine alkalinization enhances elimination of salicylate, phenobarbital, and chlorophenoxy herbicides (e.g., 2,4-dichlorophenoxyacetic acid) by mechanisms that are not clearly

understood. Urine alkalinization is a metabolically invasive procedure requiring frequent biochemical monitoring and medical and nursing expertise. Advice should be sought from those familiar with the procedure before urine alkalinization is commenced. Plasma volume depletion, electrolytes (notably plasma potassium since sodium bicarbonate exacerbates pre-existing hypokalaemia), and metabolic abnormalities should be corrected first. Sufficient bicarbonate is then administered to ensure that urine pH is over 7.5 and preferably close to 8.5. In one study, sodium bicarbonate 225 mmol was the mean amount required initially. This can be administered intravenously as 225 mL of an 8.4% solution (1 mmol bicarbonate/mL) over 1 hour but great care must be taken to avoid extravasation if given via a peripheral line.

Haemodialysis and haemodiafiltration are of little value in patients poisoned with drugs that have large volumes of distribution (e.g., tricyclic antidepressants) because the plasma contains only a small proportion of the total amount of drug in the body. These methods are indicated in patients with severe clinical features and high plasma concentrations of ethanol, ethylene glycol, isopropanol, lithium, methanol, and salicylate: that is, drugs with small volumes of distribution.

Lipid emulsion therapy involves the intravenous administration of 20% Intralipid (fractionated soya oil) 1.5 mL/kg, followed by 0.25 to 0.5 mL/kg per minute for 30 to 60 minutes to an initial maximum of 500 mL. Use of lipid emulsion as an antidote for severe poisoning with highly lipid-soluble drugs is based on experience of its efficacy in local anaesthetic poisoning. It is believed to increase the intravascular lipid phase, so creating a 'lipid sink' to reduce the amount of active drug in plasma. Other proposed mechanisms of benefit include increasing myocardial fatty acid energy substrate and improving the function of cell membrane-bound ion channels (particularly Ca^{2+} and Na^{+}). The role of Intralipid in the management of poisoning with highly lipid-soluble drugs is not well defined but case reports suggest some benefit in life-threatening cardiotoxicity that is resistant to other measures.

SPECIFIC POISONS

- All patients must be reviewed by senior medical staff.
- All patients should have relevant blood tests and monitoring when appropriate.
- The general principles of management of self-poisoning outlined earlier will always need to be applied.

In this section, only the specific treatment regimens are outlined. The first section covers the 'top ten' drug types encountered in poisoning: paracetamol, non-steroidal anti-inflammatory drugs (NSAIDs), opiates/opioids, antidepressants, benzodiazepines, stimulants (including cocaine), ethanol, neuroleptics, cannabis/synthetic cannabinoids, and anticonvulsants. All doctors working in an acute setting should be familiar with the clinical presentation and principles of management of poisoning with these agents.

The top ten

Paracetamol (acetaminophen)

In ***therapeutic doses***, paracetamol is conjugated with glucuronide and sulphate. A small amount of paracetamol is metabolized by mixed-function oxidase enzymes to form a highly reactive toxic compound (*N*-acetyl-*p*-benzoquinoneimine; NAPQI), which is then immediately conjugated with glutathione and subsequently excreted as cysteine and mercapturic conjugates.

In ***overdose***, large amounts of paracetamol are metabolized by oxidation because of saturation of the sulphate conjugation pathway. Liver glutathione stores become depleted so that the liver is unable to deactivate NAPQI. Paracetamol-induced kidney injury probably results from a mechanism similar to that responsible for hepatotoxicity.

Clinical features

Following the ingestion of an overdose of paracetamol, patients usually remain asymptomatic for the first 24 hours or, at the most, develop anorexia, nausea, and vomiting. Liver damage is not usually detectable by routine liver function tests until at least 18 hours after ingestion of the drug. Liver damage usually reaches a peak, as assessed by measurement of alanine aminotransferase (ALT) activity and prothrombin time (or INR), at 72 to 96 hours after ingestion. Without treatment, a small percentage of patients will develop fulminant hepatic failure. Acute kidney injury due to acute tubular necrosis occurs in 25% of patients who have severe hepatic damage and in a few without evidence of serious disturbance of liver function.

Management

Acetylcysteine is an effective protective agent, provided that it is administered within 8 to 10 hours of ingestion of the overdose. It acts by replenishing cellular glutathione stores, though it may also repair oxidation damage caused by NAPQI.

There are now two main approaches to treatment worldwide. In the UK, following a decision by the Medicines and Healthcare products Regulatory Agency in 2012 to abandon a detailed risk assessment in the decision to treat, it is deemed that patients with concentrations above a 'treatment line' starting at 100 mg/L 4 hours after ingestion (**Fig. 15.1**) should be given treatment with acetylcysteine. In most other countries, a parallel line starting at 150 mg/L 4 hours after ingestion is used. The treatment lines are uncertain if the patient presents 15 hours or more after ingestion or has taken a modified-release preparation of paracetamol. Although these lines are often extended to 24 hours (dotted lines), the concentrations are not based on clinical trial data.

A summary of single acute paracetamol overdose management at different times post ingestion is given in Box 15.12. Note that patients whose paracetamol overdose is staggered over more than 1 hour or who take repeated therapeutic excess are at greater risk of liver damage, and decisions to treat are generally based on the dose ingested/24 hours. In the UK, treatment with acetylcysteine is considered if more than 75 mg/kg has been ingested in 24 hours.

Up to 15% of patients treated with intravenous acetylcysteine develop a rash, angio-oedema, hypotension, and/or bronchospasm. These reactions, which are histamine-mediated and related to the initial bolus, are seldom serious, and pausing the infusion is usually all that is required. In more severe, adult cases, intravenous chlorphenamine 10 to 20 mg should be given.

The Royal College of Emergency Medicine and The National Poisons Information Service in the UK both now formally advocate use of SNAP (Scottish and Newcastle Acetylcysteine Protocol), a 12-hour, 2-bag regimen (**Box 15.11**), instead of the standard 21-hour, 3-bag regimen. SNAP has been shown to be as efficacious as the

Box 15.11 SNAP regimen for acetylcysteine

- Acetylcysteine 100 mg/kg in 200 mL 5% glucose over 2 h
- Then 200 mg/kg in 1000 mL 5% glucose over 10 h
- Total dose 300 mg/kg over 12 h

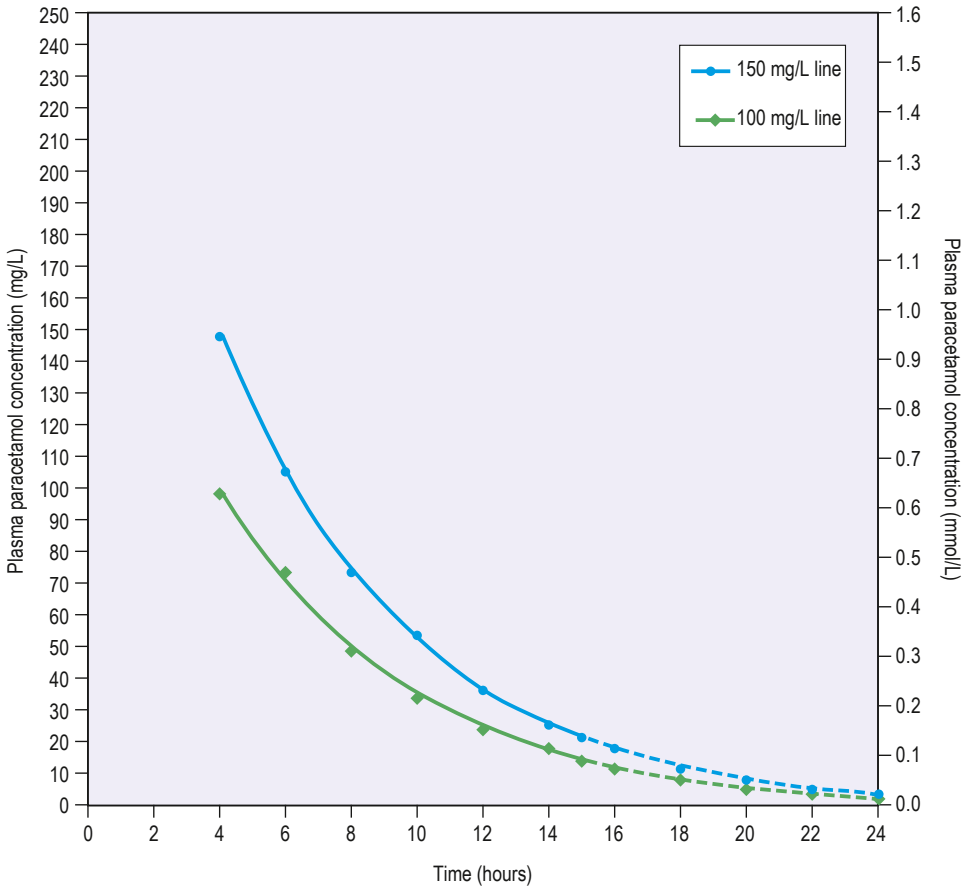


Fig. 15.1 Nomogram for treatment of paracetamol poisoning. In the UK, patients with concentrations above a 'treatment line' starting at 100 mg/L 4 hours after ingestion should be given treatment with acetylcysteine. In most other countries, a parallel line starting at 150 mg/L 4 hours after ingestion is used. Note that the treatment lines are uncertain if the patient presents 15 hours or more after ingestion or has taken a modified-release preparation of paracetamol. Although these lines are often extended to 24 hours (dotted lines), the concentrations are not based on clinical trial data.

Box 15.12 Summary of management of single acute paracetamol overdose

All patients

- Clarify date, time, formulation, and dose of paracetamol taken
- Weigh the patient
- Calculate the paracetamol dose ingested in mg/kg

Patients presenting 0–8 hours since last paracetamol ingestion

- Wait until 4 h post ingestion then take venous blood for plasma paracetamol concentration, serum creatinine and electrolytes, bicarbonate, liver function, and INR
- If the blood results will not be available until after 8 h post ingestion *AND* patient has ingested >150 mg/kg, commence acetylcysteine immediately
- When the plasma paracetamol concentration is available, assess need for antidotal treatment with acetylcysteine using the treatment nomogram (see **Fig. 15.2**)
- Administer acetylcysteine if required. If already commenced but not required, acetylcysteine can be discontinued at this stage
- Seek specialist advice if plasma paracetamol concentration is below the treatment line but liver function tests (LFTs) or clotting are abnormal
- At end of acetylcysteine course, take venous blood for serum creatinine and electrolytes, paracetamol, bicarbonate, liver function, and INR
- Consider need for further acetylcysteine if post-treatment bloods are abnormal

Patients presenting between 8 h and 24 h after paracetamol ingestion

- Take venous blood for urgent measurement of plasma paracetamol concentration, serum creatinine and electrolytes, bicarbonate, liver function, INR
- Commence acetylcysteine immediately if ingested paracetamol dose is >150 mg/kg
- If ingested dose <150 mg/kg, await blood results and assess need for acetylcysteine using treatment nomogram
- Administer acetylcysteine if required. If already commenced but not required, acetylcysteine can be discontinued at this stage
- BEWARE that for patients presenting >12 h after ingestion, the plasma paracetamol concentration at which treatment is required may be close to the laboratory's limit of detection. If borderline, treat with acetylcysteine
- Seek specialist advice if plasma paracetamol concentration is below the treatment line but LFTs or clotting are abnormal
- At end of acetylcysteine course, take venous blood for serum creatinine and electrolytes, paracetamol concentration, bicarbonate, liver function, and INR
- Consider need for further acetylcysteine if post-treatment bloods are abnormal

Patients presenting more than 24 h post overdose

- Commence acetylcysteine if the patient is jaundiced or has hepatic tenderness
- Measure the plasma paracetamol concentration (before NAC is administered if possible), serum creatinine and electrolytes, bicarbonate, LFTs, glucose, full blood count, and INR
- Treat with acetylcysteine if the ALT is above the upper limit of normal, *OR* INR is >1.3 (in the absence of another cause, e.g., warfarin) *OR* paracetamol concentration is detectable
- If the paracetamol concentration is not detectable, the INR and ALT are normal, and the patient is asymptomatic, no treatment with an antidote is indicated. If acetylcysteine has been started, it may then be discontinued
- At end of acetylcysteine course, take venous blood for serum creatinine and electrolytes, bicarbonate, liver function, and INR
- Consider need for further acetylcysteine if post treatment bloods are abnormal

21-hour regimen, has a reduced risk of anaphylactoid reactions, and is more convenient.

If liver or renal failure ensues, this should be treated conventionally, although there is evidence that a continuing infusion of acetylcysteine (continue 200 mg/kg in 1000 mL of 5% glucose infused at a rate of 100 mL/h until recovery) will improve the morbidity and mortality. Liver transplantation has been performed successfully in patients who have paracetamol-induced acute hepatic failure (see Chapter 23).

Non-steroidal anti-inflammatory drugs

Self-poisoning with NSAIDs, other than salicylates, has increased, particularly now that ibuprofen is available without prescription, over the counter, in many countries. Poisoning with salicylates is considered in Box 15.3.

Clinical features and management

In most cases, minor gastrointestinal disturbance is the only feature but in more severe cases, coma, convulsions, and acute kidney injury have occurred. Transient renal impairment is common after ibuprofen overdose. Poisoning with mefenamic acid commonly results in convulsions, though these are usually short-lived.

Treatment is symptomatic and supportive.

Opiates and opioids

Clinical features

Cardinal signs of opiate poisoning are pinpoint pupils, reduced respiratory rate, and coma. Hypothermia, hypoglycaemia, and convulsions are occasionally observed in severe cases. Non-cardiogenic pulmonary oedema has been reported in severe heroin overdose. See Chapter 50 for more information about the worldwide opioid crisis.

Management

Intravenous naloxone will reverse respiratory depression and coma, at least partially. In severe poisoning, an initial dose of 1.2 mg is likely to be required and repeat doses necessary. Lower doses (0.4–0.8 mg) may suffice in less severe cases and if precipitating opioid withdrawal is a concern. The duration of action of naloxone is often less than the drug taken in overdose, for example, methadone, which has a very long half-life. For this reason, an infusion of naloxone is often required. Non-cardiogenic pulmonary oedema should be treated with mechanical ventilation.

Antidepressants: tricyclics and selective serotonin reuptake inhibitors

Tricyclic antidepressants block the reuptake of monoamines (e.g., norepinephrine (noradrenaline) and serotonin) into peripheral and intracerebral neurones, thereby increasing the concentration of these neurotransmitters in these areas. They also have antimuscarinic actions and class 1 antiarrhythmic (quinidine-like) sodium-channel-blocking activity.

Selective serotonin reuptake inhibitors (SSRIs) (e.g., citalopram, escitalopram, fluoxetine, fluvoxamine, paroxetine, and sertraline) lack the antimuscarinic and sodium-channel-blocking actions of tricyclic antidepressants.

Clinical features

Tricyclic antidepressants. Mild poisoning commonly causes drowsiness, sinus tachycardia, dry mouth, dilated pupils, urinary retention (all antimuscarinic effects), and increased reflexes and extensor plantar responses. Severe intoxication leads to coma, convulsions, and, occasionally, divergent strabismus. Plantar, oculocephalic, and oculovestibular reflexes may be abolished temporarily. An ECG will often show a wide QRS interval (due to sodium channel blockade) and there is a reasonable correlation between the width of the QRS complex and the severity of poisoning. Life-threatening arrhythmias may ensue. Metabolic acidosis and cardiorespiratory depression are observed in severe cases.

Selective serotonin reuptake inhibitors. Even in large overdoses, SSRIs appear to be relatively safe unless potentiated by ethanol. Most patients will show no signs of toxicity but drowsiness, nausea, diarrhoea, and sinus tachycardia have been reported. Rarely, junctional bradycardia, seizures, and hypertension have been encountered and influenza-like symptoms may develop.

Serotonin syndrome occasionally occurs (see Chapter 30).

Other antidepressants. Venlafaxine is a serotonin-norepinephrine reuptake inhibitor with a similar toxicity profile to the SSRIs but with greater risk of cardiac arrhythmias and convulsions. Mirtazapine, an antidepressant that acts via presynaptic alpha blockade to increase central serotonin release, is probably the safest antidepressant in overdose, causing mainly only sedation.

Management

The majority of patients recover with supportive therapy alone (adequate oxygenation, control of convulsions, and correction of acidosis), although a small percentage who ingest a tricyclic will require assisted ventilation for 24 to 48 hours. Widening of the QRS interval, the onset of supraventricular tachycardia and ventricular tachycardia should be treated with ***intravenous sodium bicarbonate*** (8.4%) 50 mmol, even if there is no acidosis present; a second bolus may be required. The benefit of sodium bicarbonate in the setting of tricyclic antidepressant (TCA) overdose is thought to be due to both an increase in serum pH and the increase in extracellular sodium. Alkalinization increases protein binding of the drug and the high sodium load increases the electrochemical gradient across cardiac cell membranes, potentially attenuating the TCA-induced blockade of sodium channels. A serum pH of 7.5 to 7.55 should be aimed for.

Benzodiazepines

Benzodiazepines are commonly taken in overdose but rarely produce severe poisoning, except in the elderly or those with chronic respiratory disease.

Box 15.13 Clinical features of ethanol poisoning

Blood (ethanol)		Clinical features
mg/L	*mmol/L*	
500–1500	11.0–32.5	Emotional lability Mild impairment of coordination
1500–3000	32.5–65.0	Visual impairment Incoordination Slowed reaction time Slurred speech
3000–5000	65.0–108.5	Marked incoordination Blurred or double vision Stupor Occasionally, hypoglycaemia, hypothermia, and convulsions
>5000	>108.5	Depressed reflexes Respiratory depression Hypotension Hypothermia Death (from respiratory or circulatory failure, or from aspiration)

Clinical features

Benzodiazepines produce drowsiness, ataxia, dysarthria, and nystagmus. Coma and respiratory depression develop in severe intoxication.

Management

If respiratory depression is present in patients who have severe benzodiazepine poisoning, ***intravenous flumazenil*** 0.5 to 1.0 mg is given in an adult and this dose often needs repeating. Flumazenil use often avoids the need for assisted ventilation. It is relatively contraindicated in patients with mixed proconvulsant (e.g., tricyclic antidepressants)/benzodiazepine poisoning and those with a history of epilepsy because it may cause convulsions.

Stimulants

The most commonly abused stimulants are the amfetamines, cocaine, and the synthetic cathinones.

Amfetamines are central nervous system (CNS) and cardiovascular stimulants. These effects are mediated by increasing synaptic concentrations of adrenaline (epinephrine) and dopamine. The *N*-methylated derivative, metamfetamine (the crystalline form of this salt is known as 'crystal meth' or 'ice'), and 3,4-methylenedioxymetamfetamine (MDMA), known as ecstasy, are common examples.

Cocaine hydrochloride ('street' cocaine, 'coke') is a water-soluble powder or granule that can be taken orally, intravenously, or intranasally ('snorting'). 'Freebase' or 'crack' cocaine comprises crystals ('rocks') of relatively pure cocaine without the hydrochloride moiety. Crack cocaine is more suitable for smoking in a pipe or mixed with tobacco; it can also be heated on foil and the vapour inhaled (approximately 35 mg of drug per 'line' or a 'rail'). The effects of cocaine are experienced almost immediately after intravenous administration or smoking, about 10 minutes following intranasal administration, and 45 to 90 minutes after ingestion. The effects start to resolve in about 20 minutes but may last up to 90 minutes. In severe poisoning, death occurs in minutes, and death rarely occurs more than 3 hours after exposure.

Cocaine blocks the reuptake of biogenic amines:

- Inhibition of dopamine reuptake is responsible for psychomotor effects.
- Blockade of noradrenaline (norepinephrine) reuptake produces tachycardia.
- Inhibition of serotonin reuptake induces hallucinations.
- CNS arousal is enhanced by potentiating the effects of excitatory amino acids.
- Cocaine is a powerful local anaesthetic and vasoconstrictor.

Cathinone (khat) is a naturally occurring stimulant derived from the plant *Catha edulis*. The drug is released during prolonged chewing of plant leaves. An increasing number of synthetic cathinone derivatives are available, the best-known of which is mephedrone (4-methyl methcathinone). While originally sold as 'plant foods' or 'bath salts', these substances are purchased for their recreational abuse potential. Synthetic cathinones may be taken by mouth, nasal insufflation, or injection.

Clinical features

Stimulants cause dilated pupils, tachycardia and hypertension, sweating, euphoria, extrovert behaviour, and a lack of desire to eat or sleep. More severe intoxication is associated with agitation, paranoid delusions, hallucinations, and violent behaviour. Convulsions, rhabdomyolysis, hyperthermia, metabolic acidosis, and cardiac arrhythmias may develop. Rarely, dissection of the aorta, myocarditis, myocardial infarction, dilated cardiomyopathy, subarachnoid haemorrhage, or cerebral haemorrhage or infarction may occur and can be fatal. Disseminated intravascular coagulation and acute kidney injury are recognized. If a young person presents with a stroke or myocardial infarction, poisoning with a stimulant should be considered in the differential diagnosis.

Management

- Agitation is best controlled by intravenous diazepam 10 to 20 mg.
- Hypertension and tachycardia also usually respond to sedation but if hypertension persists, consider intravenous glyceryl trinitrate starting at 1 to 2 mg/h and gradually increase the dose (maximum 12 mg/h) until the blood pressure is controlled. The peripheral sympathomimetic actions of amfetamines can be antagonized by β-adrenoceptor-blocking drugs but these are rarely required. The use of beta-blockers in cocaine poisoning is controversial.
- Active external cooling should be employed for hyperthermia and, if uncontrolled, intravenous dantrolene 2.5 mg/kg body weight should be considered (see 'Hyperthermia' section).
- Early use of a benzodiazepine is often effective in relieving stimulant-associated non-cardiac chest pain. Myocardial ischaemia/infarction should be treated conventionally.

Ethanol

Ethanol is commonly ingested in beverages and deliberately with other substances in overdose. It is also present in many cosmetic and antiseptic preparations. Following absorption, ethanol is oxidized to acetaldehyde and then to acetate. Ethanol is a CNS depressant and the features of ethanol intoxication are generally related to blood concentrations (**Box 15.13**).

Clinical features

In children, in particular, severe hypoglycaemia may accompany alcohol intoxication due to inhibition of gluconeogenesis. Hypoglycaemia is also observed in those who are malnourished or who have fasted in the previous 24 hours. In severe cases of intoxication, coma and hypothermia are often present, and lactic acidosis, ketoacidosis, and acute kidney injury have been reported.

Management

Supportive care is paramount. Intravenous glucose may be required to treat hypoglycaemia. Haemodialysis should be considered if the blood ethanol concentration exceeds 5000 mg/L and if a severe metabolic acidosis is present, which has not been corrected by fluids and intravenous bicarbonate.

Neuroleptics and atypical neuroleptics

Neuroleptic (antipsychotic) drugs are thought to act predominantly by blockade of dopamine D_2 receptors. The ***first-generation neuroleptics*** include the phenothiazines (chlorpromazine), the butyrophenones (haloperidol), and the substituted benzamides (sulpiride). More ***selective second-generation (or 'atypical') antipsychotics*** include amisulpride, aripiprazole, clozapine, olanzapine, quetiapine, and risperidone.

Clinical features

These include impaired consciousness, hypotension, respiratory depression, hypothermia or hyperthermia, antimuscarinic effects

such as tachycardia, dry mouth and blurred vision, occasionally seizures, rhabdomyolysis, cardiac arrhythmias (both atrial and ventricular), and ARDS. Extrapyramidal effects, including acute dystonic reactions, occur but are not dose related. Most 'atypical' antipsychotics have less profound sedative actions than the older neuroleptics. QT interval prolongation and subsequent ventricular arrhythmias (including torsades de pointes) have occurred following overdose with the atypical neuroleptics (especially amisulpride). Unpredictable fluctuations in conscious level, with variations between agitation and marked somnolence, have been particularly associated with olanzapine overdose.

Management

Intravenous procyclidine 5 to 10 mg in an adult can occasionally be required for the treatment of dyskinesia and oculogyric crisis. After acidosis has been corrected with sodium bicarbonate, the preferred treatment for arrhythmias caused by antipsychotic drugs (including torsades de pointes) is intravenous magnesium (see Chapter 17) or cardiac pacing (Chapter 11).

Cannabis (marijuana) and synthetic cannabinoids

Cannabis is usually smoked but may be ingested as a 'cake' or made into a tea. The major psychoactive constituent is δ-9-tetrahydrocannabinol (THC). THC possesses activity at the benzodiazepine, opioid, and cannabinoid receptors.

Synthetic cannabinoids emerged onto the drug scene around 2009, initially sold as herbal incense or 'spice'. There are now hundreds of derivatives and most are considerably more potent than THC (common street name 'mamba').

Clinical features

Cannabis intoxication initially causes euphoria followed by distorted and heightened images, colours and sounds, altered tactile sensations, and sinus tachycardia. Visual and auditory hallucinations and acute psychosis are particularly likely to occur after substantial ingestion in naive users. Heavy users suffer impairment of memory and attention and poor academic performance. There is an increased risk of anxiety and depression. Regular users are at risk of dependence. Cannabis use results in an overall increase in the relative risk for later schizophrenia and psychotic episodes (see Chapter 30). Cannabis smoke is probably carcinogenic.

The clinical features observed following intoxication with synthetic cannabinoids are variable and unpredictable, reflecting the numerous derivatives available. Tachycardia and agitation are the most consistent effects but psychosis, seizures, respiratory failure, coma, acute kidney injury, stroke, and myocardial infarction have all been reported and fatalities have occurred.

Management

Reassurance is usually the only treatment required, although sedation with intravenous diazepam (10–20 mg in an adult) or intramuscular or oral haloperidol (2–5 mg in an adult) is sometimes required if agitation is severe. Occasionally, agitation is sufficiently severe for patients to require sedation in a critical care environment.

Anticonvulsants

Clinical features

The clinical features of poisoning with anticonvulsant drugs are summarized in **Box 15.14**.

Box 15.14 Clinical features of poisoning with anticonvulsant drugs

Anticonvulsant drug	Clinical features of poisoning
Carbamazepine and oxcarbazepine (a prodrug of carbamazepine)	Dry mouth, coma, convulsions, ataxia, incoordination, hallucinations (particularly in the recovery phase) Ocular: nystagmus, dilated pupils (common), divergent strabismus, complete external ophthalmoplegia (rare)
Phenytoin	Nausea, vomiting, headache, tremor, cerebellar ataxia, nystagmus, loss of consciousness (rare)
Sodium valproate	Most frequent: drowsiness, impairment of consciousness, respiratory depression Uncommon complications: liver damage, hyper-ammonaemia, metabolic acidosis Very severe poisoning: myoclonic jerks, seizures, cerebral oedema
Gabapentin and pregabalin	Lethargy, ataxia, slurred speech, gastrointestinal symptoms
Lamotrigine	Lethargy, coma, ataxia, nystagmus, seizures, cardiac conduction abnormalities
Levetiracetam	Lethargy, coma, respiratory depression
Tiagabine	Lethargy, facial grimacing, nystagmus, posturing, agitation, coma, hallucinations, seizures
Topiramate	Lethargy, ataxia, nystagmus, myoclonus, coma, seizures, metabolic acidosis (secondary to carbonic anhydrase inhibition)

Management

Elimination of carbamazepine can be significantly increased by MDAC.

If hepatotoxicity and encephalopathy develop in severe valproate poisoning, early intravenous supplementation with L-carnitine will reduce the formation of ammonia and other toxic metabolites of valproic acid. Haemodialysis increases the elimination of valproate and should be instituted if severe hyper-ammonaemia and electrolyte and acid–base disturbances occur.

Other drugs of importance in poisoning

For ease of reference, these are arranged in alphabetical order and include antidiabetic drugs, antimalarial drugs, beta-blockers, calcium-channel blockers, digoxin, iron salts, lithium salts, nitrous oxide, and salicylates.

Antidiabetic drugs

Insulin (if injected but not if ingested) and ***sulphonylureas*** cause hypoglycaemia. This is not seen with ***metformin***, since its mode of action is to increase glucose utilization, but lactic acidosis is a potentially serious complication of metformin poisoning.

Clinical features

Features of severe hypoglycaemia include drowsiness, coma, convulsions, depressed limb reflexes, extensor plantar responses, and cerebral oedema. Hypokalaemia also occurs. Cranial (neurogenic) diabetes insipidus and persistent vegetative states are possible long-term complications if hypoglycaemia is prolonged.

Management

The blood or plasma glucose concentration should be measured urgently and intravenous glucose given, if necessary. Glucagon produces only a slight rise in blood glucose, although it can reduce the amount of glucose required (see Chapter 11).

Severe insulin poisoning

A continuous infusion of 10 to 20% glucose (with K^+ 10–20 mmol/L) is required, together with carbohydrate-rich meals, though there may be difficulty in maintaining normoglycaemia. For refractory cases, corticosteroids can be used (e.g., i.v. hydrocortisone 100 mg up to 4 times a day).

Sulphonylurea poisoning

The administration of glucose increases already high circulating insulin concentrations. Intravenous octreotide (50 μg), which inhibits insulin release, should be given as well as glucose.

Antimalarials

Chloroquine

Severe poisoning may present with hypotension, cardiac failure, pulmonary oedema, and cardiac arrest. Agitation, acute psychosis, convulsions, and coma may occur. Hypokalaemia is common and is due to chloroquine-induced potassium-channel blockade. Bradyarrhythmias and tachyarrhythmias are common, and ECG conduction abnormalities are similar to those seen in quinine poisoning.

Quinine

Cinchonism (tinnitus, deafness, vertigo, nausea, headache, and diarrhoea) is common. In more severe poisoning, convulsions, hypotension, pulmonary oedema, and cardiorespiratory arrest are seen (due to ventricular arrhythmias that are often preceded by ECG conduction abnormalities, particularly QT prolongation). Quinine cardiotoxicity is due to sodium channel blockade. Patients may also develop ocular features, including blindness, which can be permanent.

Primaquine

The main concern is primaquine's propensity to cause methaemoglobinaemia and haemolytic anaemia.

Management

- MDAC increases quinine and probably chloroquine clearance.
- Hypokalaemia should be corrected.
- Intravenous sodium bicarbonate 50 to 100 mmol is given if the ECG shows QRS prolongation but it will exacerbate hypokalaemia, which should be corrected first.
- Mechanical ventilation, the administration of an inotrope (see Chapter 13), and high doses of diazepam (1 mg/kg as a loading dose and 0.25 to 0.4 mg/kg per hour maintenance) may reduce the mortality in severe chloroquine poisoning.
- Overdrive pacing may be required if torsades de pointes occurs in quinine poisoning and does not respond to intravenous magnesium sulphate (see Chapter 17).
- If clinically significant methaemoglobinaemia (generally above 30%) develops in primaquine poisoning, methylthioninium (methylene blue) 1 to 2 mg/kg body weight should be administered.

Beta-adrenoceptor–blocking drugs

In mild poisoning, sinus bradycardia is the only feature, but if a substantial amount has been ingested, coma, convulsions, and hypotension develop. Less commonly, delirium, hallucinations, and cardiac arrest supervene. Bronchospasm and hypoglycaemia are rare complications.

Management

- ***Glucagon*** 50 to 150 μg/kg (typically 5–10 mg in an adult), followed by an infusion of 5 to 10 mg/h, is the most effective agent. It acts by bypassing the blocked beta-receptor, thus activating adenylyl cyclase and promoting formation of cyclic adenosine monophosphate from adenosine triphosphate; cyclic adenosine monophosphate, in turn, exerts a direct beta-stimulant effect on the heart.
- ***Intravenous atropine*** 0.6 to 1.2 mg can be used to treat bradycardia but is usually less effective.
- ***High-dose insulin*** (initial intravenous bolus of 1.0 IU/kg, followed by 1–10 IU/kg per hour) with hypertonic glucose to avoid hypoglycaemia has been shown to improve myocardial contractility and systemic perfusion.

Calcium-channel blockers

Calcium-channel blockers all act by blocking voltage-gated calcium channels. Dihydropyridines (e.g., amlodipine, felodipine, nifedipine) are predominantly peripheral vasodilators, while verapamil (a phenylalkylamine) and diltiazem (a benzothiazepine) also have significant cardiac effects. Poisoning, particularly with verapamil and diltiazem, causes heart block and hypotension; in severe poisoning, there is a substantial fatality rate. When a sustained-release preparation has been ingested, the onset of severe features is delayed, sometimes for more than 12 hours. Overdose with even small amounts can have profound effects.

Clinical features

Hypotension occurs due to peripheral vasodilation, myocardial depression, and conduction block. The ECG may progress from sinus bradycardia through first, then higher, degrees of block to asystole. Cardiac and non-cardiac pulmonary oedema may ensue in severely poisoned patients. Other features include nausea, vomiting, seizures, and a lactic acidosis.

Management

- ***Intravenous atropine*** 0.6 to 1.2 mg, repeated as required, should be given for bradycardia and heart block. The initial dose can be repeated every 3 to 5 min, but if there is no response in pulse rate or blood pressure after three doses, it is unlikely that further boluses will be helpful. The response to atropine is sometimes improved following intravenous 10% calcium chloride, 5 to 10 mL (at 1–2 mL/min). If there is an initial response to calcium, a continuous infusion is warranted as 10% calcium chloride, 1 to 10 mL/h.
- ***Cardiac pacing*** has a role if there is evidence of atrioventricular conduction delay but failure to capture occurs.
- Hypotension that persists despite volume replacement may respond to ***i.v. glucagon***, which activates myosin kinase independent of calcium. An initial bolus of 50 to 150 μg/kg can be repeated as necessary or an infusion of 5 to 10 mg/h can be commenced.
- Insulin–glucose euglycaemia has been shown to improve myocardial contractility and systemic perfusion and should be considered in cases of resistant hypotension. ***Insulin*** is given as a bolus dose of 1 IU/kg, followed by an infusion of 1 to 10 IU/kg per hour with hypertonic glucose and frequent monitoring of blood glucose and potassium.

Digoxin

Toxicity occurring during chronic administration is common, though acute poisoning is infrequent.

Clinical features

These include nausea, vomiting, dizziness, anorexia, and drowsiness. Rarely, confusion, visual disturbances, and hallucinations occur. Initial sinus bradycardia may be followed by supraventricular arrhythmias with or without heart block, ventricular premature beats, and ventricular tachycardia. Hyperkalaemia occurs due to inhibition of the sodium–potassium activated ATPase pump.

Management

- ***Intravenous atropine*** 1.2 to 2.4 mg is given to reduce sinus bradycardia, atrioventricular block, and sinoatrial standstill.
- ***Digoxin-specific antibody fragments*** (digoxin-Fab) should be given intravenously for significant hyperkalaemia, marked arrhythmias (usually severe bradycardia compromising the cardiac output), and asystole. In both acute and chronic poisoning, only half the estimated dose required for full neutralization (calculated from amount of drug taken or serum digoxin concentration) needs to be given initially; a further dose is given if clinically indicated.

Iron salts

Unless more than 20 mg of elemental iron per kg of body weight is ingested (a ferrous sulphate tablet contains 60 mg of iron), features are unlikely to develop. As a result, poisoning is seldom severe but deaths still occur. Iron salts have a direct corrosive effect on the upper gastrointestinal tract.

Clinical features

Initial features include nausea, vomiting (the vomit may be grey or black in colour), abdominal pain, and diarrhoea. Severely poisoned patients develop haematemesis, hypotension, coma, and shock at an early stage. Usually, however, most patients suffer only mild gastrointestinal symptoms. A small minority deteriorate 12 to 48 hours after ingestion and develop shock, metabolic acidosis, acute tubular necrosis, and hepatocellular necrosis. Rarely, up to 6 weeks after ingestion, intestinal strictures occur due to corrosive damage. The serum iron concentration should be measured some 4 hours after ingestion; if the concentration exceeds the predicted normal iron-binding capacity (usually >5 mg/L; 90 μmol/L), free iron is circulating and treatment with desferrioxamine should be considered.

Management

The majority of patients ingesting iron do not require desferrioxamine therapy. If a patient develops coma or shock, ***desferrioxamine*** should be given without delay in an intravenous dose of 15 mg/kg per hour (the total amount of infusion usually not to exceed 80 mg/kg in 24 h). If the recommended rate of administration is continued for several days, adverse effects, including pulmonary oedema and ARDS (see Chapter 13), have been reported.

Lithium salts

Lithium poisoning usually results from impaired renal elimination in an individual on lithium therapeutically; that is, chronic toxicity rather than deliberate self-poisoning (acute toxicity). Lithium has a narrow therapeutic index (target plasma range 0.8–1.2 mmol/L).

Clinical features

Features of intoxication include diarrhoea and vomiting, tremor, thirst, polyuria, and, in more serious cases, impairment of consciousness, hypertonia, and convulsions; irreversible neurological damage may occur. Measurement of the serum lithium concentration confirms the diagnosis. However, in chronic poisoning the plasma lithium concentration may not be markedly above the upper end of the therapeutic range despite the presence of severe neurological features (which are due to accumulation of lithium in the CNS). Conversely, acute massive overdose may produce high plasma concentrations with few toxic features.

Management

The plasma lithium concentration must always be interpreted in association with the clinical features present. Since lithium is eliminated unchanged via the kidneys, diuresis with sodium chloride 0.9% is effective in increasing clearance. Haemodialysis is far superior and is used if neurological features are present and/or if renal function is impaired. It is employed particularly in cases of chronic toxicity.

Nitrous oxide

Nitrous oxide (N_2O), also known as 'laughing gas', is a general anaesthetic agent with a long history of recreational misuse preceding use as an anaesthetic. It is usually inhaled from balloons filled with N_2O-containing canisters. The small steel canisters are also known as 'whippets'.

Clinical features

Inhalation produces a short-lived feeling of euphoria or relaxation. Hypoxia, asphyxiation, and injuries secondary to acute intoxication are possible. Cold burns to the lips or thighs can occur as the canister becomes ice cold as it discharges. Chronic misuse over weeks and months can lead to N_2O-induced myeloneuropathy, a syndrome ranging from mild damage to peripheral nerves to severe spinal cord damage, known as subacute combined degeneration of the cord. N_2O irreversibly inactivates vitamin B12, resulting in functional B12 deficiency, which leads to a failure in the production of myelin. An increasing number of cases of N_2O-induced myeloneuropathy have been reported in recent years. Patients usually present with paraesthesia in the upper and lower limbs and gait ataxia.

Management

There is no specific toxicological management for acute intoxication and patients should be treated with symptomatic and supportive care. N_2O-induced myeloneuropathy should be treated with ***intramuscular B12 injections*** (e.g., 1 mg hydroxocobalamin i.m. initially on alternate days) regardless of the measured B12 concentration in the blood, which may be normal. Treatment duration depends on symptom improvement and can also be guided by biomarkers of functional B12 deficiency such as methylmalonic acid and homocysteine, which are often raised at the point of diagnosis.

Salicylates

Aspirin is metabolized to salicylic acid (salicylate) by esterases present in many tissues, especially the liver, and subsequently to salicyluric acid and salicyl phenolic glucuronide (**Fig. 15.2**); these two pathways become saturated, with the consequence that the renal excretion of salicylic acid increases after overdose; this excretion pathway is extremely sensitive to changes in urinary pH.

Clinical features

Salicylates stimulate the respiratory centre, increase the depth and rate of respiration, and induce a respiratory alkalosis. Compensatory mechanisms, including renal excretion of bicarbonate and

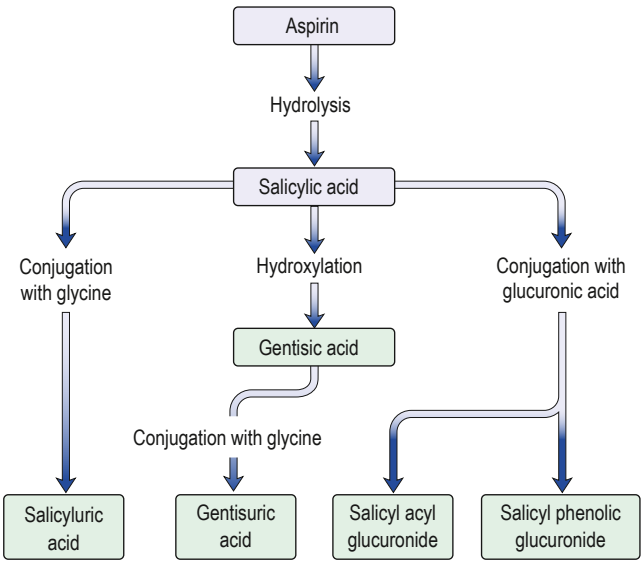


Fig. 15.2 The metabolism of aspirin.

potassium, result in a metabolic acidosis. Salicylates also interfere with carbohydrate, fat, and protein metabolism and disrupt oxidative phosphorylation, producing increased concentrations of lactate, pyruvate, and ketone bodies, all of which contribute to the acidosis.

Common features of poisoning include tachypnoea, sweating, vomiting, epigastric pain, tinnitus, and deafness. Respiratory alkalosis and metabolic acidosis supervene and a mixed acid–base disturbance is typical. Rarely, in severe poisoning, non-cardiogenic pulmonary oedema, coma, and convulsions ensue.

The severity of salicylate toxicity is dose related.

Management

Fluid and electrolyte replacement are required with prompt correction of dehydration and hypokalaemia.

- Mild cases of salicylate poisoning are managed with parenteral fluid and electrolyte replacement only.
- Patients whose plasma salicylate concentrations are in excess of 500 mg/L (3.6 mmol/L) should receive urine alkalinization (see 'Increasing poison elimination' section).
- ***Haemodialysis*** is the treatment of choice for severely poisoned patients (plasma salicylate concentration >700 mg/L; >5.1 mmol/L), particularly those with coma and metabolic acidosis.

Poisoning with chemicals

Included in this section, in alphabetical order, are arsenic, carbamate insecticides, carbon monoxide, copper sulphate, corrosives, cyanide, ethylene glycol, household agents, lead, mercury, methanol, nerve agents, OP insecticides, and phosphides.

Arsenic

Arsenic poisoning worldwide is commonly caused by contamination of the ground water by inorganic arsenates, particularly in Asia. Occasionally, arsenic is found in Chinese and Indian traditional remedies.

Clinical features

Acute ingestion causes abdominal pain, vomiting, and diarrhoea. Hypovolaemic shock and acute tubular necrosis occur in severe cases.

Chronic exposure to lower doses produces gastrointestinal effects, accompanied by skin changes (hyperkeratosis of palms and soles, 'raindrop' pattern of hyperpigmentation, and alopecia), neurological features (headache, sensorimotor neuropathy), abnormal liver biochemistry (non-cirrhotic portal hypertension is recognized), peripheral vascular arteriosclerosis, and haematological abnormalities (pancytopenia).

Management

The identification of, and removal from, the source of exposure to arsenic is vital. Chelation therapy with oral DMSA 30 mg/kg per day or intravenous DMPS 30 mg/kg per day may be indicated.

Carbamate insecticides

Carbamate insecticides inhibit acetylcholinesterase but the duration of this inhibition is comparatively short-lived in comparison with OP insecticides (see 'Organophosphorus insecticides' section), since the carbamate–enzyme complex tends to dissociate spontaneously.

Clinical features

Although carbamate insecticide poisoning is generally less severe than OP insecticide poisoning, acute poisoning with a carbamate can be severe and fatal. Cholinergic symptoms usually develop within a few minutes. In the most severe cases, muscle twitching, profound weakness, profuse sweating, incontinence, mental confusion, and progressive cardiac and respiratory failure ensue. In less severe cases, cholinergic symptoms are usually evident within 2 hours and typically resolve within 24 hours. Seizures are relatively uncommon, since carbamate penetration into the CNS is limited.

Management

Mild cases require no specific treatment other than the removal of soiled clothing. ***Intravenous atropine*** 2 mg should be given every 3 to 5 minutes, if necessary, to reduce increased secretions, rhinorrhoea, and bronchorrhoea. If this measure fails, the patient should be intubated and mechanical ventilation instituted. Since carbamates have a shorter duration of action than OP insecticides, pralidoxime should be used only rarely in carbamate poisoning. If intoxication is life-threatening, give ***intravenous pralidoxime chloride*** 30 mg/kg body weight over 5 to 10 minutes, followed by an infusion of 8 to 10 mg/kg per hour.

Carbon monoxide

The most common source of carbon monoxide is an improperly maintained and poorly ventilated heating system. In addition, inhalation of methylene chloride (found in paint strippers) may also lead to carbon monoxide poisoning, as methylene chloride is metabolized *in vivo* to carbon monoxide. Carbon monoxide has a greater affinity for haemoglobin and forms carboxyhaemoglobin (COHb), thereby reducing the oxygen-carrying capacity. The affinity of the remaining haem groups for oxygen is increased. This shifts the oxyhaemoglobin dissociation curve to the left, impairing liberation of oxygen to the cells and leading to tissue hypoxia. In addition, carbon monoxide also inhibits cytochrome oxidase a_3.

Clinical features

Symptoms of mild to moderate exposure to carbon monoxide may be mistaken for a viral illness.

- A peak COHb concentration of <10% is not normally associated with symptoms.
- A peak COHb concentration of 10% to 30% usually causes headache and mild exertional dyspnoea.

- Higher concentrations of COHb are associated with coma, convulsions, and cardiorespiratory arrest. Metabolic acidosis, myocardial ischaemia, hypertonia, extensor plantar responses, retinal haemorrhages, and papilloedema also occur.

Neuropsychiatric features may develop after apparent recovery from carbon monoxide exposure.

Management

In addition to removing the patient from carbon monoxide exposure, high-flow oxygen should be administered using a tightly fitting face mask (see Chapter 13). Endotracheal intubation and mechanical ventilation are required in those who are unconscious. Several controlled studies of hyperbaric oxygen have been published but none has shown long-term clinical benefit.

Copper sulphate

Copper sulphate is used in fungicides, algicides, electroplating, dyes, inks, disinfectants, and wood preservatives.

Clinical features

Ingestion causes vomiting, abdominal pain, diarrhoea, headache, dizziness, and a metallic taste. Gastrointestinal haemorrhage, intravascular haemolysis, methaemoglobinaemia, rhabdomyolysis, coma, convulsions, and hepatorenal failure may ensue and fatalities have occurred. Body secretions may be blue/green.

Management

Treatment is supportive with replenishment of lost fluids/blood. Blood copper concentrations correlate with the severity of poisoning; a copper concentration of >8 mg/L is indicative of severe poisoning. Early endoscopy (or CT scan with contrast if endoscopy is not possible) is recommended if corrosive damage is suspected. Methaemoglobinaemia of >30% should be treated with intravenous methylthioninium chloride 1 to 2 mg/kg. Renal failure may require extracorporeal support.

Corrosive agents

Strong acids and alkalis cause chemical burns. Inorganic acids such as hydrochloric and sulphuric acid are generally more toxic than organic acids such as acetic acid. Hydrofluoric acid is considered separately (see 'Hydrofluoric acid and hydrogen fluoride' section). Some household products, such as water-sterilizing tablets, are strong alkalis. Following corrosive ingestion, both CT scan and endoscopy are useful in assessing the severity of injury, risk of mortality, and risk of subsequent stricture formation. Prompt surgical intervention may be required to remove necrotic tissue, if indicated.

Cyanide

Cyanide and its derivatives are used widely in industry. Hydrogen cyanide is also released during the thermal decomposition of polyurethane foams, such as that in mattresses. Cyanide reversibly inhibits cytochrome oxidase a_3, so that cellular respiration ceases.

Clinical features

Inhalation of hydrogen cyanide produces symptoms within seconds and death within minutes. By contrast, the ingestion of a cyanide salt may not produce features for 1 hour. After exposure, initial symptoms are non-specific and include a feeling of constriction in the chest and dyspnoea. Coma, convulsions, and metabolic acidosis may then supervene.

Management

Oxygen should be provided and an antidote administered. ***Dicobalt edetate*** was historically used and is highly efficacious but manufacturing has been discontinued. Dicobalt edetate (and the free cobalt contained in the preparation) complexes free cyanide. An alternative and more widely available antidote is ***intravenous hydroxocobalamin*** (5 g); a second dose may be required in severe cases. Each hydroxocobalamin molecule can bind one cyanide ion to form cyanocobalamin, which is then excreted in the urine. Hydroxocobalamin is now considered first line in cases of moderate and severe cyanide poisoning. ***Intravenous sodium thiosulphate*** 12.5 g, which acts by enhancing endogenous detoxification, can be used as an adjunct in severe cases, or alone in mild or moderate poisoning.

Ethylene glycol

Ethylene glycol is found in a variety of common household products, including antifreeze, windshield washer fluid, brake fluid, and lubricants. The features observed in poisoning are due to metabolites predominantly, not the parent chemical. Ethylene glycol (**Fig. 15.3**) is metabolized to glycolate, the cause of the acidosis. A small proportion of glyoxylate is metabolized to oxalate. Calcium ions chelate oxalate to form insoluble calcium oxalate, which is responsible for renal toxicity.

Clinical features

Initially, the features of ethylene glycol poisoning are similar to those of ethanol intoxication (though there is no ethanol on the breath). Coma and convulsions follow and a variety of neurological abnormalities, including nystagmus and ophthalmoplegias, are seen. Severe metabolic acidosis, hypocalcaemia, and acute kidney injury are well-recognized complications.

Management

If the patient presents early after ingestion, the priority is to inhibit metabolism using either intravenous fomepizole or ethanol; the former does not require monitoring of blood concentrations.

- ***Fomepizole:*** 15 mg/kg body weight should be administered, followed by four 12-hourly doses of 10 mg/kg, then 15 mg/kg every 12 h until glycol concentrations are not detectable. Following a substantial ingestion, haemodialysis or haemodiafiltration should be employed to remove the glycol and metabolites. If dialysis is used, fomepizole should be administered as an infusion

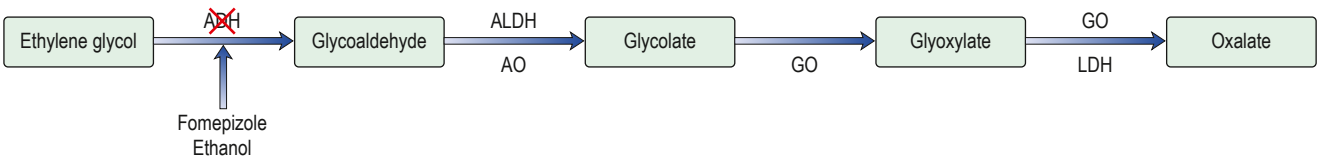


Fig. 15.3 The metabolism of ethylene glycol. Fomepizole and ethanol inhibit ADH. *ADH,* Alcohol dehydrogenase; *ALDH,* aldehyde dehydrogenase; *AO,* aldehyde oxidase; *GO,* glycolate oxidase; *LDH,* lactate dehydrogenase.

of 1 mg/kg/h, or the frequency of fomepizole dosing should be increased to 4 hourly because fomepizole is dialysable.

- ***Ethanol:*** alternatively, a loading dose of ethanol 50 g can be administered, followed by an intravenous infusion of ethanol 10 to 12 g/h to produce blood ethanol concentrations of 500 to 1000 mg/L (11–22 mmol/L). The infusion is continued until the glycol is no longer detectable in the blood. If haemodialysis is employed, the rate of ethanol administration will need to be increased to 17 to 22 g/h, as ethanol is dialysable.

Supportive measures to combat shock, hypocalcaemia, and metabolic acidosis should be instituted.

Household products

The agents most commonly involved are bleach, cosmetics, toiletries, detergents, disinfectants, and petroleum distillates such as paraffin and white spirit. Ingestion of household products is usually accidental and is most common among children less than 5 years of age.

Clinical features

If ingestion is accidental, features very rarely occur, except in the case of petroleum distillates where aspiration is a recognized complication because of their low surface tension. Powder detergents, sterilizing tablets, denture cleaning tablets, and industrial bleaches (which contain high concentrations of sodium hypochlorite) are corrosive to the mouth and pharynx if ingested. Nail polish and nail polish remover contain acetone, which may produce a coma if ingested in substantial quantities. Inhalation by small children of substantial quantities of talcum powder has occasionally given rise to severe pulmonary oedema and death.

Hydrofluoric acid and hydrogen fluoride

Hydrofluoric acid, a solution of hydrogen fluoride in water, is a colourless, fuming liquid that is widely used in industry. It is particularly dangerous because of its unique ability among acids to penetrate tissue.

Clinical features

Dermal exposure to hydrofluoric acid results in rapid liquefactive necrosis and erosion of bone. If more than 1% of the body surface area is contaminated with a 50% or higher solution of hydrofluoric acid, there is a high risk of hypocalcaemia from the formation of calcium fluoride. This may lead to cardiac conduction disturbances, notably QT interval prolongation and an increased risk of ventricular arrhythmias, particularly torsades de pointes (see Chapter 17).

Ingestion of hydrofluoric acid causes severe corrosive injury to the gastrointestinal tract.

Inhalation of hydrogen fluoride causes irritation of the eyes and nose, with sore throat, cough, chest tightness, headache, ataxia, and confusion. Dyspnoea and stridor due to laryngeal oedema may follow, depending on the concentration of hydrogen fluoride. Haemorrhagic pulmonary oedema with increasing breathlessness, wheeze, hypoxia, and cyanosis may take up to 36 hours to develop.

Management

- ***Dermal exposure.*** Immediate irrigation of the skin with copious volumes of water is the priority, followed by the prompt application of calcium gluconate gel to reduce pain and limit skin damage.
- ***Ingestion.*** This is a medical emergency with the need for immediate assessment for surgery after endoscopy or CT.
- ***Inhalation.*** If the patient has clinical features of bronchospasm after inhalation, then treat conventionally with nebulized bronchodilators and steroids. Treat pulmonary oedema and/or ARDS (see Chapter 13) with continuous positive airway pressure or, in severe cases, with intermittent positive pressure ventilation.

Lead

Exposure to lead occurs occupationally, children may eat lead-painted items in their homes (pica), and the use of lead-containing cosmetics or ‘drugs’ has also resulted in poisoning.

Clinical features

Mild intoxication may result in no more than lethargy and occasional abdominal discomfort, though abdominal pain, vomiting, constipation, and encephalopathy (seizures, delirium, coma) may develop in more severe cases. Encephalopathy is more common in children than in adults but is rare in the developed world. Typically, though very rarely, lead poisoning results in foot drop attributable to peripheral motor neuropathy.

Anaemia (normally normochromic normocytic) occurs due both to inhibition by lead of several enzymes involved in haem synthesis and to haemolysis. The latter results from damage to the red cell membrane by aggregates of RNA that accumulate owing to inhibition by lead of pyrimidine-5-nucleotidase, causing characteristic ‘basophilic stippling’ of erythrocytes.

Management

The social and occupational dimensions of lead poisoning must be recognized. Simply giving patients chelation therapy and then returning them to a contaminated environment is of no value.

The decision to use chelation therapy is based not only on the blood lead concentration but also on the presence of symptoms. Parenteral ***sodium calcium edetate*** 75 mg/kg per day or oral ***succimer (DMSA)*** 30 mg/kg per day is of similar efficacy. At least 5 days’ treatment is usually required. As chelation of zinc may occur with sodium calcium edetate, serum zinc concentrations should be checked.

Mercury

Mercury is the only metal that is liquid at room temperature. It exists in three oxidation states (elemental/metallic Hg^0, mercurous Hg_2^{2+}, and mercuric Hg^{2+}) and can form inorganic (e.g., mercuric chloride) and organic (e.g., methylmercury) compounds. Metallic mercury is very volatile; when spilled, it has a large surface area so that high atmospheric concentrations may be produced in enclosed spaces, particularly when environmental temperatures are high. Thus great care should be taken in clearing up a spillage. If ingested, metallic mercury will usually be eliminated per rectum, though small amounts may settle in the appendix and remain there for years. Mercury salts are well absorbed following ingestion, as is mercury vapour following inhalation.

Clinical features

Acute inhalation of mercury vapour causes headache, nausea, cough, chest pain, and, occasionally, a chemical pneumonitis. Proteinuria and nephrotic syndrome are observed rarely.

Ingestion of inorganic or organic mercury compounds causes an irritant gastroenteritis. Mercurous (Hg_2^{2+}) compounds are less corrosive and less toxic than mercuric (Hg^{2+}) salts.

Systemic accumulation of mercury from any source and by any route of exposure leads to characteristic neurological features, including a fine tremor, lethargy, memory loss, insomnia,

personality changes, and ataxia. Peripheral nerve damage has also been observed, as has renal tubular damage.

Management

Unithiol (DMPS) is the antidote of choice and is given in an intravenous dose of 30 mg/kg per day. At least 5 days' treatment is usually required.

Methanol

Methanol is used widely as a solvent and is found in antifreeze solutions. Methanol is metabolized to formaldehyde and formate (**Fig. 15.4**). The concentration of formate increases greatly and is accompanied by accumulation of hydrogen ions, leading to metabolic acidosis.

Clinical features

Methanol causes inebriation and drowsiness. After a latent period, coma supervenes. Blurred vision and diminished visual acuity occur due to formate accumulation. The presence of dilated pupils that are unreactive to light suggests that permanent blindness is likely to ensue. A severe metabolic acidosis may develop and be accompanied by hyperglycaemia and a raised serum amylase activity. A blood methanol concentration of 500 mg/L (15.6 mmol/L) confirms severe poisoning. The mortality correlates well with the severity and duration of metabolic acidosis. Survivors may show permanent neurological sequelae, including parkinsonian-like signs as well as blindness.

Management

Treatment is similar to that of ethylene glycol poisoning (see 'Ethylene glycol' section) with the addition of intravenous folinic acid 30 mg 6 hourly for 48 hours, which accelerates formate metabolism, thereby reducing ocular toxicity.

Nerve agents

Nerve agents are related chemically to OP insecticides (see below) and have a similar mechanism of toxicity but a much higher mammalian acute toxicity, particularly via the dermal route. In addition to inhibition of acetylcholinesterase, a chemical reaction known as 'ageing' also occurs rapidly and more completely than in the case of insecticides. This makes the enzyme resistant to spontaneous reactivation or treatment with oximes (pralidoxime, obidoxime or HI-6).

Several classes of nerve agents are recognized: G agents (named after Gerhard Schrader, who synthesized the first agents) and V agents (V purportedly standing for venomous). G agents include tabun, sarin, soman, and cyclosarin. The V agents, for example, VX, were introduced later. The G agents are both dermal and respiratory hazards, whereas the V agents, unless aerosolized, are contact poisons. The latest class of nerve agents to receive notoriety are the Novichok agents, which are potent contact poisons.

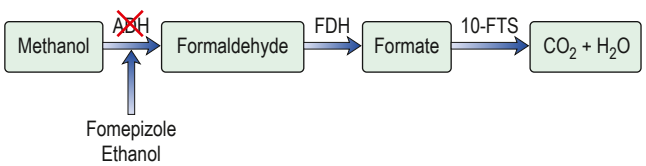


Fig. 15.4 The metabolism of methanol. Fomepizole and ethanol inhibit ADH. *10-FTS,* 10-Formyl tetrahydrofolate synthetase; *ADH,* alcohol dehydrogenase; *FDH,* formaldehyde dehydrogenase.

Clinical features

Systemic features include increased salivation, rhinorrhoea, bronchorrhoea, miosis (which is characteristic) and eye pain, abdominal pain, nausea, vomiting and diarrhoea, involuntary micturition and defecation, muscle weakness and fasciculation, tremor, restlessness, ataxia, and convulsions. Bradycardia, tachycardia, and hypotension occur, depending on whether muscarinic or nicotinic effects predominate. Death occurs from respiratory failure within minutes but individuals with mild or moderate exposure usually recover completely. Diagnosis is confirmed by measuring the erythrocyte cholinesterase activity.

Management

- The administration of ***intravenous atropine*** 2 mg repeated every 3 to 5 min as necessary to patients presenting with rhinorrhoea and bronchorrhoea may be lifesaving.
- An oxime should be given to all those requiring atropine as soon as possible after exposure before 'ageing' has occurred; for example, ***intravenous pralidoxime*** chloride 30 mg/kg, followed by an infusion of pralidoxime chloride 8 to 10 mg/kg per hour. Alternatively, boluses of pralidoxime chloride 30 mg/kg may be given 4 to 6 hourly.
- ***Intravenous diazepam*** 10 to 20 mg, repeated as required, is useful in controlling apprehension, agitation, fasciculation, and convulsions.

Organophosphorus insecticides

OP insecticides are used widely throughout the world and are a common cause of poisoning, leading to thousands of deaths annually in the developing world. Intoxication may follow ingestion, inhalation, or dermal absorption. OP insecticides inhibit acetylcholinesterase, causing accumulation of acetylcholine at central and peripheral cholinergic nerve endings, including neuromuscular junctions. Many OP insecticides require biotransformation before becoming active and so the features of intoxication may be delayed.

Clinical features

Poisoning is characterized by anxiety and restlessness, which is typically followed by nausea, vomiting, abdominal colic, diarrhoea (particularly if exposure is by ingestion), tenesmus, sweating, hypersalivation, and chest tightness. Respiratory failure will ensue in severe cases and is exacerbated by the development of bronchorrhoea and pulmonary oedema. Miosis is characteristic. Muscle fasciculation and flaccid paresis of limb muscles and, occasionally, paralysis of extraocular muscles are observed. Coma and convulsions occur in severe poisoning.

Diagnosis is confirmed by measuring the erythrocyte acetylcholinesterase activity; plasma cholinesterase activity is less specific but may also be depressed.

The ***intermediate syndrome*** usually becomes established 1 to 4 days after exposure, when the symptoms and signs of the acute cholinergic syndrome are no longer obvious. The characteristic features of the syndrome are weakness of the muscles of respiration (diaphragm, intercostal muscles, and accessory muscles, including neck muscles) and of proximal limb muscles. Accompanying features often include weakness of muscles innervated by some cranial nerves.

Delayed polyneuropathy is a rare complication of acute exposure to some OP insecticides not marketed in most countries. It presents with glove and stocking paraesthesiae some 1 to 4 weeks after exposure, followed by an ascending motor polyneuropathy in

many cases. It is initiated by phosphorylation, and subsequent ageing, of at least 70% of an esterase—neuropathy target esterase—in peripheral nerves. Most mild cases improve with time but severe cases can be left with an upper motor neurone syndrome and permanent disability.

Management

Mild cases require no specific treatment other than the removal of soiled clothing. ***Intravenous atropine*** 2 mg should be given every 3 to 5 minutes if necessary, to reduce increased secretions, rhinorrhoea, and bronchorrhoea.

Symptomatic patients should also be given an oxime (pralidoxime, obidoxime) to reactivate inhibited acetylcholinesterase; for example, ***pralidoxime chloride*** 30 mg/kg by slow intravenous injection, followed by an infusion of pralidoxime chloride 8 to 10 mg/kg per hour. There is no specific treatment for the intermediate syndrome apart from supportive care, including prolonged ventilation, though a prospective trial suggested that early treatment with oximes could reduce the likelihood of the syndrome developing. Most patients recover in 2 to 3 weeks.

Phosphides

Aluminium and zinc phosphides are used as rodenticides and insecticides. They react with moisture in the air (and the gastrointestinal tract) to produce phosphine, the active pesticide. Acute poisoning with these compounds may be direct, due to ingestion of the salts, or indirect, from accidental inhalation of phosphine generated during their approved use.

Clinical features

Ingestion causes vomiting, epigastric pain, peripheral circulatory failure, severe metabolic acidosis, acute kidney injury, and disseminated intravascular coagulation, in addition to the features induced by phosphine.

Exposure to phosphine causes lacrimation, rhinorrhoea, cough, breathlessness, chest tightness, dizziness, diplopia, headache, nausea, drowsiness, intention tremor, and ataxia. Acute pulmonary oedema, hypertension, cardiac arrhythmias, convulsions, and jaundice have been described in severe cases.

Management

Treatment is symptomatic and supportive. Gastric lavage should not be used, as it can speed up the rate of disintegration of the product ingested and increase toxicity. Activated charcoal may bind metal phosphides. The mortality is high, despite supportive care.

Poisons in the natural world

Numerous biological organisms produce or contain substances that are toxic to humans, including animals (snakes, insects, spiders, and marine creatures), and some species of plants and fungi. **An online supplement covers these naturally occurring poisons in detail**.

Venomous snakes

Approximately 15% of the 3000 species of snake found worldwide are considered to be dangerous to humans. Snakebite is common in some tropical countries; rural areas of West Africa, South-east Asia, the Indian subcontinent, New Guinea, and the Amazon region are particularly affected. Bites by venomous snakes cause more than 100 000 deaths and many permanent sequelae each year (some 46 000 people are killed each year in India alone).

Clinical features

The main effects of envenoming are:

- local swelling, bruising, blistering, regional lymph node enlargement and necrosis
- anti-haemostatic defects: consumption coagulopathy and spontaneous systemic bleeding from gums, nose, skin, gut, genitourinary tract, and intracranial haemorrhage
- shock (hypotension) and myocardial damage
- descending paralysis: progressing from ptosis and external ophthalmoplegia to bulbar, respiratory muscle, and total flaccid paralysis
- generalized rhabdomyolysis with myoglobinuria
- intravascular haemolysis
- acute kidney injury.

A more detailed account of mechanisms and clinical features of snakebite envenoming, including discussion of the main groups of venomous snakes, can be located in the online supplement.

Management

Following a snakebite, all efforts should be made to transport the patient quickly to a hospital or dispensary. Traditional methods should be discouraged, as they are often ineffective and may harm the patient. Arterial tourniquets should *not* be used, and incision or excision of the bite area should *not* be performed.

As a first aid measure, a firm pressure bandage should be placed over the bite and the limb immobilized, as this may delay the spread of the venom. Local wounds often require little treatment. If necrosis is present, antibiotics should be given. Skin grafting may be required later. Anti-tetanus prophylaxis must be given. The type of snake should be identified, if possible.

In about 50% of cases, no venom has been injected by the bite but, nevertheless, careful observation for 12 to 24 hours is necessary in case envenomation develops. General supportive measures should be carried out, as necessary. These include intravenous fluids with volume expanders for hypotension and diazepam for anxiety. Treatment of acute respiratory, cardiac, and kidney injury is instituted as necessary.

Antivenoms are not generally indicated unless envenomation is present, as they can cause severe allergic reactions. Antivenoms can rapidly neutralize venom, but only if an amount in excess of the amount of venom is given. Large quantities of antivenom may be required; as antivenoms cannot reverse the effects of the venom, they must be given early to minimize some of the local effects and may prevent necrosis at the site of the bite. Antivenoms should be administered intravenously by slow infusion, the same dose being given to children and adults.

Allergic reactions are frequent, and adrenaline (epinephrine) 1 in 1000 solution should be available. In severe cases, the antivenom infusion should be continued even if an allergic reaction occurs, with subcutaneous injections of adrenaline being given as necessary. Some forms of neurotoxicity, such as those induced by the death adder, respond to anticholinesterase therapy with neostigmine and atropine.

Further reading

Hill SL, Thanacoody R. Poisoning. *Medicine* 2020; 48(3):153–224.
Ravikar R, et al. Managing snakebite. *BMJ* 2022; 376:e057926.

See Elsevier eBooks+ for additional online content on naturally occurring poisons.

Questions for this chapter can be found on Elsevier eBooks+.

POISONS IN THE NATURAL WORLD

This online-only content supplements the more limited information on snake envenomation in the print edition and includes in-depth descriptions of different forms of snake envenomation alongside a range of other toxins present in the natural world. Many of these have particular patterns of geographic distribution, making local knowledge invaluable in assessing regional risks.

Contents:

Marine animals
- Shellfish poisoning
- Poisonous fish
- Jellyfish stings

Venomous animals
- Insects
- Spiders
- Snakes

Poisonous plants

Poisonous mushrooms

Marine animals

Amnesic shellfish (domoic acid) poisoning

The syndrome should be known more accurately as domoic acid poisoning because amnesia is not always present. In one outbreak, the first symptoms were experienced between 15 minutes and 38 hours after mussel consumption. There are now legal limits on the domoic acid concentrations allowed in mussels.

Clinical features and management

The most common symptoms are nausea, vomiting, abdominal cramps, headache, diarrhoea, and short-term memory loss. Axonal sensory motor neuropathy, seizures, coma, and death have also been reported. Treatment is symptomatic and supportive.

Diarrhoeic shellfish (okadaic) poisoning

Okadaic poisoning occurs worldwide, often after eating bivalve molluscs such as mussels and scallops. Okadaic acid is produced by dinoflagellates belonging to the genus *Dinophysis*. It inhibits the activity of the protein phosphatases 1 and 2a. As a result, increased phosphorylation of intestinal transport proteins, with increased permeability to solutes, leads to diarrhoea.

Clinical features and management

The predominant symptoms are diarrhoea, nausea, vomiting, and abdominal pain. Symptoms tend to occur between 30 minutes and a few hours after shellfish consumption, with patients recovering within 2 to 3 days. Treatment is symptomatic and supportive.

Neurotoxic shellfish (brevetoxin) poisoning

Neurotoxic shellfish poisoning is caused by brevetoxins produced by the dinoflagellate *Gymnodinium breve*. Brevetoxins open voltage-gated sodium ion channels in cell walls and enhance the inward flow of sodium ions into the cell.

Clinical features and management

The symptoms of neurotoxic shellfish poisoning occur within 30 minutes to 3 hours and last a few days; they include nausea, vomiting, diarrhoea, chills, sweats, reversal of temperature sensation, hypotension, arrhythmias, numbness, tingling, paraesthesiae of the lips, face and extremities, cramps, bronchoconstriction, paralysis, seizures, and coma. Treatment is symptomatic and supportive.

Paralytic shellfish (saxitoxin) poisoning

This is caused by bivalve molluscs being contaminated with neurotoxins, including saxitoxin, produced by toxic dinoflagellates on which the molluscs graze. Saxitoxin blocks voltage-gated sodium channels in nerve and muscle cell membranes, thereby blocking nerve signal transmission.

Clinical features and management

Symptoms develop within 30 minutes. The illness is characterized by paraesthesiae of the mouth, lips, face, and extremities and is often accompanied by nausea, vomiting, and diarrhoea. In more severe cases, dystonia, dysphagia, muscle weakness, paralysis, ataxia, and respiratory depression occur. In one outbreak involving 187 cases, there were 26 deaths. Treatment is symptomatic and supportive.

Ciguatera fish poisoning

Over 400 fish species have been reported as ciguatoxic (*cigua* is Spanish for poisonous snail), though barracuda, red snapper, amberjack, and grouper are most commonly implicated. Ciguatera fish contain ciguatoxin, maitotoxin, and scaritoxin, which are lipid-soluble, heat-stable compounds that are derived from dinoflagellates such as *Gambierdiscus toxicus*. Ciguatoxin opens voltage-sensitive sodium channels at the neuromuscular junction and maitotoxin opens calcium channels of the cell plasma membrane.

Clinical features and management

The onset of symptoms occurs from a few minutes to 30 hours after ingestion of toxic fish. Typically, features appear between 1 and 6 hours and include abdominal cramps, nausea, vomiting, and watery diarrhoea. In some cases, numbness and paraesthesiae of the lips, tongue, and throat occur. Other features described include malaise, dry mouth, metallic taste, myalgia, arthralgia, blurred vision, photophobia, and transient blindness. In more severe cases, hypotension, cranial nerve palsies, and respiratory paralysis have been reported. Treatment is symptomatic and supportive. Recovery takes from 48 hours to 1 week in the mild form and from 1 to several weeks in the severe form. The mortality in severe cases may be as high as 12%. To prevent ciguatera fish poisoning, consumption of reef fish should be limited and parts of the fish that concentrate ciguatera toxin (e.g., the head, intestines, liver, roe) should be avoided. Ingestion of high-risk fish such as barracuda and moray eel should be avoided altogether.

Scombroid fish poisoning

This is due to the action of bacteria such as *Proteus morgani* and *Klebsiella pneumoniae* in the decomposing flesh of fish such as tuna, mackerel, mahi-mahi, bonito, and skipjack stored at insufficiently low temperatures. The spoiled fish can contain excessively high concentrations of histamine (muscle histidine is broken down by the bacteria to histamine), though the precise role of histamine in the pathogenesis of the clinical syndrome is uncertain.

Clinical features and management

Clinically, the mean incubation period is 30 minutes. The illness is characterized by flushing, headache, sweating, dizziness, burning of the mouth and throat, abdominal cramps, nausea, vomiting, and diarrhoea; it is usually short-lived, the mean duration being 4 hours. Treatment is symptomatic and supportive. Antihistamines may alleviate the symptoms.

Stings from marine animals

Several species of fish have venomous spines in their fins. These include the weaver fish, short-spine cottus, spiny dogfish, and stingray. Bathers and fishermen may be stung if they tread on or handle these species. The immediate result of a sting is intense local pain, swelling, bruising, blistering, necrosis, and, if the poisoned spine is not removed, chronic sepsis (although this is uncommon). Occasionally, systemic symptoms, including vomiting, diarrhoea, hypotension, and tachycardia, occur. Treatment by immersing the affected part in hot water may relieve local symptoms, as this denatures the thermolabile toxin.

Jellyfish stings

Most of the jellyfish found in North European coastal waters are non-toxic, as their stings cannot penetrate human skin. A notable exception is the 'Portuguese man-o'-war' *(Physalia physalis)*, whose sting contains a toxic peptide, phospholipase A, and a histamine-liberating factor. Toxic jellyfish are found more frequently in Australia and some, notably the box jellyfish, *Carukia barnesi*, cause the Irukandji syndrome (see below).

Clinical features and management

Local pain occurs, followed by myalgia, nausea, griping abdominal pain, dyspnoea, and even death. The cluster of severe systemic symptoms that constitute the Irukandji syndrome occur some 30 minutes after the jellyfish sting. The symptoms include severe low back pain, excruciating muscle cramps in all four limbs, abdomen, and chest, sweating, anxiety, restlessness, nausea, vomiting, headache, palpitations, life-threatening hypertension, and cardiogenic pulmonary oedema.

Adhesive tape may be used to remove any tentacles still adherent to the bather. Local application of 5% acetic acid is said to prevent stinging cells adherent to the skin discharging. Local analgesia and antihistamine creams provide symptomatic relief. Other features should be treated symptomatically and supportively.

Further reading

Berling I, Isbister G. Marine envenomations. *Aust Fam Physician* 2015; 44:28–32.

Venomous animals

Insect stings and bites

Insect stings from wasps and bees, and bites from ants, produce pain and swelling at the puncture site. Following the sting or bite, patients should be observed for 2 hours for any signs of evolving urticaria, pruritus, bronchospasm, or oropharyngeal oedema. The onset of anaphylaxis requires urgent treatment (see Chapter 11).

Scorpions

Scorpion stings are a serious problem in North Africa, the Middle East, and the Americas. Scorpion venoms stimulate the release of acetylcholine and catecholamines, causing both cholinergic and adrenergic symptoms.

Clinical features and management

Severe pain occurs immediately at the site of puncture, followed by swelling. Signs of systemic involvement, which may be delayed for 24 hours, include vomiting, sweating, piloerection, abdominal colic, and diarrhoea. In some cases, depending on the species, cardiogenic shock, respiratory depression, and pulmonary oedema may develop.

Local infiltration with anaesthetic or a ring block will usually alleviate local pain, though systemic analgesia may be required. Specific antivenom, if available, should be administered as soon as possible.

Spiders

The black widow spider *(Latrodectus mactans)* is found in North America, Australia, the tropics, and, occasionally, in Mediterranean countries.

Clinical features and management

The bite quickly becomes painful, and generalized muscle pain, sweating, headache, and shock may occur. No systemic treatment is required except in cases of severe systemic toxicity, when specific antivenom should be given, if this is available.

Venomous snakes

Contributed by C. Ariaranee Gnanathasan
University of Colombo, Sri Lanka

Venomous snakes are found in most parts of the world except in very cold regions such as Antarctica and islands such as Ireland, Iceland, New Zealand and most islands of the Western Mediterranean, Caribbean, and Pacific. Sea snakes (**Fig. E15.1**) inhabit the Indian and Pacific Oceans between latitudes 30°N and 30°S.

Two families of venomous snakes are of major medical importance:

- ***Elapidae:*** African and Asian cobras (**Fig. E15.2**), Asian kraits (Figs. E15.3 and E15.4), African mambas, American coral snakes, most Australo-Papuan venomous snakes and sea snakes
- ***Viperidae:*** that is, vipers (**Fig. E15.5**), adders, pit vipers (**Fig. E15.6**), and rattlesnakes.

Snakebite envenoming is a category A neglected tropical disease of high global impact. The burden of snakebite envenoming has the highest impact on populations and healthcare systems in sub-Saharan Africa, South and South-east Asia, Latin America, and Oceania. The World Health Organization (WHO) estimated that 4.5 to 5.4 million people are bitten by snakes annually, resulting in 1.8 to 2.7 million envenomings, 81 000 to 138 000 deaths, at least 400 000 people suffering from physical or psychological sequelae, and a staggering 400 000 limb amputations. A large number of victims survive with permanent sequelae of envenoming, both physical and psychological (e.g., post-traumatic stress disorder).

Snake venom is not a single toxin. It is a complex mixture of more than 100 different components. More than 90% of the dry weight is protein, comprising a variety of enzymes, non-enzymatic polypeptide toxins, and non-toxic proteins. Non-protein components include carbohydrates and metals (often part of glycoprotein

Fig. E15.1 **Sea snake: *Pelamis platurus*.**

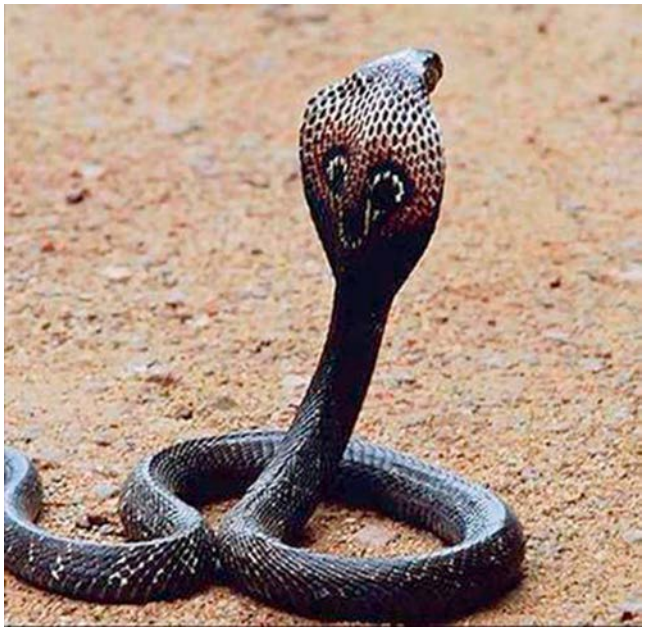


Fig. E15.2 Sri Lankan cobra: *Naja naja*.

Fig. E15.3 Indian krait: *Bungarus caeruleus*.

Fig. E15.4 Ceylon krait: *Bungarus ceylonicus*.

Fig. E15.5 Russell's viper: *Daboia russelli*.

Fig. E15.6 Hump-nosed viper: *Hypnale hypnale*.

metalloproteinase enzymes), lipids, free amino acids, nucleosides, and biogenic amines such as serotonin and acetylcholine.

- Proteinases act as cytotoxins, causing severe local swelling and tissue damage in the region of the bite.
- Neurotoxins interfere with neuromuscular transmission.
- Some phospholipase A2 toxins also directly damage myocytes and red blood cells (causing rhabdomyolysis and haemolysis).
- Some venoms contain haemorrhagins (metalloproteinases) that damage the vascular endothelium, causing systemic haemorrhage.
- Components of venom can act at many different sites in the coagulation pathways, resulting in anticoagulant or procoagulant effects.
- Venom components can also act directly on platelets.

Snake venoms are capable of producing adverse effects in one or more systems of the body. Most have multisystem effects. There are qualitative or quantitative variations in the chemistry of venoms in different snake species of the same family. Such differences may also occur within a species, depending on their geographical distribution.

Clinical features

The clinical effects of snakebite envenoming will vary with the species of snake, its age, size and geographical origin, the quantity of venom injected, the route of injection, the age, size and previous health of the victim, and past exposure to venom. Accurate identification of the biting species, ideally by examination of the dead snake, is crucial. Care should be taken; however, when examining a dead snake, as a reflex bite can still occur. Careful clinical evaluation of the patient is necessary before treatment is administered because bites by venomous species do not always result in envenoming ('dry bite').

Envenoming may cause both local and systemic effects. Elapid snakes (kraits, cobras) most commonly cause neurotoxicity whereas viperine snakes (vipers, pit vipers, rattlesnakes, adders) cause local tissue damage and coagulopathy. However, this distinction may be blurred; for example, Australasian elapids may cause both neurotoxicity and coagulopathy.

Local effects

- ***Oedema*** at the site of the bite, which may spread to involve the whole limb.
- ***Tissue necrosis*** at the site of the bite (Fig. E15.7).
- ***Painful regional lymph node enlargement***—a common early sign of envenoming.

Systemic effects

- ***Coagulopathy and haemorrhage*** include bleeding from the gums and other sites (Fig. E15.8). Bleeding may occasionally be life-threatening.
- ***Shock*** may occur as a result of loss of fluid into a limb or haemorrhage.
- ***Neurological involvement*** classically causes progressive descending paralysis. Ptosis (Fig. E15.9) occurs initially; this progresses with involvement of bulbar muscles and, ultimately, paralysis of the diaphragm.
- ***Skeletal muscle necrosis*** leads to painful myopathy, myoglobinuria, and paralysis. This complication is especially seen following envenoming by sea snakes.
- ***Renal failure*** occurs due to envenoming by several species, such as Russell's vipers.

Management

First aid

- Reassurance should be given, as even the most venomous species often bite without injecting enough venom to be harmful.
- Immobilization of the patient, especially of the bitten limb, to slow venom spread is the most important component of first aid.
- Pressure immobilization can reduce systemic spread of large-molecular-weight neurotoxins and delay the onset of life-threatening neurotoxicity. The technique requires training and the appropriate equipment in order to avoid aggravating soft tissue injury from local envenoming. The entire limb is bandaged 'as tightly as for a sprained ankle', starting around the fingers or toes and finishing at the axilla or groin using stretch bandages, crepe, and a splint. Pressure immobilization should be used in all cases unless the snake has not been identified or it is clear that the patient was bitten by a large viper, as development of severe local swelling may compromise the circulation.
- Transport of the victim to a medical centre should take place as quickly, safely, and passively as possible.
- Traditional methods, such as incision, suction, tourniquet, electric shock, cryotherapy, instillation of chemicals, and snake stones, are useless and harmful.
- Avoidance of aspirin and NSAIDs is important, as they may aggravate bleeding.
- The biting snake is valuable evidence but should not be pursued or handled carelessly.

Hospital treatment

Hospital treatment involves rapid clinical assessment, resuscitation, and administration of antivenom, the only specific treatment for envenoming. Wherever possible, the priority is to assess critically and stabilize vital systems. Attention to airway, breathing, and circulation should take priority.

Diagnosis

There are three principal diagnostic questions:

- Is this a snakebite?
- If it is, is there significant envenoming?
- What snake species was responsible?

In the history, three preliminary questions should be asked of the victim:

- Which part of your body was bitten?
- How long ago were you bitten?
- Have you brought the snake along, and if not, can anyone describe it?

Look for circumstantial evidence if the answers are not clear.

Immunodiagnostic tests that can identify the offending species by detecting specific snake venom antigens in wound swabs, serum, urine, or other body fluids are not widely available.

Fig. E15.7 **Severe local effects of envenomation.**

Fig. E15.8 **Bleeding following snakebite.**

Fig. E15.9 **Neurological symptoms following snakebite.**

Investigations

The white blood count is commonly elevated and thrombocytopenia is often seen, particularly in patients with coagulopathy. Renal function should be assessed. Creatine kinase levels are elevated when myotoxicity occurs. The 20-minute whole-blood clotting test (20 WBCT) is a good bedside indicator of envenoming by species that cause coagulopathy. Blood left in a clean, dry, glass tube is examined after 20 minutes to see whether clotting has occurred. The test may also be used to assess response to treatment with antivenom.

Antivenom

Antivenom is indicated for the treatment of systemic or severe local envenoming. A delay in administration reduces its efficacy, but antivenom may remain useful for as long as signs of systemic envenoming persist. The antivenom used depends on the biting species. In many parts of the world, polyspecific antivenoms covering locally occurring snake species are employed. This is because the biting species is only rarely identified. Monospecific/monovalent antivenoms can be used only if the snake has been positively identified, either by direct identification if the offending snake is brought to hospital or by detection of venom antigens in body fluids using immunodiagnostic tests.

Indications for antivenom include:

- spontaneous systemic bleeding
- incoagulable blood based on the 20 WBCT or other laboratory tests (e.g., international normalized ratio, INR)
- shock, indicated by low or falling blood pressure or cardiac arrhythmia
- ptosis, diplopia, neck muscle weakness, paralysis of limb muscles, respiratory paralysis, altered sensorium
- black urine positive for blood/haemoglobin/myoglobin (indicating rhabdomyolysis or intravascular haemolysis)
- local swelling involving more than half the bitten limb, swelling after bites on the fingers and toes, or swelling after bites by species likely to cause necrosis.

Antivenom should be administered by slow intravenous infusion. The appropriate dose is dictated by local experience. The initial dose of antivenom depends on the type of antivenom, the offending snake species, and the severity of envenoming. Children require the same dose as adults. Reactions to some preparations of antivenoms are common and include life-threatening anaphylaxis. Test doses of antivenom do not reliably predict these reactions and have no role in management. Adrenaline (epinephrine) must be kept available whenever antivenom is administered, to treat possible anaphylactic reactions.

Antivenom only neutralizes circulatory venom and will not neutralize or combine with venom once the venom is attached to target organs: that is, platelets, red blood cells, vascular endothelium, renal tubules, muscles, or neuromuscular receptors. Effective antivenom treatment should halt the progression of symptoms and reverse coagulopathy within 6 hours of administration. The dose of antivenom should be repeated if life-threatening bleeding, shock, or paralysis persists a few hours after the first infusion, or if the blood remains incoagulable when retested after 6 hours. Most postsynaptic neuromuscular junction neurotoxins (cobra and krait) are neutralized by antivenom, but presynaptic neurotoxins (krait) are less responsive and may need supportive therapy.

Supportive treatment and treatment of complications

- ***Shock*** usually results from hypovolaemia (extravasation of blood into the bitten limb) and should be corrected with intravenous fluids.
- ***Respiratory failure*** from respiratory muscle paralysis should be treated with assisted ventilation (common in krait bite).
- ***Paralysis*** due to postsynaptic neurotoxicity may respond to anticholinesterases such as neostigmine.
- ***Tetanus prophylaxis*** should be given as per requirement.
- ***Acute renal failure*** may need dialysis (common after Russell's viper bites).
- ***Surgical complications*** include necrotic tissue, which should be debrided and split skin grafts should be applied. Compartment syndrome is uncommon. Fasciotomy is seldom necessary, even when there is severe oedema, the only indication being raised intra-compartmental pressure. Surgery should not be performed until coagulopathy has been corrected.

Prevention

Avoid contact with snakes, even harmless or apparently dead ones. Wear protective clothing (boots not open sandals, socks, long

trousers) when walking in the undergrowth or deep sand and carry a light at night. Avoid blindly putting hands or feet into termite holes and crevices in trees, bushes, and under logs. Sleep off the ground and ideally under a mosquito net to prevent nocturnal bites (kraits). Fishermen should not handle sea snakes trapped in fishing nets.

Further reading

Isoardi KZ, Isbister GK. Poisoning by venomous animals. *Medicine* 2020; 48(3):220–223.
Ravikar R, et al. Managing snakebite. *BMJ* 2022; 376:e057926.

Plants

Life-threatening poisoning from plant ingestion is rare, though many plants contain potentially toxic substances. These include antimuscarinic agents, calcium oxalate crystals, cardiogenic glycosides, pro-convulsants, cyanogenic compounds, mitotic inhibitors, nicotine-like alkaloids, alkylating agent precursors, sodium channel activators, and toxic proteins (toxalbumins). While many plants contain gastrointestinal toxins, these rarely give rise to life-threatening sequelae. In contrast, other botanical poisons may cause specific organ damage, and death may occur from only small ingestions of yew (genus: *Taxus*), oleander (*Thevetia peruviana* and *Nerium oleander*), and cowbane *(Cicuta)*.

Atropa belladonna

Atropa belladonna (deadly nightshade) contains hyoscyamine (atropine) and hyoscine. It causes antimuscarinic effects—a dry mouth, nausea, and vomiting—leading to blurred vision, hallucinations, confusion, and hyperpyrexia.

Cicuta species

Cicuta spp. (water hemlock) and the related genus *Oenanthe* contain cicutoxin, a potent CNS stimulant that produces violent seizure activity. The CNS effects of cicutoxin are similar to those of picrotoxin, a known inhibitor of gamma-aminobutyric acid (GABA). Severe gastrointestinal symptoms, diaphoresis, salivation, and skeletal muscle stimulation may precede the seizure activity.

Conium maculatum

Conium maculatum (poison hemlock) contains a variety of volatile piperidine alkaloids, which have a toxic activity similar to that of nicotine. Large doses produce non-polarizing neuromuscular blockade, which may result in respiratory depression and death.

Datura stramonium

Datura stramonium (jimsonweed) and other *Datura* spp. contain hyoscyamine (atropine) and hyoscine. These alkaloids are potent antagonists of acetylcholine at muscarinic receptors and produce the anticholinergic syndrome. While morbidity is significant, fatalities are rare and are the consequence of hyperthermia, seizures, and/or arrhythmias.

Digitalis purpurea, Nerium oleander, Thevetia peruviana (yellow oleander)

Ingestion of *Digitalis purpurea* (Foxglove) or the common oleander *(Nerium oleander)* or yellow oleander *(Thevetia peruviana)* can produce a syndrome similar to digoxin poisoning (see 'Digoxin' section). A randomized controlled trial has shown that digoxin-specific antibody fragments can rapidly and safely reverse yellow oleander-induced arrhythmias, restore sinus rhythm, and rapidly reverse bradycardia and hyperkalaemia. The administration of multiple doses of activated charcoal is used but the effect on survival is debated.

Further reading

Singh M, Routledge PA. Poisoning by toxic plants and fungi. *Medicine* 2020; 48(3):218–219.

Mushrooms

Poisoning due to mushrooms is usually accidental, though ingestion of hallucinogenic ('magic') mushrooms is invariably intentional.

Cytotoxic mushrooms

Cytotoxic mushroom poisoning is caused by the amatoxins and orellanin. Amatoxins are found in *Amanita phalloides*, *A. virosa*, and *A. verna*, and in some *Galerina* and *Lepiota* spp. Amatoxins inhibit transcription from DNA to messenger RNA (mRNA) by the blockade of nuclear RNA polymerase II; this results in impaired protein synthesis and cell death.

Clinical features and management

Intense watery diarrhoea starts 8 to 24 hours after ingestion and persists for 24 hours or longer. Patients often become severely dehydrated. Signs of liver damage appear during the second day and hepatic failure may ensue. Impaired kidney function is often seen both because of fluid loss and as a result of direct kidney injury. In all patients, fluid, electrolyte, and acid–base disturbances should be corrected and renal and hepatic function supported. The value of silibinin and benzylpenicillin is not proven. Occasionally, liver transplantation is necessary.

Gyromitrin poisoning

Gyromitrin is found in *Gyromitra* spp., including in particular the false morel *(Gyromitra esculenta)* and *Cudonia circinans*. Gyromitrin decomposes in the stomach to form hydrazines that inhibit pyridoxine kinase and thus produce functional pyridoxine deficiency. GABA deficiency ensues since pyridoxine is an essential co-factor in GABA synthesis. Seizures may result. Hydrazines also cause haemolysis, methaemoglobin formation, and hepatorenal toxicity.

Clinical features and management

Vapours from the mushrooms are irritating to the eyes and respiratory tract. Gastrointestinal symptoms appear 5 to 8 hours after exposure. Vertigo, sweating, diplopia, headache, dysarthria, incoordination, ataxia, and seizures may follow. Symptomatic and supportive care is required. Pyridoxine 25 mg/kg as an infusion over 30 minutes should be given if severe CNS toxicity develops; repeat doses may be required.

Hallucinogenic mushroom poisoning

Psilocybin produces pharmacological effects similar to those of LSD and is found in *Psilocybe* and *Panaeolus* spp.

Clinical features and management

Symptoms occur within 20 to 60 minutes. Effects include altered time and space sense, depersonalization, hallucinations, derealization, and euphoria. Symptoms are usually maximal within 2 hours and disappear within 4 to 6 hours, though 'flashbacks' may recur after weeks or months. Anxiety and agitation should be treated with diazepam, 10 to 20 mg i.v., repeated as necessary.

Isoxazole poisoning

Isoxazoles (e.g., ibotenic acid, muscimol, muscazone) occur in *Amanita muscaria* and *A. pantherina*, and act as GABA agonists.

Clinical features and management

Nausea, vomiting, inebriation, euphoria, confusion, anxiety, visual disturbances, and hallucinations occur often within 30 minutes. Drowsiness is common and a coma-like state may ensue. Severe agitation and violent behaviour are seen occasionally. Other features include myoclonic jerks, muscle fasciculation, and seizures. Symptomatic and supportive care should be given as necessary. Diazepam 10 to 20 mg i.v., repeated as required, should be administered for anxiety, agitation, and seizures.

Neurotoxic mushroom poisoning

Muscarine is found in, for example, *Inocybe* spp., *Clitocybe* spp., and *Mycena pura*. Muscarine stimulates cholinergic receptors in the autonomic nervous system.

Clinical features and management

Diarrhoea, abdominal pain, diaphoresis, salivation, lacrimation, miosis, bronchorrhoea, bronchospasm, bradycardia, and hypotension occur. Atropine 0.6 to 2 mg i.v. should be given to manage the cholinergic syndrome.

Orellanin poisoning

Orellanin is a potent nephrotoxin found in, for example, *Cortinarius orellanus* and *C. speciosissimus*. A metabolite of orellanin inhibits protein synthesis in the kidneys.

Clinical features and management

Symptoms are typically delayed for 2 to 4 days. Some patients suffer a mild gastrointestinal disturbance before developing signs of renal impairment, headache, fatigue, intense thirst, chills, myalgia, and abdominal, lumbar, and flank pain. Transient polyuria with proteinuria, haematuria, and, characteristically, leucocyturia is followed by oliguria and then anuria. Renal function may recover only partially; chronic kidney disease is reported in about 10 to 40% of cases. Management involves careful monitoring and haemodialysis/haemofiltration if renal failure supervenes.

16 Environmental medicine

Michael Greene and Peter David Hodkinson

Clinical skills 371
Aviation and space 372
Fitness to fly 372
High altitude 374
Altitude illness 374
Cold environments 376
Hypothermia 376
Peripheral cold injury 378
Hot environments 378
Heat stroke 379
Water environment 379
Drowning 380
Diving 380
The physical world 381
Ionizing radiation 381
Lightning strike 382
Electric shock 383

CORE SKILLS AND KNOWLEDGE

Environmental emergency medicine studies the challenge to the human body when it interacts with the physical world. Healthcare professionals operating in these situations should be trained to perform in stressful conditions with minimal support and limited resources. Doctors specializing in this field often have backgrounds in emergency medicine, pre-hospital care, expedition, and military medicine. This chapter explores the physiology, pathophysiology, and management of environmental emergencies. It will consider when adaptation is possible and the nature of illness when the body is subjected to excessive environmental stress. Climate change is causing more extreme weather events, exposing whole populations to heat, cold, and flooding. Many of those most affected live in deprived communities. Ease of global transportation has brought people into environments with which they are unfamiliar.

Key learning outcomes include:

- Understanding how the physical environment impacts normal human physiology.
- Explaining how the body can acclimatize to different environments.
- Recognizing common illness and injury caused by excessive environmental stress.
- Formulating management plans for common environment medical emergencies.

The life-threatening conditions described in this chapter occur with different frequency. For example, drowning is a common worldwide problem, whereas acute high-altitude cerebral oedema (HACE) is rare. **Box 16.1** describes 10 fatal conditions in environmental emergency medicine.

Box 16.1 Fatal conditions in environmental emergency medicine

Environment	Life-threatening condition
Altitude	High-Altitude Cerebral Oedema High-Altitude Pulmonary Oedema
Water	Drowning Arterial Gas Embolus—in divers
Cold	Hypothermia
Heat	Heat Stoke Hyponatraemia—due to excessive water intake
Physical	Lightning Strike
Ionizing Radiation	Acute Radiation Sickness—in high dose exposure Increased risk of malignancy—medical radiation

CLINICAL SKILLS

The conditions seen in environmental medicine affect many body systems and require broad expertise in clinical medicine. Much of the initial management takes place in the pre-hospital outdoor environment. The practitioner must be able to assess the situation for danger and manage the associated risk. Training to work in different austere pre-hospital situations is essential (**Fig. 16.1**).

History

The diagnosis depends on a careful, detailed history. This must consider the environment and the patient's activity, as well as relevant features of the patient's previous medical problems and physical fitness. In a challenging pre-hospital setting, history-taking may be difficult; however (as in other fields of medicine), it is almost always more helpful for diagnosis than examination, and if curtailed by circumstances, it should be completed as soon as possible once the situation allows.

Fig. 16.1 **Rescue of a hypothermic patient.**

Box 16.2 Patient assessment in environmental medicine

Assessment of the scene
- Danger to yourself, others, casualty
- Weather
- Terrain

Primary Survey
- The aim is to identify life-threatening conditions, start initial essential treatment, and obtain baseline observations.
- Safety
- Patient Response: conscious/unconscious
- C—Catastrophic life-threatening bleeding
- A—Airway
- B—Breathing
- C—Circulation
- D—Conscious level
- E—(In this context) Environment protection.

History Taking
- A history of the events leading up to the injury or illness with a chronological account of progress
- Activity profile of the casualty, for example, dive profile, ascent profile
- Nutritional status—recent food and fluids
- Clothing and protective equipment
- Result of any recent acclimatization, for example, heat, altitude
- Previous exposure to the conditions, for example, altitude
- Current medication
- Immunization in foreign countries
- Medical insurance—evacuation, in-country treatment, repatriation
- The examination will focus on the signs most relevant to the case and the context.

Examination

This can be difficult and limited by the situation. Wind masks subtle breath sounds, vasoconstriction inhibits some cardiovascular signs, altitude alters physiological observations, and removing the patient's clothing for examination may not be possible in some situations. Box 16.2 outlines special considerations for patient assessment in environmental medicine.

There are often no investigations or imaging to support initial diagnosis or treatment decisions. Resources may be limited and successful management requires a flexible and pragmatic approach. The time to reach definite care can be hours or days. Good communication and team work are essential.

The integration of skills required for successful management in the outdoors is illustrated in Figs. 16.2 and 16.3.

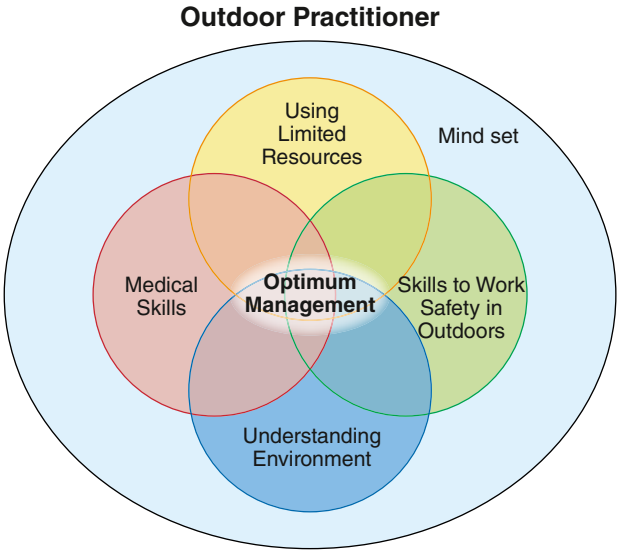


Fig. 16.2 **Skills required for environmental emergency medicine.**

AVIATION AND SPACE

Aviation and space medicine takes a holistic approach to consider the health, performance, and safety of those who travel in air or space. At the heart of aviation and space medicine is the management of human risk in the extreme environments of aviation and spaceflight. The main areas of activity are providing clinical medicine in a flight environment, understanding how flight or space travel may affect human performance, and human factors engineering (designing the human into the system).

The challenges vary between operational settings but include reduced barometric pressure (hypoxia, decompression sickness, gas volume changes), increased forces of acceleration, increased radiation exposure, thermal stress, motion sickness, noise, vibration, spatial disorientation, and circadian dysrhythmia. Additionally, in space, there are challenges associated with weightlessness that affect all body systems and drive a requirement for various countermeasures including regular resistance and aerobic exercise.

Environmental challenges in flight may cause new medical issues or exacerbate pre-existing conditions. As well as managing medical problems in individual patients, an assessment must be made of how a medical emergency might affect others on the flight: ***Flight safety critical emergencies*** describe situations where medical incapacitation of the individual might impact flight safety. They require the definition of acceptable medical risk and standards depending on the individual's role, for instance, as pilots, cabin crew, air traffic controllers, and drone or remotely piloted aerial system operators. ***Non–flight safety–critical emergencies*** affect passengers, aeromedical evacuation patients, and their clinical teams, where there is no threat to the safety of the whole flight.

Fitness to fly

Billions of people fly by air as passengers every year. Therefore, it is useful for all doctors to have some awareness of fitness-to-fly considerations for passengers and patients, although the ultimate decision on fitness to fly is the responsibility of the

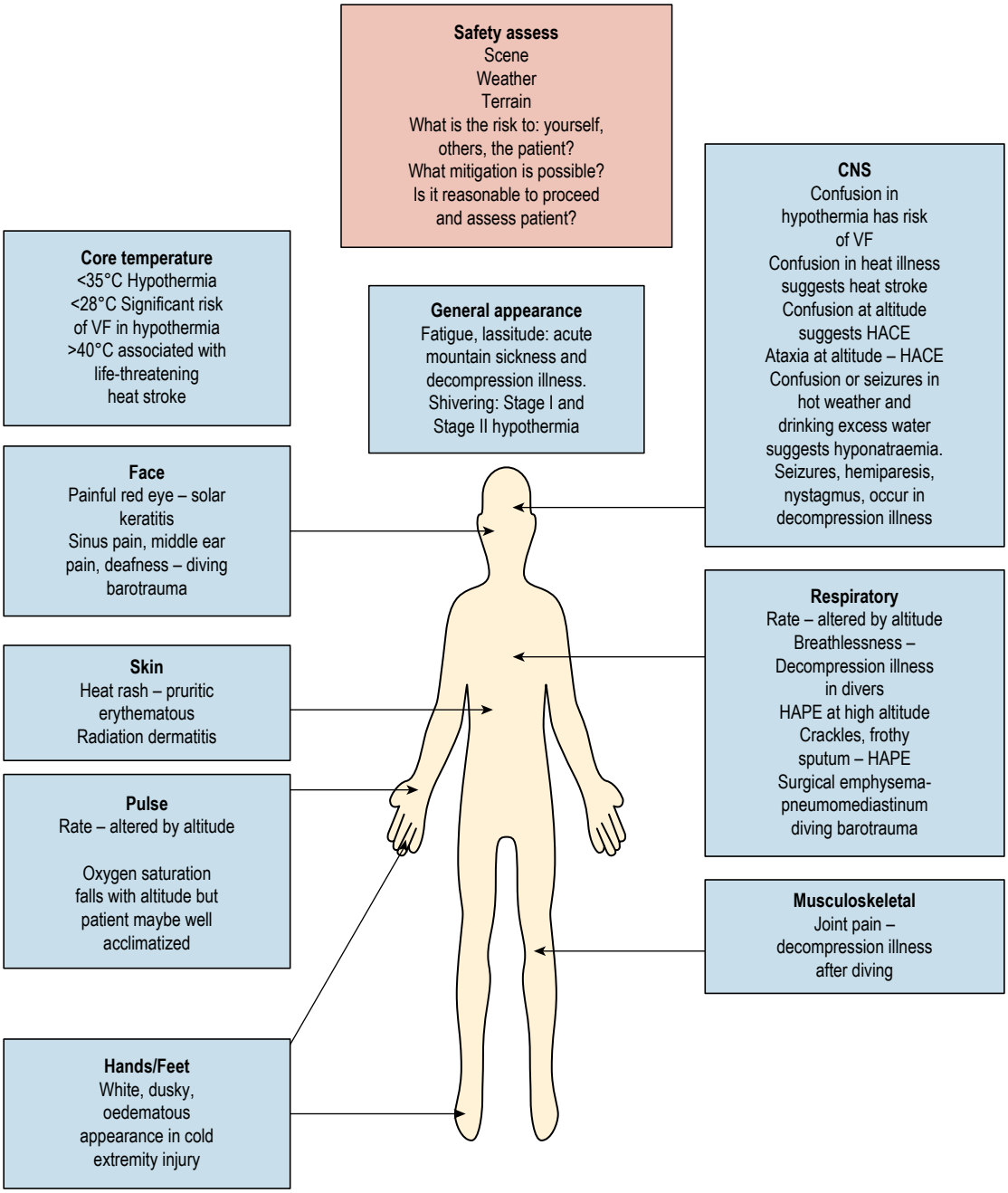


Fig. 16.3 **The examination of a patient in environmental medicine.**

airline, and patients with questions about fitness-to-fly should be directed to contact the airline medical officer. A patient's treating doctor may be asked, subject to patient consent, to complete a Medical Information Form (MEDIF) by the airline medical officer. This form is specific to each airline but describes the patient's clinical condition, course, treatment, and complications and is used to inform the fitness-to-fly decision. The treating doctor does not require any aviation medicine knowledge to fill it in. More information can be obtained by consulting their local aviation authority advice (e.g., UK Civil Aviation Authority website).

Pilots and cabin crew fitness-to-fly considerations are dealt with by doctors with specialist training in aviation medicine, called Aeromedical Examiners (AMEs), who work on behalf of the aviation regulatory authorities.

Specific considerations for air travel

- Standard travel health considerations such as vaccination or malaria prophylaxis should be considered, depending on destination and planned activities.
- Commercial aircraft expose people to an atmospheric pressure equivalent to approximately 1500 to 2500 m (5000–8000 feet). This induces a mild degree of hypoxia, meaning that patients with cardiorespiratory compromise may need to consider flying with supplemental oxygen, and some patients will need to travel with their non-invasive ventilation equipment. It is critical the

battery supply is sufficient to last 1.5 times the flight duration in case of delay or diversion.
- Deep vein thrombosis risk should be considered in passengers in the context of prolonged immobility on long haul flights.
- Patients should carry sufficient medication for the flight and a supply for use on arrival in case their hold luggage is lost. This may require that a doctor's letter be provided to accompany the medication at security checks.
- All travellers should be encouraged to have travel insurance to cover overseas medical costs in the event of a medical problem and to cover the cost of flight home, which might need the inclusion of an aeromedical escort team. Patients must declare pre-existing medical conditions for many of these travel insurance policies to be valid. Travel insurance companies have medical teams that can provide advice and help to arrange flights home.
- Medical conditions may restrict pilots from flying. Some of these will be similar to driving restrictions, for example, around epilepsy or diabetes. It is the pilot's responsibility to discuss new or ongoing medical conditions and treatments with regulatory authorities regarding their fitness to fly.

Astronauts are dealt with by flight surgeons with specialist space medicine training at government agencies such as the European Space Agency or in commercial companies such as SpaceX.

Further reading

British Thoracic Statement on Air Travel for Passengers with Respiratory Disease: https://www.brit-thoracic.org.uk/document-library/clinical-statements/air-travel/bts-clinical-statement-on-air-travel/.
Gradwell D, Rainford D. *Ernsting's Aviation and Space*. 5th edn. CRC Press Taylor and Francis Group. 2016.
Hodkinson PD, Anderton RA, Posselt BN and Fong KJ. An overview of space medicine. *Br J Anaesth* 2017; 119 (S1):i143–i153.
UK Civil Aviation Authority Website for Passengers: https://www.caa.co.uk/passengers/before-you-fly/am-i-fit-to-fly/guidance-for-health-professionals/.
Uk Civil Aviation Authority Website For Pilots, Cabin Crew, Air Traffic Controllers: https://www.caa.co.uk/aeromedical-examiners/.

HIGH ALTITUDE

Physiology

The challenges of high altitude are caused by the low partial pressure of gases in the atmosphere, cold, and increased exposure to solar radiation.

The partial pressure of oxygen decreases with a fall in barometric pressure. This is dependent on latitude, the season, and local weather conditions. Although the concentration of oxygen throughout the troposphere is 21%, barometric pressure directly determines the inspired PO_2. At higher altitudes as barometric pressure falls, there are fewer oxygen molecules in each breath, so less available to bind to haemoglobin, and tissues become hypoxic. This decreases exercise performance and increases risk of altitude illness. The general effects on the human body are summarized in **Box 16.3**, which lists the effects of ascent in altitude.

At the summit of Mount Everest (Chomolungma, Sagarmatha) (8848 m) the barometric pressure is 34 kPa (253 mmHg). Arterial blood sampled from climbers near the summit in 2007 demonstrated a PO_2 of 3.3 kPa (24.6 mmHg) in acclimatized climbers and an alveolar PO_2 of 4.0 to 4.7 kPa (30–35 mmHg)—near the limit of human physiology.

Box 16.3 Effects of ascent in altitude

Altitude above sea level	Effect on human body
High altitude, 2500–3500 m (European Alps)	Illness occurs with rapid ascent Decreased exercise performance
Very high altitude, 3500–5800 m (Mt Blanc—4809 m)	Illness will occur without acclimatization Decreased exercise performance Extreme hypoxia possible with exercise and sleep
Extreme altitude, >5800 m Mt Everest (Chomolungma, Sagarmatha)—88849 m	Complete acclimatization is generally not possible Rapid ascent causes severe and life-threatening disease Progressive physical deterioration

Box 16.4 Physiological changes with altitude acclimatization

System	Start duration	Adaptive change during acclimatization
Respiratory	Immediate Days	Increase in ventilation (hypoxic ventilatory response) with a fall in $PaCO_2$ and rise in PaO_2
Renal	24–48 h Days	The kidneys excrete bicarbonate to normalize the pH over 6–8 days
Cardiac	Immediate Days	There is an increased cardiac output; this returns to normal with acclimatization
Haemopoietic	Days Weeks	Increase in red cells and haemoglobin over 2–3 weeks, increase in 2,3 diphosphoglycerate with right shift of oxygen dissociation curve

Acclimatization

There are adaptations in many body systems during acclimatization to altitude that mitigate the effects of hypoxia caused by a fall in the partial pressure of oxygen. Despite this improvement, individuals still do not achieve the level of performance they could achieve at sea level. **Box 16.4** describes the physiological changes with altitude acclimatization.

Accent profiles

Acclimatization is supported by a suitable rate of ascent. The rate should be guided by the principle of ascending to a sleeping height of no more than 300 to 500 m per day (above previous sleeping height) when over 3000 m. Greater height can be gained during the day if the sleeping height is not exceeded. A rest day each 1000 m or 3 to 4 days should be added to allow for acclimatization.

Altitude illness

Clinical features

Acute mountain sickness (AMS), high-altitude cerebral oedema (HACE), and high-altitude pulmonary oedema (HAPE) are the three conditions associated with acute exposure to altitude.

People who normally reside at low altitudes and who ascend for leisure or work can be affected. The incidence depends on many factors including the rate of ascent, previous exposure to high-altitude environments, duration of stay, and genetic factors. Altitude illness usually occurs at more than 2500 m but can affect those with pre-existing conditions at lower altitudes. All these conditions respond to descent in altitude.

Acute mountain sickness

AMS is a clinical syndrome characterized by the symptoms of headache, plus fatigue, nausea, anorexia, lassitude, or poor sleep. This generally starts within 4 to 6 hours of arrival at altitudes above 3000 m and occurs in 10% to 25% of unacclimatized people travelling to areas above 2500 m. Treatment is rest, simple analgesia, and no further ascent. Symptoms usually resolve within 24 to 36 hours. These symptoms can mimic other conditions such as dehydration, carbon monoxide poisoning, and migraine, but at altitude, they should be considered to be AMS until proven otherwise. If symptoms persist or progress, descent is indicated. Prophylactic treatment using acetazolamide has been shown to reduce symptoms but is not a substitute for slow ascent and physiological acclimatization. Acetazolamide is a carbonic anhydrase inhibitor that promotes ventilation by inducing a mild metabolic acidosis through increasing urinary bicarbonate excretion. Dexamethasone prevents AMS and can reduce symptoms but does not facilitate acclimatation. Supplementary oxygen relieves symptoms.

Acute high-altitude cerebral oedema

This is a life-threatening illness with a similar pathophysiology to AMS. It is the least common of the three conditions, but death can occur within hours of onset. There is cerebral dysfunction often with ataxia and a change in the level of consciousness; coma and death may follow. Immediate descent is indicated and is lifesaving. Dexamethasone and oxygen should be administered if available. If descent is impossible (weather, terrain), a portable hyperbaric chamber may be used while rescue is organized. Long-term effects such as ataxia and cognitive impairment have been reported.

Acute high-altitude pulmonary oedema

This is a life-threatening condition of the lungs with raised pulmonary venous pressure and impaired alveolar gas transfer. Some individuals are HAPE susceptible. Untreated HAPE has a 50% mortality. It typically occurs 2 to 5 days after ascent and is characterized by breathlessness at rest, decreased exercise tolerance, tachypnoea, and crackles in the lungs. There may be a dry cough, followed by the production of frothy sputum. Immediate descent is indicated. Oxygen should be given if available. The use of nifedipine, sildenafil, and beta-agonists in the acute phase has been shown to be beneficial.

The portable high-altitude chamber

A portable chamber is made from an air-tight material and fitted with a pressure release valve that can be inflated to higher than external atmospheric pressure (Fig. 16.4). It can be used in high-altitude emergencies as a temporary life saving measure while evacuation to a lower altitude is organized. Pumping the chamber generates a pressure of 13.7 kPa above ambient pressure and simulates a descent of about 900 to 1200 m.

Fig. 16.4 **A portable high-altitude chamber in use in Nepal.**

Other effects of high altitude

Solar keratitis

'Snow blindness' is a painful burn of the cornea caused by exposure to UV light without adequate protection (Fig. 16.5). It renders sufferers 'snow blind' and makes movement impossible on difficult terrain. Recovery is spontaneous over 2 to 3 days and analgesia is likely to be required. Prevention is by wearing sunglasses.

Retinal haemorrhage

Small flame haemorrhages within the retinal nerve fibre layer are common above 5000 m and often asymptomatic. Rarely will a haemorrhage cover the macula, causing painless loss of central vision. Recovery is usual.

Physical deterioration

Prolonged residence between 6000 and 7000 m leads to weight loss, anorexia, and listlessness after several weeks. Over 8000 m, the effects of deterioration become apparent over several days.

Chronic mountain sickness

Chronic mountain sickness occurs in long-term residents at high altitudes, usually after several decades, and is seen in individuals living in the Andes and in Central Asia.

Polycythaemia, lassitude, cyanosis, finger clubbing, congested cheeks and ear lobes, and right ventricular enlargement develop. Chronic mountain sickness is gradually progressive.

Further reading

Luks AM, Auerbach PS, Freer L et al. Wilderness medical society clinical practice guidelines for the prevention and treatment of acute altitude illness: 2019 update. *Wilderness Environ Med* 2019; 30(4):S3–S18.
Luks AM, Swenson ER, Bärtsch P. Acute high-altitude sickness. *Eur Respir Rev*: 201 Luks A. 7;26(143):160096.
Ward, Milledge and West's High Altitude Medicine and Physiology. 2020.

Fig. 16.5 **A porter with solar keratitis in Nepal.**

COLD ENVIRONMENTS

Physiology

The effects of cold can be divided into hypothermia (cooling of the whole body) and peripheral cold injury. There is limited physiological adaptation to cold, and prevention depends on appropriate clothing, food, and behaviour. Heat is lost to the environment by conduction, convection, radiation, and evaporation. The lowest temperature from which an adult has recovered from accidental hypothermia with good neurological outcome is a core body temperature (T_{Core}) of 13.7°C.

Hypothermia

Hypothermia is a fall in T_{Core} to less than 35°C below which normal muscular and brain function are impaired. It occurs when the cold challenge is greater than heat production and retention. Shivering is effective and increases temperature by 3–4°C per hour. However, it uses energy reserves and oxygen and stops below a T_{Core} of 30°C. Once shivering stops, the patient cannot independently restore their temperature. Brain metabolism falls by approximately 10% for each 1°C decrease in temperature. The faster this occurs, the greater the chance of a good neurological outcome. Hypothermia occurs in both urban and austere environments.

Urban

Hypothermia can occur if there is poor heating, inadequate clothing, and poor nutrition. Depressant drugs, alcohol, or intercurrent illness (such as hypothyroidism) may also contribute. It is commonly seen in social deprivation, as well as in the frail and the elderly. Older people have a diminished sense of cold and may have little insulating fat. Neonates and infants become hypothermic rapidly because of a proportionally larger surface area and less subcutaneous fat.

Austere environments

Accidental hypothermia

This occurs in challenging environmental conditions and is a danger for climbers, skiers, the military, and workers in the outdoors. Wet, cold conditions, with wind chill, physical exhaustion, injuries, and inadequate clothing, are contributory.

Cold water immersion

Hypothermia can develop following immersion for over 30 minutes in temperatures of 15 to 20°C. In water below 12°C the limbs rapidly become numb and weak. Cooling of the muscles and nerves causes weakness and loss of coordination. Drowning occurs from swim failure.

Clinical features of hypothermia

Mild hypothermia (T_{Core} 32–35°C) causes shivering, which can be painful. As the T_{Core} continues to fall, moderate hypothermia (28–32°C) causes impaired judgement, including lack of awareness of cold and drowsiness. In severe cases the risk of ventricular fibrillation increases significantly below T_{Core} of 28°C. Coma follows and death is due to cardiorespiratory arrest from dysrhythmia.

Pulse rate and blood pressure fall in hypothermia, but vital signs change at different absolute temperatures in different individuals. Cardiac output and cerebral blood flow fall, and respiration becomes shallow and slow. Muscles are stiff and there is a loss of coordination. In severe hypothermia, as coma ensues, pupillary and other brainstem reflexes are lost, and pupils become fixed and dilated. There is a shift of the oxygen dissociation curve to the left because of the reduction in the temperature of haemoglobin. An arterial blood sample from a hypothermic patient must be corrected for the patient's temperature prior to analysis. Dysrhythmias are common; atrial fibrillation, prolongation of the PR and QT intervals, and ventricular dysrhythmias all occur. Bradycardia with 'J' waves are pathognomonic of hypothermia but is uncommon.

Diagnosis

In rescue situations, the diagnosis of hypothermia is usually clinical because low-reading thermometers are rarely available and can be unreliable in field conditions. The Swiss Hypothermia Staging method describes clinical features that guide pre-hospital decisions and treatment. In the hospital a low-reading thermometer such as temperature probe placed in the lower third of the oesophagus should be used to measure core temperature. **Box 16.5** outlines the Swiss staging of hypothermia.

Field management

Swiss stage I

The patient is managed in a sheltered environment. Protection from wind and rain must be provided, wet clothing should be removed (if practical), and extra dry clothing should be added for insulation. Food and fluid should be provided, and the patient allowed to rewarm. Warm drinks do not warm the patient but boost morale and provide calories. There is little heat loss from radiation, because of cold vasoconstriction; therefore foil blankets may provide some shelter but do not rewarm the patient. Such patients may be able to self-evacuate to a safer situation.

Swiss stage II

Patients should be encouraged to stop exercise. Without shivering, these people cannot maintain a normal core

Box 16.5 The Swiss staging of accidental hypothermia

Swiss stage	Clinical features	Temperature range	Risk of hypothermic cardiac arrest
Stage I (Mild)	Conscious (GCS = 15, AVPU = A). Shivering	35–32°C	Low
Stage II (Moderate)	Impaired conscious level (GCS < 15, AVPU = Confusion, V or P). Shivering may or may not be present	<32–28°C	Moderate
Stage III (Severe)	Unconscious. Vital signs present on 1 min check	<28°C	High
Stage IV (Severe)	Apparently dead. No vital sign present on 1 min check	Variable <24°C	Very High

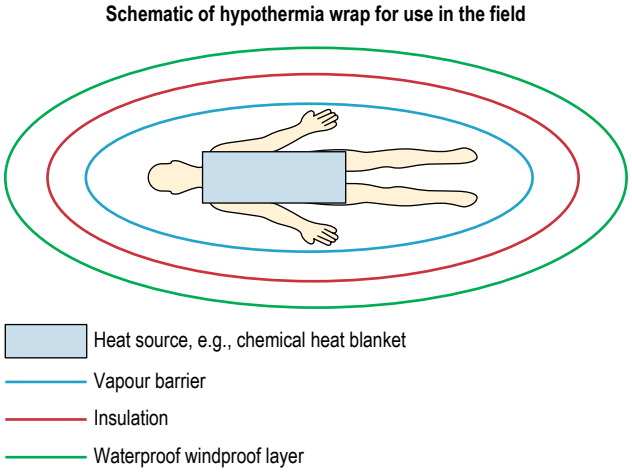


Fig. 16.6 Hypothermia wrap.

temperature. They should be evacuated in a multilayer wrap that provides shelter from the elements, insulation, and a vapour barrier (Fig. 16.6). The latter will minimize the energy loss from the heat used to evaporate water. Heat sources can be added to the wrap to minimize the temperature drop that occurs after insulation.

Swiss stage III

In severe hypothermia (T_{Core} < 28°C), people look dead with poor, difficult-to-detect vital signs. These patients are at risk of cardiopulmonary arrest. They should be transported in a hypothermia wrap with gentle movement to a hospital, preferably with rewarming and intensive care unit facilities.

Swiss stage IV

These patients have no discernible vital signs after 1 min of assessment. Start standard CPR and continue to hospital, preferably with extracorporeal cardiac life support (ECLS) facilities. The use of a mechanical cardiopulmonary resuscitation device will facilitate prolonged CPR during transfer. Survival is possible in some patients after prolonged resuscitation and active rewarming to good neurological outcome. The longest period of CPR for a hypothermic patient with full recovery is 6.5 hours in the context of an initial T_{Core} of 23.2°C.

Hospital treatment

In the hospital, core temperature should be measured. Patients should be rewarmed in a monitored environment. Passive rewarming with blankets, forced warm air, etc., is effective in mild and moderate cases. In severe hypothermia, rewarming has been successfully undertaken with invasive techniques such as peritoneal, thoracic, or bladder lavage when resources are limited. However, ECLS such as cardiac bypass or extracorporeal membrane oxygenation (ECMO) are preferred and improve survival. The Hypothermia Outcome Prediction after Extracorporeal life support (HOPE) score calculates the probability of survival for patients in hypothermic cardiac arrest and guides decision-making. Give warm intravenous fluids slowly. Correct metabolic abnormalities and check for underlying medical causes.

Further reading

HOPE Score: https://www.hypothermiascore.org/.
Paal P, Pasquier M, Darocha T et al. Accidental Hypothermia: 2021 Update. *Int J Environ Res Public Health*. 2022; 19(1):501.

Avalanche medicine

In Europe and North America, avalanche kills 150 people each year. Many are the result of recreational activities (skiing, climbing, snowmobiles), but avalanches also bury people on roads and in villages.

A person is said to be buried if the head and chest are under the snow. Breathing may still be possible if the airway remains open, the chest can expand, and there is sufficient gas diffusion through the snowpack for oxygenation and CO_2 disbursal. Gas diffusion depends on snow density. Death in the first few minutes is due to asphyxia or trauma, and survival depends on immediate rescue by bystanders. Survival falls rapidly after 10 to 15 minutes, and these survivors often have poor neurological outcome because of hypoxia. As time progresses, hypoxia, hypercapnia, and hypothermia become increasingly important. However, if the victim has been able to breathe under the snow and become hypothermic before cardiac arrest occurs, then survival with good neurological outcome has been recorded more than an hour after burial.

Management

If the victim is in cardiac arrest in the first 60 minutes, asphyxia is the likely cause and standard CPR methods should be employed. After 60 minutes hypothermia is possible, and in cardiac arrest, continuous resuscitation and transfer to a facility with ECLS should be considered.

Further reading

Pasquier M, Strapazzon G, Kottmann A et al. On-site treatment of avalanche victims. Scoping review and 2023 recommendations of the international commission for mountain emergency medicine (ICAR MedCom). *Resuscitation* 2023; 184:10970.

Peripheral cold injury

Extremity injury due to cold exposure is often associated with hypothermia.

Non-freezing cold injury

Non-freezing cold injury (NFCI) describes extremity injury caused by exposure to wet and cold conditions above freezing temperatures. It is seen in outdoor sportspersons and rough sleepers. In military personnel it has been recognized ('trench foot') from the time of Napoleon's campaigns (1812). The long-term sequelae can be very disabling.

Phases of injury

- Pre-hyperaemic: hours or days, when extremities are cold, pale, cyanotic, and oedematous, with sensory changes.
- Hyperaemic: 10 days, when there is sensory disturbance, weakness, atrophy, and anhidrosis.
- Post-hyperaemic: months or years, characterized by cold sensitivity, hyperhidrosis, infection, and trauma.

Chilblains (pernio) occur on exposed skin as blue plaques which may ulcerate and last for 1 to 2 months.

Management

- Treat underlying hypothermia
- Elevate limb
- Provide a cool, dry air flow
- Analgesia
- Long-term foot care and chronic pain management.

Freezing cold injury

This occurs in tissues exposed to freezing temperatures below 0°C.

Frostnip

A superficial injury with transient pallor, numbness, and pain that resolves on rewarming with no significant long-term damage.

Frostbite

The phases of injury are characterized as follows.

- Pre-freeze: cooling, vasoconstriction, and ischaemia
- Freeze/thaw: intra- or extracellular ice crystal formation causing cellular damage. Thawing initiates a reperfusion injury and inflammatory reaction
- Vascular stasis: sludging, coagulation, and ischaemia
- Late ischaemic: infarction, with an inflammatory cascade leading to microcirculatory damage and cell death

Prevention

Human tissue freezes at –5°C. Injury is determined by the type and duration of cold insult.

At-risk individuals should maintain their body core temperature, adequate nutrition, and hydration. Avoid restriction to blood flow caused by tight clothing or footwear and insulate all exposed skin. Supplementary oxygen can protect high-altitude climbers.

Clinical features

Frostbite is most common in the hands and feet (90%); the remainder is mostly on the face, buttocks, and thighs. The extremity is white and firm with numbness (75%) and sensory changes (10%). There is severe pain with reperfusion. Superficial clear fluid vesicles appear in 20% of frostbite and haemorrhagic bullae in 30%. Deep frostbite involves subcutaneous connective tissues, muscle, and bone.

Classification

Clinical classification describes the depth of injury. Plain radiographs will only show abnormalities after several weeks. Bone scintigraphy can help to predict outcome and monitor progress.

Management

- Treat hypothermia.
- Do not rewarm in the field if there is any possibility of refreezing, in order to prevent reperfusion injury. In a suitable environment, rewarm in a water bath with moving water at 37 to 39°C.
- Analgesia.
- Elevate.
- Dry, sterile, non-constricting dressings.
- Aloe vera topical cream decreases prostaglandin formation and improves superficial injury.
- Non-steroidal anti-inflammatory drugs block the arachidonic acid pathway and reduce inflammation; early low-molecular-weight dextran may decrease microthrombi; antibiotics are indicated if there are signs of infection.
- Tissue plasminogen activator (tPA) and the prostacyclin analogue iloprost have been used with some success and must be supervised by a frostbite expert.

Surgery

In the absence of life- or limb-endangering infection, surgery should be delayed. Tissue salvage often exceeds initial estimates. Consult with a surgeon with expertise in frostbite.

Further reading

Brugger H, Zafren K, Festi L. *Mountain Emergency Medicine.* 2021. Edra. S.p.A. Chapter 21.

Imray C, Grieve A, Dhillon S. Cold damage to the extremities: frostbite and non-freezing cold injuries. *Postgrad Med J* 2009; 85(1007):481–8.

HOT ENVIRONMENTS

Physiology

Body core temperature is maintained at T_{Core} 37°C by the thermoregulatory centre in the hypothalamus, which integrates information from skin temperature sensors with core temperatures from receptors in the brain and in the walls of large blood vessels.

Heat is produced by cellular metabolism, is dissipated through the skin by vasodilatation and sweating, and is expired in air via the alveoli. Heat is lost to the environment by conduction, convection, radiation, and evaporation. Sweat evaporation is the principal mechanism for controlling temperature following exercise or in response to an increase in environmental temperature. The elderly and those with chronic disease, whose ability to adjust physiologically to heat is reduced due to diminished thermoregulatory capacity, are more susceptible. The rising global temperature has given rise to heat waves and increased mortality in these groups.

Heat acclimatization

Acclimatization occurs over several weeks and should be guided by workload. Sweat volume increases and the salt content falls. Cardiovascular and renal compensations also occur.

Heat illness

Exertional heat illness occurs when the body cannot regulate temperature in the prevailing environmental conditions or during heat-producing activity (e.g., sporting, work, and military events). In tropical areas other causes for fever, for example, malaria, may need to be excluded.

Prickly heat rash (miliaria rubra)

This is characterized by a pruritic, erythematous rash. Sweating may be limited by plugging or damage to the sweat ducts. It may be prevented by using loose cotton clothing and frequent cool showers. Treatment includes cool bathing and talcum powder.

Heat cramps

These are painful muscle cramps, usually in the legs, that often occur in fit people when they exercise excessively, especially in hot weather. Two theories have been proposed for heat cramps. The serum electrolyte and dehydration theory is traditional, but recent evidence supports the altered neuromuscular control theory. In a hot environment, fatigue and pain increase muscle spindle activity. Such cramps are self-limiting and respond to muscle stretching, massage, and rehydration.

Heat syncope

Peripheral vasodilatation can reduce central venous return and lead to syncope on standing or with effort. This recovers with rehydration and rest in the supine position.

Heat exhaustion

This is mild to moderate heat illness due to exposure to a high environmental temperature or strenuous exercise. Clothing that inhibits sweating, such as a wetsuit or military uniform, can cause an elevation in T_{Core} in less than 15 minutes. Clinical features are weakness, dizziness, syncope, nausea, and vomiting. The temperature can be normal or moderately elevated. Untreated heat exhaustion may progress to heat stroke which is a medical emergency.

Management

Reduce environmental temperature (T_{Env}) if possible, and cool the patient with sponging and fans. Oral rehydration is indicated, but in severe heat exhaustion, intravenous fluids may be needed. Monitor serum electrolytes.

Heat stroke

Heat stroke is an acute life-threatening condition in which neurological features are the main diagnostic features. T_{Core} usually rises above 40°C. There is headache, nausea, vomiting, and weakness, progressing to confusion, seizures, coma, and death. The skin feels intensely hot to the touch. Sweating is often but not invariably absent. Heat stroke can develop in unacclimatized people in hot, humid, windless conditions, even without exercise. Excessive exercise in inappropriate clothing can lead to heat injury in temperate climates. Diabetes, alcohol, and drugs such as antimuscarinics, diuretics, and phenothiazines can contribute. Heat stroke can lead to a fall in cardiac output, lactic acidosis and intravascular coagulation, and multiorgan failure. Morbidity and mortality are directly related to the duration of the high T_{Core}.

Prompt treatment is essential and can be curative, even with a T_{Core} higher than 40.5°C.

Complications are hypovolaemia, intravascular coagulation, cerebral oedema, rhabdomyolysis, and renal and hepatic failure.

Prevention

Acclimatization, fluids, appropriate clothing, and modifying activity are all required.

Management

The priority of treatment is to reduce the T_{Core} below 40.5°C as soon as possible.

No pharmacological agents are effective in lowering core temperature:

- Reduce the environmental temperature, if possible, by removing the patient to a cool and shady location.
- Reduce T_{Core} urgently—arrange cold water immersion if available. Otherwise, cool the patient with sponging, ice packs, or fanning.
- Move the patient to a medical facility. Manage in intensive care: monitor cardiac output and respiratory, renal, and hepatic function. Measure biochemistry, clotting, and muscle enzymes.
- Give fluids intravenously so that the intravascular volume remains normal.

Hyponatremia

This is recognized in endurance athletes and others who overhydrate in hot climates. Arginine and vasopressin promote fluid retention and there is a loss of sodium from natriuretic peptide. The clinical features are headache, nausea, and fatigue, followed by confusion, fitting, and encephalopathy. Early diagnosis and treatment with hypertonic fluid can be life saving. Prevention is helped by drinking to thirst.

Further reading

Ballester, J., Quijal-Zamorano, M., Méndez Turrubiates, R.F. et al. Heat-related mortality in Europe during the summer of 2022. *Nat Med* 2023.
Hew-Butler T, Rosner MH, Fowkes-Godek S et al. Statement of the Third International Exercise-Associated Hyponatremia Consensus Development Conference, Carlsbad, California, 2015. 2015; 25(4):303–20.
Lipman GS, Gaudio FG, Eifling KP, Ellis MA, Otten EM, Grissom CK. Wilderness medical society clinical practice guidelines for the prevention and treatment of heat illness: 2019 update. *Wilderness Environ Med* 2019; 30(4): S33–S46.

WATER ENVIRONMENT

Oceans, lakes, and rivers are complex environments. Human interaction occurs at the water's edge or in the water. Water-borne infections such as schistosomiasis, onchocerciasis, and gastrointestinal infections affect millions of people around the world. Marine and bankside wildlife such as sharks, snakes, and crocodiles injure or kill people each year. This section reviews drowning, immersion, and diving emergencies.

Immersion

The first phase of immersion in cold water is the cold shock response. This is mediated by skin temperature receptors and causes involuntary gasping breaths, a loss of control of breathing, and an increased cardiac load. This lasts for about 90 seconds. Many victims inhale enough water (1.5–3 L of water) to die from early asphyxiation. In this phase, victims should be encouraged to 'float to survive'. Early rescue efforts should be by reaching or throwing buoyant objects and not by entering the water.

The second phase lasts 20 to 30 minutes and is the result of continuing superficial cooling of the peripheral nerves and muscles. There is a loss of muscle dexterity and strength so that the victim cannot help themselves. This period ends with swim failure.

The third phase is after 30 minutes and is the onset of hypothermia. There is progressive disorientation, a fall in level of consciousness, and ventricular fibrillation below 28°C.

Drowning

Drowning is defined as respiratory impairment from submersion in a liquid. The outcome is survival, with or without morbidity, or death. It is a global health problem, with 1000 deaths a day, mainly of children and adults less than 30 years old.

The majority of drowning deaths are due to asphyxia rather than hypothermia. Resuscitation from hypoxic drowning arrest is unlikely after 10 minutes and has not been recorded after 20 minutes. However, if the brain has cooled to below 30°C before cardiac arrest outcome this can improve the hypoxic survival time. The longest recorded underwater survival time with full recovery was recorded in a 7-year-old child with a temperature of 13.8°C who was immersed in icy water for 83 minutes.

A rescued patient may appear well but have continuing injury to the lungs for 4 to 6 hours after an event. In the presence of cough, chest pain, or any sign of aspiration, hospital observation for at least 6 hours post-event is recommended.

The difference between saltwater and freshwater immersion events does not alter the initial treatment. All patients are hypoxic and hypercapnic. Early rescue breaths and oxygen should be administered. Ventricular fibrillation only occurs in 10% of drowning events, and oxygenation and ventilation are essential initial treatment.

Of rescued victims, 20% die around the time of rescue, called circum-rescue collapse. If the victim is lifted vertically from the water, the relative hypovolaemia caused by cold water immersion can cause a collapse in arterial blood pressure and cardiac arrest. Victims should be lifted into the rescue boat or helicopter horizontally and transported in an orientation to assist venous return.

Further reading

Bierens JJLM, Lunetta P, Tipton M, Warner DS. Physiology of drowning: A review. *Physiology (Bethesda, Md)* 2016; 31(2):147–66.
Schiermeier Q. Increased flood risk linked to global warming. *Nature (London)*. 2011; 470(7334):316.
Szpilman D, Bierens JJLM, Handley AJ, Orlowski JP. Drowning. *N Engl J Med* 2012; 366(22):2102–10.
Szpilman D, Morgan PJ. Management for the drowning patient. *Chest* 2021; 159(4):1473–83.

Diving

Diving below the surface of the water can be classified as follows.

- Free diving: this involves breath-holding usually to a depth of about 5 m; this has been practiced for more than 500 years.
- Snorkel diving: limited to 0.5 m by the inspiratory effort required to breathe through a tube reaching the surface at that depth.
- Open-circuit scuba diving: usually to depths above 30 m, breathing oxygen or oxygen–nitrogen mix from tanks. Mortality rate 1.5–9 per 100 000 dives.
- Technical diving: deeper or commercial diving involving complex gas mixtures or rebreathing circuits that allow deeper dives and longer time at depth.

Divers can suffer from heat illness, barotrauma, decompression illness, trauma, and drowning.

Decompression illness

Decompression illness (DCI) results from the volume–pressure changes within the air-filled cavities in the body and the dissolution of gases in the tissues. Water is denser than air, and each 0.3 m of seawater exerts an additional 0.021 bar on the diver. The gases in a diver's breathing mixture are dissolved in proportion to their partial pressures. The quantity is determined by the length of the dive and gas solubility. During descent, nitrogen and helium become dissolved in the tissues. If the diver ascends too rapidly, gas bubbles form in the intravascular or extravascular spaces. The bubbles initiate an inflammatory cascade, thrombosis, ischaemia, and hypoxia. This presents as clinical syndromes depending on the tissues affected. Decompression tables indicate the time required to return safety from a given depth to the surface. In general, decompression is unnecessary for dives above 30 m. The incidence of DCI is 2.8 cases per 10 000 dives. Classification is usually descriptive, although some countries classify Type I as affecting the musculoskeletal system, skin, and lymphatics and Type II as affecting all other systems. **Box 16.6** highlights the signs and symptoms of decompression illness.

Box 16.6 Signs and symptoms of decompression illness

Body system	Symptoms and signs of decompression problem
Musculoskeletal	Limb or joint pain: dull pain ('toothache') improved with activity
Lungs	Breathlessness
Central nervous	Seizures, hemiparesis, loss of consciousness (cerebral arterial gas embolism) Vertigo, nystagmus, vomiting
Spinal cord	Paraplegia, lower limb paralysis or weakness, a sensory level or bladder dysfunction
Cognitive	Behavioural and personality changes
Skin	Cutaneous itchy rash (cutis marmorata)
General systemic features	Fatigue, malaise, nausea, anorexia

Symptoms may occur within minutes or several hours after a dive. During the 24 hours after a dive, further ascent such as air travel can provoke DCI.

Treatment

- Basic resuscitation following an ABCDE approach.
- Give oxygen as close to 100% as possible.
- Lie flat.
- Keep cool but avoid hypothermia.
- Give oral or i.v. fluids for rehydration.
- Take expert advice.
- Arrange for emergency recompression as required.

Barotrauma

The volume of a gas-filled space varies inversely with pressure (Boyle's gas law). On descent, air-containing spaces such as the middle ear and facial sinuses will be compressed unless the pressure between the body cavity and the pressured gases breathed in can be equalized. **Box 16.7** describes signs and symptoms of barotrauma.

Arterial gas embolism

This severe form of barotrauma is the second leading cause of death in sports divers (30%).

Gas bubbles are forced across the alveolar-capillary membrane and into the pulmonary venous circulation. Embolization to the coronary arteries causes cardiac ischaemia, infarction, and dysrhythmias. Cerebral emboli cause symptoms and signs similar to acute stroke.

Gas problems

Hypoxia: This is usually caused by a failure of oxygen supply or during technical dives using rebreathers. Treatment is with oxygen.

Nitrogen narcosis: at depths below 30 m, nitrogen exerts a high partial pressure and has an effect similar to an anaesthetic. Symptoms include euphoria ('rapture of the deep'), loss of coordination, loss of motor control, loss of judgment, and loss of consciousness. This may be avoided by replacing air with a helium and oxygen mixture for deeper dives. It resolves on ascent to surface.

Box 16.7 Signs and symptoms of barotrauma

Anatomical area	Symptoms and signs of barotrauma
The middle ear	Pain due to failure to equalize pressures through the eustachian tube. The tympanic membrane may perforate
Injury to the inner ear	Injury to the vestibular apparatus gives rise to vertigo ('the staggers'), nausea, and tinnitus. Cochlea damage causes deafness
Facial sinus	Pain and epistasis
Dental	Pain beneath teeth that have cavities
Pulmonary	Pneumothorax (chest pain and breathlessness) and pneumomediastinum (chest pain, neck swelling, voice change). Gas entering the pulmonary vasculature causes coronary artery or cerebral gas embolus (see below)

Hypercapnia: breathing is normally controlled through the partial pressure of CO_2. Divers who are exercising and trying to conserve their oxygen supply can become hypercapnic, causing headache, palpitations, and drowsiness.

Oxygen toxicity: Oxygen breathed at high partial pressures for an extended time can be toxic to the central nervous system and lungs; therefore pure oxygen is not used for diving. Toxicity is rare if diving on air because the threshold for toxicity is 66 m. Pathologically, there is lung atelectasis, endothelial cell damage, and pulmonary oedema when alveolar pressure exceeds 1.5 atmospheres. Symptoms include visual problems, tinnitus, nausea, agitation, confusion, muscle twitching, and convulsions.

Recompression therapy

Advice is available from specialized diving emergency centres that give clinical advice and help to organize recompression. Recompression should be performed as soon as possible. Recompression follows standardized tables and is guided by clinical response. Recovery is usual. Rare long-term problems such as avascular necrosis of the hip from infarction can occur; some chronic focal neurological damage has been reported.

Further reading

Divining Diseases Research Centre. https://www.ddrc.org/ Plymouth. UK.
Simon J Mitchell, Michael H Bennett, Phillip Bryson et al. Pre-hospital management of decompression illness: expert review of key principles and controversies. *Diving Hyperb Med*, 2018; 48(1).

THE PHYSICAL WORLD

Ionizing radiation

Naturally occurring radon gas is produced from rocks and soil as the result of the breakdown of Uranium-238 and has been implicated in some lung cancers. Nuclear energy has many uses in medicine and industry. There is a good safety record, but when accidents occur, there is the potential for contamination and illness of large numbers of people, for example, Chernobyl (Soviet Union 1986) and Fukushima (Japan 2001). The devastating effects of nuclear warfare were witnessed in Hiroshima and Nagasaki (1945), and the possibility of a rogue nation or terrorist group using a nuclear weapon remains a concern.

Physics and physiology

Radiation is the transfer of energy by particles, waves, or rays. Non-ionizing radiation has a long wavelength, low frequency, and low energy, for example, microwaves, light, and radio. The main adverse effects are related to heat. Ionizing radiation has a short wavelength, high frequency, and high energy. Injury occurs at the cellular level and to DNA. Cellular repair is less effective when high doses are delivered rapidly. Tissues with a high rate of cellular division are more sensitive to radiation damage, for example, hematopoietic and gastrointestinal cells.

Ionizing radiation is either non-penetrating (α- or β-particles) or penetrating (X-rays, γ-rays, or neutrons). Non-penetrating radiation affects the skin alone, while penetrating radiation also affects the deeper tissues.

The unit of radiation absorbed in living tissues is the Gray (Gy). 1 Gy = 1 joule per kilogram (J/kg). Assuming good medical care, the lethal dose 50 (the dose given all at once that causes the death of 50% of subjects) is 4.5 Gy, and the maximal survivable dose is 10 Gy. Ionizing radiation is emitted by a radioactive substance (radionucleotide or radioisotope). When such radiation passes through a person, they have been irradiated but pose no hazard to other people.

Contamination

If a radioactive substance contaminates an exposed surface, its presence is of no immediate threat to the life of the patient or healthcare provider. The contaminated material emits radiation with an effect that depends on the type of substance, time of exposure, and distance from the source. This should not prevent treatment for associated illness or injury. The initial responders must take universal precautions and use respiratory protection if airborne contamination is a danger. The patient's clothes are removed and the skin is washed. All materials are collected in sealed containers and labelled for disposal as radioactive waste. Skin surfaces should be monitored using a suitable instrument.

Internal exposure

Radioactive material can be ingested, be inhaled, or enter through a wound. **Box 16.8** explains the management of internal radiation exposure.

External exposure

The individual is exposed to whole- or partial-body ionizing radiation. Even those with lethal dose exposure present no danger to other people.

Acute radiation sickness

Acute radiation sickness (ARS) is a syndrome that occurs after internal or external exposure. It varies with dose, dose rate, distribution, and tissue susceptibility. In mild ARS, nausea, vomiting, and malaise follow doses of approximately 1 Gy. Lymphopenia occurs within several days, followed 2 to 3 weeks later by a fall in all white cell and platelet counts. In more severe forms multiple organ systems are affected.

Haematopoietic syndrome. Absorption of 2 to 10 Gy is followed by transient vomiting in some individuals, followed by a period of improvement. Lymphocytes are particularly sensitive to radiation damage; severe lymphopenia develops over several days. A decrease in granulocytes and platelets follows 2 to 3 weeks later, because no new cells are formed in the marrow. Complications include bleeding and overwhelming infections, and there is a high mortality.

Gastrointestinal syndrome. Doses higher than 6 Gy cause vomiting several hours after exposure. This recurs some 4 days later, accompanied by diarrhoea. The villous lining of the intestine becomes denuded. Intractable bloody diarrhoea follows, with dehydration, secondary infection, and sometimes death.

Central nervous system syndrome. Exposures of more than 30 Gy are followed rapidly by nausea, vomiting, disorientation, and coma. Death due to cerebral oedema can follow, usually within 36 hours.

Radiation dermatitis. Skin erythema, purpura, blistering, and secondary infection occur. Total loss of body hair is a bad prognostic sign and usually follows an exposure of more than 5 Gy.

Management of ARS

Management is largely supportive: prevention and treatment of infection, haemorrhage, and fluid loss. The haemopoietic system can be supported using cytokine therapy and blood products.

Late effects of radiation exposure

Risks of acute myeloid leukaemia and cancer, particularly of skin, thyroid, and salivary glands, increase. Infertility, teratogenesis, and cataract are also late sequelae, developing years after exposure.

Medical radiation

There is a demonstrated risk of cancer from the use of therapeutic and diagnostic radiation.

Radiation differs in the density of ionization it causes. Therefore, a dose-equivalent called Sievert (Sv) is used. This is the absorbed dose weighted for the damaging effect of the radiation. The annual natural background radiation is approximately 2.5 mSv. A chest X-ray delivers 0.02 mSv, and CT of the abdomen/pelvis about 10 mSv. A dose of 100 mSv is known to have a significant long-term risk for solid tumours.

Focusing techniques are used to target therapeutic radiation towards the field being treated. Radiosensitive organs, such as the ovaries, are protected by shielding.

The risk from diagnostic imaging can be reduced by the use of imaging guided by valid decision-making rules, considering the benefit versus the risk for each patient, using alternative non-ionizing imaging, and consultation with radiological colleagues.

Children, younger adults, and pregnant women have a greater risk. Healthcare professionals should protect themselves with shielding and aprons when using ionizing radiation.

Box 16.8 Management of internal radiation exposure

Method of entry	Management
Ingestion	The use of a blocking or chelating agent may reduce uptake, e.g., Lugol's iodine
Inhalation	Bronchopulmonary lavage and chelating agents may be of benefit
Wounds	Decontamination and excision. Followed by tissue monitoring. Normal wound closure

Further reading

Bui E, Joseph B, Rhee P, Diven C, Pandit V, Brown CVR. Contemporary management of radiation exposure and injury. *J Trauma Acute Care Surg* 2014; 77(3):495–500.

Calder A, Bland S. Chemical, biological, radiological and nuclear considerations in a major incident. *Surgery* 2015; 33(9):442–8.

Lightning strike

This describes cloud-to-ground lightning originating in thunderstorms. 1 in 10 victim strikes are fatal. Human tissues are directly damaged by the high-voltage direct current (DC) of more than 10 million volts that lasts only for a few milliseconds. Tissue resistance

determines the pathway through the body and the potential to generate heat. The heart is at risk of dysthymias and cardiac arrest is due to asystole. Respiratory muscles can be paralyzed. Blunt trauma occurs when the victim is thrown away from the strike.

Fern-shaped patterns seen on the skin are not thermal burns and require no treatment.

Management

Scene safety must be ensured and bystander cardiopulmonary resuscitation initiated for cardiopulmonary arrest. Respiratory muscle paralysis may persist for a considerable time and recover with support. Mortality is high and those who survive are left with variable CNS damage.

Electric shock

This describes injury from man-made electricity. Typically, this involves alternating current (AC), which is more dangerous than a similar DC current.

Electric shock can produce:

Minor Shock. The common domestic electric shock is typically painful but rarely fatal. A brief, immediate jerking episode can occur, which is not a seizure. A small burn is often present. There is usually no lasting neurological, cardiac, or skin damage.

Serious effects. These are rare following accidents in the home or in industry.

- **Cardiac, neurological, and muscle damage.** Ventricular fibrillation, muscular contraction, and spinal cord damage can follow a major shock.
- **Electrical burns.** These are commonly restricted to the skin. Muscle necrosis and spinal cord damage can also occur.

Further reading

Zafren K, Durrer B, Herry J-P, Brugger H. Lightning injuries: Prevention and on-site treatment in mountains and remote areas: Official guidelines of the International Commission for Mountain Emergency Medicine and the Medical Commission of the International Mountaineering and Climbing Federation (ICAR and UIAA MEDCOM). *Resuscitation* 2005; 65(3):369–72.

Questions for this chapter can be found on Elsevier eBooks+.

17 Cardiology

Robin Ray and Hitesh C. Patel

Clinical skills 385

Anatomy, physiology, and embryology of the heart 388
Conduction system of the heart 388
Starling's law of the heart 390
Nerve supply of the myocardium 390
Cardiac cycle 390
Coronary circulation 391
Blood vessel control and functions of the vascular endothelium 391

Clinical approach to the patient with heart disease 393
Clinical features of heart disease 393
Examination of the cardiovascular system 394
Cardiac investigations 397

Therapeutic procedures 409

Cardiac arrhythmias 411

Bradycardias and heart block 413
Bradycardia 413
Heart block 414

Tachycardias 418
Supraventricular tachycardias 418
Atrioventricular junctional tachycardias 419
Atrial tachyarrhythmias 420
Ventricular tachyarrhythmias 423

Heart failure 429
Pathophysiology 429
Clinical syndromes of heart failure 431
Management of heart failure 432
Advanced heart failure 437

Coronary artery disease 437
Pathophysiology of coronary atherosclerosis 438
Diagnosis 439

Angina 440

Acute coronary syndromes 444
NSTEMI and unstable angina 445
ST-elevation myocardial infarction 448
Post-ACS lifestyle modification 451

Valvular heart disease 452

Mitral valve 452
Mitral stenosis 452
Mitral regurgitation 455

Aortic valve 456
Aortic stenosis 456
Aortic regurgitation 460

Tricuspid valve 462
Tricuspid stenosis 462
Tricuspid regurgitation 462

Pulmonary valve 463
Pulmonary stenosis 463
Pulmonary regurgitation 463

Prosthetic valves 464

Infective endocarditis 465

Congenital heart disease 469
Adult congenital heart disease 470
Marfan's syndrome 475

Pulmonary heart disease 477
Pulmonary hypertension 477

Myocardial and endocardial disease 480
Cardiac tumours 480
Myocardial disease 480
Myocarditis 480
Cardiomyopathy 481

Pericardial disease 487

Peripheral vascular disease 490
Peripheral arterial disease 490
Aneurysmal disease 492

Peripheral venous disease 494

CORE SKILLS AND KNOWLEDGE

Heart disease remains the most common cause of death in the developed world and the third most common in low-income countries. Reducing this burden requires close collaboration between general practitioners, public health doctors, hospital specialists, and allied healthcare professionals, with interventions including reducing smoking rates, encouraging healthy eating choices, promoting aerobic exercise, and targeting hypertension (Chapter 18) and dyslipidaemia (Chapter 19).

In the outpatient setting patients with stable cardiac symptoms (e.g., exertional chest pain, palpitations, or peripheral oedema) are evaluated, and those with established cardiological disease are followed up in order to optimize lifestyle factors, titrate medications, and discuss timely percutaneous or surgical interventions.

Some cardiologists are trained in various interventional techniques (including direct current cardioversion, percutaneous coronary intervention (PCI), transcatheter valve procedures, cardiac ablation, or implantation of cardiac rhythm devices). Working closely with cardiac surgeons is pivotal to obtaining excellent outcomes. Cardiologists also look after patients with emergency cardiac presentations (e.g., acute coronary syndromes or heart block), often on coronary care units.

Key skills in cardiology include:

- Becoming familiar with the management of cardiac emergencies, including acute coronary syndrome (ACS) and life-threatening rhythm disorders

- Understanding cardiovascular risk reduction and the management of chronic cardiological conditions such as heart failure and atrial fibrillation
- Confidently interpreting electrocardiograms (ECGs) and echocardiogram reports and understanding the role of different forms of invasive cardiac intervention.

Cardiology can be best learned by observing cardiology outpatient clinics, attending ward rounds on the coronary care unit, observing invasive procedures such as coronary angiography or DC cardioversion, and seeing patients in the community with general practitioners or specialist cardiology nurses.

CLINICAL SKILLS

History

A full cardiological history covers the relevant symptoms and risk factors for cardiac disease in detail and is focused on establishing the cause of symptoms. Is this patient's chest pain likely to originate in the heart? Is their breathlessness more likely to be due to heart failure, lung disease, or something else? A fuller description of cardiac history-taking begins at the 'Clinical approach to the patient with heart disease' section, but **Box 17.1** provides a suggested structure to ensure that the relevant material is covered systematically.

Box 17.1 History-taking in cardiac disease

Symptoms

Chest pain

Think anatomically, about the type of pain caused by disease in different structures within the chest. The rapidity of onset is also informative.

- ***Heart***—central, squeezing or crushing in nature, radiation to left arm or jaw, with associated sweating or nausea
- ***Pericardium***—central, sharp in character, retrosternal, may radiate to back, worse sitting upright, associated viral symptoms or causative medication
- ***Aorta***—sudden onset, severe, tearing pain, can radiate to the back or legs (dissection)
- ***Oesophagus***—central, burning or tight in nature, worse on lying flat, worse after eating, with associated burping, epigastric discomfort or altered taste
- ***Lung/pleural***—localized and often peripheral, sharp in nature, worse on deep inspiration or coughing, with associated cough, fever, breathlessness, or haemoptysis (pleurisy, pneumonia, pulmonary embolus pneumothorax)
- ***Musculoskeletal***—often localized, intermittent, and related to movement; possible trigger event (e.g., abnormal activity); reproducible on palpation or movement
- ***Skin***—superficial pain often associated with a rash, for example, shingles
- ***Abdomen***—gall bladder, stomach, and pancreas pathology may present with upper abdominal or lower chest pain.

Breathlessness

Identify severity (see **Box 17.20**) and likely cause:

- ***Heart disease***—worse on lying flat, episodes of PND, peripheral oedema
- ***Lung disease***—associated respiratory symptoms, for example, cough, wheeze, pleuritic pain, fever, calf pain, or haemoptysis.

Lung disease and heart disease often coexist not least due to shared risk factors such as smoking.

Palpitations

These are usually episodic. Identify:

- ***Nature of symptoms***—fast or slow, feelings of 'missing a beat', associated chest pain or tightness, breathlessness, and dizziness
- ***Pattern of symptoms***—duration of problem, frequency of attacks, associations with exercise, caffeine or alcohol intake
- ***Severity of symptoms***—chest pain, dizziness, pre-syncope or loss of consciousness.

Dizziness or collapse

A full event history is required, ideally correlated by an eye-witness or video clip, if there was a loss of consciousness. Causes may include:

- ***Vasovagal syncope***—situational (intercurrent illness, dehydration, pain, emotional, postprandial, post micturition, post cough), antihypertensive medication, preceding pre-syncopal symptoms, rapid recovery
- ***Cardiac cause*** (malignant dysrhythmia or valve disease)—preceding chest pain or palpitations, sudden loss of consciousness, symptoms on exertion, lack of alternative explanation, family history of collapse or sudden death
- ***Neurological cause***—neurological aura, abnormal movements, urinary incontinence, tongue-biting, post-ictal phase.

Risk factors

- ***Previous cardiovascular disease***—enquire about or look up previous investigations and interventions
- ***Smoking***—quantified in pack years
- ***Diabetes***—time since diagnosis, degree of glycaemic control
- ***Hypertension***—time since diagnosis, degree of control
- ***Hyperlipidaemia***
- ***Chronic inflammatory disorders*** (e.g., rheumatoid arthritis)
- ***Family history of early-onset cardiovascular disease***
- ***Recreational drug use***—amfetamines, cocaine, cannabis.

Functional baseline and disability

- ***Social history***—who are the patient's social supports? What work do they do? What type of driving license do they have?
- ***Symptom-related disability***—exertional chest pain or breathlessness (try to quantify—how far can they walk? Can they climb a flight of stairs?). Does it limit their work performance? Are they house- or bed-bound?
- ***Self-care***—ranging from full functional independence to carer visits or full residential or nursing care.

Examination

Many features of cardiovascular disease can be elicited on physical examination (see figure opposite). They include:

- Evidence of life-threatening cardiac disease requiring emergency resuscitation: shock, dysrhythmias, pulmonary oedema
- Evidence of underlying conditions that contribute to cardiovascular disease (e.g., hyperlipidaemia or Marfan's syndrome)
- Evidence of heart failure
- Murmurs and specific signs associated with particular forms of valvular heart disease.

ECG interpretation

The ECG is an essential bedside test in the investigation of all patients with cardiac disease. A full guide to understanding the science behind it and its interpretation begins at the 'Electrocardiography' section, and **Box 17.2** offers a systematic approach to ensure that abnormalities are not missed.

Box 17.2 A systematic approach to interpreting the ECG

Step 1: Preliminaries

- Correct patient?
- Correct date and time?
- Standard speed (25 mm/sec)?
- Standard amplitude (10 mm/mV)?

Step 2: Heart structure

- Axis (left axis deviation if <–30 degrees, right axis deviation if >90 degrees)?
- Left ventricular hypertrophy (various criteria, see **Fig. 17.74**)?
- Evidence of atrial enlargement?

Step 3: Rate and rhythm

- Heart rate?
- Regular or irregular?
- Sinus rhythm (P–QRS–T)?
- Atrioventricular node delay/block (PR duration should be 120–200 ms)?
- Intraventricular conduction delay (QRS duration should be <120 ms)?
- Repolarization delay (QTc should be <450 ms)?

Step 4: Ischaemia

- Features include:
  - Q waves
  - ST elevation or depression
  - T-wave inversion.
- Anatomical regions:
  - Anterior (V_1/V_2)
  - Septal (V_3/V_4)
  - Lateral (V_5/V_6)
  - Interior (II, III, AVF)
  - High lateral (I, AVL).

Step 5: Global features

- Pericarditis (global saddle-shaped ST elevation)?
- Electrolyte abnormalities, for example, hyperkalaemia (peaked T waves)?
- Pre-excitation (upstroke before QRS complexes)?
- Rarer causes of global abnormalities include hypothermia, thyroid disease, and drug effects.

General
Breathless
Position - sitting up?
Colour - cyanosed (central)
- pale
- malar flush (mitral stenosis)
Pain
Obesity/cachexia
Signs of other conditions, e.g., Marfan's
Temperature
Sweating

Jugular venous pulsation
Height
Character

Carotid pulse
- Character
- Bruits

Blood pressure

Radial pulse
Rate
Rhythm
Volume

Hands
Splinter haemorrhages
Clubbing
Peripheral cyanosis
Tendon xanthomata (hyperlipidaemia)

Splinter haemorrhages

Xanthelasma

Mouth
Oral hygiene
Poor dentition

Precordium
Inspect - look for scars
Palpate - apex beat
- thrills
- ventricular heaves
Auscultate
- added sounds
- murmurs

Chest
Basal crackles

Abdomen
Hepatomegaly (pulsatile in tricuspid incompetence)
Ascites
Aortic aneurysm

Legs
Oedema
Feel for peripheral pulses

Feet
Look for ischaemia
- Feel for pulses
- Temperature
Vasculitis rash

Infective endocarditis, septic embolization

(From Johnson RJ, Feehally J. Comprehensive Clinical Nephrology, 2nd edn. St. Louis, Mosby; 2000, with permission.)

ANATOMY, PHYSIOLOGY, AND EMBRYOLOGY OF THE HEART

The heart is made of four main cell types: cardiomyocytes, endothelial cells, fibroblasts, and smooth muscle cells. Cardiomyocytes (CMs) are designed to perform two fundamental functions: initiation and conduction of electrical impulses, and contraction. Although most CMs are able to perform both these functions, the vast majority are predominantly contractile cells and a small number are specifically designed as electrical cells. The latter, collectively known as the conducting system of the heart, are not nervous tissue but modified CMs lacking in myofibril components. They have the ability to generate electrical impulses, which are then conducted to the myocytes, leading to contraction by a process known as excitation–contraction coupling. The rate of electrical impulse generation and the force of myocardial contraction are modified by numerous factors, including autonomic input and stretch.

Three epicardial coronary arteries supply blood to the myocardium, and a more complex network of veins is responsible for drainage. In the face of continuous arterial pressure fluctuations blood vessels maintain constant tissue perfusion by a process known as ***'autoregulation'***; blood vessel control is, however, complex, involving additional local and central mechanisms.

Conduction system of the heart

Sinus node (sinoatrial node)

The sinus node is a complex, spindle-shaped structure that lies in the lateral and epicardial aspects of the junction between the superior vena cava and the right atrium (Fig. 17.1). Physiologically, it generates impulses automatically by spontaneous depolarization of its membrane at a rate quicker than any other cardiac cell type. It is therefore the ***natural pacemaker of the heart***.

A number of factors are responsible for the spontaneous decay of the sinus node cell membrane potential ('pacemaker potential'), the most significant of which is a small influx of sodium ions into the cells. This small sodium current has two components: the background inward (I_b) current and the 'funny' (I_f) current (or pacemaker current) (Fig. 17.2). The term 'funny' current denotes ionic flow through channels activated in ***hyperpolarized*** cells (–60 mV or greater), unlike other time- and voltage-dependent channels activated by ***depolarization***. The rate of depolarization of the sinus node membrane potential is modulated by autonomic tone (i.e., sympathetic and parasympathetic input), stretch, temperature, hypoxia, and blood pH and responds to other hormonal influences (e.g., tri-iodothyronine and serotonin).

Atrial and ventricular myocyte action potentials

Action potentials in the sinus node trigger depolarization of the atrial myocytes and subsequently the ventricular ones. These cells have a different action potential from that of sinus node cells (see Fig. 17.2). Their resting membrane potential is a consequence of a small flow of potassium ions into the cells through open 'inward rectifier' channels (I_{Kl}); at this stage, sodium and calcium channels are closed. The arrival of adjacent action potentials triggers the opening of voltage-gated, fast, self-inactivating sodium channels, resulting in a sharp depolarization spike. This is followed by a partial repolarization of the membrane due to activation of 'transient outward' potassium channels.

The plateau phase that follows is unique to myocytes and results from a small but sustained inward calcium current through L-type calcium channels (I_{CaL}) lasting 200 to 400 ms. This calcium influx is caused by a combined increase in permeability of the cell, especially the sarcolemmal membranes to calcium (Fig. 17.3). This plateau (or refractory) phase in myocyte action potential prevents early reactivation of the myocytes and directly determines the strength of contraction. The gradual inactivation of the calcium channels activates delayed rectifier potassium channels (I_K), repolarizing the membrane. Atrial tissue is activated like a 'forest fire', but the activation peters out when the insulating layer between the atrium and the ventricle—the annulus fibrosus—is reached. There is debate about whether impulses from the sinoatrial (SA) node travel over specialized conducting 'pathways' in the atrium or over ordinary atrial myocardium.

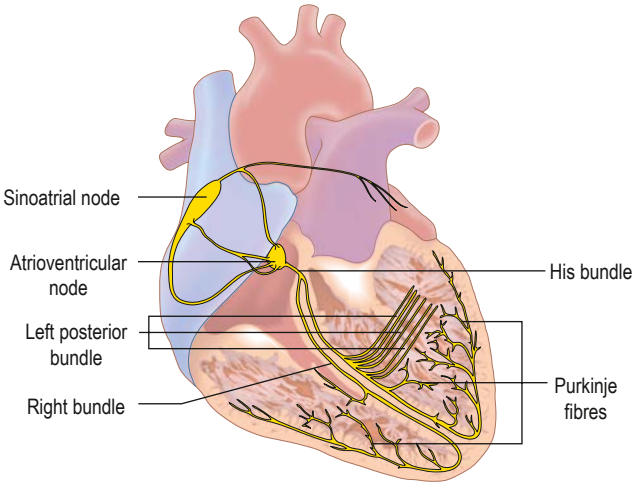


Fig. 17.1 The conducting system of the heart.

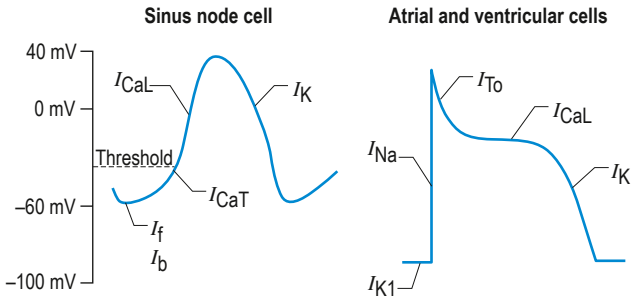


Fig. 17.2 Myocardial action potentials. I_b, Background inward sodium current; I_{CaL}, long-lasting (or 'L'-type) calcium channel; I_{CaT}, transient (or 'T'-type) calcium channel; I_f, 'funny' current; I_K, delayed rectifier potassium current; I_{K1}, inward rectifier potassium current; I_{Na}, inward sodium current; I_{To}, transient outward potassium current.

Atrioventricular node, His bundle, and Purkinje fibres

The depolarization continues to conduct slowly through the atrioventricular (AV) node. This is a small, bean-shaped structure that lies beneath the right atrial endocardium within the lower interatrial septum. The AV node continues as the His bundle, which penetrates the annulus fibrosus and conducts the cardiac impulse rapidly towards the ventricle. The His bundle reaches the crest of the interventricular septum and divides into the right bundle branch and the main left bundle branch.

The ***right bundle branch*** continues down the right side of the interventricular septum to the apex; from here, it radiates and divides to form the Purkinje network, which spreads throughout the subendocardial surface of the right ventricle. The ***main left bundle branch*** is a short structure, which fans out into many strands on the left side of the interventricular septum. These strands can be grouped into an anterior superior division (the anterior hemi-bundle) and a posterior inferior division (the posterior hemi-bundle), though there is large anatomical variation between patients. The anterior hemi-bundle supplies the subendocardial Purkinje network of the anterior and superior surfaces of the left ventricle, and the inferior hemi-bundle supplies the inferior and posterior surfaces. Impulse conduction through the AV node is slow and depends on action potentials, largely produced by slow transmembrane calcium flux. In the atria ventricles and His–Purkinje system, conduction is rapid and is due to action potentials generated by rapid transmembrane sodium diffusion.

Cellular basis of myocardial contraction–excitation–contraction coupling

Each myocyte, approximately 100 μm long, branches and interdigitates with adjacent cells. An intercalated disc permits electrical conduction to adjacent cells. Myocytes contain bundles of parallel myofibrils. Each myofibril is made up of a series of sarcomeres (Fig. 17.4). A sarcomere (the basic unit of contraction) is bound by two transverse Z lines, to each of which is attached a perpendicular filament of the protein actin. The actin filaments from each of the two Z bands overlap with thicker parallel protein filaments known as myosin. Actin and myosin filaments are attached to each other

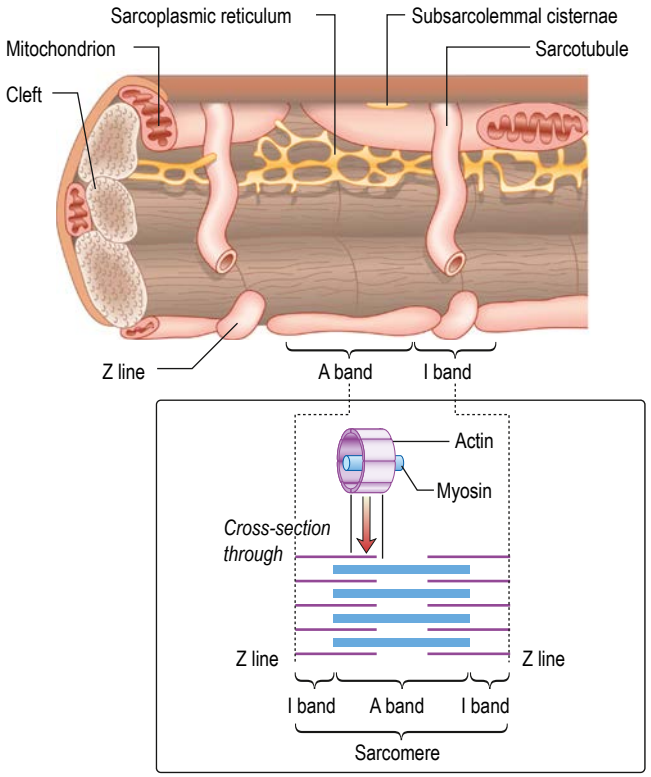


Fig. 17.4 The structure of a myofibril within a myocyte. The myofibrils are made up of a series of sarcomeres joined at the Z line.

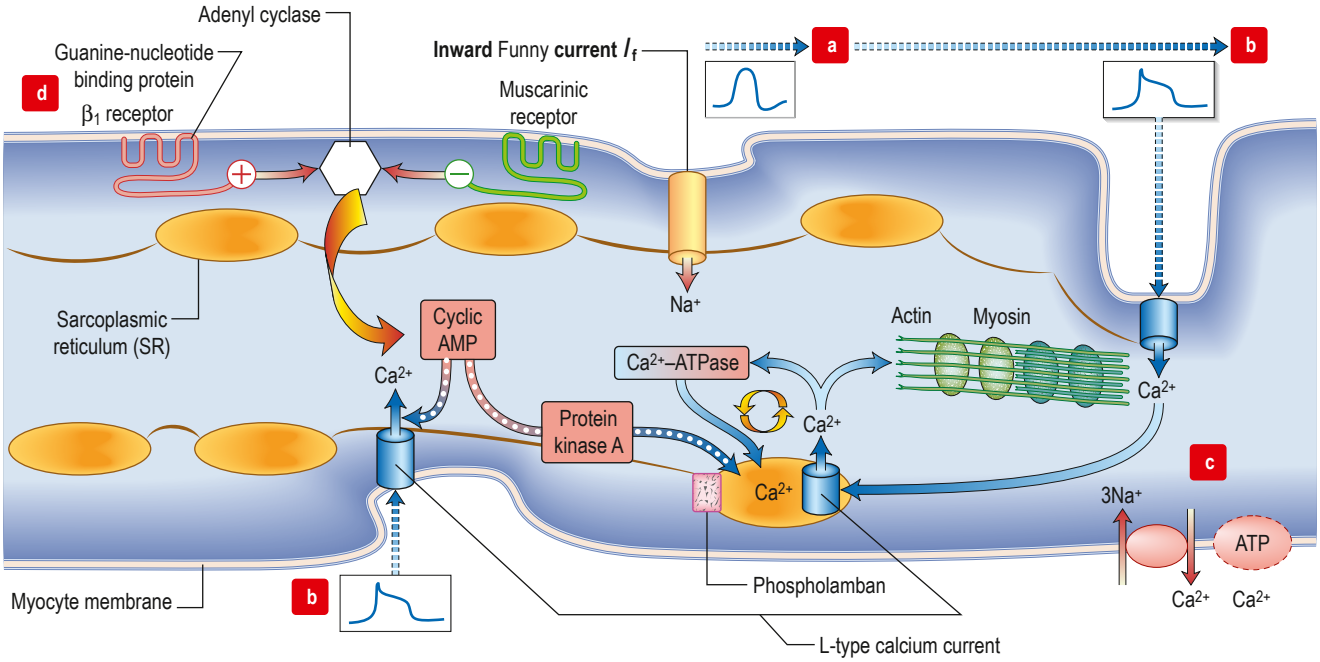


Fig. 17.3 The 'complete' cardiac cell. (a) *Spontaneous depolarization* in sinus node cells, due to sodium (Na) influx through the 'funny' current, generates the 'pacemaker' potential. **(b)** This activates other atrial and ventricular myocytes, triggering ***action potentials*** and activating L-type calcium (Ca) channels in the surface and transverse tubule membranes (at the top and bottom of the figure). **(c)** The resulting ***Ca influx*** acts on Ca-induced Ca release channels (RyR2) on the sarcoplasmic reticulum (SR); this causes the release of stored Ca, which acts on actin and myosin fibrils, leading to contraction. Ca reuptake pumps in the SR, regulated by phospholamban, replenish the stores; various exchange pumps also expel Ca from the cell. **(d)** ***Autonomic input*** has either a positive chronotropic/inotropic effect (β_1 receptors) or a negative chronotropic/inotropic effect (muscarinic receptors). *AMP,* Adenosine monophosphate; *ATPase,* adenosine triphosphatase.

by cross-bridges that contain adenosine triphosphatase (ATPase), which breaks down adenosine triphosphate (ATP) to provide the energy for contraction.

Two chains of actin molecules form a helical structure with another molecule, tropomyosin, in the grooves of the actin helix; a further molecule, troponin, is attached to every seven actin molecules. During cardiac contraction, the length of the actin and myosin monofilaments does not change. Rather, the actin filaments slide between the myosin filaments when ATPase splits a high-energy bond of ATP. To supply the ATP, the myocyte (which cannot stop for a rest) has a very high mitochondrial density (35% of the cell volume). As calcium ions bind to troponin C, the activity of troponin I is inhibited, which induces a conformational change in tropomyosin. This event unlocks the active site between actin and myosin, enabling contraction to proceed.

Calcium is made available during the plateau phase of the action potential when calcium ions enter the cell and are mobilized from the sarcoplasmic reticulum (SR) through the ryanodine receptor (RyR2) calcium-release channel. RyR2 activity is regulated by the protein calstabin 2 and nitric oxide (NO). The force of cardiac muscle contraction ('inotropic state') is thus regulated by the influx of calcium ions into the cell through calcium channels (see Fig. 17.3). T (transient) calcium channels open when the muscle is more depolarized, whereas L (long-lasting) calcium channels require less depolarization. The extent to which the sarcomere can shorten determines the stroke volume of the ventricle. It is maximally shortened in response to powerful inotropic drugs or severe exercise.

Starling's law of the heart

The contractile function of an isolated strip of cardiac tissue can be described by the relationship between the velocity of muscle contraction, the load that is moved by the contracting muscle, and the extent to which the muscle is stretched before contracting. As with all other types of muscle, the velocity of contraction of myocardial tissue is reduced by increasing the load against which the tissue must contract. However, in the non-failing heart, pre-stretching of cardiac muscle improves the relationship between the force and velocity of contraction (Fig. 17.5).

This phenomenon was described in the intact heart as an increase of stroke volume (ventricular performance) with an enlargement of the diastolic volume (preload) and is known as 'Starling's law of the heart' or the 'Frank–Starling relationship'. It has been transcribed into more clinically relevant indices. Thus, stroke work (aortic pressure × stroke volume) is increased as ventricular end-diastolic volume is raised. Alternatively, within certain limits, cardiac output rises as left atrial pressure increases (this is assessed invasively by measuring the pulmonary capillary wedge pressure). This clinical relationship is described by the ventricular function curve (see Fig. 17.5), which also shows the effect of sympathetic stimulation.

Nerve supply of the myocardium

Adrenergic nerves supply atrial and ventricular muscle fibres, as well as the conduction system. Beta$_1$ receptors predominate in the heart, with both adrenaline (epinephrine) and noradrenaline (norepinephrine) having positive inotropic and chronotropic effects. Cholinergic nerves from the vagus supply mainly the SA and AV nodes via M_2 muscarinic receptors. The ventricular myocardium is sparsely innervated by the vagus. Under basal conditions, vagal inhibitory effects predominate over the sympathetic excitatory effects,

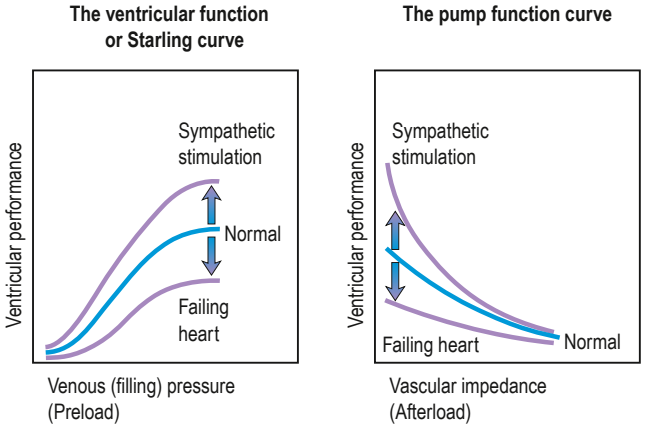


Fig. 17.5 The Frank–Starling mechanism. The effect on ventricular contraction of alteration in filling pressures and outflow impedance in the normal, failing, and sympathetically stimulated ventricle is shown.

resulting in a slow heart rate. Interestingly, post cardiac transplant, following which both sympathetic and parasympathetic nervous input to the heart is abrogated, resting heart rate increases to 90 to 100 beats per minute.

Adrenergic stimulation and cellular signalling

Beta$_1$-adrenergic stimulation enhances Ca^{2+} flux in the myocyte and thereby strengthens the force of contraction (see Fig. 17.3). Binding of catecholamines, such as adrenaline, to the myocyte β_1-adrenergic receptor stimulates membrane-bound adenylate kinases. These enzymes enhance production of cyclic adenosine monophosphate (cAMP), which activates intracellular protein kinases; these, in turn, phosphorylate cellular proteins, including L-type calcium channels within the cell membrane. Beta$_1$-adrenergic stimulation of the myocyte also enhances myocyte relaxation.

The return of calcium from the cytosol to the SR is regulated by phospholamban (PL), a low-molecular-weight protein in the SR membrane. In its dephosphorylated state PL inhibits Ca^{2+} uptake by the SR ATPase pump (see Fig. 17.3). However, β_1-adrenergic activation of protein kinase phosphorylates PL and blunts its inhibitory effect. The subsequently greater uptake of Ca^{2+} ions by the SR hastens Ca^{2+} removal from the cytosol and promotes myocyte relaxation.

The increased cAMP activity also results in phosphorylation of troponin I, an action that inhibits actin–myosin interaction and further enhances myocyte relaxation.

Cardiac cycle

The cardiac cycle (Fig. 17.6) consists of precisely timed rhythmic electrical and mechanical events that propel blood into the systemic and pulmonary circulations. The first event in the cardiac cycle is atrial depolarization (a P wave on the surface ECG), followed by right atrial and then left atrial contraction. Ventricular activation (the QRS complex on the ECG) follows after a short interval (the PR interval). Left ventricular contraction starts and, shortly thereafter, right ventricular contraction begins. The increased ventricular pressures exceed the atrial pressures and close first the mitral and then the tricuspid valves.

Until the aortic and pulmonary valves open, the ventricles contract with no change of volume (isovolumetric contraction). When ventricular pressures rise above the aortic and pulmonary artery

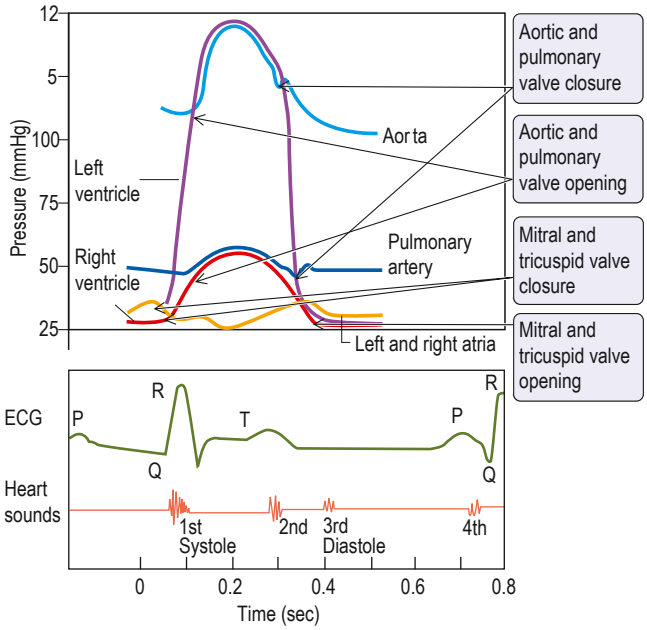


Fig. 17.6 The cardiac cycle.

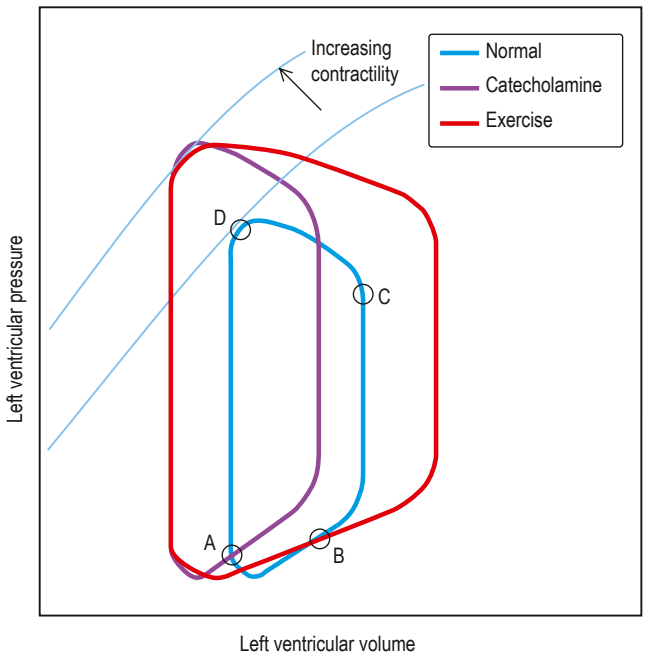


Fig. 17.7 Pressure–volume loop. AB, Diastole: ventricular filling; BC, systole: isovolumetric ventricular contraction; CD, systole: ventricular emptying; DA, diastole: isovolumetric ventricular relaxation.

pressures, the pulmonary valve and then the aortic valve open and ventricular ejection occurs. As the ventricles begin to relax, their pressures fall below the aortic and pulmonary arterial pressures, and aortic valve closure is followed by pulmonary valve closure. Isovolumetric relaxation then occurs. After the ventricular pressures have fallen below the right atrial and left atrial pressures, the tricuspid and mitral valves open. The cardiac cycle can be depicted graphically as the relationship between the pressure and volume of the ventricle (Fig. 17.7), illustrating the changing pressure–volume relationships in response to increased contractility and to exercise.

Coronary circulation

The coronary arterial system (Fig. 17.8) consists of the right and left coronary arteries. These arteries branch from the aorta, arising immediately above two cusps of the aortic valve. The right and left coronary arteries are unique in that they fill during diastole, when not occluded by valve cusps, and when not squeezed by myocardial contraction. The right coronary artery arises from the right coronary sinus and courses through the right side of the AV groove, giving off vessels that supply the right atrium and the right ventricle. The vessel usually continues as the posterior descending coronary artery, which runs in the posterior interventricular groove and supplies the posterior part of the interventricular septum and the posterior left ventricular wall.

Within 2.5 cm of its origin from the left coronary sinus the left main coronary divides into the left anterior descending (LAD) artery and the circumflex artery. The LAD artery runs in the anterior interventricular groove and supplies the anterior septum and the anterior left ventricular wall. The left circumflex artery travels along the left AV groove and gives off branches to the left atrium and the left ventricle (marginal branches).

The sinus node and the AV node are supplied by the right coronary artery in about 60% and 90% of people, respectively. Therefore disease in this artery may cause sinus bradycardia and AV nodal block. The majority of the left ventricle is supplied by the left coronary artery and disease in this vessel can cause significant myocardial dysfunction.

Some blood from the capillary beds in the wall of the heart drains directly into the cavities of the heart via tiny veins, but the majority returns by veins that accompany the arteries, to empty into the right atrium via the coronary sinus. An extensive lymphatic system drains into vessels that travel along the coronary vessels and then into the thoracic duct.

Blood vessel control and functions of the vascular endothelium

In functional terms the tunica intima with the vascular endothelium and the smooth muscle cell–containing tunica media are the main constituents of blood vessels. These two structures are closely interlinked by a variety of mechanisms to regulate vascular tone. The central control of blood vessels is achieved via the neuroendocrine system. Sympathetic vasoconstrictor and parasympathetic vasodilator nerves regulate vascular tone in response to daily activity. Where neural control is impaired or in various pathological states, such as haemorrhage, endocrine control of blood vessels, mediated through adrenaline (epinephrine), angiotensin, and vasopressin, takes over.

At a local level, tissue perfusion is maintained automatically and by the effect of various factors synthesized and/or released in the immediate vicinity. In the face of fluctuating arterial pressures blood vessels vasoconstrict independently of nervous input when blood pressure drops, and vice versa. This process of ***autoregulation*** is a consequence of:

- the ***Bayliss myogenic*** response—the ability of blood vessels to constrict when distended
- the ***vasodilator washout*** effect—the vasoconstriction triggered by a decrease in the concentration of tissue metabolites.

The vascular endothelium is a cardiovascular endocrine organ, which occupies a strategic interface between blood and other tissues. It produces various compounds (e.g., NO, prostacyclin [PGI_2], endothelin, endothelial-derived hyperpolarizing factor [EDHF],

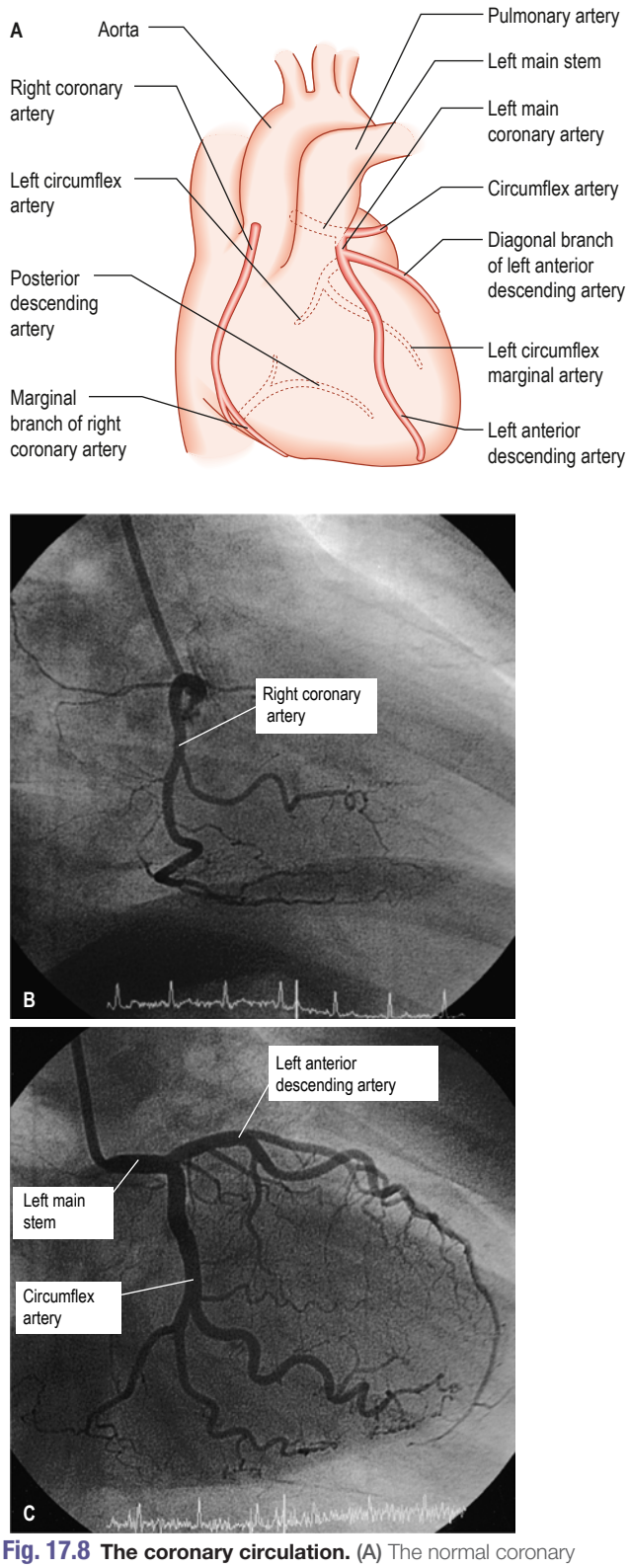


Fig. 17.8 The coronary circulation. (A) The normal coronary arterial anatomy. (B) Angiogram of the dominant right coronary system. (C) Angiogram of the left coronary system from the same patient (right anterior oblique projection).

adhesion molecules, vascular endothelial growth factor [VEGF]) and has enzymes located on the surface, controlling the levels of circulating compounds such as angiotensin, bradykinin, and serotonin. It has many regulatory roles.

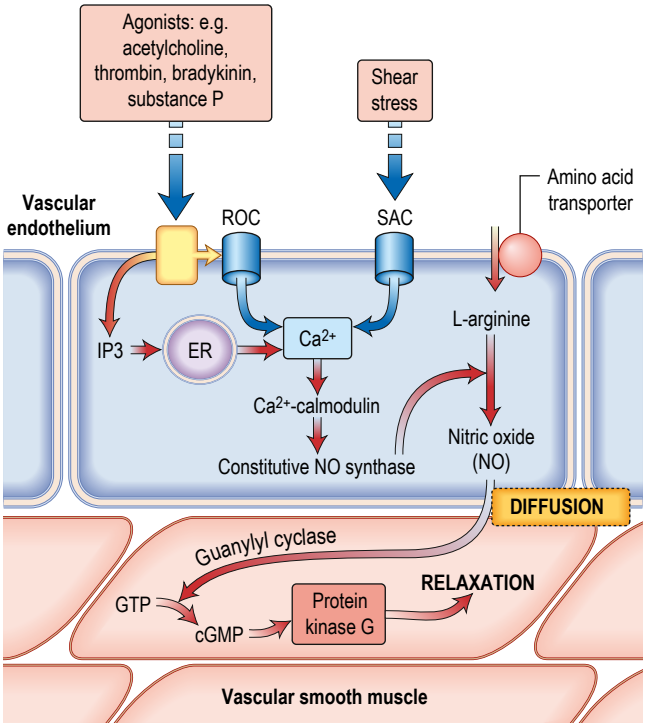


Fig. 17.9 The stimulus for production and function of nitric oxide (NO). Various stimuli lead to the production of NO via cytoplasmic calcium/calmodulin. NO triggers smooth muscle relaxation via the activation of guanylyl cyclase. *cGMP,* Cyclic guanine monophosphate; *ER,* endoplasmic reticulum; *GTP,* guanine triphosphate; *IP3,* inositol triphosphate; *ROC,* receptor-operated Ca^{2+} channel; *SAC,* stretch-activated Ca^{2+} channel.

Vasomotor control

- ***Nitric oxide*** is a diffusible gas with a very short half-life; it is produced in endothelial cells from the amino acid L-arginine via the action of the enzyme NO synthase (NOS), which is controlled by cytoplasmic calcium/calmodulin (Fig. 17.9). It is produced in response to various stimuli (Box 17.3), triggering vascular smooth muscle relaxation through activation of guanylate cyclase; this leads to an increase in the intracellular levels of cyclic 3,5-guanine monophosphate (cGMP). Its cardiovascular effects protect against atherosclerosis, high blood pressure, heart failure, and thrombosis. NO is also the neurotransmitter in various 'nitrergic' nerves in the central and peripheral nervous systems and may play a role in the central regulation of vascular tone. The class of drugs used to treat erectile dysfunction, the phosphodiesterase (PDE_5) inhibitors, prevent the breakdown of cGMP and promote vasodilation.
- ***PGI_2*** is synergistic with NO and also plays a role in the local regulation of vasomotor tone.
- ***Endothelin (ET)*** is a 21-amino-acid peptide that counteracts the effects of NO. Its production is inhibited by shear stress—that is, the stress exerted on the vessel wall by the flowing blood—and it causes profound vasoconstriction and vascular smooth muscle hypertrophy. It is thought to play a role in the genesis of hypertension and atheroma.
- ***Angiotensin-converting enzyme (ACE)***, located on the endothelial cell membrane, converts circulating angiotensin I (synthesized by the action of renin on angiotensinogen) to angiotensin II, which has vasoconstrictor properties and leads to aldosterone release (see Fig. 17.6). Aldosterone promotes sodium

Box 17.3 Some of the products and functions of the vascular endothelium

Endothelial product	Function(s)	Stimulus
Nitric oxide	Vasodilation Inhibition of platelet aggregation Inhibition of transcription of adhesion molecules Inhibition of vascular smooth muscle proliferation	Shear stress, e.g., induced by exercise Agonists: thrombin, acetylcholine, endothelin, bradykinin, serotonin, substance P Inflammation/ endotoxin shock
Prostacyclin (PGI_2)	Vasodilation Inhibition of platelet aggregation	Agonist: thrombin Inflammation
Prostanoids	Vasoconstriction	Hypoxia
Endothelin	Vasoconstriction	Thrombin, angiotensin II, vasopressin Hypoxia *Note:* Inhibited by shear stress
Endothelial-derived hyperpolarizing factor	Vasodilation	Agonists: bradykinin, acetylcholine
Angiotensin-converting enzyme	Vasoconstriction	Expressed naturally
von Willebrand factor	Promotion of platelet aggregation Stabilization of factor VIII	Agonists: thrombin, adrenaline (epinephrine)
Adhesion molecules		
P, L, E selectins ICAM, VCAM, PECAM	Margination of white blood cells Binding and diapedesis of WBCs into vessel wall	Inflammatory mediators: histamine, thrombin, TNF, IL-6
Vascular endothelial growth factor (VEGF)	Angiogenesis Vasodilation Increase in vascular permeability	Pregnancy Hypoxia Inflammation, e.g., rheumatoid arthritis Trauma Tumours

ICAM, Intracellular adhesion molecule; *IL,* interleukin; *PECAM,* platelet/endothelial cell adhesion molecule; *TNF,* tumour necrosis factor; *VCAM,* vascular cell adhesion molecule; *WBC,* white blood cell.

absorption from the kidney and, together with the angiotensin-induced vasoconstriction, provides haemodynamic stability.

- ***Other factors*** that influence vasomotor tone include histamine (released by mast cells), bradykinin (synthesized from kininogen by the action of coagulation factor XIIa), and serotonin (released by platelets).

Anti- and prothrombotic mechanisms

PGI_2, produced from arachidonic acid in the endothelial cell membrane by the action of the enzyme cyclo-oxygenase, inhibits platelet aggregation. Low-dose aspirin prevents activation of the cyclo-oxygenase pathway in platelets but only to a degree that does not affect PGI_2 synthesis, unlike higher doses. Other antithrombotic agents, such as clopidogrel (an adenosine diphosphate [ADP] receptor antagonist) or abciximab and tirofiban (glycoprotein IIb/IIIa inhibitors), achieve their effects by acting directly on platelet receptors. The antithrombotic effect of PGI_2 is aided by NO, affecting platelets via activation of guanylate cyclase. The endothelial cell membrane also produces other anticoagulant molecules, such as thrombomodulin, heparin sulphate, and various fibrinolytic factors. Clinically used, fast-acting heparin preparations are identical to this naturally occurring molecule.

In addition to their ability to prevent clotting, endothelial cells also aid thrombosis. They are responsible for the production of von Willebrand factor through a unique organelle called the Weibel–Palade body, which not only acts as a carrier for factor VIII but also promotes platelet adhesion by binding to exposed collagen.

Further reading

Levick JR. *An Introduction to Cardiovascular Physiology*, 6th edn. London: Arnold; 2018.

CLINICAL APPROACH TO THE PATIENT WITH HEART DISEASE

Clinical features of heart disease

The following symptoms occur with heart disease:

- chest pain
- dyspnoea (breathlessness)
- palpitations
- syncope
- fatigue
- peripheral oedema.

The severity of cardiac symptoms or fatigue is classified according to the New York Heart Association (NYHA) grading of cardiac status (see **Box 17.20**). The differential diagnosis of chest pain is given in **Box 17.4**.

Central chest pain

This is the most common symptom associated with heart disease. The pain of angina pectoris and myocardial infarction (MI) is due to myocardial hypoxia.

Types of pain include:

- a retrosternal heavy or gripping sensation with radiation to the left arm or neck that is provoked by exertion and eased with rest or nitrates—***angina*** (see 'Angina' section)
- similar pain at rest—***acute coronary syndrome*** (see 'Acute coronary syndromes' section)
- severe, tearing chest pain radiating through to the back—***aortic dissection*** (see 'Aneurysmal disease' section)
- sharp, central chest pain that is worse with movement or respiration but relieved with sitting forward—***pericarditis pain*** (see 'Pericardial disease' section)
- sharp, stabbing, left submammary pain associated with anxiety—***da Costa's syndrome.***

Dyspnoea

Left ventricular failure causes dyspnoea due to oedema of the pulmonary interstitium and alveoli. This makes the lungs stiff (less compliant), thus increasing the respiratory effort required to ventilate

Box 17.4 Differential diagnosis of chest pain

Central

Cardiac
- Ischaemic heart disease (infarction or angina)
- Coronary artery spasm
- Pericarditis/myocarditis
- Mitral valve prolapse
- Aortic aneurysm/dissection

Non-cardiac
- Pulmonary embolism
- Oesophageal disease
- Mediastinitis
- Costochondritis (Tietze's disease)
- Trauma (soft tissue, rib)

Lateral/peripheral

Pulmonary
- Infarction
- Pneumonia
- Pneumothorax
- Lung cancer
- Mesothelioma

Non-pulmonary
- Bornholm disease (epidemic myalgia)
- Herpes zoster
- Trauma (ribs/muscular)

them. ***Tachypnoea*** (increased respiratory rate) is often present owing to stimulation of pulmonary stretch receptors.

Orthopnoea refers to breathlessness on lying flat. Blood is redistributed from the legs to the torso, leading to an increase in central and pulmonary blood volume. The patient uses an increasing number of pillows to sleep.

Paroxysmal nocturnal dyspnoea (PND) is when a patient is woken from sleep fighting for breath. It is caused by the same mechanisms as orthopnoea. However, as sensory awareness is reduced during sleep, pulmonary oedema can become quite severe before the patient is awoken.

Platypnea-orthodeoxia is paradoxical, worsening shortness of breath and hypoxia when the patient sits up. This is usually seen with intracardiac shunting, pulmonary arteriovenous connections, or hepatopulmonary syndrome.

Bendopnea is shortness of breath when bending over at the waist. This is typically seen in patients with reduced cardiac output and elevated right or left atrial pressure.

Hyperventilation with alternating episodes of apnoea ***(Cheyne–Stokes respiration)*** occurs in severe heart failure.

If hypopnoea occurs rather than apnoea, the phenomenon is termed 'periodic breathing', but the two variations are known together as ***central sleep apnoea syndrome (CSAS)***. This occurs due to malfunctioning of the respiratory centre in the brain, caused by poor cardiac output with concurrent cerebrovascular disease. The symptoms of CSAS, such as daytime somnolence and fatigue, are similar to those of obstructive sleep apnoea syndrome (OSAS, see Chapter 20, 'Obstructive sleep apnoea' section), and there is considerable overlap with the symptoms of heart failure. CSAS is believed to lead to myocardial hypertrophy and fibrosis, deterioration in cardiac function, and complex arrhythmias, including non-sustained ventricular tachycardia, hypertension, and stroke. Patients with CSAS have a worse prognosis than similar patients without CSAS.

Palpitations

These represent an increased awareness of the normal heartbeat or the sensation of slow, rapid, or irregular heart rhythms. The most common arrhythmias felt as palpitations are premature ectopic beats and paroxysmal tachycardias. A useful trick is to ask patients to tap out the rate and rhythm of their palpitations, as the different arrhythmias have different characteristics:

- ***Premature beats (ectopics)*** are felt by the patient as a pause followed by a forceful beat. This is because premature beats are usually followed by a pause before the next normal beat, as the heart resets itself. The next beat is more forceful, as the heart has had a longer diastolic period and therefore is filled with more blood before this beat.
- ***Paroxysmal tachycardias*** (see 'Tachycardias' section) are felt as a sudden, racing heartbeat.
- ***Bradycardias*** (see 'Bradycardia' section) may be appreciated as slow, regular, heavy, or forceful beats. Most often, however, they are simply not sensed.

Syncope

Syncope is a transient loss of consciousness due to inadequate cerebral blood flow. This is detailed further in Chapter 11 ('Dizziness and blackouts' section).

Fatigue

Fatigue may be a symptom of inadequate systemic perfusion in heart failure. Other contributing factors may include:
- poor sleep
- side-effects of medication, particularly beta-blockers
- electrolyte imbalance caused by diuretic therapy
- a systemic manifestation of infection (endocarditis) or anaemia.

Peripheral oedema

Heart failure results in salt and water retention due to renal underperfusion and consequent activation of the renin–angiotensin–aldosterone system (see Chapter 27). This leads to dependent pitting oedema. Non-cardiac causes of oedema include renal dysfunction, hepatic dysfunction, and medications such as calcium-channel blockers.

Examination of the cardiovascular system

General examination

General features of the patient's wellbeing should be noted, as well as the presence of conjunctival pallor, obesity, jaundice, and cachexia.
- Cardiac causes of ***clubbing*** (see Chapter 20, 'Examination of the respiratory system' section) include congenital cyanotic heart disease, particularly Fallot's tetralogy; less commonly, it is seen in patients with subacute infective endocarditis (IE) and atrial myxoma.
- ***Splinter haemorrhages*** are small, subungual linear haemorrhages that are frequently due to trauma but are also seen in IE.
- ***Cyanosis*** is a dusky blue discoloration of the skin (particularly at the extremities) or of the mucous membranes when the capillary oxygen saturation is below 85%. ***Central cyanosis*** is seen with shunting of deoxygenated venous blood into the systemic circulation, as in the presence of a right-to-left heart shunt, which is associated with congenital heart disease. ***Peripheral cyanosis*** is seen in the hands and feet, which are cold. It occurs in conditions associated with peripheral vasoconstriction and stasis of blood in the extremities, leading to increased peripheral oxygen extraction. Such conditions include congestive heart failure, circulatory shock, exposure to cold temperatures, and abnormalities of the peripheral circulation, such as Raynaud's phenomenon (see Chapter 32, 'Raynaud's phenomenon' section).

Arterial pulse

The first pulse to be examined is the right radial pulse. Timing should be synchronous with other peripheral pulses; delayed femoral

pulsation occurs because of a proximal stenosis, particularly of the aorta (coarctation).

Rate

The pulse rate should be between 60 and 80 beats per minute (b.p.m) when an adult patient is lying quietly in bed.

Rhythm

The rhythm is regular, except for a slight quickening in early inspiration and a slowing in expiration (sinus arrhythmia).

- ***Premature beats*** occur as occasional or repeated irregularities superimposed on a regular pulse rhythm. Similarly, intermittent heart block is revealed by occasional beats dropped from an otherwise regular rhythm.
- ***Atrial fibrillation*** produces an 'irregularly irregular' pulse. This irregular pattern persists when the pulse quickens in response to exercise, in contrast to pulse irregularity due to ectopic beats, which usually disappears on exercise.

Character

- ***Carotid pulsations*** are not normally apparent on inspection of the neck but may be visible (Corrigan's sign) in conditions associated with a large-volume pulse, including high-output states (such as thyrotoxicosis, anaemia, or fever) and in aortic regurgitation.
- ***A 'collapsing' or 'water hammer' pulse*** (Fig. 17.10) is a large-volume pulse characterized by a short duration with a brisk rise and fall. This is best appreciated by palpating the radial artery with the palmar aspect of four fingers while elevating the patient's arm above the level of the heart. A collapsing pulse is characteristic of aortic valvular regurgitation or a persistent ductus arteriosus.
- ***A small-volume pulse*** is seen in cardiac failure, shock, and obstructive valvular or vascular disease. It may also be present during tachyarrhythmias.
- ***A plateau pulse*** (see Fig. 17.10) is small in volume and slow in rising to a peak; it is due to aortic stenosis.
- ***An alternating pulse (pulsus alternans)*** (see Fig. 17.10) is characterized by regular alternate beats that are weak and strong. It is a feature of severe myocardial failure and is due to the prolonged recovery time of damaged myocardium; it indicates a very poor prognosis. It is easily noticed when taking the blood pressure because the systolic pressure may vary from beat to beat by as much as 50 mmHg.
- ***A bigeminal pulse (pulsus bigeminus)*** (see Fig. 17.10) is caused by a premature ectopic beat following every sinus beat. The rhythm is not regular because every weak pulse is premature.
- ***Pulsus bisferiens*** (see Fig. 17.10) is found in hypertrophic cardiomyopathy (HCM) and in mixed aortic valve disease (regurgitation combined with stenosis). The first systolic wave is the 'percussion' wave, produced by transmission of the left ventricular pressure in early systole. The second peak is the 'tidal' wave, caused by recoil of the vascular bed. This normally happens in diastole (the dicrotic wave), but when the left ventricle empties slowly or is obstructed and cannot empty completely, the tidal wave occurs in late systole. The result is a palpable double pulse.
- ***A dicrotic pulse*** (see Fig. 17.10) results from an accentuated dicrotic wave. It can occur in sepsis, hypovolaemic shock, tamponade, and in the immediate postoperative period after aortic valve replacement.
- ***Paradoxical pulse*** is a misnomer, as it is actually an exaggeration of the normal pattern. In normal subjects the systolic pressure and the pulse pressure (the difference between the systolic and diastolic blood pressures) fall during inspiration. The normal fall of systolic pressure is less than 10 mmHg, and this can be measured using a sphygmomanometer. It is due to increased pulmonary intravascular volume during inspiration. In severe airflow limitation (especially severe asthma) there is an increased negative intrathoracic pressure on inspiration, which enhances the normal fall in blood pressure. In patients with cardiac tamponade the fluid in the pericardium increases the intrapericardial pressure, thereby impeding diastolic filling of the heart. The normal inspiratory increase in venous return to the right ventricle is at the expense of the left ventricle, as both ventricles are confined by the accumulated pericardial fluid within the pericardial space. Paradox can occur via a similar mechanism in constrictive pericarditis but is less common.

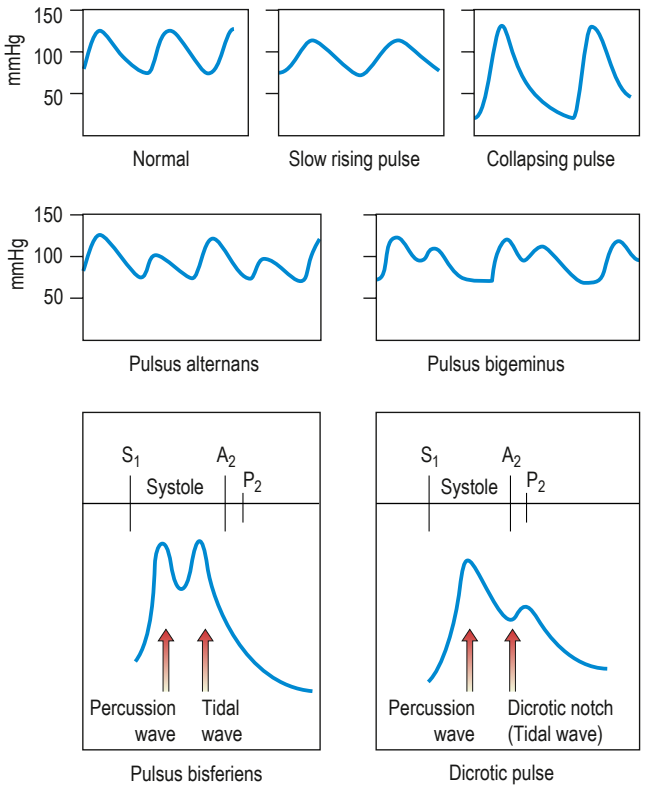


Fig. 17.10 Arterial waveforms. A_2, Aortic component of the second heart sound; P_2, pulmonary component of the second heart sound; S_1, first heart sound.

Blood pressure

The peak systemic arterial blood pressure is produced by transmission of left ventricular systolic pressure. Vascular tone and an intact aortic valve maintain the diastolic blood pressure. Instructions for taking the blood pressure using the auscultatory method are outlined in Box 17.5. See Chapter 18 for further details.

Jugular venous pressure

There are no valves between the internal jugular vein and the right atrium. Observation of the column of blood in the internal jugular system is therefore a good measure of right atrial pressure. The external jugular cannot be relied on because of its valves and because it may be obstructed by the fascial and muscular layers through which it passes; it can be used only if typical venous pulsation is seen, indicating no obstruction to flow.

Box 17.5 Taking the blood pressure

Use a properly calibrated machine.

1. Take the blood pressure in the (right) arm with the patient relaxed and comfortable.
2. Wrap the sphygmomanometer cuff around the upper arm and place the inflation bag over the brachial artery.
3. Inflate the cuff until the pressure exceeds the arterial pressure—when the radial pulse is no longer palpable.
4. Position the diaphragm of the stethoscope over the brachial artery just below the cuff.
5. Slowly reduce the cuff pressure until sounds (Korotkoff sounds) can be heard ***(phase 1).*** This is the systolic pressure.
6. Allow the pressure to fall further until the Korotkoff sounds suddenly become muffled ***(phase 4).***
7. Allow the pressure to fall still further until the sounds disappear ***(phase 5).***
8. The diastolic pressure is usually taken as ***phase 5*** because this phase is more reproducible and nearer to the intravascular diastolic pressure. The Korotkoff sounds may disappear ***(phase 2)*** and reappear ***(phase 3)*** between the systolic and diastolic pressures. Do not mistake ***phase 2*** for the diastolic pressure or ***phase 3*** for the systolic pressure.

Measurement of jugular venous pressure

See Box 17.6.

Elevation of the jugular venous pressure (JVP) occurs in ***heart failure***. An elevated JVP also occurs in:

- constrictive pericarditis and cardiac tamponade (increases in inspiration—***Kussmaul's sign***)
- renal disease with salt and water retention
- over-transfusion or excessive infusion of fluids
- superior vena cava obstruction.

A reduced JVP occurs in hypovolaemia.

Jugular venous pressure wave

This consists of three peaks and two troughs (Fig. 17.11). The peaks are described as ***a***, ***c***, and ***v*** waves and the troughs are known as ***x*** and ***y*** descents. The ***a*** wave and ***v*** wave are the only visible positive waves.

- ***The a wave*** is produced by atrial systole and is increased with right ventricular hypertrophy secondary to pulmonary hypertension or pulmonary stenosis. Giant cannon waves occur in complete heart block and ventricular tachycardia.
- ***The x descent*** occurs when atrial contraction finishes.
- ***The c wave*** occurs during the ***x*** descent and is due to transmission of right ventricular systolic pressure before the tricuspid valve closes.
- ***The v wave*** occurs with venous return filling the right atrium. Giant ***v*** waves occur in tricuspid regurgitation.
- ***The y descent*** follows the ***v*** wave when the tricuspid valve opens. A steep ***y*** descent is seen in constrictive pericarditis and tricuspid incompetence.

Precordium

- With the patient at 45 degrees, the cardiac apex is located in the fifth intercostal space mid-clavicular line. Left ventricular dilation will displace the apex downwards and laterally. It may be impalpable in patients with emphysema, obesity, or pericardial or pleural effusions.
- ***A tapping apex*** is a palpable first sound and occurs in mitral stenosis.
- ***A vigorous apex*** may be present in diseases with volume overload, for example, aortic regurgitation.

Box 17.6 Measurement of jugular venous pressure (JVP)

- The patient is positioned at about 45 degrees to the horizontal (between 30 and 60 degrees), wherever the top of the venous pulsation can be seen in a good light.
- The JVP is measured as the vertical distance between the manubriosternal angle and the top of the venous column.
- The normal JVP is usually <3 cmH_2O, which is equivalent to a right atrial pressure of 8 cmH_2O when measured with reference to a point midway between the anterior and posterior surfaces of the chest.
- The venous pulsations are not usually palpable (except for the forceful venous distension associated with tricuspid regurgitation).
- Compression of the right upper abdomen causes a temporary increase in venous pressure and makes the jugular venous pulse more visible (hepatojugular reflux).

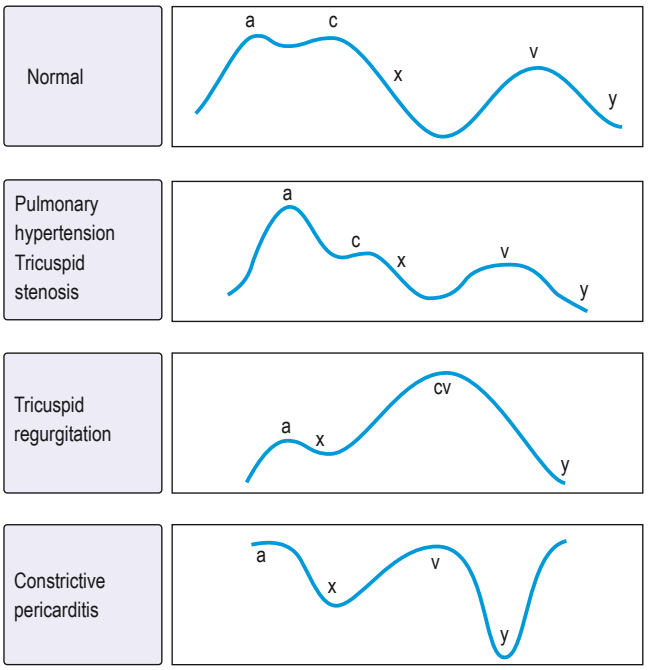


Fig. 17.11 Jugular venous waveforms.

- ***A heaving (sustained) apex*** may occur with left ventricular hypertrophy—aortic stenosis, systemic hypertension, and HCM.
- ***A double pulsation*** may occur in HCM.
- ***A sustained left parasternal heave*** occurs with right ventricular hypertrophy or left atrial enlargement.
- ***A palpable thrill*** may be felt overlying an abnormal cardiac valve, for example, systolic thrill with aortic stenosis.

Auscultation

The ***bell*** of the stethoscope is used for low-pitched sounds (heart sounds and mid-diastolic murmur in mitral stenosis). The ***diaphragm*** is used for high-pitched sounds (systolic murmurs, aortic regurgitation, ejection clicks, and opening snaps). The areas of auscultation are shown in Fig. 17.12. Left-sided valve murmurs may be more prominent in expiration and right-sided in inspiration. Mitral murmurs may be more audible with the patient reclining to the left. Left heart valve closure sounds usually precede their corresponding right heart valve closure sounds.

First heart sound (S1)

This is due to mitral and tricuspid valve closure. A loud S1 occurs in thin people, hyperdynamic circulation, tachycardias, and mild

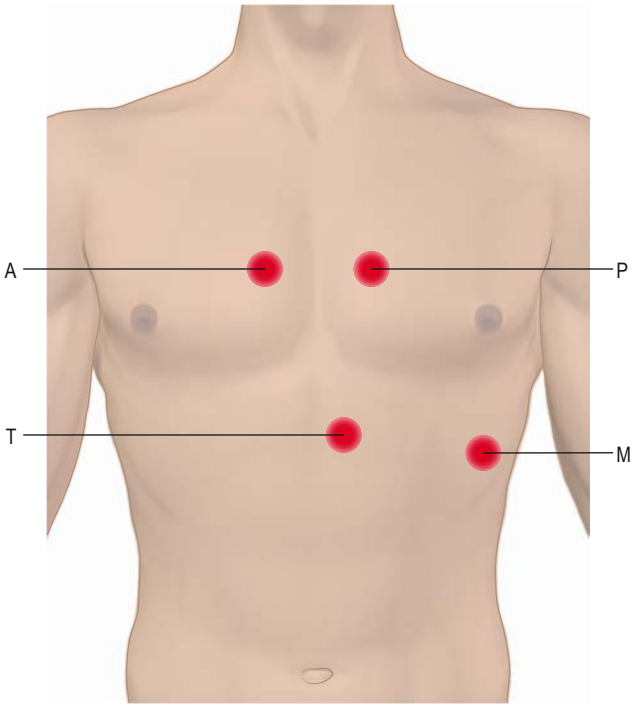


Fig. 17.12 Heart auscultation. Aortic area second intercostal space (ICS) right sternal edge **(A)**. Pulmonary area second ICS left sternal edge **(P)**. Mitral area fifth ICS mid-clavicular line **(M)**. Tricuspid area third to fifth ICS left sternal edge **(T)**.

to moderate mitral stenosis. A soft S1 occurs in obesity, emphysema, pericardial effusion, severe calcific mitral stenosis, mitral or tricuspid regurgitation, heart failure, shock, bradycardias, and first-degree block.

Second heart sound (S2)

This is due to aortic and pulmonary valve closure. Physiological splitting of S2 occurs during inspiration in children and young adults. Keen auscultators may identify additional abnormalities of S2, which may be described as:

- widely split (mitral regurgitation, ventricular septal defect (VSD), right bundle branch block)
- reversed split (aortic stenosis, HCM, left bundle branch block, patent ductus arteriosus)
- fixed split (secundum atrial septal defect)
- single (pulmonary stenosis, Eisenmenger's syndrome).

Third and fourth heart sounds

These are pathological:

- ***A third heart sound ('volume overload')*** is due to rapid ventricular filling and is present in significant heart failure. Reducing left atrial pressure with diuresis can diminish this sound.
- ***A fourth heart sound (pressure overload)*** occurs in late diastole and is associated with atrial contraction. Causes include aortic stenosis, severe systemic hypertension, and left ventricular outflow obstruction, as in HCM, that is, causes of significant left ventricular hypertrophy.

Singly or together, especially in a tachycardic state, they will produce a gallop rhythm.

Heart murmurs

These are due to turbulent blood flow and occur in hyperdynamic states or with abnormal valves. Audio recordings of common murmurs are available to be listened to online via the Elsevier eBooks+. Innocent or flow murmurs are soft, early systolic, short, and non-radiating. They are heard frequently in children and young adults. Heart murmurs can be graded:

- 1/6: very faint; not always heard in all positions
- 2/6: quiet but not difficult to hear
- 3/6: moderately loud
- 4/6: loud with or without thrills
- 5/6: very loud with or without thrills; heard with the stethoscope partly off the chest
- 6/6: with or without thrills; heard with the stethoscope completely off the chest.

The individual murmurs are discussed under valve disease (see 'Valvular heart disease' section).

Cardiac investigations

Blood tests

A panel of routine blood tests is typically carried out in patients with heart disease to exclude other conditions (e.g., kidney disease or anaemia) which may be relevant. Specialist blood tests in cardiology include:

- cardiac enzymes (troponin is used in the diagnosis of ACS, see 'Acute coronary syndromes' section; creatine kinase, aspartate aminotransferase [AST], and lactate dehydrogenase [LDH] may also be elevated in this context).
- B-type natriuretic peptide (BNP), which is used in the diagnosis of heart failure.

Chest radiograph

This allows gross inspection of cardiac size and associated lung pathology such as pulmonary oedema or effusions. This is discussed in more detail in Chapter 8 ('Chest radiography [chest X-ray]' section).

Electrocardiography

The ECG is a recording of the electrical activity of the heart. It is the vector sum of the depolarization and repolarization potentials of all myocardial cells (see **Fig. 17.2**). At the body surface, these generate potential differences of about 1 mV, and fluctuations in these potentials create the familiar P–QRS–T pattern. At rest, the intracellular voltage of the myocardium is polarized at –90 mV compared with that of the extracellular space. This diastolic voltage difference occurs because of the high intracellular potassium concentration, which is maintained by the sodium–potassium pump despite the free membrane permeability to potassium. Depolarization of cardiac cells occurs when there is a sudden increase in the permeability of the membrane to sodium. Sodium rushes into the cell and the negative resting voltage is lost (phase 0 on **Fig. 17.13**). The depolarization of a myocardial cell causes the depolarization of adjacent cells, and in the healthy heart, the entire myocardium is depolarized in a coordinated fashion. During repolarization, cellular electrolyte balance is slowly restored (phases 1, 2, and 3). Slow diastolic depolarization (phase 4) follows until the threshold potential is reached. Another action potential then follows.

The ECG is recorded from two or more simultaneous points of skin contact (electrodes). When cardiac activation proceeds towards the positive contact, an upward deflection is produced on the ECG. Correct representation of a three-dimensional spatial vector requires recordings from three mutually perpendicular (orthogonal) axes. The shape of the human torso does not make

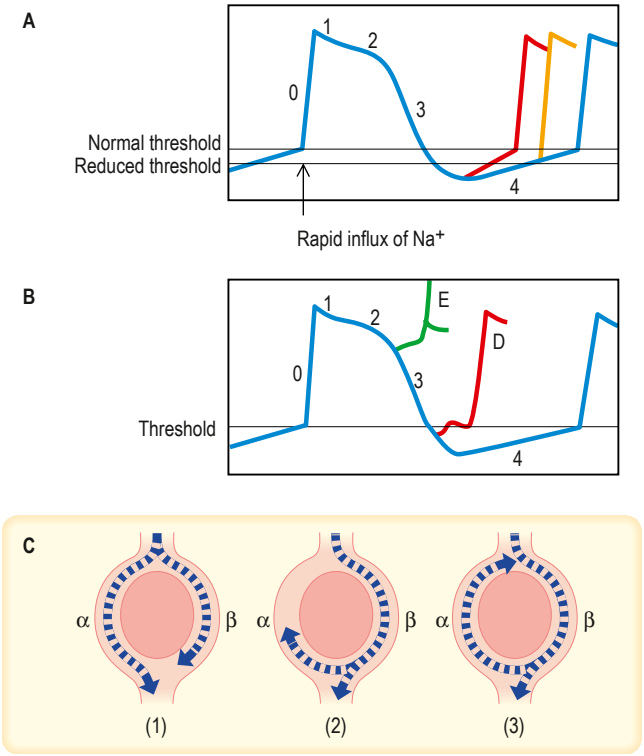


Fig. 17.13 Mechanisms of arrhythmogenesis. **(A–B)** Action potentials (i.e., the potential difference between intracellular and extracellular fluid) of ventricular myocardium after stimulation. **(A)** Increased (accelerated) automaticity due to reduced threshold potential or an increased slope of phase 4 depolarization. **(B)** Triggered activity due to early (E) or delayed (D) 'after-depolarizations' reaching threshold potential. **(C)** Mechanism of circus movement or re-entry. In panel 1 the impulse passes down both limbs of the potential tachycardia circuit. In panel 2 the impulse is blocked in one pathway (α) but proceeds slowly down pathway β, returning along pathway α until it collides with refractory tissue. In panel 3 the impulse travels so slowly along pathway β that it can return along pathway α and complete the re-entry circuit, producing a circus movement tachycardia.

this easy, so the practical ECG records 12 projections of the vector, called 'leads' (Fig. 17.14).

Six of the leads are obtained by recording voltages from the limbs (I, II, III, AVR, AVL, and AVF). The other six leads record potentials between points on the chest surface and an average of the three limbs: RA, LA, and LL. These are designated V_1–V_6 and aim to select activity from the right ventricle (V_1–V_2), interventricular septum (V_3–V_4), and left ventricle (V_5–V_6). Note that leads AVR and V_1 are oriented towards the cavity of the heart; leads II, III, and AVF face the inferior surface; and leads I, AVL, and V_6 face the lateral wall of the left ventricle. A V_4 on the right side of the chest (V_4R) is occasionally useful (e.g., for the diagnosis of right ventricular infarction).

Most ECG machines are simultaneous three-channel recorders, with output given either as a continuous strip or with automatic channel switching. Many modern ECG machines provide automatic interpretation which can be helpful, but many arrhythmias still defy automatic analysis.

ECG waveform

The shape of the normal ECG waveform (Fig. 17.15) has similarities, regardless of the orientation. The first deflection is caused by atrial depolarization; it is a low-amplitude, slow deflection called a ***P wave***. The ***QRS complex*** reflects ventricular activation or depolarization and is sharper and larger in amplitude than the P wave. An initial downward deflection is called the ***Q wave***. An initial upward deflection is called an ***R wave***. The ***S wave*** is the last part of ventricular activation. The ***T wave*** is another slow and low-amplitude deflection that results from ventricular repolarization.

The ***PR interval*** is the length of time from the start of the P wave to the start of the QRS complex. It is the time taken for activation to pass from the sinus node, through the atrium, AV node, and His–Purkinje system, to the ventricle.

The ***QT interval*** extends from the start of the QRS complex to the end of the T wave. This interval represents the time taken to depolarize and repolarize the ventricular myocardium. The QT interval varies greatly with heart rate and is often represented as a corrected QT interval (or QTc) for a given heart rate. There are a number of formulae for derivation of QTc, but the most widely accepted are Bazett's formula and Fridericia's correction (Box 17.7).

An abnormally prolonged QTc can predispose to a risk of dangerous ventricular arrhythmias. Prolongation of the QT interval may be congenital or can occur in many acquired conditions (see Box 17.15).

The ***ST segment*** is the period between the end of the QRS complex and the start of the T wave. In the normal heart all cells are depolarized by this phase of the ECG: that is, the ST segment represents ventricular repolarization.

A normal 12-lead ECG is shown in Fig. 17.16, and the normal values for the electrocardiographic intervals are indicated in Box 17.7. Leads that face the lateral wall of the left ventricle have predominantly positive deflections, and leads looking into the ventricular cavity are usually negative. Detailed patterns depend on the size, shape, and rhythm of the heart and the characteristics of the torso.

Cardiac vectors ('axis')

At any point in time during depolarization and repolarization, electrical potentials are being propagated in different directions. Most of these cancel each other out and only the net force is recorded. This net force in the frontal plane is known as the cardiac vector.

The mean QRS vector can be calculated from the six standard leads (Fig. 17.17):

- ***Normal:*** between –30 degrees and +90 degrees
- ***Left axis deviation:*** between –30 degrees and –90 degrees
- ***Right axis deviation:*** between +90 degrees and +150 degrees.

Calculation of this vector is useful in the diagnosis of some cardiac disorders.

Exercise electrocardiography

This is now rarely used to detect coronary ischaemia as the test has a low sensitivity. The ECG is recorded while the patient walks or runs on a motorized treadmill. The test is based on the principle that exercise increases myocardial demand on coronary blood supply, which may be inadequate during exercise and, at peak stress, can result in relative myocardial ischaemia. Most exercise tests are performed according to a standardized method, such as the Bruce protocol. Normally, there is little change in the T wave or ST segment during exercise.

The patient's exercise capacity (the total time achieved) will depend on many factors; however, individuals who can exercise for less than 6 minutes generally have a poorer prognosis.

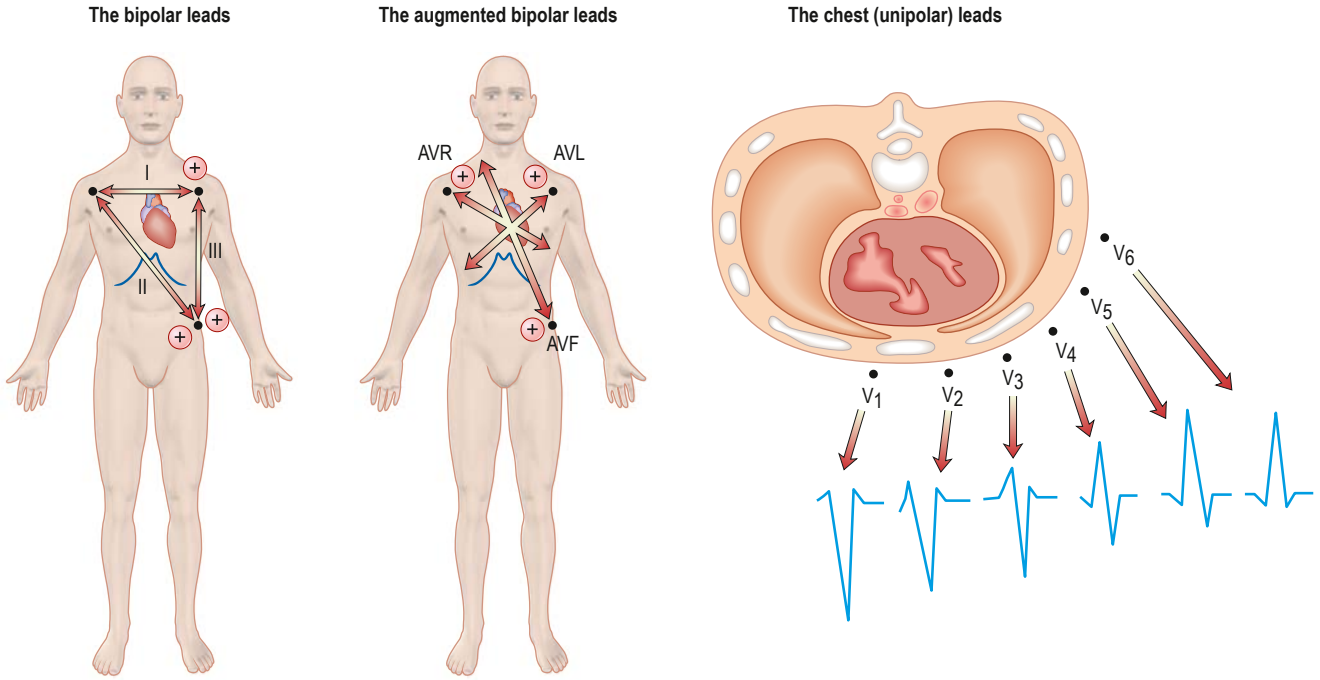


Fig. 17.14 **The connections or directions that comprise the 12-lead electrocardiogram.**

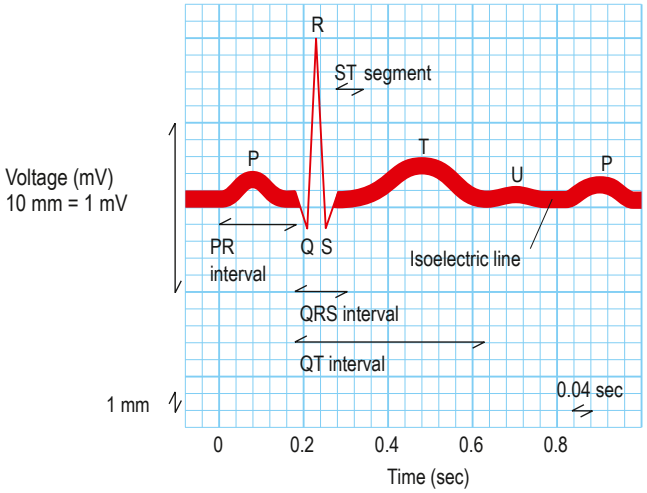


Fig. 17.15 **The waves and elaboration of the normal electrocardiogram.**

Box 17.7 Normal ECG intervals

Parameter	Interval
P wave duration	≤0.12 sec
PR interval	0.12–0.20 sec
QRS complex duration	≤0.11 sec
Corrected QT (QTc)	≤0.44 sec in males
	≤0.46 sec in females
$QT_{cB} = QT\sqrt{}^{2}$ (R–R)	Bazett's square root formula
$QT_{cF} = QT\sqrt{}^{3}$ (R–R)	Fridericia's cube root formula

Myocardial ischaemia provoked by exertion can result in ST-segment depression (>1 mm) in leads facing the affected area. The form of ST-segment depression provoked by ischaemia is characteristic: it either is planar or shows down-sloping depression (Fig. 17.18). Up-sloping depression is a non-specific finding. The degree of ST-segment depression is positively correlated to the degree of myocardial ischaemia.

ST-segment elevation during an exercise test is induced much less frequently than ST depression. When it does occur, it reflects transmural ischaemia caused by coronary spasm or critical stenosis.

Although most abnormalities are detected in lead V_5 (anterior and lateral ischaemia) or AVF (inferior ischaemia), it is best to record a full 12-lead ECG. During an exercise test, the blood pressure and rhythm responses to exercise are also assessed. Exercise normally causes an increase in heart rate and blood pressure. A sustained fall in blood pressure usually indicates severe coronary artery disease (CAD) or cardiomyopathy. A slow recovery of the heart rate to basal levels has also been reported to be a predictor of mortality.

Frequent premature ventricular depolarizations in the recovery stage (post exercise) are associated with an increased risk of death following adjustment for baseline risk factors.

Ambulatory monitoring of heart rhythm

Ambulatory taped electrocardiography

This records transient changes, such as a brief paroxysm of tachycardia, an occasional pause in the rhythm, or intermittent ST-segment shifts (Fig. 17.19). A conventional 12-lead ECG records 10 seconds of cardiac activity and usually samples fewer than 10 complexes. In a 24-hour period over 100 000 complexes are recorded. Such a large amount of data must be analyzed by automatic or semi-automatic methods. This technique is called ***Holter electrocardiography***, named after its inventor; 72-hour monitoring increases sensitivity for less frequent abnormalities and may be the standard monitoring period in some units, and occasionally longer period of event recording may be carried out where there is a high index of suspicion.

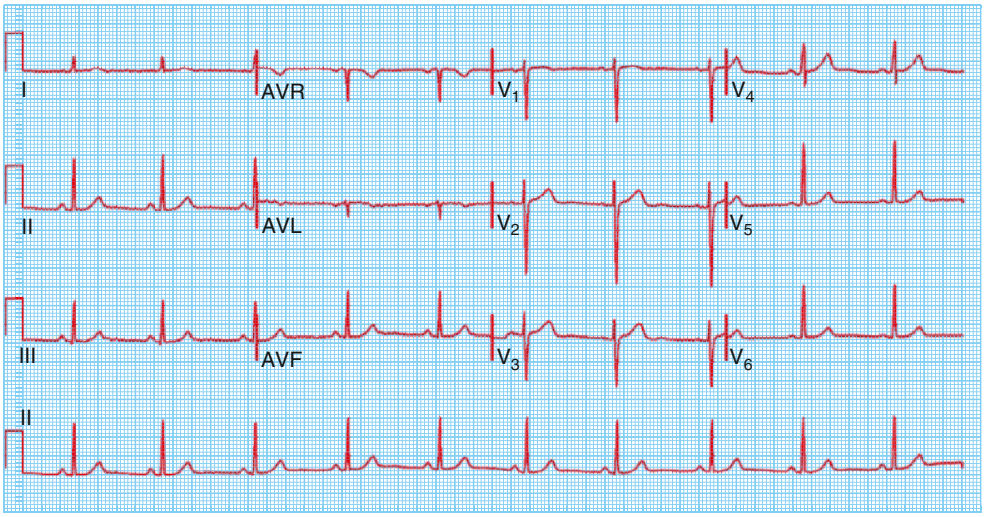


Fig. 17.16 **A normal 12-lead electrocardiogram.**

Implantable loop recording

A very small event recorder, known as an implantable loop recorder (ILR), can be implanted subcutaneously. It is useful in the management of infrequent syncope and has data supporting its use in the identification of paroxysmal atrial fibrillation in patients with unexplained embolic stroke.

Cardiac data from patient devices

Heart rate and single-lead ECG data are now increasingly available from personal, wearable, smart technology; integrating this into care for the individual patient requires a personalized approach.

Analysis of ambulatory heart rhythm recordings

Ambulatory cardiac rhythm recordings using any of the techniques above may reveal evidence of transient arrhythmias such as atrial fibrillation or heart block. In addition, the following techniques have been used to detect signs of cardiac disease from subtle abnormalities of the ECG trace:

- ***Heart rate variability (HRV)*** can be automatically assessed from a 24-hour ECG. HRV is decreased in some patients following MI and represents an abnormality of autonomic tone or cardiac responsiveness. It is not utilized in mainstream cardiology.
- ***Signal-averaged ECG (SAECG)*** is a technique that requires amplification and averaging of abnormal low-amplitude signals that occur beyond the end of the QRS complex and extend well into the ST segment. These signals are therefore also known as late potentials and are too small to be detected on a surface ECG. They arise in areas of slow conduction in the myocardium, such as the border zone of an infarct, where re-entrant ventricular arrhythmias can originate or from scar in patient with arrhythmogenic cardiomyopathy (ACM).
- ***T wave alternans (TWA)*** is a niche technique used as a non-invasive marker of susceptibility to ventricular arrhythmias and sudden cardiac death. TWA represents microvolt level changes in the morphology of the T waves in every other beat and can be detected during acute myocardial ischaemia using amplification techniques. Visible TWA on an ordinary surface ECG is quite a rare phenomenon, except in patients with long QT syndromes, particularly during emotion or exercise.

Tilt testing

Patients with suspected neurocardiogenic (vasovagal) syncope could be further investigated by upright tilt testing. The patient is secured to a table that is tilted to +60 degrees to the vertical for 45 minutes or more. The ECG and non-invasive blood pressure are continuously monitored throughout. If neither symptoms nor signs develop, isoprenaline may be slowly infused or glyceryl trinitrate administered and the tilt repeated. A positive test results in hypotension, sometimes bradycardia (**Fig. 17.20**), and pre-syncope/syncope, and supports the diagnosis of neurocardiogenic syncope. If symptoms and signs appear, placing the patient flat can quickly reverse them. The effect of treatment can be evaluated by repeating the tilt test, but it is not always reproducible. The overall sensitivity, specificity, and reproducibility are low. Tilt table testing can also support a diagnosis of postural orthostatic tachycardia syndrome (POTS, see 'Postural orthostatic tachycardia syndrome' section) if the heart rate increased more than 30 beats/min or increased to more than 120 beats/min after standing, especially if accompanied by symptoms.

Carotid sinus massage

Carotid sinus massage (**Box 17.8**) may lead to asystole (>3 sec) and/or a fall in systolic blood pressure (>50 mmHg). This hypersensitive response occurs in many of the normal (especially elderly) population but may also be responsible for loss of consciousness in some patients with carotid sinus syndrome (see 'Neurally mediated syndromes' section). In one-third of cases, carotid sinus massage is positive only when the patient is standing. Atherosclerosis can cause narrowing and stenosis of carotid arteries. Carotid sinus massage should thus be avoided in patients with carotid bruits.

Echocardiography

Echocardiography is a non-invasive diagnostic technique that is widely employed in clinical cardiology. It involves the use of ultrasound (either alone or with a contrast agent) to assess cardiac structure and function. A physician or technician performs the studies and a comprehensive examination takes 15 to 45 minutes. The ultrasound machines are either mobile on wheels or handheld.

Physics

Echocardiography uses transmitted ultrasound wavelengths of 1 mm or less, which correspond to frequencies of approximately 2

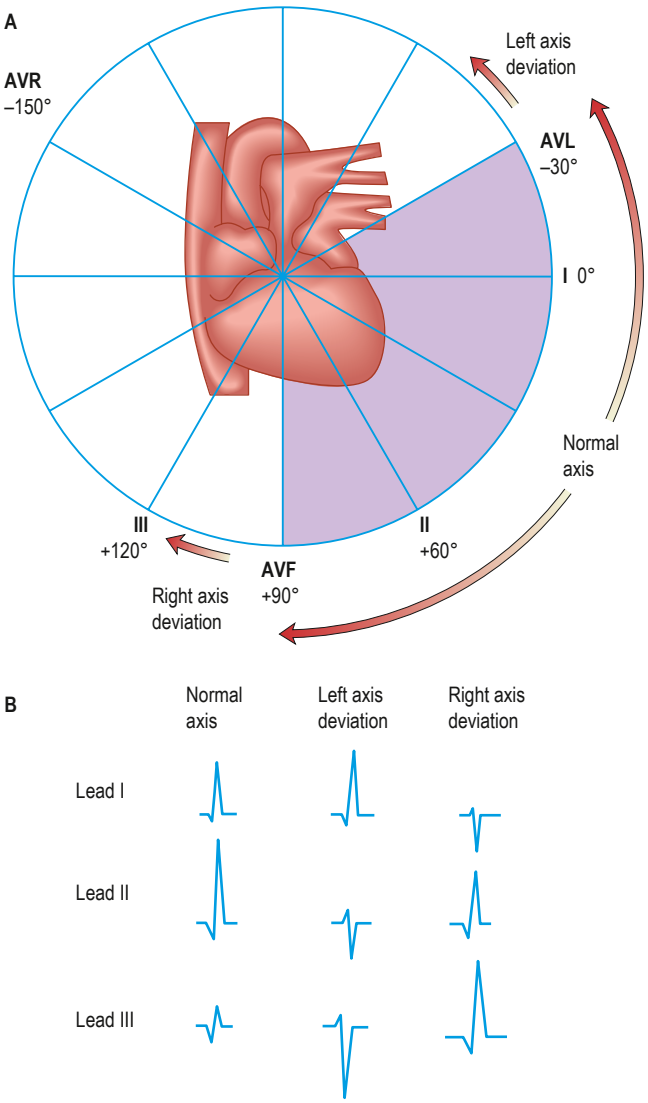


Fig. 17.17 Cardiac vectors. (A) The hexaxial reference system illustrating the six leads in the frontal plane; for example, lead I is 0 degrees, lead II is +60 degrees, and lead III is +120 degrees. **(B)** Calculating the direction of the cardiac vector. In the first column the QRS complex with zero net amplitude (i.e., when the positive and negative deflections are equal) is seen in lead III. The mean QRS vector is therefore perpendicular to lead III and is either −150 degrees or +30 degrees. Lead I is positive, so the axis must be +30 degrees, which is normal. In left axis deviation (second column) the main deflection is positive (R wave) in lead I and negative (S wave) in lead III. In right axis deviation (third column) the main deflection is negative (S wave) in lead I and positive (R wave) in lead III. The frontal plane QRS axis is normal only if the QRS complexes in leads I and II are predominantly positive.

MHz or more (2 million cycles per second or more). At such high frequencies, the ultrasound waves can be focused into a 'beam' and aimed at a particular region of the heart. The waves are generated in very short bursts or pulses a few microseconds long by a crystal transducer, which also detects returning echoes and converts them into electrical signals.

When the handheld crystal transducer is placed on the body surface, the emitted ultrasound pulses encounter interfaces between various body tissues as they pass through the body. In crossing each interface some of the wave energy is reflected, and if the beam

Fig. 17.18 Electrocardiographic changes during exercise testing. The upper traces show significant horizontal ST-segment depression during exercise. The lower traces show ST elevation during exercise.

path is approximately at right angles to the plane of the interface, the reflected waves return to the transducer as an echo. Since the velocity of sound in body tissues is almost constant (1550 m/s), the time delay for the echo to return measures the distance of the reflecting interface. Thus, if a single ultrasound pulse is transmitted, a series of echoes return, the first of which is from the closest interface.

Echocardiographic modalities

M-mode and two-dimensional echocardiography

M-mode echocardiography is a technique that details the changing motion of structures along the ultrasound beam with time. Thus the motion of the interventricular septum during the cardiac cycle (either towards or away from the transducer placed on the chest wall) can be assessed and quantified. Stationary structures thus generate horizontal straight lines; the distances of these lines from the top of the screen indicate the depth of the structures, and movements, such as those of heart valves, are indicated by zigzag lines (Fig. 17.21C). Alternatively, a series of views from different positions can be obtained in the form of a two-dimensional image (cross-sectional two-dimensional echocardiography; Fig. 17.21A and B). This method is useful for delineating anatomical structures and for quantifying volumes of the cardiac chambers.

Ejection fraction (EF) is best calculated from the difference in LV volumes from systole to diastole, derived from planimetered measurements of LV area in at least two planes. A normal EF is over 55%. This method is helpful in diagnosing heart failure as well as assessing the response of the patient to therapy. It also permits estimation of LV mass.

Three-dimensional echocardiography

Three-dimensional echocardiography is increasingly used in clinical practice and allows acquisition of a volumetric dataset using a multiplane probe rotating around a fixed axis. This allows more accurate volumetric analysis of ventricular function and mass, assessment of mitral and aortic valve disease, and assessment of adult congenital heart disease (Fig. 17.22).

Doppler echocardiography

Echocardiography utilizes echoes from tissue interfaces. Using high amplification, it is also possible to detect weak echoes scattered by small targets, including those from red blood cells. Blood velocity in

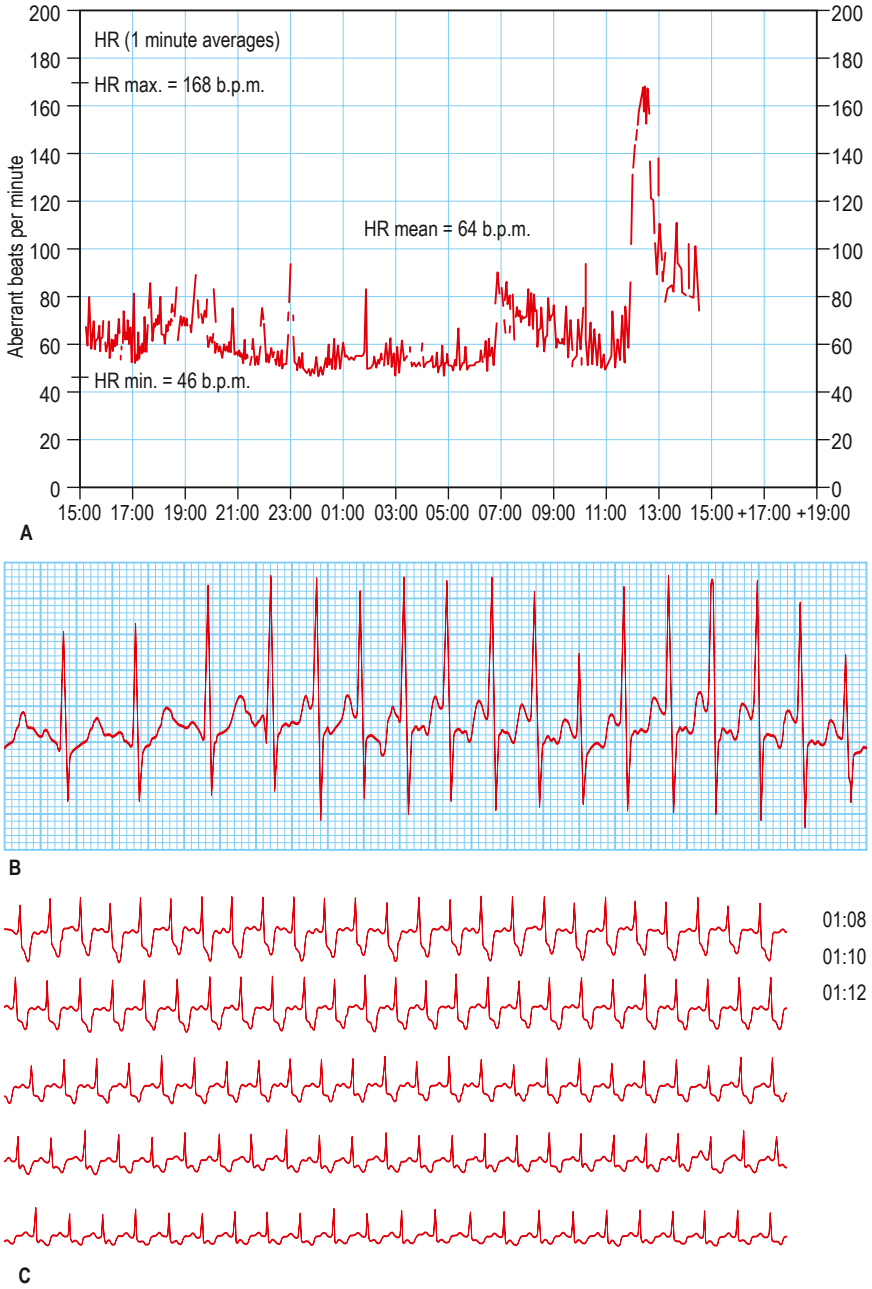


Fig. 17.19 Examples of Holter recordings. **(A)** A tachograph showing a sudden increase in heart rate (HR) between 11:00 and 13:00 hours. **(B)** An ambulatory ECG record of the same patient, revealing supraventricular tachycardia. **(C)** An ambulatory ECG showing ST depression in a patient with silent myocardial ischaemia.

the heart chambers is typically much more rapid (>1 m/s) than the movement of myocardial tissue. If the blood is moving in the same direction as the direction of the ultrasound beam, the frequency of the returning echoes will be changed according to the Doppler phenomenon. The Doppler shift frequency is directly proportional to the blood velocity. Blood velocity data can be acquired and displayed in several ways.

Pulsed-wave (PW) Doppler extracts velocity data from the pulse echoes used to form a two-dimensional image and gives useful qualitative information. PW echoes can be specified from locations within an image identified by a sample volume cursor placed on the screen. Such information from the left ventricular outflow tract (LVOT) and right ventricular outflow tract provides the stroke distance and is used to estimate cardiac output and also to quantify intracardiac shunts. Cardiac output can then be derived using the formula:

$$\text{CO} = \text{stroke volume} \times \text{heart rate}$$

Stroke volume is the stroke distance multiplied by the area of the LVOT, which can also be measured echocardiographically. PW Doppler of the flow across the mitral valve and into the left atrium through the pulmonary veins can be used as an element of the estimation of left ventricular filling pressure.

Colour flow Doppler imaging uses one colour (red) for blood flowing towards the transducer and another colour (blue) for blood flowing away. This technique allows the direction, velocity, and timing of the flow to be measured with a simultaneous view of cardiac

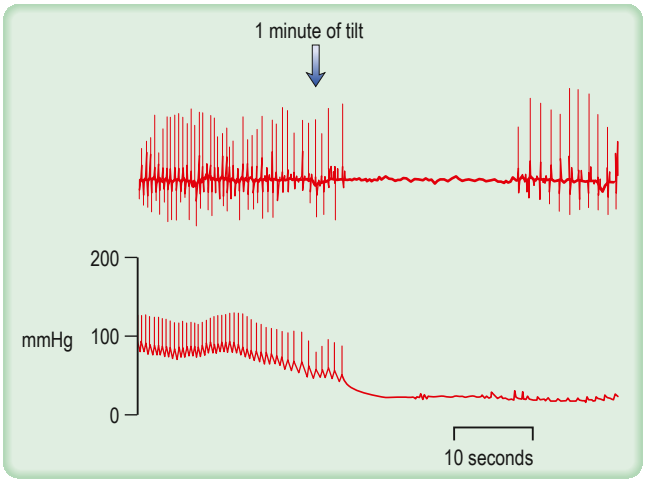


Fig. 17.20 Tilt test. The ECG *(top trace)* and arterial blood pressure *(bottom trace)* recorded during a tilt test. After 1 minute of tilt, hypotension, bradycardia, and syncope occur.

Box 17.8 Carotid sinus massage

- Explain the procedure to the patient and obtain informed consent.
- Ensure that there is no significant carotid artery disease (carotid bruits).
- Provide continuous electrocardiographic monitoring.
- Place the patient in the supine position with the head slightly extended.
- Start with right carotid sinus massage.
- Apply firm rotary pressure to the carotid artery at the level of the third cervical vertebra for 5 sec.
- Alternatively, apply steady pressure.
- If there is no response, massage the left carotid sinus.
- Generally, right carotid sinus massage decreases the sinus node discharge, and left carotid sinus massage slows atrioventricular conduction.
- Do not massage both carotid sinuses at the same time.
- Single application of carotid sinus pressure may be effective in about 20% to 30% of patients with paroxysmal supraventricular tachycardias; multiple applications can terminate tachycardia in about 50% of patients.
- Asystole is a potential but rare complication.

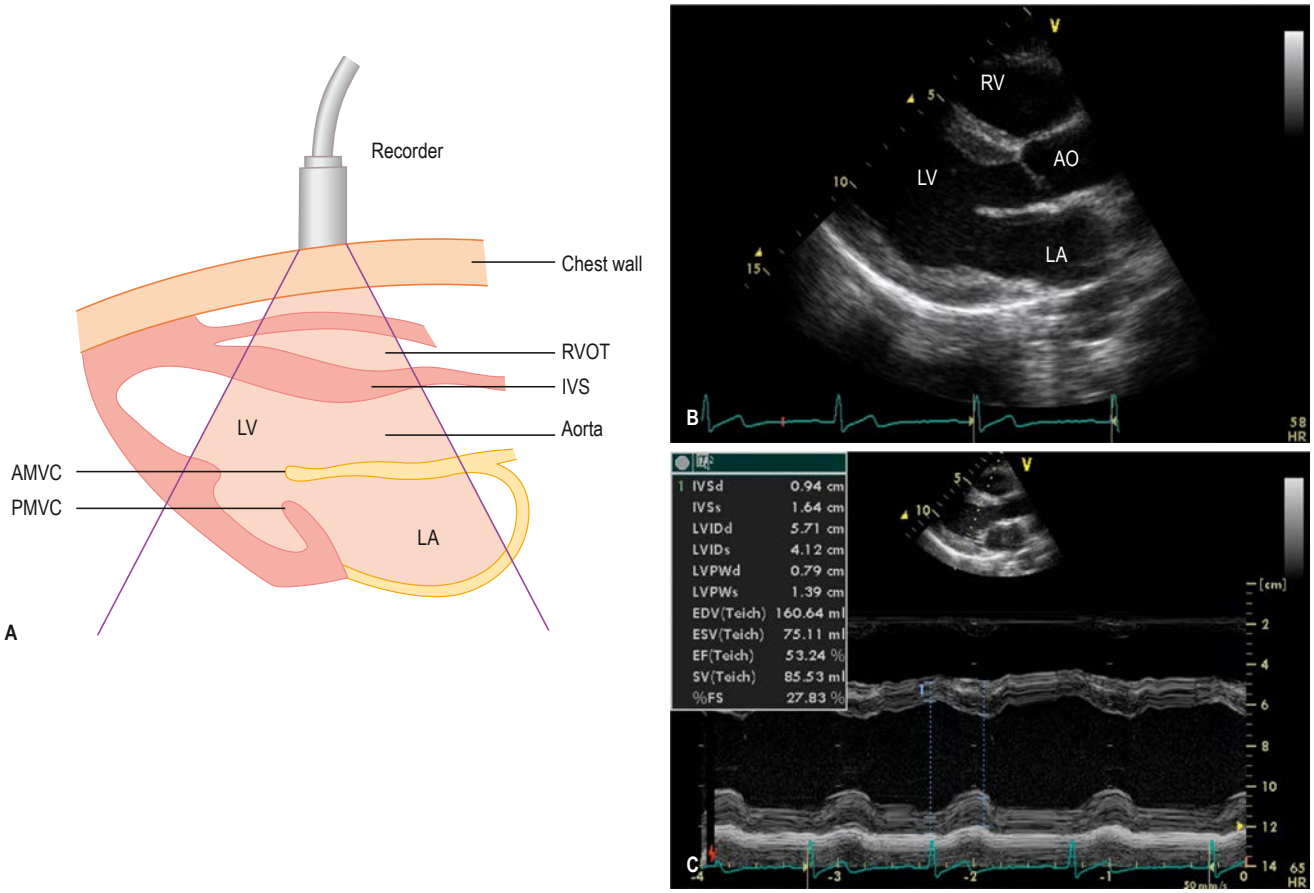


Fig. 17.21 Echocardiography. (A) Diagrammatic representations of the anatomy of the area scanned and of the echocardiogram. **(B)** Two-dimensional long-axis view from a normal subject. **(C)** M-mode recording from a normal subject with the ultrasound beam directed across the left ventricle, just below the mitral valve. *AMVC,* Anterior mitral valve cusp; *AO,* aorta; *IVS,* interventricular septum; *LA,* left atrium; *LV,* left ventricle; *PMVC,* posterior mitral valve cusp; *RV,* right ventricle; *RVOT,* right ventricular outflow tract.

structure and function. Colour flow Doppler is used to help assess valvular regurgitation (Fig. 17.23) and may be helpful in the assessment of coronary blood flow.

Continuous-wave (CW) Doppler collects all the velocity data from the path of the beam and analyzes it to generate a spectral display. This is unlike PW Doppler, which provides information from a particular sample volume at one location along a line. Thus CW Doppler does not provide any depth information.

The outline of the envelope of the spectral display is used to estimate the value of peak velocity throughout the cardiac cycle. CW Doppler is typically used to assess valvular obstruction, which then causes increased velocities. For example, normal flow velocities are

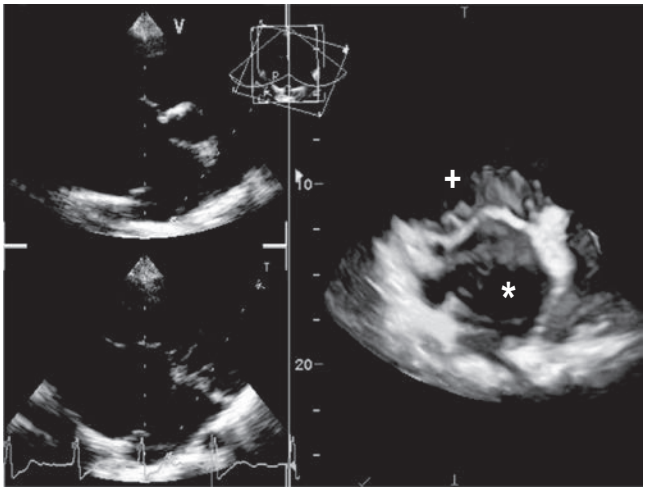


Fig. 17.22 Three-dimensional echocardiograph in a patient with an atrial septal defect (*). The mitral valve (+) and the left atrial anatomy are well delineated.

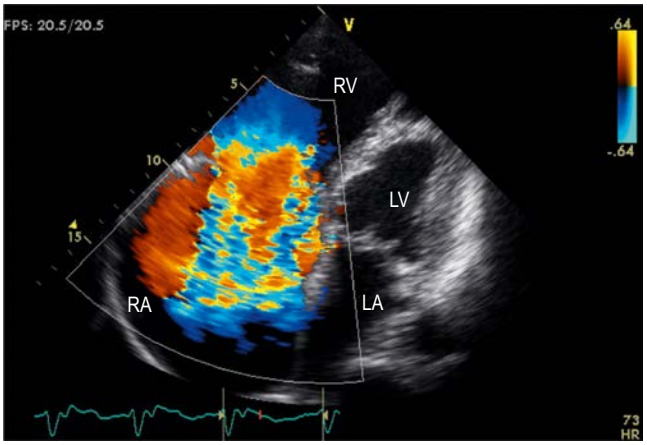


Fig. 17.23 Colour Doppler. Blood is shown flowing away from the echocardiography probe as a blue signal and towards the probe as a red signal. In this patient with tricuspid regurgitation blood leaks from the right ventricle (RV) to the right atrium (RA) during cardiac systole. *LA,* Left atrium; *LV,* left ventricle.

of the order of 1 m/s across the normal aortic valve, but if there is a severe obstructive lesion, such as a severely stenotic aortic valve, velocities of 4 m/s or more can occur. These velocities are generated by the pressure gradient that exists across the lesion.

According to the Bernoulli equation, the pressure difference between two chambers is calculated as: 4 multiplied by the square of the CW Doppler velocity between chambers. Thus a velocity of 5 m/s across the aortic valve suggests a peak gradient of 4 × 5 × 5 = 100 mmHg between the ascending aorta and the left ventricle. This equation has been validated in a wide variety of clinical situations, including valve stenoses, VSDs, and intraventricular obstruction (as in HCM). It is often clinically unnecessary to resort to invasive methods such as cardiac catheterization to measure intracardiac pressure gradients.

Similarly, pulmonary artery (PA) systolic pressure and right ventricular diastolic pressure can be calculated using the Bernoulli equation. In this case CW Doppler tracing of the tricuspid regurgitant jet is used to estimate the pressure gradient between the right ventricle and the right atrium. The pulmonary artery systolic pressure is then calculated by adding the estimated right atrial pressure to the pressure gradient between the right ventricle and the right atrium.

Tissue Doppler is similar to PW Doppler. It measures myocardial tissue velocities within a particular sample volume placed on the image. Such velocities are of the order of 1 cm/s. Currently, tissue Doppler of the mitral annulus is used as part of the estimation of left ventricular filling pressure.

Other ultrasound modalities include harmonic power Doppler, pulse inversion Doppler, and ultra-harmonics, which are used to detect and amplify microsphere-specific signals as part of the echocardiographic assessment of myocardial perfusion.

Transoesophageal echocardiography

In transoesophageal echocardiography (TOE, see **Fig. 17.75**), a transducer mounted on a flexible tube is placed into the oesophagus. This involves the use of local anaesthesia to suppress the gag reflex and, frequently, intravenous sedation. High-resolution images can be obtained because of the close proximity of the heart to the transducer in the oesophagus, and also because of the higher frequencies that are used relative to transthoracic imaging. TOE is most commonly used to assess valve structure and function (to assess for reparability of mitral valve prolapse or suitability of transcatheter repair), the features and complications of IE, and the aorta (for aortic dissection); it is also useful for seeking a cardiac source of embolus.

Stress echocardiography

Echocardiography can be used clinically to evaluate the patient for the presence of myocardial scars and reversible ischaemia. Since ultrasound cannot directly detect red blood cells in capillaries, myocardial wall motion is used as a surrogate for perfusion. Myocardial segments that demonstrate a change in function (defined as a change or reduction in thickening) from rest to stress can be assumed to be supplied by a flow-limiting stenosis in the epicardial artery or graft.

Stress for this indication needs to be inotropic to induce true ischaemia. Physiological stress includes treadmill exercise, which requires rapid image acquisition, at high heart rates as the patient comes off the treadmill, to maintain sensitivity of the test. Alternatively, pharmacological stress can be induced with dobutamine at graded doses. This is relatively safe but complications such as ventricular arrhythmia have been reported. This technique can also assess viability of the myocardium and hibernating or stunned myocardium (see 'Hibernating myocardium and myocardial stunning' section).

Myocardial perfusion 'contrast' echocardiography

In order to assess myocardial perfusion by echocardiography (MPE), microspheres of similar size to red blood cells are used as an intravenous contrast agent. Microsphere-specific ultrasound modes, such as harmonic power Doppler, can be used for detection. Intravenous contrast infusion is employed to fill the myocardium. A pulse of ultrasound then destroys microspheres within the capillaries (and not the left ventricular cavity), and the time taken to replenish the capillaries is a measure of myocardial blood flow. The time taken to fill should be significantly shorter at stress than at rest.

Intravascular coronary ultrasound

Intravascular ultrasound (IVUS) probes can be used to obtain images of proximal coronary arteries as part of a percutaneous transluminal coronary angioplasty (PTCA) procedure, for example, to assess adequacy of deployment of intracoronary stents.

Nuclear imaging

Nuclear imaging is used to detect MI or to measure myocardial function, perfusion, or viability, depending on the radiopharmaceutical agent employed and the imaging technique chosen. These data are particularly valuable when used in combination. All involve a significant radiation dose.

Image type

Gamma cameras produce a planar image in which structures are superimposed, as in a standard X-ray. Nowadays, single-photon emission computed tomography (SPECT) imaging is preferred and uses similar raw data to construct tomographic images, just as a computed tomography (CT) image is reconstructed from X-rays. This gives finer anatomical resolution. These methods may be used with any of the radiopharmaceutical agents.

Myocardial perfusion and viability

Thallium-201 is rapidly taken up by the myocardium, so an image taken immediately after injection reflects the distribution of blood flow to the myocardium. Areas of ischaemia or infarction receive less ^{201}Tl and appear dark. Between 2 and 24 hours after injection, ^{201}Tl is redistributed so that all cardiac myocytes contain a comparable concentration. Images obtained at this time show dark areas where the myocardium has infarcted but normal density in ischaemic areas. Comparison of early and late images is one method of predicting whether an ischaemic area of myocardium contains enough viable tissue to justify coronary bypass or angioplasty. Technetium-99-labelled tetrofosmin (Fig. 17.24) is also taken up rapidly by cardiac myocytes but does not undergo redistribution. When this substance is injected during exercise, its distribution in the myocardium reflects the distribution of blood at the time of the exercise, even if the image is taken several hours later. This is a sensitive method of detecting myocardial viability. Images produced following injection of [^{99m}Tc]tetrofosmin during exercise can be compared to images produced following injection at rest to decide which areas of ischaemia are reversible. In patients unable to exercise the heart can be stressed with drugs, such as dipyridamole or dobutamine.

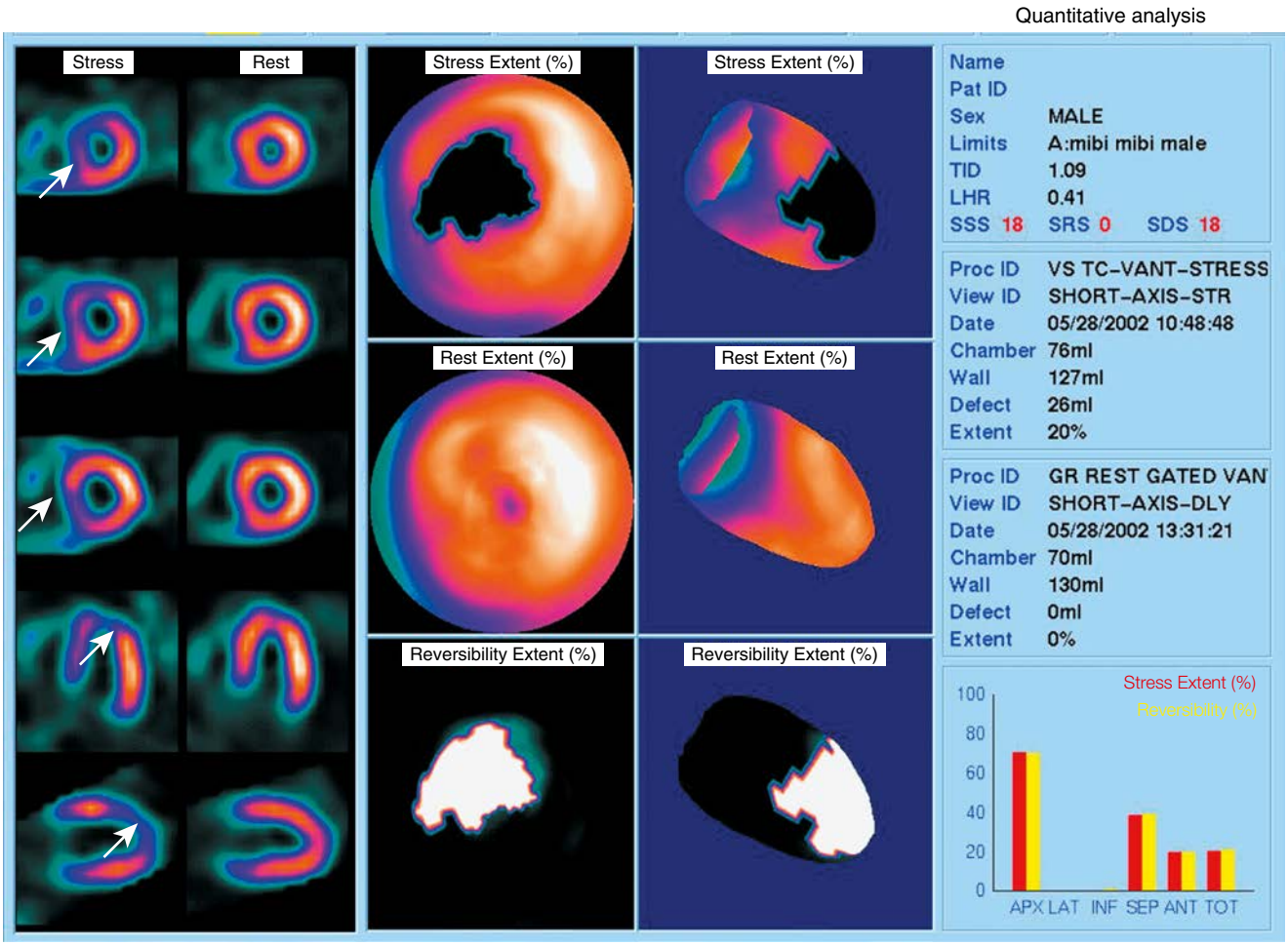


Fig. 17.24 Myocardial single-photon emission computed tomography (SPECT) study acquired with [^{99m}Tc]tetrofosmin tracer. ***Left panel:*** Three short-axis slices and a horizontal and a vertical axis plane of the left ventricle after stress and following a resting re-injection of tracer. The rest images demonstrate normal tracer uptake (orange signal) in the whole of the left ventricle. The stress images demonstrate reduced tracer uptake (*arrows*; purple-blue signal) in the anterior and septal walls, consistent with a significant stenosis in the left anterior descending artery. ***Middle panel:*** Polar maps of the whole myocardium can localize the ischaemic territory. ***Right panel:*** Quantitative analysis can help define the extent and reversibility of the ischaemia.

Infarct imaging

Perfusion images produced using compounds labelled with ^{201}Tl or [^{99m}Tc]sestamibi show a MI as a perfusion defect or 'cold spot'. These methods are sensitive for detecting and localizing the infarct but give no information about its age.

Imaging for amyloid disease

[^{99m}Tc]Pyrophosphate scintigraphy (DPD scan) is now frequently used in the diagnosis of cardiac amyloidosis, in particular transthyretin-type, obviating the need for a cardiac biopsy.

Cardiac computed tomography

CT is useful for assessment of the thoracic aorta and mediastinum. The development of 512-slice multidetector CT (MDCT) scanners has enabled accurate, non-invasive imaging of the coronary arteries.

Coronary artery calcification

Calcium is absent in normal coronary arteries but is present in atherosclerosis and increases with age. Studies have demonstrated a positive correlation between calcification and the presence of coronary artery stenoses, although the relationship is non-linear. MDCT scanners and electron beam CT (EBCT) are used to obtain multiple thin axial slices through the heart, allowing a calcium score to be calculated. The calcium score is based on the X-ray attenuation coefficient or CT number measured in Hounsfield units. Meta-analyses have demonstrated that a higher calcium score is associated with a higher event rate and higher relative risk ratios, although no study to date has shown a net effect on health outcomes of calcium scoring. The calcium score can be used to reclassify intermediate-risk patients to either low-risk or higher-risk. US guidelines support the initiation of statin therapy if the calcium score is more than 100. Guidelines on how to incorporate calcium score in primary prevention do vary between the USA, Europe, the UK, and Australasia.

CT coronary angiography

CT coronary angiography (CTCA) is performed with a supine patient connected to a three-lead ECG for cardiac synchronization. Image quality is optimal with a slow and steady heart rate (<65 b.p.m), which can be obtained with the use of oral or intravenous beta-blockers. Sublingual nitroglycerin (0.4–0.8 mg dose) may improve visualization of the coronary artery lumen. A volume dataset containing the whole heart is acquired during a single breath-hold with the injection of 60 to 80 mL of iodinated contrast agent at 4 to 6 mL/sec. The radiation dose during the scan can consistently be reduced to less than 5 milli-Sieverts (mSv) with ECG-controlled dose modulation; this compares favourably with 2.5 to 5.0 mSv for diagnostic coronary angiography and 15 to 20 mSv for SPECT. The volume dataset is then analyzed with multiplanar reformatting for the presence of coronary artery stenoses (Fig. 17.25). Studies have reported high sensitivity (>85%) and specificity (>90%) for the detection of CAD, with a very high negative predictive value (NPV) (>95%). CTCA has been adopted as an investigation for acute chest pain service in the emergency medicine department to exclude aortic dissection, pulmonary embolism, and CAD.

CT has also established its role in the planning of transcatheter valve procedures and left atrial ablation.

Cardiovascular magnetic resonance

Cardiovascular magnetic resonance (CMR) is a non-invasive imaging technique that does not involve potentially harmful radiation.

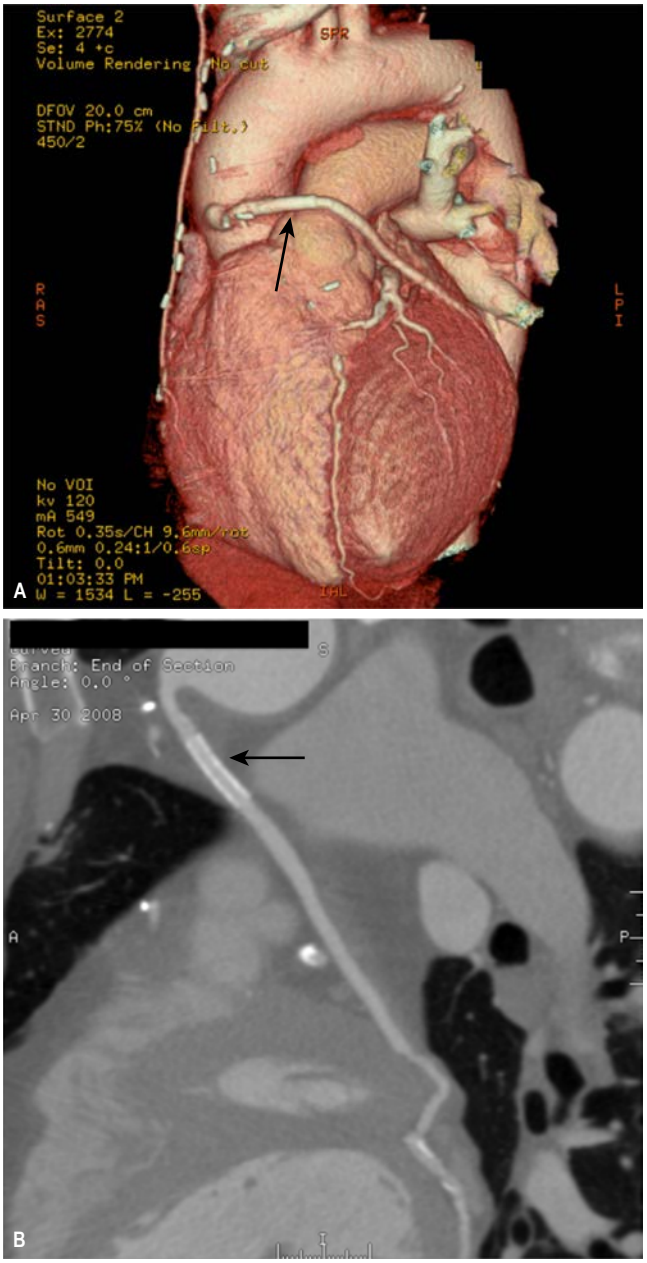


Fig. 17.25 Multislice computed tomography: coronary artery stent. (A) A volume-rendered image in a patient with coronary artery stent insertion *(arrowed)* in a saphenous vein graft to the native circumflex coronary artery. **(B)** The multiplanar reformatted image demonstrates stent patency *(arrowed)* and insertion of the graft to the circumflex coronary artery.

CMR is usually performed with multiple breath-holds to minimize respiratory motion artefacts and cardiac gating to reduce blurring during the cardiac cycle. Several different sequences provide anatomical and functional information. Most sequences do not require a contrast agent but intravenous gadolinium may be needed for magnetic resonance angiography and myocardial perfusion, infarct, and fibrosis imaging. The major contraindications are intracerebral clips, residual metallic fragments (e.g., foreign bodies), and significant claustrophobia. Previously, pacemakers or defibrillators were a strong contraindication to magnetic resonance imaging; however, newer devices are now compatible. Patients with coronary stents and prosthetic valves can be safely scanned.

Clinical use

The current indications for CMR are summarized in Box 17.9 and include:

- ***Congenital heart disease:*** CMR provides additional and complementary information to echocardiography in patients with congenital heart disease. Cine imaging can accurately assess systemic and non-systemic ventricular function and mass. Extracardiac conduits, anomalous pulmonary venous return, and aortic coarctations before and after repair can be studied by CMR and the studies repeated for long-term follow-up without the risk of ionizing radiation.
- ***Cardiomyopathies, pericardial disease, and cardiac masses:*** In HCM, CMR accurately defines the extent and distribution of myocardial hypertrophy and can be used in patients with suboptimal echocardiograms. Intravenous gadolinium can be utilized to demonstrate regional myocardial fibrosis, which is associated with an adverse prognosis. In patients with suspected arrhythmogenic right ventricular cardiomyopathy CMR is the imaging investigation of choice to detect global and regional wall motion abnormalities of the right ventricle and right ventricular outflow tract and to detect fatty or fibro-fatty infiltration of the right and left ventricles. In constrictive pericarditis and restrictive cardiomyopathy (RCM), CMR can demonstrate the effects of the impaired ventricular filling common to both conditions (dilated right atrium and inferior vena cava) and can also determine the thickness of the pericardium (usually 4 mm in normal individuals) (Fig. 17.26). In patients with dilated cardiomyopathy (DCM) CMR can accurately quantify biventricular function and, with gadolinium, can demonstrate myocardial fibrosis. In inflammatory and infiltrative conditions of the myocardium, such as myocarditis, sarcoidosis, and amyloidosis, CMR is increasingly used as a diagnostic investigation due to the different patterns of signal enhancement seen with gadolinium. In patients with thalassaemia CMR can detect iron deposition within the myocardium and guide chelation therapy. CMR can be useful in patients with cardiac masses to differentiate benign from malignant tumours and to identify thrombus not visualized on echocardiography.
- ***Diseases of the aorta:*** CMR is an excellent technique for assessing patients with aortic dissection and can detect the clinical features of an aortic dissection: the intimal flap, thrombosis in a false lumen, aortic regurgitation, pericardial effusion, and aortic dilation. As it does not involve radiation or need contrast, CMR is an ideal method of surveillance of patients with dilated thoracic aortas or repaired coarctation.
- ***Valvular heart disease:*** Valvular stenosis produces signal void on gradient-echo CMR. CMR can quantify the velocity across a stenosed valve using phase-contrast velocity mapping. Valvular regurgitation can be accurately quantified using phase-contrast velocity mapping across the valve, or by calculating the stroke volumes of the left and right ventricles, which are equal in the absence of significant regurgitation. However, in most patients transthoracic and TOE should provide sufficient information.
- ***Coronary artery disease:*** CMR can be used to evaluate coronary artery origin, left ventricular function, myocardial perfusion, and viability in a 'one-stop' approach to the assessment of patients with CAD. Coronary artery origin can be identified with ultra-fast breath-hold or respiratory-gated sequences with high accuracies. Global left ventricular function and wall motion abnormalities can be detected with cine-imaging performed at rest and during dobutamine stress. Myocardial perfusion can be assessed with gadolinium and first-pass imaging; ischaemia can be demonstrated with adenosine for coronary vasodilation. Myocardial viability can be determined using gadolinium and 'delayed enhancement' images. With these techniques, CMR is increasingly used to assess both ischaemia in patients with suspected coronary disease, and myocardial viability prior to revascularization in individuals with impaired cardiac function (Figs. 17.27 and 17.28). The CE-MARC study suggested a higher diagnostic yield for CMR over SPECT in the diagnosis of coronary disease.

Box 17.9 Indications for cardiovascular magnetic resonance (CMR)

- Congenital heart disease (CHD):
  - Anatomical assessment, particularly in patients with complex CHD
  - Follow-up/surveillance studies, particularly post procedures
  - Anomalous pulmonary or systemic venous return
  - Assessment of right or left ventricular function/dilation
- Cardiomyopathies/cardiac infiltration/pericardial disease/myocarditis
  - Disease of the aorta, including aortic dissection, aneurysm, coarctation
  - Valvular heart disease (quantification of stenosis and regurgitation)
  - Cardiac masses (tumours, thrombus, normal structures)
- Coronary artery disease:
  - Ischaemia (wall motion/perfusion assessment)
  - Coronary artery origins
  - Myocardial infarct imaging and viability assessment with delayed enhancement

Positron emission tomography

Positron emission tomography (PET) is based on detection of high-energy emissions caused by annihilation of positrons released from unstable isotopes. PET has several advantages over other techniques, such as improved spatial resolution, accurate quantification,

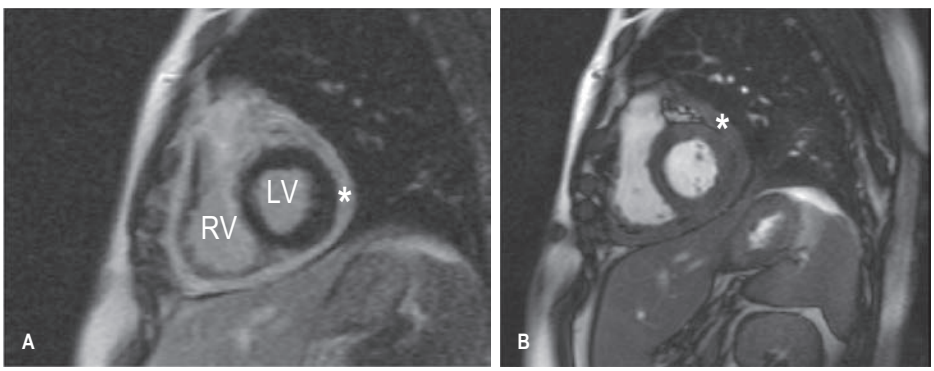


Fig. 17.26 Short-axis cardiac magnetic resonance (CMR) in a patient with constrictive pericarditis. Both images show thickening of the pericardium (*). *LV,* Left ventricle; *RV,* right ventricle.

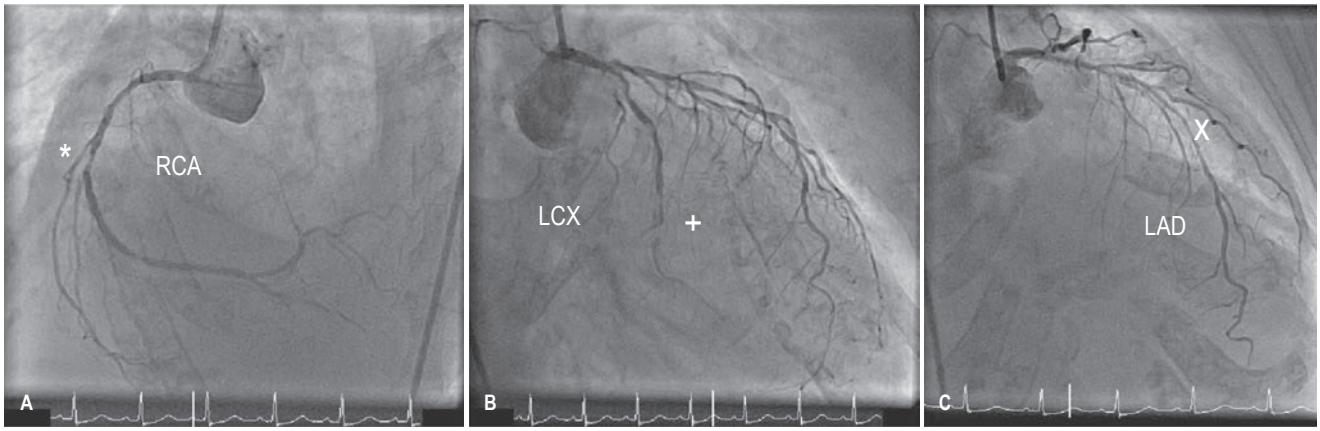


Fig. 17.27 Coronary angiogram (CMR) in a patient with type 2 diabetes mellitus and exertional breathlessness. **(A)** Diagnostic coronary angiography demonstrates a severe stenosis (*) in the right coronary artery (RCA). **(B)** There is a subtotally occluded (+) circumflex coronary artery (LCX). **(C)** A long diseased segment (X) is also seen in the left anterior descending (LAD) coronary artery.

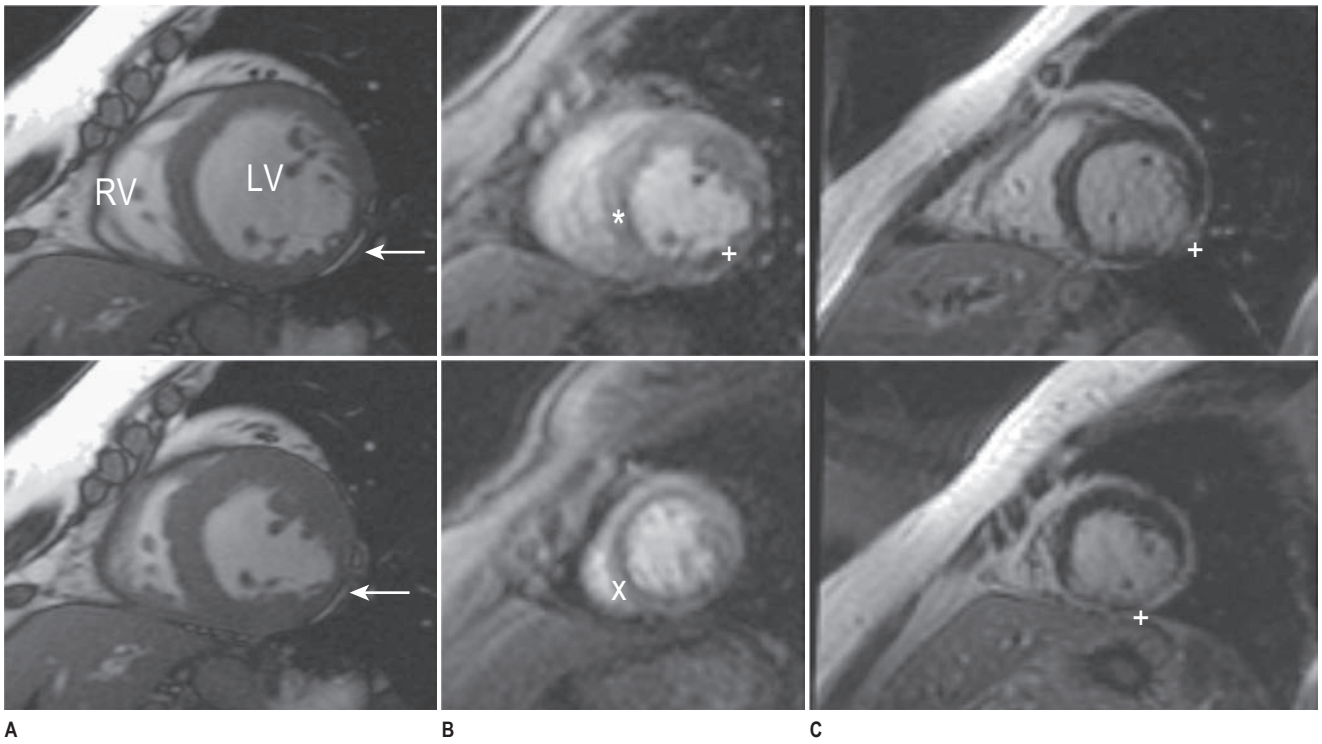


Fig. 17.28 Cardiac magnetic resonance (CMR). **(A)** Short-axis CMR FIESTA cine-imaging (diastole top, systole bottom) demonstrates thinning and hypokinesia of the inferolateral wall *(arrowed)* but preserved function of the rest of the myocardium. *LV,* Left ventricle; *RV,* right ventricle. **(B)** First-pass perfusion CMR (mid-top, apex bottom) demonstrates subendocardial perfusion defects in the inferolateral wall (+), infero-septum (*), and apical segments (X). **(C)** Delayed-enhancement CMR confirms a transmural myocardial infarction of the mid-apical inferolateral wall (+) secondary to a previous infarction from the circumflex coronary artery and demonstrates inducible ischaemia in the right and left coronary territories.

and use of biological isotopes of carbon, nitrogen, and oxygen. However, it is expensive and requires a cyclotron to produce the short-lived tracers. PET/CT has become a useful investigation in the detection of viable myocardium in patients who are suitable for revascularization.

Myocardial perfusion and ischaemia can be determined using PET with ^{13}N-ammonia or ^{15}oxygen with greater sensitivity than SPECT. Myocardial metabolism and viability can be detected with the use of ^{18}F-fluorodeoxyglucose (FDG), which the cardiac myocyte utilizes for energy production in the presence of reduced oxygen supply and blood flow. There may be reduced perfusion to infarcted or fibrotic myocardium, but also reduced FDG uptake. In hibernating myocardium, with viable but dysfunctional myocardium, PET can demonstrate reduced myocardial perfusion but with preserved or increased FDG uptake. FDG PET has a specific role in the diagnosis of active cardiac sarcoidosis and patient evaluation when infection of pacemaker leads or prosthetic valves is thought possible.

Cardiac catheterization

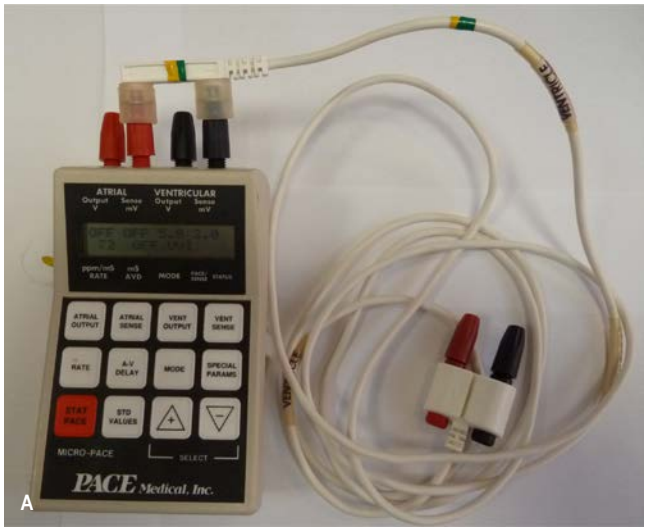
Cardiac catheterization is the introduction of a thin radio-opaque tube (catheter) into the circulation. The right heart is catheterized by introducing the catheter into a peripheral vein (usually the right femoral or internal jugular vein) and advancing it through the right atrium and ventricle into the pulmonary artery. The pressures in the right heart chambers and in the pulmonary artery can be measured directly. An indirect measure of left atrial pressure can be obtained by 'wedging' a balloon-tipped catheter into the distal pulmonary artery (see Fig. 13.13). In this position the pressure from the right ventricle is obstructed by the catheter, and only the pulmonary venous and left atrial pressures are recorded.

Left heart catheterization is usually performed via the right femoral or radial artery. A pigtail catheter is advanced up the aorta and manipulated through the aortic valve into the left ventricle. Pressure tracings are taken from the left ventricular cavity. The end-diastolic pressure is invariably elevated in patients with left ventricular dysfunction. A power injection of radio-opaque contrast material opacifies the left ventricular cavity (left ventriculography) and thereby assesses left ventricular systolic function. The catheter is then withdrawn across the aortic valve into the aorta and the 'pullback' gradient across the valve is measured. Aortography (a power injection into the aortic root) can be performed to assess the aortic root and the presence and severity of aortic regurgitation. Specially designed catheters are then used to engage the left and right coronary arteries selectively, and contrast cine-angiograms are taken in order to define the coronary circulation and identify the presence and severity of any CAD. During the procedure, intracoronary nitrate or adenosine may be used to dilate the coronary arteries. During cardiac catheterization, blood samples may be withdrawn to measure the oxygen content. These estimations are used to quantify intracardiac shunts and measure cardiac output.

If a significant coronary artery stenosis is discovered during a coronary angiogram this can be treated with a stent as an adjunct procedure.

Further reading

Douglas P, Carbello B, Lang R, et al. 2019 ACC/AHA/ASE Key data elements and definitions for transthoracic echocardiography. *Circulation Cardio Vasc Imag* 2019; 12:e000027.
Golub I, Termeie O, Kristo S, et al. Major global coronary artery calcium guidelines. *J Am Coll Cardiol Img* 2023; 16:98–117.
Gongora CA, Bavishi C, Uretsky S et al. Acute chest pain evaluation using coronary computed tomography angiography compared with standard of care: a meta-analysis of randomised clinical trials. *Heart* 2018; 104:215.
Hundley WG, Bluemke DA, Finn JP et al.; American College of Cardiology Foundation Task Force on Expert Consensus Documents. ACCF/ACR/AHA/NASCI/SCMR 2010 expert consensus document on cardiovascular magnetic resonance: a report of the American College of Cardiology Foundation Task Force on Expert Consensus Documents. *Circulation* 2010; 121:2462–2508.



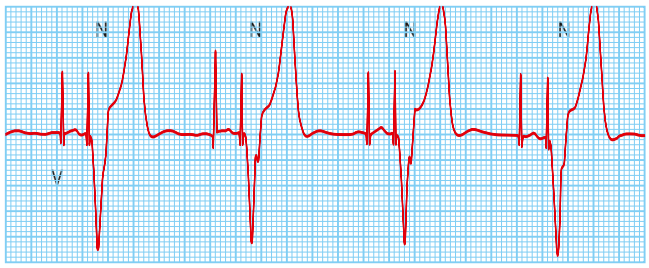


Fig. 17.29 Pacemakers. **(A)** A temporary pacemaker unit. **(B)** A permanent pacemaker with atrial and ventricular leads. **(C)** An electrocardiogram showing dual (atrial and ventricular) chamber pacing. A wide QRS complex results from abnormal activation of the ventricles from the right ventricular apex.

THERAPEUTIC PROCEDURES

Resuscitation, defibrillation, direct current cardioversion, and temporary cardiac pacing are discussed in the Acute Medicine section (see Chapter 11, 'Palpitations' section).

Permanent pacing

Permanent pacemakers are fully implanted in the body and connected to the heart by one or three electrode leads (Fig. 17.29B). The pacemaker is powered by solid-state lithium batteries, which usually last 5 to 10 years. Pacemakers are 'programmable', in that their operating characteristics (e.g., the pacing rate) can be changed by a programmer that transmits specific electromagnetic signals through the skin. The pacemaker leads are passed transvenously to the right heart chambers. Leadless pacemakers have been developed and can be considered in patients without upper extremity venous access or at high risk of device infection (Fig. 17.30). They are implanted via the femoral vein. The uptake of this technology is limited due to cost, single chamber pacing only, and the need to abandon existing device in the right ventricle at the time of battery expiration.

Implantation

Permanent pacemakers are inserted under local anaesthetic using fluoroscopy to guide the insertion of the electrode leads via the cephalic, axillary, or subclavian vein. Perioperative prophylactic antibiotics are routinely prescribed. The pacemaker is usually positioned subcutaneously on top of the pectoral muscle. Following surgery, which usually takes 60 to 90 minutes, the patient rests in bed for 6 to 12 hours before being discharged. Patients may not drive for at least 1 week after implantation and must inform the licensing authorities and their motor insurers.

Complications are rare and can be challenging to manage. Patients with these devices should be periodically reviewed in a pacemaker clinic. Complications include the following:

- Early: Pneumothorax, tamponade, lead displacement, haematoma, infection
- Late: Infection, erosion, lead failure, battery depletion.

Types of pacemaker

Pacemakers are designed both to pace and to sense either the ventricles, the atria, or, more commonly, both chambers. A single-chamber ventricular pacemaker is described as a 'VVI' unit because it paces the ventricle (V), senses the ventricle (V), and is inhibited (I) by a spontaneous ventricular signal. Occasionally (e.g., in symptomatic sinus bradycardia), an atrial pacemaker (AAI) may be implanted. Pacemakers that are connected to both the right atrium and ventricle ('dual-chamber' pacemakers) are used to simulate the natural pacemaker and activation sequence of the heart. This form of pacemaker is called DDD because it paces the two (dual) chambers, senses both (D), and reacts in two (D) ways: pacing in the same chamber is inhibited by spontaneous atrial and ventricular signals, and ventricular pacing is triggered by spontaneous atrial events (Fig. 17.29C).

In addition, pacemakers may be 'rate-responsive' (R). A rate-responsive pacemaker detects motion (level of vibration or acceleration), respiration, or changes in QT interval; by employing one or more biosensors, it changes its rate of pacing so that it is appropriate to the level of exertion.

The choice of pacemaker mostly depends on the underlying rhythm abnormality and the general condition of the patient. For example, complete heart block in patients with sinus rhythm should be treated with a dual-chamber device in order to maintain AV synchrony, whereas inactive or infirm patients may not benefit from the most sophisticated units.

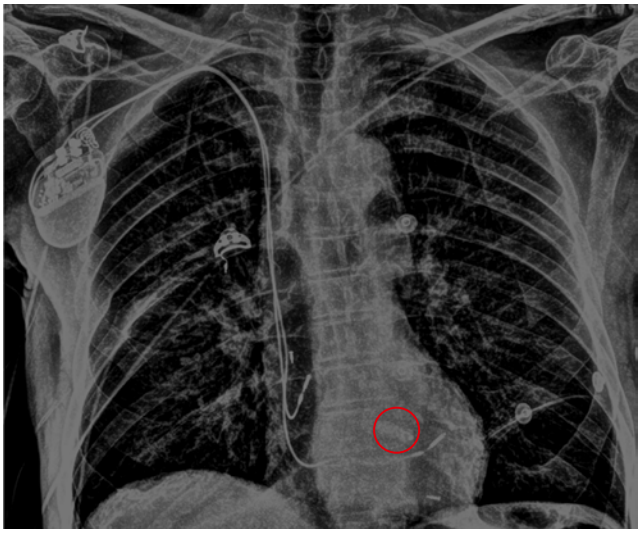


Fig. 17.30 Leadless pacemaker. Chest radiograph showing a right-sided dual chamber pacemaker. The *red circle* shows a leadless pacemaker implanted into the right ventricle that, in this case, was implanted as the transvenous pacemaker has failed leads and the superior vena cava was occluded.

Cardiac resynchronization therapy

In patients with heart failure and, in particular, left bundle branch block (LBBB) another ventricular lead may be implanted via the coronary sinus to the epicardial surface of the left ventricle. More recently, conduction system pacing has become vogue, whereby a lead is torqued deeper into the RV septum such that the electrode can directly capture a branch of the left bundle, thereby effecting a narrower paced QRS on ECG and a more synchronized ventricular activation pattern compared to standard RV endocardial pacing (Fig. 17.31). The longer-term implications of pacing using this technique are currently under evaluation.

Implantable cardiac defibrillators

An implantable cardioverter–defibrillator (ICD) may be considered in a survivor of a ventricular fibrillation or ventricular tachycardia arrest, or as a primary prevention device to abort sudden cardiac death in patients at high risk of developing ventricular arrhythmia (see 'Implantable cardioverter–defibrillator' section).

Pericardiocentesis

A pericardial effusion is an accumulation of fluid between the parietal and visceral layers of pericardium. Fluid is removed for relief of symptoms that are due to haemodynamic embarrassment, or for diagnostic purposes. This can be a technically difficult procedure,

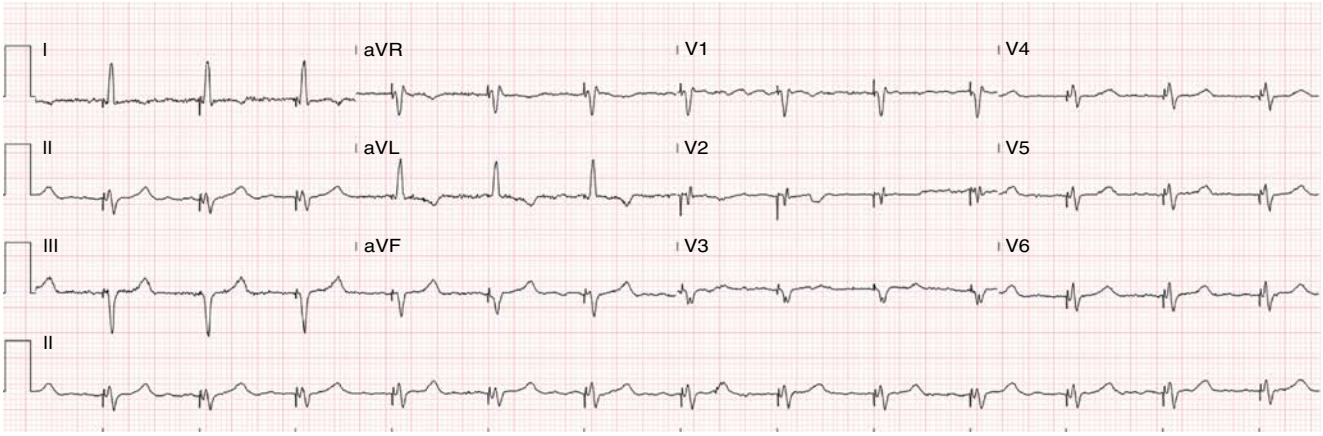


Fig. 17.31 An ECG showing atrial fibrillation and a 'narrower' or more physiologic ventricular paced complex. The pacing lead is capturing the conduction system of the heart.

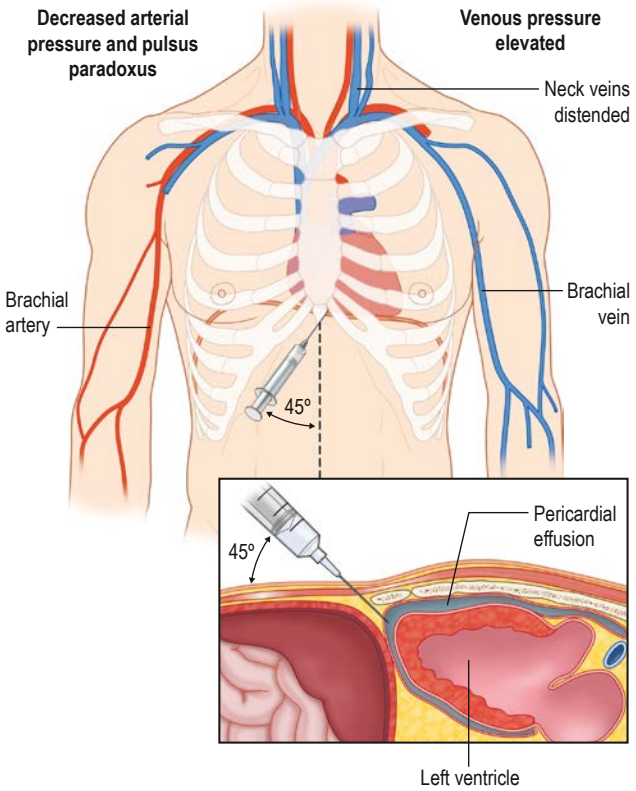


Fig. 17.32 Pericardiocentesis.

particularly in the acute setting, and should be performed in a cardiac laboratory. In an emergency it can be performed at the bedside.

Pericardial aspiration or pericardiocentesis (Fig. 17.32) is performed by inserting a needle into the pericardial space, usually via a subxiphisternal route, under ultrasound guidance. Certain effusions, particularly posterior ones, require surgical drainage under a general anaesthetic. If a large volume of fluid is to be removed, a wide-bore needle and cannula are inserted. The needle may be removed and the cannula left *in situ* to drain the fluid. Fluid that is removed is sent for chemical analysis, microscopy, including cytology, Gram stain, and culture. If reaccumulation of pericardial fluid is anticipated, the cannula may be left in place for several days, or an operation can be performed to cut a window in the parietal pericardium (fenestration) or to remove a large section of the pericardium.

Temporary mechanical cardiac support

Patients with cardiogenic shock have an inpatient mortality of 50%. Definitions of shock vary but include persistent hypotension (systolic blood pressure ≤90 mmHg), with elevated filling pressures, reduced cardiac output, and evidence of end-organ hypoperfusion. Temporary mechanical support should not be used when there is no remediable cause of cardiac dysfunction. It is also unsuitable in patients with moderate or severe aortic regurgitation, aortic dissection, and severe peripheral vascular disease.

Techniques used to provide temporary mechanical cardiac support in the cardiac-intensive care unit setting include:

- ***Miniature cardiac pumps.*** The Impella® device is a miniaturized axial flow device that is delivered across the aortic valve usually from the femoral artery and drives blood from the left ventricle to the aorta. Flow rates of up to 5 L/min can be achieved while the heart is revascularized and recovers from an acute event (see Fig. 17.33). It cannot be used if there is left ventricle thrombus.
- ***Extracorporeal membrane oxygenation (ECMO)*** provides biventricular support in the veno-arterial configuration. Blood is removed from a large (>15 Fr), usually femoral, venous cannula via an external centrifugal pump and, after oxygenation, is returned via a large femoral arterial cannula capable of delivering flows of 4 to 6 L/min. Physiologically, ECMO increases afterload and increases LV end-diastolic pressure and can potentiate pulmonary oedema.
- ***Intra-aortic balloon pumping*** is now infrequently used to assist the failing left ventricle. A catheter with a long, sausage-shaped balloon at its tip is introduced percutaneously into the femoral artery and manipulated under X-ray control so that the balloon lies in the descending aorta just below the aortic arch (Fig. 17.34). The balloon is rhythmically deflated and inflated with carbon dioxide gas. Using the ECG or intra-aortic pressure changes, the inflation is timed to occur during ventricular diastole to increase diastolic aortic pressure and, consequently, to improve coronary and cerebral blood flow. During systole the balloon is deflated, resulting in a reduction in the resistance to left ventricular emptying. Complications occur in about 20% of patients and include aortic dissection, leg ischaemia, emboli from the balloon, and balloon rupture. Embolic complications are reduced by anticoagulation with heparin.

Fig. 17.33 Chest radiograph showing an Impella cardiac pump. The pump propels blood from an entry point near the tip of the left ventricle across the aortic valve and into the aorta.

Further reading

Balthazar T, Vandenbriele C, Verbrugge F, et al. Managing patients with short-term mechanical circulatory support. *J Am Coll Cardiol* 2021; 77(9):1243–1256.

CARDIAC ARRHYTHMIAS

An abnormality of the cardiac rhythm is called a cardiac arrhythmia. Arrhythmias may cause sudden death, syncope, heart failure, chest pain, dizziness, palpitations, or no symptoms at all. There are two main types of arrhythmia:

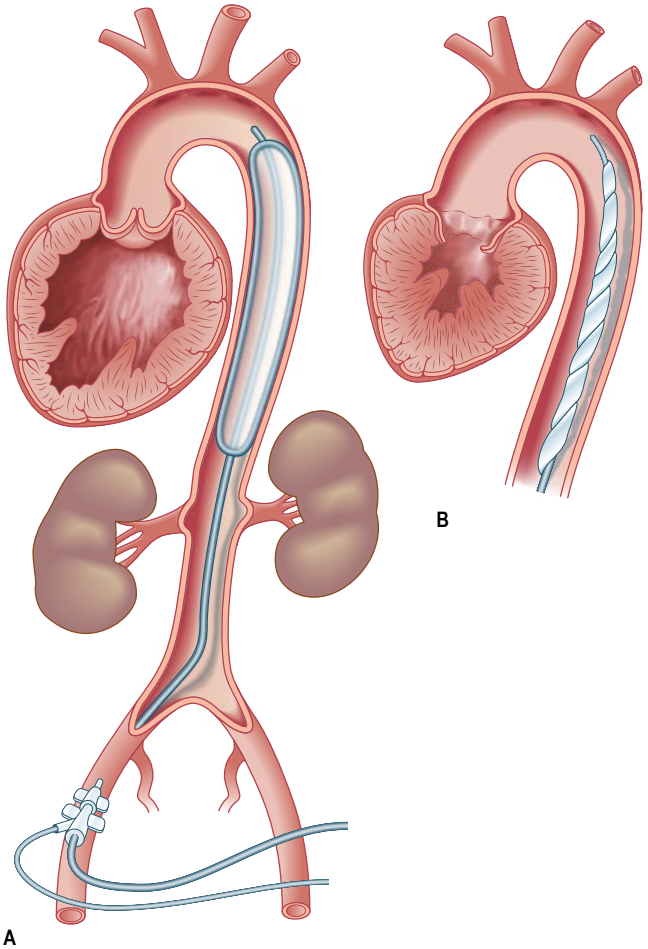


Fig. 17.34 Intra-aortic balloon pump. **(A)** The pump is inflated in the aorta during diastole. **(B)** The balloon is deflated during systole.

- ***Bradycardia:*** the heart rate is slow (<60 b.p.m during the day or <50 b.p.m at night).
- ***Tachycardia:*** the heart rate is fast (>100 b.p.m).

Tachycardias are more symptomatic when the arrhythmia is fast and sustained. Tachycardias are subdivided into ***supraventricular tachycardias (SVTs)***, which arise from the atrium or the AV junction, and ***ventricular tachycardias***, which arise from the ventricles.

Some arrhythmias occur in patients with apparently normal hearts; in others arrhythmias originate from diseased tissue, such as scar, as a result of underlying structural heart disease. When myocardial function is poor, arrhythmias are more symptomatic and are potentially life-threatening.

Sinus node function

The biological cardiac pacemaker is the sinus node (see 'Sinus node [sinoatrial node]' section), and, like most cardiac tissue, it depolarizes spontaneously.

The rate of sinus node discharge is modulated by the autonomic nervous system. Normally, the parasympathetic system predominates, resulting in slowing of the spontaneous discharge rate from approximately 100 to 70 b.p.m. A reduction of parasympathetic tone or an increase in sympathetic stimulation leads to tachycardia; conversely, increased parasympathetic tone or decreased sympathetic stimulation produces bradycardia. The sinus rate in women is slightly faster than in men. Normal sinus rhythm is characterized by P waves that are upright in leads I and II of the ECG (see Fig. 17.16) but inverted in lead AVR.

Sinus arrhythmia

Fluctuations of autonomic tone result in phasic changes of the sinus discharge rate. During inspiration, parasympathetic tone falls and the heart rate quickens; on expiration, the heart rate falls. This variation is normal, particularly in children and young adults. Typically, sinus arrhythmia results in predictable irregularities of the pulse.

Sinus bradycardia

A sinus rate of less than 60 b.p.m (European guidelines) or less than 50 b.p.m (US guidelines) is known as sinus bradycardia. It is usually asymptomatic unless the rate is very slow. Sinus bradycardia is normal in athletes owing to increased vagal tone. Other causes may be divided into systemic or cardiac and are discussed in the 'Bradycardia' section.

Sinus tachycardia

Sinus rate acceleration to more than 100 b.p.m is known as sinus tachycardia. Again, causes may be divided into systemic or cardiac and are discussed in the 'Inappropriate sinus tachycardia' section.

Mechanisms of arrhythmia production

Abnormalities of automaticity, which could arise from a single cell, and abnormalities of conduction, which require abnormal interaction between cells, account for both bradycardia and tachycardia. Sinus bradycardia is a result of abnormally slow automaticity, while bradycardia due to AV block is caused by abnormal conduction within the AV node or the intraventricular conduction system. The mechanisms generating tachycardia are shown in Fig. 17.13.

Accelerated automaticity

The normal mechanism of spontaneous cardiac rhythmicity is slow depolarization of the transmembrane voltage during diastole until the threshold potential is reached and the action potential of the pacemaker cells takes off. This mechanism may be accelerated by increasing the rate of diastolic depolarization or changing the threshold potential (Fig. 17.13A). For example, sympathetic stimulation releases adrenaline (epinephrine), which enhances automaticity. Such changes are thought to produce sinus tachycardia, escape rhythms, and accelerated AV nodal (junctional) rhythms.

Triggered activity

Myocardial damage can result in oscillations of the transmembrane potential at the end of the action potential. These oscillations, called 'after-depolarizations', may reach threshold potential and produce an arrhythmia. If they occur before the transmembrane potential reaches its threshold (at the end of phase 3 of the action potential), they are called 'early after-depolarizations' (E in Fig. 17.13B). When they develop after the transmembrane potential is completed, they are called 'delayed after-depolarizations' (D in Fig. 17.13B).

The abnormal oscillations can be exaggerated by pacing, catecholamines, electrolyte disturbances, hypoxia, acidosis, and some medications, which may then trigger an arrhythmia. The atrial tachycardias produced by digoxin toxicity are due to triggered activity. The initiation of ventricular arrhythmia in the long QT syndrome (see 'Long QT syndrome' section) may be caused by this mechanism.

Re-entry (or circus movements)

The mechanism of re-entry (Fig. 17.13C) occurs when a 'ring' of cardiac tissue surrounds an inexcitable core (e.g., in a region of scarred

myocardium). Tachycardia is initiated if an ectopic beat finds one limb refractory (α), resulting in unidirectional block and the other limb excitable. Provided conduction through the excitable limb (β) is slow enough, the other limb (α) will have recovered and will allow retrograde activation to complete the re-entry loop. If the time to conduct around the ring is longer than the recovery times (refractory periods) of the tissue within the ring, circus movement will be maintained, producing a run of tachycardia. The majority of regular paroxysmal tachycardias are produced by this mechanism.

BRADYCARDIAS AND HEART BLOCK

Bradycardias may be due to failure of impulse formation (sinus bradycardia) or failure of impulse conduction from the atria to the ventricles (AV block).

Bradycardia

Sinus bradycardia

Sinus bradycardia is due either to extrinsic factors that influence a relatively normal sinus node or to intrinsic sinus node disease. The mechanism can be acute and reversible, or chronic and degenerative.

Common ***extrinsic causes*** of sinus bradycardia include:

- hypothermia, hypothyroidism, cholestatic jaundice, and raised intracranial pressure
- drug therapy including beta-blockers, digitalis, non-dihydropyridine calcium-channel blockers
- neurally mediated syndromes (see next section).

Common ***intrinsic causes*** include:

- acute ischaemia and infarction of the sinus node (as a complication of acute MI)
- chronic degenerative changes, such as fibrosis of the atrium and sinus node (sick sinus syndrome).

Sick sinus syndrome or ***sinoatrial disease*** is usually caused by idiopathic fibrosis of the sinus node. Other causes of fibrosis, such as ischaemic heart disease, cardiomyopathy, or myocarditis, can also cause the syndrome. Patients develop episodes of sinus bradycardia or sinus arrest (Fig. 17.35B) and commonly experience paroxysmal atrial tachyarrhythmias (tachy-brady syndrome) owing to diffuse atrial disease.

Neurally mediated syndromes

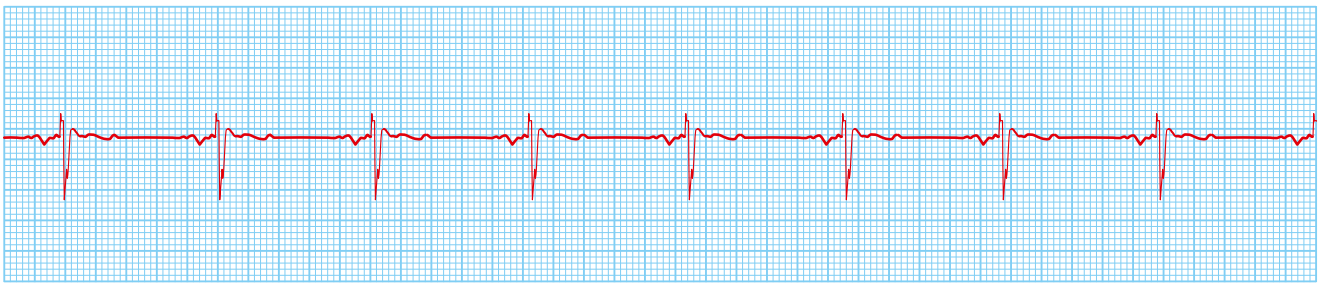
Neurally mediated syndromes are due to a reflex (Bezold–Jarisch) that may result in bradycardia (sinus bradycardia, sinus arrest, and AV block), hypotension, and peripheral vasodilation. These syndromes usually present as syncope or pre-syncope (dizzy spells).

- ***Carotid sinus syndrome*** occurs in the elderly and mainly leads to bradycardia.
- ***Neurocardiogenic (vasovagal) syncope*** usually presents in young adults but may present for the first time in elderly patients. It results from a variety of situations (physical and emotional) that affect the autonomic nervous system. The efferent output may be predominantly bradycardic, predominantly vasodilatory, or mixed.

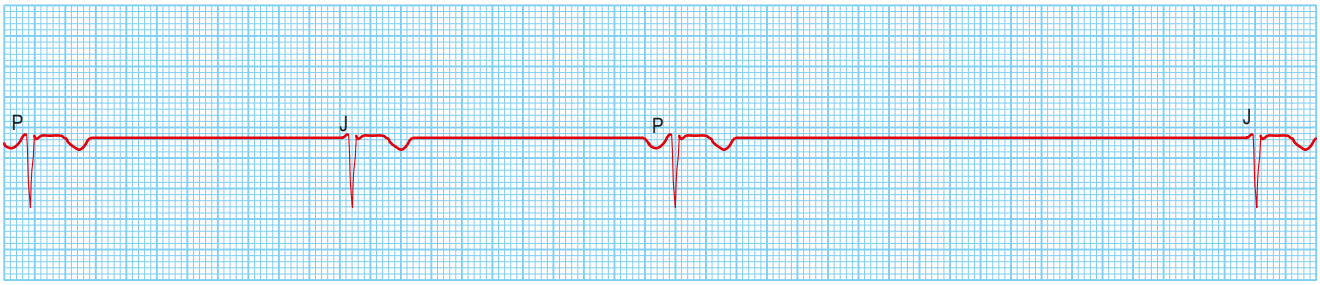
Many medications, such as antihypertensives, tricyclic antidepressants, and neuroleptics, can be the cause of syncope, particularly in the elderly. Careful dose titration and avoidance of combination of two agents with the potential to cause syncope help to prevent iatrogenic syncope.

Management

The management of sinus bradycardia is first to identify and then, if possible, remove any extrinsic causes. Temporary pacing may be employed in patients with reversible causes until a normal sinus rate is restored, and in individuals with chronic degenerative conditions until a permanent pacemaker is implanted.



A



B

Fig. 17.35 Sinus node. **(A)** An ECG showing normal sinus rhythm (PR interval <0.2 sec). A P wave precedes each QRS complex. **(B)** A patient with sick sinus syndrome. The ECG shows sinus arrest (only occasional sinus P waves) and junctional escape beats (J). P waves are inverted in the cavity leads shown.

Chronic symptomatic sick sinus syndrome may require permanent dual chamber pacing (DDD). In patients afflicted by both tachycardias and bradycardias (tachy-brady syndrome), a pacemaker can allow the safe introduction of antiarrhythmic drugs (or ablation therapy) to manage the tachycardic element.

Patients with carotid sinus hypersensitivity (who experience asystole >3 sec), especially if symptoms are reproduced by carotid sinus massage, may also benefit from pacemaker implantation.

Treatment options in vasovagal attacks include avoidance, if possible, of situations known to cause syncope in a particular patient and sitting/lying down and applying counter-pressure manoeuvres (pushing the palms together or crossing the legs) if an attack threatens. Increased salt intake, compression of the lower legs with compression stockings, and drugs such as beta-blockers, alpha-agonists (e.g., midodrine), or myocardial negative inotropes (e.g., disopyramide) may be helpful.

In selected patients, particularly those older than 40 years of age, with symptomatic neurocardiogenic syncope, permanent pacemaker therapy with a dual-chamber pacemaker is helpful. Some devices have features that can help in neurocardiogenic syncope, such as 'rate drop response', which, once activated, paces the heart at a fast rate for a set period of time in order to prevent syncope, or 'closed loop stimulation', which is a proprietary algorithm that increases pacing rate if a reduction in estimated right ventricle contractility is sensed.

Heart block

Heart block or conduction block may occur at any level in the conducting system. Block in either the AV node or the His bundle results in AV block; block lower in the conduction system produces a bundle branch block pattern on ECG.

Atrioventricular block

There are three forms.

First-degree AV block

This is simple prolongation of the PR interval to more than 0.20 seconds. Every atrial depolarization is followed by conduction to the ventricles but with delay (Fig. 17.36).

Second-degree AV block

This occurs when some P waves conduct and others do not. There are several forms (Fig. 17.37):

- ***Mobitz I block*** (Wenckebach block phenomenon) is progressive PR interval prolongation until a P wave fails to conduct. The PR interval before the blocked P wave is much longer than the PR interval after the blocked P wave.
- ***Mobitz II block*** occurs when a dropped QRS complex is not preceded by progressive PR interval prolongation. Usually, the QRS complex is wide (>0.12 sec).
- ***2:1 AV block*** occurs when every second P wave conducts to the ventricles. This form of second-degree block is neither Mobitz I nor Mobitz II.

Advanced or high-degree AV block occurs when only every third P wave (or higher ratio) conducts to the ventricles. ***Wenckebach AV block***, in general, is due to block in the AV node, whereas Mobitz II block signifies block at an infranodal level, such as the His bundle. The risk of progression to complete heart block is greater and reliability of the resultant escape rhythm is less with Mobitz II block. Therefore pacing is usually indicated urgently in Mobitz II block, whereas patients with Wenckebach AV block are usually monitored.

Acute MI may produce ***second-degree, 2:1***, or ***higher-degree heart block***. In inferior MI close monitoring and transvenous temporary backup pacing are all that is required. In anterior MI second-degree heart block is associated with a high risk of progression to complete heart block, and permanent pacemaker implantation is usually indicated. If cardiac function is severely depressed, biventricular pacing may be employed (see 'Cardiac resynchronization therapy or implantable cardioverter–defibrillator' section).

Third-degree (complete) AV block

Complete heart block occurs when all atrial activity fails to conduct to the ventricles (Fig. 17.38). In patients with complete heart block the aetiology needs to be established (Box 17.10). In this situation life is maintained by a spontaneous escape rhythm.

A narrow-complex escape rhythm (<0.12 sec QRS complex) originates from the His bundle and therefore implies that the region of block lies more proximally in the AV node. The escape rhythm occurs with an adequate rate (50–60 b.p.m.) and is relatively reliable. Treatment depends on the aetiology. Recent-onset,

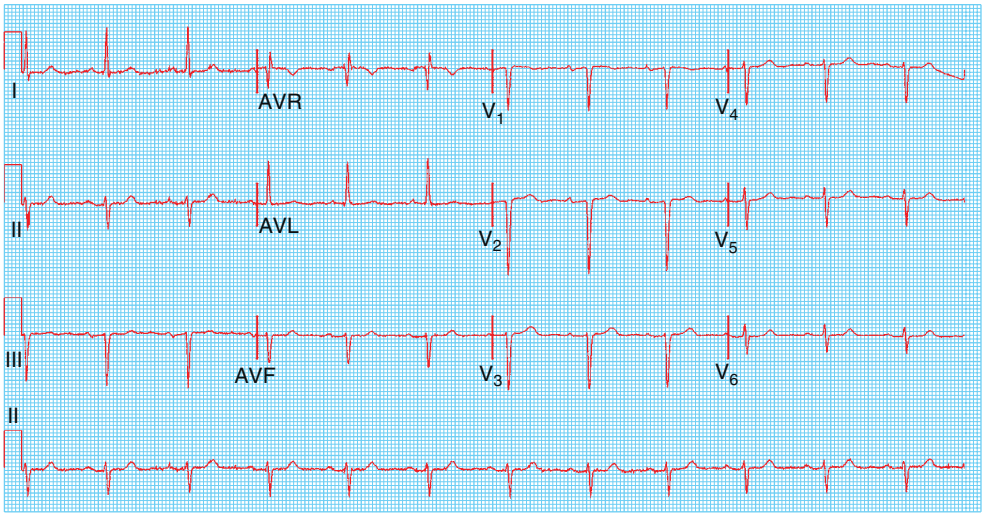


Fig. 17.36 **An ECG showing first-degree atrioventricular block with a prolonged PR interval.**

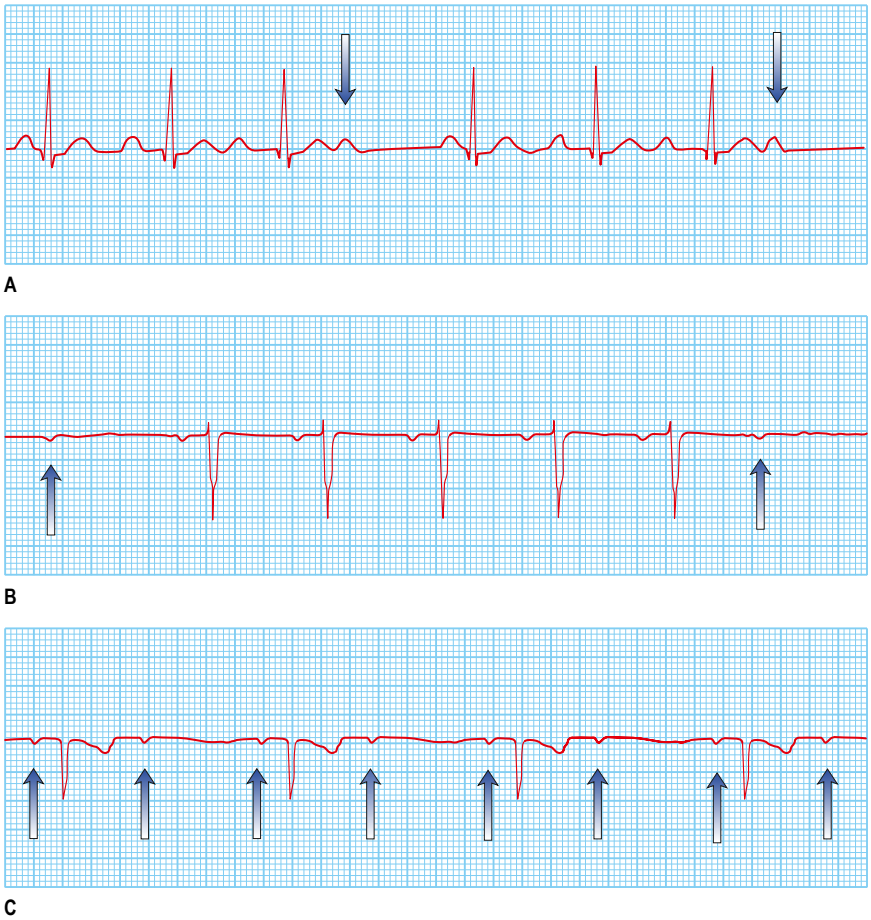


Fig. 17.37 Three varieties of second-degree atrioventricular (AV) block. **(A)** Wenckebach (Mobitz type I) AV block. The PR interval gradually prolongs until the P wave does not conduct to the ventricles *(arrowed)*. **(B)** Mobitz type II AV block. The P waves that do not conduct to the ventricles *(arrowed)* are not preceded by gradual PR interval prolongation. **(C)** Two P waves to each QRS complex. The PR interval prior to the dropped P wave is always the same. It is not possible to define this type of AV block as type I or type II Mobitz block, and it is therefore a third variety of second-degree AV block (*arrows* show P waves).

narrow-complex AV block that has transient causes may respond to intravenous atropine, but temporary pacing facilities should be available for the management of these patients. Chronic narrow-complex AV block requires permanent pacing (dual-chamber; see 'Permanent pacing' section) if it is symptomatic or associated with heart disease. Pacing can be considered in isolated, congenital AV block, even if asymptomatic.

Broad-complex escape rhythm (>0.12 sec) implies that the escape rhythm originates below the His bundle and therefore that the region of block lies more distally in the His–Purkinje system. The resulting rhythm is slow (15–40 b.p.m.) and relatively unreliable. Dizziness and blackouts ***(Stokes–Adams attacks)*** often occur. It is usually caused by degenerative fibrosis and calcification of the distal conduction system ***(Lenegre–Lev's disease)***. Broad-complex AV block may also be caused by ischaemic heart disease, myocarditis, cardiomyopathy, or genetic mutations. Permanent pacemaker implantation (see 'Permanent pacing' section) is indicated, as pacing considerably reduces the mortality. An ICD (see 'Implantable cardioverter–defibrillator' section) may be indicated in those at higher risk of ventricular arrhythmia, such as with severe left ventricular dysfunction or pathogenic gene mutations.

Bundle branch block

The His bundle gives rise to the right and left bundle branches. The left bundle subdivides into the anterior and posterior divisions of the left bundle (though it is debated whether anatomically there are two discrete divisions or just arborisation of the system). Various conduction disturbances can occur.

Bundle branch conduction delay

This produces slight widening of the QRS complex (up to 0.12 sec). It is known as incomplete bundle branch block.

Complete block of a bundle branch

This is associated with a wider QRS complex (≥0.12 sec). The shape of the QRS depends on whether the right or the left bundle is blocked.

Right bundle branch block (Fig. 17.39A) produces late activation of the right ventricle. This is seen as deep S waves in leads I and V_6 and as a tall late R wave in lead V_1 (late activation moving towards right-sided leads and away from left-sided leads).

LBBB (Fig. 17.40) produces the opposite: a deep S wave in lead V_1 and a tall late R wave in leads I and V_6. Because left bundle branch conduction is normally responsible for the initial ventricular activation, LBBB also produces abnormal Q waves.

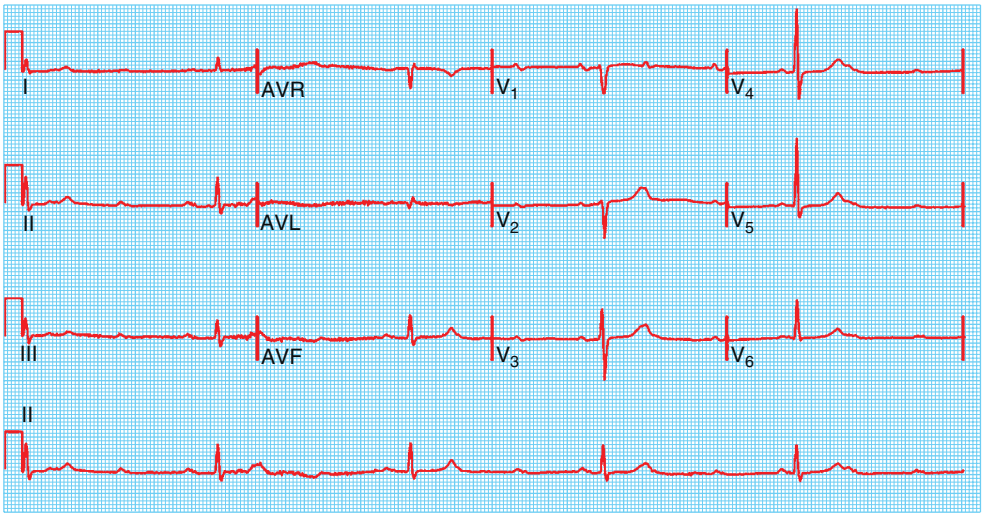


Fig. 17.38 Complete heart block. In this example, the QRS complexes are narrow (0.08 sec) and there is complete dissociation of atrial and ventricular complexes.

? Box 17.10 Causes of complete heart block

Congenital
- Autoimmune (e.g., maternal SLE)
- Structural heart disease (e.g., transposition of great vessels)

Idiopathic fibrosis
- Lev's disease (progressive fibrosis of distal His–Purkinje system in elderly patients)
- Lenegre's disease (proximal His–Purkinje fibrosis in younger patients)

Ischaemic heart disease
- Acute myocardial infarct
- Ischaemic cardiomyopathy

Non-ischaemic heart disease
- Calcific aortic stenosis
- Idiopathic dilated cardiomyopathy
- Infiltrations (e.g., amyloidosis, sarcoidosis, neoplasia)

Cardiac surgery
- For example, following aortic valve replacement, CABG, VSD repair

Iatrogenic
- Radiofrequency AV node ablation and pacemaker implantation

Drug-induced
- e.g., Digoxin, beta-blockers, non-dihydropyridine calcium-channel blockers, amiodarone

Infections
- Endocarditis
- Lyme disease
- Chagas' disease

Autoimmune rheumatic disease
- For example, SLE, rheumatoid arthritis

Neuromuscular diseases
- For example, Duchenne muscular dystrophy

Hemiblock

Delay or block in the divisions of the left bundle branch produces a swing in the direction of depolarization (electrical axis) of the heart. When the anterior division is blocked (left anterior hemiblock), the left ventricle is activated from inferior to superior. This produces a superior and leftward movement of the axis (left axis deviation). Delay or block in the postero-inferior division swings the QRS axis inferiorly to the right (right axis deviation).

Bifascicular block

Bifascicular block (see Fig. 17.39B) is a combination of a block of any two of the following: the right bundle branch, the left antero-superior division, and the left postero-inferior division. Block of the remaining fascicle will result in complete AV block.

Trifascicular block

This term is commonly and incorrectly used when there is a combination of bifasicular block and first-degree block (prolongation of the PR interval). The correct definition of this term is when all three fascicles are blocked, resulting in complete heart block with right bundle branch block and either left anterior fascicular block or left posterior fascicular block. Trifasicular block is also present where there is alternating left and right bundle branch block. The rate of progression to complete heart block in a patient with bifasicular and bifasciular block with prolonged PR interval are similar at 1%/year.

Clinical features of heart blocks

Bundle branch blocks are usually asymptomatic. Right bundle branch block causes wide but physiological splitting of the second heart sound. LBBB may cause reverse splitting of the second sound. Patients with intraventricular conduction disturbances may complain of syncope. This is due to intermittent complete heart block or ventricular tachyarrhythmias. ECG monitoring (which can include implantable loop monitors) and electrophysiological studies can determine the cause of syncope in these patients.

Aetiology

Right bundle branch block occurs as an isolated congenital anomaly, present in 5% of healthy individuals, or is associated with cardiac or pulmonary conditions. Conditions commonly associated with right bundle branch block include congenital cardiac disorders, such as atrial and ventricular septal defects, pulmonary stenosis and Fallot's tetralogy, pulmonary embolism, pulmonary hypertension, MI, fibrosis of conduction tissue, and Chagas' disease. Block in the right bundle alone does not tend to alter the electrical axis of the heart unless accompanied by right ventricular hypertrophy (RV overload) or coexistent fascicular block. The combination of right bundle branch block with left axis deviation is seen in patients with ostium primum atrial septal defects but more often signifies diffuse

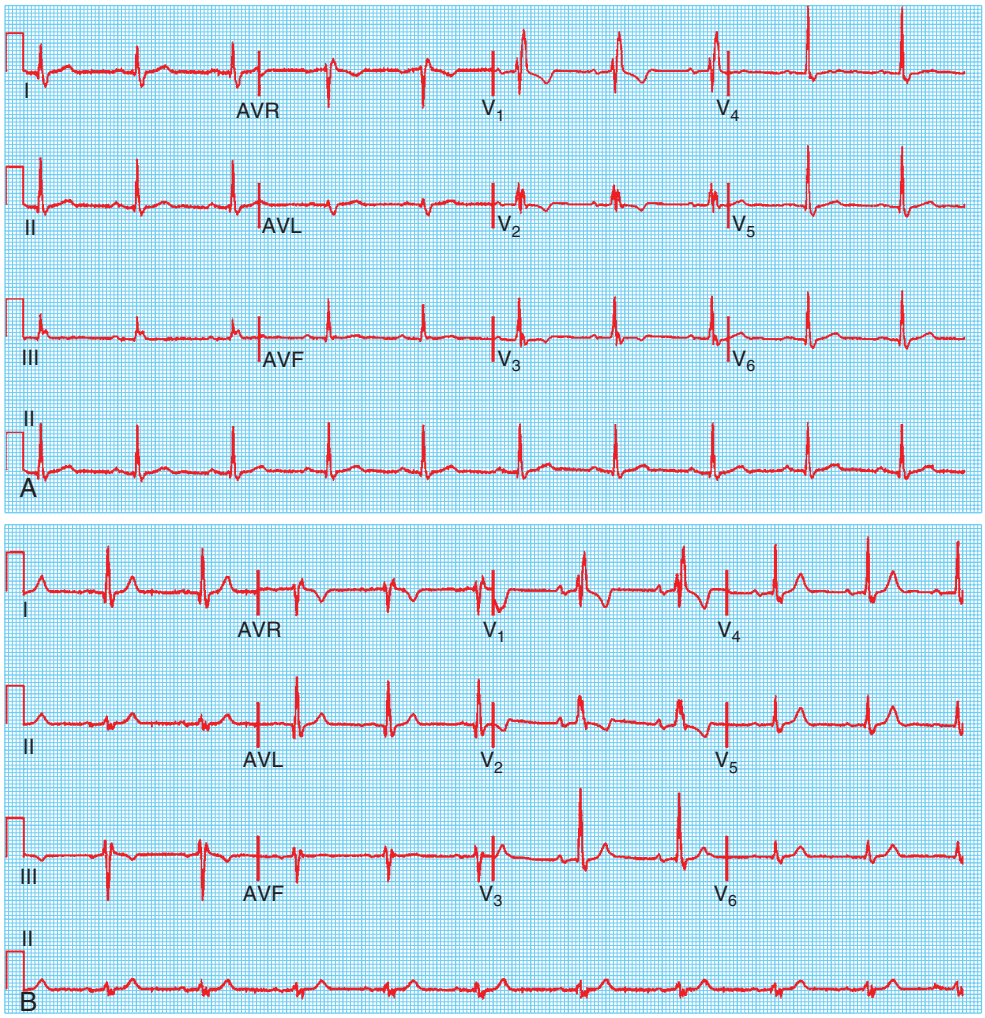


Fig. 17.39 Right bundle branch block versus bifascicular block. (A) A 12-lead ECG showing right bundle branch block. Note an rsR′ pattern with the tall R′ in lead V_1–V_2 and the broad S waves in leads I and V_5–V_6. **(B)** Compare with this ECG showing bifascicular block. In addition to right bundle branch block, note the left axis deviation and deep S waves in leads III and AVF, which are typical of left anterior hemiblock.

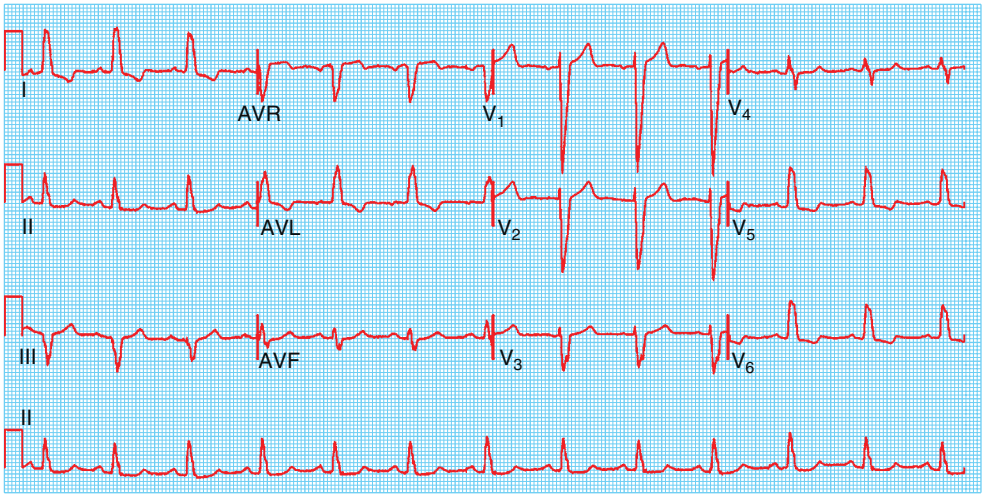


Fig. 17.40 A 12-lead ECG showing left bundle branch block. The QRS duration is more than 0.12 seconds. Note the broad, notched R waves with ST depression in leads I, AVL, and V_6, and the broad QS waves in V_1–V_3.

conduction tissue disease affecting the right bundle and the left anterior fascicle.

Complete LBBB is often associated with extensive left ventricular disease. The most common causes include aortic stenosis, hypertension, MI, and severe coronary disease, similar to the causes of complete heart block.

Management

Any patient with evidence of conduction delay should be assessed for the likely aetiology. Drugs contributing to bradycardia should be stopped if it is safe to do so. Echocardiography is helpful to assess the structure and function of the heart.

The management of different ECG findings is set out in **Box 17.11** and varies considerably between simple reassurance and urgent admission to a monitored area for cardiac pacing.

Box 17.11 The management of various ECG abnormalities frequently associated with bradycardia

Rhythm	Management
First-degree AV block	Reassurance, generally no further monitoring unless symptomatic
Second-degree AV block, Mobitz I	Address reversible causes, outpatient monitoring, permanent pacemaker if syncope
Second-degree AV block, Mobitz II, or higher-degree block	Close monitoring in a high dependency area; temporary pacing if likely short-term (reversible) cause; otherwise permanent pacemaker
Complete (third-degree) block	Close monitoring in a high dependency area; permanent pacemaker
Right bundle branch block	Exclude pathological causes; reassurance
Left bundle branch block	Identify cause; if associated with pre-syncope or syncope consider an extended period of rhythm monitoring (e.g., 24 h Holter or implantable loop device)
Higher-degree block with LV dysfunction	Cardiac resynchronization therapy with a biventricular pacemaker/defibrillator

TACHYCARDIAS

Supraventricular tachycardias

Supraventricular tachycardias (SVTs) arise from the atrium or the AV junction. Conduction is via the His–Purkinje system; therefore the QRS shape during tachycardia is usually similar to that seen in the same patient during baseline rhythm. A classification of SVT is given in **Box 17.12**. Some of these tachycardias are discussed in more detail in the sections that follow.

Inappropriate sinus tachycardia

Inappropriate sinus tachycardia is a persistent increase in resting heart rate unrelated to, or out of proportion with, the level of physical or emotional stress. It is found predominantly in young women. Sinus tachycardia due to intrinsic sinus node abnormalities, such as enhanced automaticity, or abnormal autonomic regulation of the heart with excess sympathetic and reduced parasympathetic input, is extremely rare.

In general, sinus tachycardia is a secondary phenomenon and the underlying causes need to be actively investigated. Depending on the clinical setting, acute causes include exercise, emotion, pain, fever, infection, acute heart failure (AHF), acute pulmonary embolism, and hypovolaemia. Chronic causes include pregnancy, anaemia, hyperthyroidism, and catecholamine excess. The underlying cause should be found and treated rather than treating the compensatory physiological response. If necessary, beta-blockers may be used to slow the sinus rate—in hyperthyroidism, for example (see **Box 34.30**); ivabradine, an I_F (pacemaker current) blocker, may be useful when beta-blockade cannot be tolerated.

Postural orthostatic tachycardia syndrome

This is a sudden and significant increase in heart rate (usually >130 beats per minute) associated with normal or mildly reduced blood pressure and produced by standing. There is sinus tachycardia with a structurally normal heart. The underlying mechanism is a failure of the peripheral vasculature to constrict appropriately in response to orthostatic stress, which is compensated by an excessive increase in heart rate. Management begins with supportive measures aimed

Box 17.12 Causes of supraventricular tachycardia (SVT)

Tachycardia	ECG features	Comment
Sinus tachycardia	P wave morphology similar to sinus rhythm; P waves always precede QRS	Need to determine underlying cause
Atrioventricular nodal re-entrant tachycardia (AVNRT)	No visible P wave, or inverted P wave immediately before or after QRS complex	Most common cause of palpitations in patients with normal hearts
Atrioventricular re-entrant tachycardia (AVRT) complexes	P wave visible between QRS and T wave	Due to an accessory pathway; if pathway conducts in both directions, ECG during sinus rhythm may be pre-excited
Atrial fibrillation	'Irregularly irregular' RR intervals and absence of organized atrial activity	Most common tachycardia in patients >65 years
Atrial flutter	Visible flutter waves at 300 b.p.m. (sawtooth appearance), usually with 2:1 AV conduction	Suspect in any patient with regular SVT at 150 b.p.m.
Atrial tachycardia	Organized atrial activity with P wave morphology different from sinus rhythm preceding QRS	Usually occurs in patients with structural heart disease or following extensive ablation within atria
Multifocal atrial tachycardia	Multiple P wave morphologies (≥3) and irregular RR intervals	Rare arrhythmia; most commonly associated with significant chronic lung disease
Accelerated junctional tachycardia	ECG similar to that in AVNRT	Rare in adults

at identifying triggers and avoiding situations (such as suddenly rising from sitting) that are associated with pre-syncope. Medications including beta-blockers and serotonin-selective reuptake inhibitors are sometimes used.

Atrioventricular junctional tachycardias

AV nodal re-entrant and AV re-entrant tachycardias are usually referred to as paroxysmal SVTs and are often seen in young patients with no or little structural heart disease, although congenital heart abnormalities can coexist in a small proportion of patients with these arrhythmias. The first presentation is commonly between ages 12 and 30, and the prevalence is approximately 2.5/1000.

In these tachycardias the AV node is an essential component of the re-entry circuit.

Atrioventricular nodal re-entrant tachycardia

Atrioventricular nodal re-entrant tachycardia (AVNRT) is twice as common in women. Clinically, the tachycardia often strikes suddenly without obvious provocation, but exertion, emotional stress, coffee, tea, and alcohol may aggravate or induce it. An attack may stop spontaneously or may continue indefinitely until medical intervention.

In AVNRT there are two functionally and anatomically different pathways predominantly within the AV node: one is characterized by a short effective refractory period and slow conduction, and the other has a longer effective refractory period and conducts faster. In sinus rhythm the atrial impulse that depolarizes the ventricles usually conducts through the fast pathway. If the atrial impulse (e.g., an atrial premature beat) occurs early when the fast pathway is still refractory, the slow pathway takes over in propagating the atrial impulse to the ventricles. It then travels back through the fast pathway, which has already recovered its excitability, thus initiating the most common 'slow–fast', or typical, AVNRT.

The rhythm is recognized on ECG from normal regular QRS complexes, usually at a rate of 140 to 240/min (Fig. 17.41A). Sometimes, the QRS complexes will show typical bundle branch block. P waves either are not visible or are seen immediately before or after the QRS complex because of simultaneous atrial and ventricular activation. It is less common (5%–10%) to observe a tachycardia when the atrial impulse conducts anterogradely through the fast pathway and returns through the slow pathway, producing a long RP′ interval ('fast–slow').

Atrioventricular re-entrant tachycardia

This large circuit comprises the AV node, the His bundle, the ventricle, and an abnormal connection of myocardial fibres from the ventricle back to the atrium. It is called an accessory pathway or bypass tract and results from an incomplete separation of the atria and the ventricles during fetal development.

In contrast to AVNRT, atrioventricular re-entrant tachycardia (AVRT) is due to a macro re-entry circuit and each part of the circuit is activated sequentially. As a result, atrial activation occurs after ventricular activation and the P wave is usually seen clearly between the QRS and T waves (Fig. 17.41B).

Accessory pathways are most commonly situated on the left but may occur anywhere around the AV groove. The most common accessory pathways, known as ***Kent bundles***, are in the free wall or septum. In about 10% of cases multiple pathways occur. ***Mahaim fibres*** are atriofascicular or nodofascicular fibres that enter the ventricular myocardium in the region of the right bundle branch. Accessory pathways that conduct from the ventricles to the atria only are not visible on the surface ECG during sinus rhythm and are therefore 'concealed'. Accessory pathways that conduct bidirectionally usually are 'manifest' on the surface ECG. If the accessory pathway conducts from the atrium to the ventricle during sinus rhythm, the electrical impulse can conduct quickly over this abnormal connection to depolarize part of the ventricles abnormally (pre-excitation). A pre-excited ECG is characterized by a short PR interval and a wide QRS complex that begins as a slurred part known as the ***δ wave*** (Fig. 17.41C). Patients with a history of palpitations and a pre-excited ECG have a condition known as ***Wolff–Parkinson–White (WPW) syndrome***.

During AVRT, the AV node and ventricles are activated normally ***(orthodromic AVRT)***, usually resulting in a narrow QRS complex. Less commonly, the tachycardia circuit can be reversed, with activation of the ventricles via the accessory pathway, and atrial activation via retrograde conduction through the AV node ***(antidromic AVRT)***. This results in a broad-complex tachycardia.

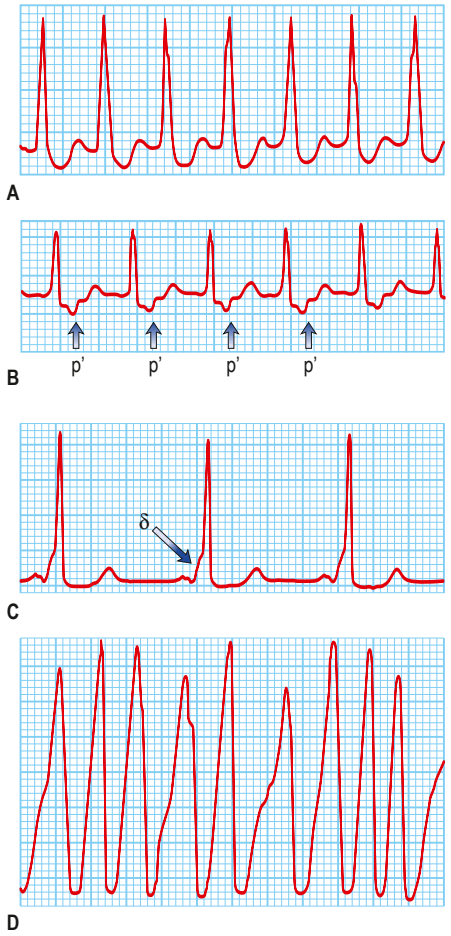


Fig. 17.41 Atrioventricular junctional tachycardia. (A) Atrioventricular nodal re-entrant tachycardia (AVNRT). The QRS complexes are narrow and the P waves cannot be seen. **(B)** Atrioventricular re-entrant tachycardia (AVRT) (Wolff–Parkinson–White syndrome, WPW). The tachycardia P waves *(arrowed)* are clearly seen after narrow QRS complexes. **(C)** An ECG taken in a patient with WPW syndrome during sinus rhythm. Note the short PR interval and the δ wave *(arrowed)*. **(D)** Atrial fibrillation in the WPW syndrome. Note the tachycardia with broad QRS complexes and the fast and irregular ventricular rate.

These patients are also prone to atrial fibrillation. During atrial fibrillation, the ventricles may be depolarized by impulses travelling over both the abnormal and the normal pathways. This results in pre-excited atrial fibrillation, a characteristic tachycardia that is typified by irregularly irregular broad QRS complexes (Fig. 17.41D). If an accessory pathway has a short antegrade effective refractory period (<250 ms), it may conduct to the ventricles at an extremely high rate and may cause ventricular fibrillation. The incidence of sudden death is 0.15% to 0.39% per patient-year and it may be a first manifestation of the disease in younger individuals. Verapamil and digoxin may allow a higher rate of conduction over the abnormal pathway and precipitate ventricular fibrillation. Therefore neither verapamil nor digoxin should be used to treat atrial fibrillation associated with WPW syndrome.

Acute management

This is reviewed in Chapter 11 ('Atrial fibrillation/Acute management' section).

Long-term management

Patients with suspected cardiac arrhythmias should be referred to a cardiologist for electrophysiological evaluation and long-term management. Both pharmacological and non-pharmacological treatments, including ablation of an accessory pathway, are readily available. Verapamil, diltiazem, and beta-blockers have proven efficacy in 60% to 80% of patients. Sodium-channel blockers (flecainide and propafenone), potassium repolarization current blockers (sotalol, dofetilide, azimilide) and the multichannel blocker amiodarone may also prevent the occurrence of tachycardia.

Refinement of catheter ablation techniques has rendered many AV junctional tachycardias entirely curable. Modification of the slow pathway is successful in 96% of patients with AVNRT, although a 1% risk of AV block is present. In AVRT the target for catheter ablation is the accessory pathway(s). The success rate of ablation of a single accessory pathway is approximately 95%, with a recurrence rate of 5%, requiring a repeat procedure.

Atrial tachyarrhythmias

Atrial tachyarrhythmias, including atrial fibrillation, atrial flutter, atrial tachycardia, and atrial ectopic beats, all arise from the atrial myocardium. They share similar aetiologies, of which the most commonly encountered in clinical practice are increasing age, MI, hypertension, obesity, diabetes mellitus, HCM, heart failure, valvular heart disease, myocarditis, pericarditis, cardiothoracic surgery, electrolyte imbalance, alcohol use, obstructive airway disease, chest infections, and hyperthyroidism.

Atrial fibrillation

This is the commonest sustained arrhythmia, occurring in 1% to 2% of the general population and 5% to 15% of patients over 75 years of age. It also occurs, particularly in a paroxysmal form, in younger patients.

Aetiology

Any condition resulting in raised atrial pressure, increased atrial muscle mass, atrial fibrosis, or inflammation and infiltration of the atrium may cause atrial fibrillation.

Hypertension and heart failure are the most common causes in the developed world. Hyperthyroidism may provoke atrial fibrillation, sometimes as virtually the only feature of the disease, and thyroid function tests are mandatory in any patient with atrial fibrillation that is unaccounted for. Atrial fibrillation occurs in one-third of patients after cardiac surgery. Alcohol, but not caffeine, is also a trigger. Mitral stenosis caused by rheumatic heart disease is a major cause in the developing world.

A genetic predisposition to AF has been proposed. About 30% to 40% of patients with AF, especially those who present at a young age, have at least one parent with the arrhythmia, and genes associated with the sodium channel, the potassium channel, gap junction proteins, and right–left isomerism have been implicated. Gene defects linked to chromosomes 10, 6, 5, and 4 have been associated with familial atrial fibrillation.

Pathophysiology

Atrial fibrillation is maintained by continuous, rapid (300–600/min) activation of the atria by multiple meandering re-entry wavelets, often driven by rapidly depolarizing automatic foci, located predominantly within the pulmonary veins. The atria respond electrically at this rate but there is no coordinated mechanical action, and only a proportion of the impulses are conducted to the ventricles. The ventricular response depends on the rate and regularity of atrial activity, particularly at the entry to the AV node, the refractory properties of the AV node itself, and the balance between sympathetic and parasympathetic tone.

Clinical features

Symptoms attributable to atrial fibrillation are highly variable. In some patients (about 30%) it is an incidental finding, while others attend hospital as an emergency with rapid palpitations, dyspnoea, and/or chest pain following the onset of atrial fibrillation. Most patients with on-going atrial fibrillation experience some deterioration of exercise capacity or wellbeing, but this may be appreciated only once sinus rhythm is restored. When caused by rheumatic mitral stenosis, the onset of atrial fibrillation results in considerable worsening of cardiac failure.

The patient has an 'irregularly irregular' pulse, as opposed to a basically regular pulse with an occasional irregularity (e.g., extrasystoles) or recurring irregular patterns (e.g., Wenckebach block). The irregular nature of the pulse in atrial fibrillation is maintained during exercise.

The ECG shows fine oscillations of the baseline (so-called fibrillation or *f* waves) and no clear P waves (Fig. 17.42A). The QRS rhythm is rapid and irregular. Untreated, the ventricular rate is usually 120 to 180/min.

The clinical classification of atrial fibrillation includes:

- ***first diagnosed episode***—irrespective of duration or severity of symptoms
- ***paroxysmal***—stops spontaneously or with intervention within 7 days
- ***persistent***—continuous for more than 7 days, including episodes terminated by cardioversion
- ***longstanding persistent***—continuous for more than 1 year, where the long-term treatment goal remains to achieve sinus rhythm
- ***permanent***—continuous, with a joint decision between the patient and the physician to cease further attempts to regain sinus rhythm.

The classification is helpful in choosing between rhythm restoration and rate control. Atrial fibrillation may be asymptomatic and the 'first diagnosed episode' should not necessarily be regarded as the true onset.

Management

Acute management

This is reviewed in Chapter 11 ('Atrial fibrillation' section).

Fig. 17.42 Twelve-lead ECGs showing atrial fibrillation and atrial flutter. **(A)** Atrial fibrillation. Note the absolute rhythm irregularity. The ECG shows fine oscillations of the baseline (so-called fibrillation or *f* waves) and no clear P waves. **(B)** Atrial flutter, showing regular sawtooth-like atrial flutter waves (F waves) between QRS complexes. The flutter frequency is 300/min. Every fourth flutter wave is transmitted to the ventricles and the ventricular rate is therefore 75/min.

Long-term management

An integrated approach to the management of AF is recommended, which prioritizes stroke prevention, symptom control with rate, and/or rhythm control and optimization of co-morbidities.

Stroke prevention

Atrial fibrillation increases the risk of stroke fivefold and anticoagulation reduces this risk by two-thirds. However, the risk of stroke is not uniform, and it is only patients who are at higher risk of stroke who are likely to benefit overall from anticoagulation, since treatment increases the risk of haemorrhage. A scoring system known as CHA_2DS_2VASc is used (**Box 17.13**) as the first step in determining the need for anticoagulation. The Australian guidelines have removed the sex (male or female) component of this score for simplicity.

The HAS-BLED score as recommended by European, Canadian, and UK (NICE) guidelines is used to identify possible modifiable risk factors for bleeding (e.g., hypertension control) rather than a tool to deny anticoagulation to patients at high risk of stroke (**Box 17.14** and **Fig. 17.43**)

When oral anticoagulation is required, a direct oral anticoagulant (DOAC) is first-line treatment. Warfarin may still be preferred where AF is secondary to valvular abnormalities (e.g., in rheumatic heart disease or following mitral valve replacement).

DOACs fall into two classes: direct thrombin inhibitors (e.g., dabigatran) and oral direct factor Xa inhibitors (e.g., rivaroxaban, apixaban, edoxaban). Their mechanism of action is described in Chapter 40 ('Anticoagulant agents' section). In comparison with warfarin, the DOACs have a rapid onset of action, shorter half-life, and fewer food and drug interactions and do not require INR blood test monitoring. Trial data have shown them to be equally effective as warfarin but safer, with a 52% reduction in intracerebral haemorrhage.

Antiplatelet agents should not be used to reduce stroke risk. Percutaneous left atrial appendage occlusion (LAAO) may be offered where anticoagulation is contraindicated or not tolerated.

Rate control

Ventricular rate control can improve symptoms of AF in the majority of patients. As a primary strategy, this is appropriate in patients who:

Box 17.13 CHA_2DS_2-VASc scoring system for non-valvular atrial fibrillation

	Risk factors	Score/points
C	Congestive heart failure Includes patients with heart failure with reduced ejection fraction; recently decompensated heart failure with preserved ejection fraction; hypertrophic cardiomyopathy	1
H	Hypertension or on treatment	1
A_2	Age ≥75	2
D	Diabetes mellitus or fasting glucose >7 mmol/L	1
S_2	Stroke/TIA/thromboembolism	2
V	Vascular disease (aorta, coronary or peripheral arteries)	1
A	Age 65–74	1
Sc	Sex category: female	1
Annual risk of stroke		
0 points = 0% risk: No prophylaxis 1 point = 1.3% risk: Consider anticoagulation in men 2+ points = 2.2% risk: Oral anticoagulant A score of 1 (for being female alone) is not enough to warrant anticoagulation.		

TIA, Transient ischaemic attack.

Box 17.14 HAS-BLED score for bleeding risk on oral anticoagulation in atrial fibrillation

Clinical characteristic	Score/points
Hypertension (systolic ≥ 160 mmHg)	1
Abnormal renal function (creatinine >200 μmol/L)	1
Abnormal liver function (liver enzymes >×3 upper limit of normal)	1
Stroke in past	1
Previous major bleeding	1
Labile INRs (in range <60% of time)	1
Elderly: age ≥65years	1
Drugs as well (antiplatelets/non-steroidal anti-inflammatory drugs)	1
Alcohol intake at same time (>14 units/week)	1

Yearly risk of developing major bleeding with anticoagulation rises with increasing score, from 1% with a score of 0 (low risk) to around >5% with a score of 3 or more (high risk).
INR, International Normalized Ratio.
(Adapted From European Society of Cardiology Clinical Practice Guidelines. Eur Heart J *2020; 42:373–498.)*

- have the permanent form of AF associated with no symptoms, or symptoms that have improved by slowing heart rate;
- are older than 65 years with recurrent atrial tachyarrhythmias ('accepted' atrial fibrillation); and
- have persistent tachyarrhythmias and have failed cardioversion(s) and serial prophylactic antiarrhythmic drug therapy, and in whom the risk/benefit ratio from using specific antiarrhythmic agents is shifted towards increased risk.

Rate control is usually achieved with a combination of ***digoxin***, ***beta-blockers***, or ***non-dihydropyridine calcium-channel blockers*** (***verapamil*** or ***diltiazem***). Beta-blockers are the preferred therapy. Digoxin monotherapy is sometimes sufficient for elderly, non-ambulant patients (the catecholamine effect of ambulation easily overwhelms the vagotonic effect of digoxin).

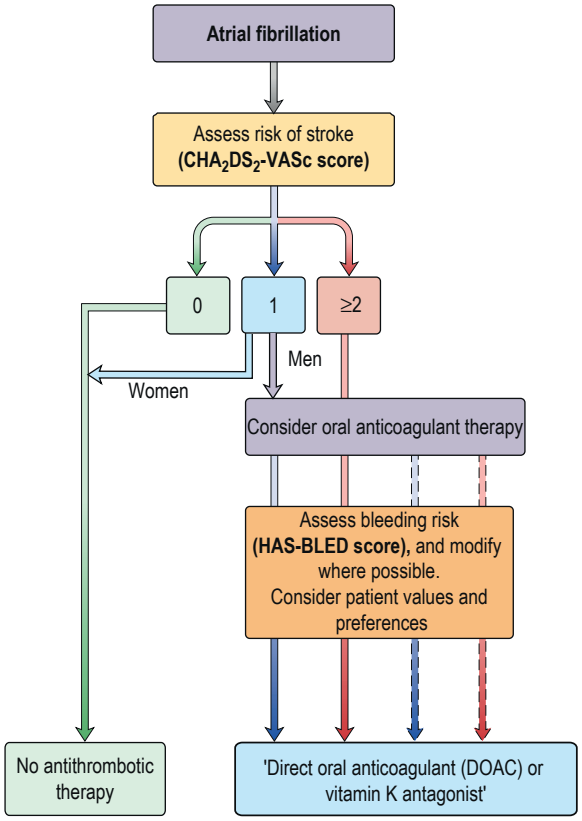


Fig. 17.43 Anticoagulation in atrial fibrillation.

The ventricular rate response is generally considered to be controlled if the resting heart rate is below 110 b.p.m., but stricter control (between 60 and 80 b.p.m. at rest and below 110 b.p.m. during moderate exercise) may be needed if symptoms persist. To assess the adequacy of rate control, an ECG rhythm strip may be sufficient but ambulatory 24-hour Holter monitoring or an exercise stress test are more informative. Older patients with poor rate control despite optimal medical therapy should be considered for pacemaker implantation followed by AV node ablation ***('ablate and pace' strategy)***.

Rhythm control

This refers to a strategy that seeks to maintain sinus rhythm. Historic randomized studies in patients predominantly over the age of 65 years (AFFIRM) or in patients with heart failure (AF-CHF) have shown that there is no net mortality benefit to be gained from rhythm control strategy compared with rate control. Newer data support the early use of rhythm control in all patients with a new diagnosis of AF, particularly when there is a concurrent diagnosis of heart failure. Which strategy to adopt needs to be assessed for each individual patient. Factors to consider include the likelihood of maintaining sinus rhythm, symptoms of AF, and the safety/tolerability of antiarrhythmic drugs or cardiac ablation.

Antiarrhythmic medications

Recurrent paroxysms may be prevented with oral medication. In general, patients with no significant heart disease can be treated with any class Ia, Ic, or III antiarrhythmic drug, although it is recommended that amiodarone (because of its substantial extracardiac adverse effect profile) should be reserved until other drugs have

failed. For patients with ***heart failure*** or ***left ventricular hypertrophy*** only amiodarone is recommended. Patients with ***coronary artery disease*** may be treated with sotalol or amiodarone.

Ablation

Patients with ***paroxysmal atrial fibrillation*** or with early ***persistent atrial fibrillation***, who have little left atrial dilation, may be treated with left atrial ablation. The ectopic triggers for atrial fibrillation are generally found in the pulmonary veins, which can be isolated from the atria using radiofrequency, cryothermal, or pulsed-field energy. Occasionally, more extensive ablation within the left atrium is needed. These techniques are more successful than antiarrhythmic drugs and may represent a 'cure' in some patients. However, the procedure is invasive and carries hazard of serious complications such as stroke and bleeding in about 2% of cases. In the long term, recurrence is not uncommon, and an apparently successful ablation does not remove the obligation for appropriate anticoagulation. Ablation has not been shown to improve long-term cardiovascular outcome but it does successfully treat symptoms due to atrial fibrillation.

Atrial flutter

Atrial flutter is often associated with atrial fibrillation and frequently requires a similar initial therapeutic approach. Atrial flutter is usually an organized atrial rhythm with an atrial rate typically between 250 and 350 b.p.m. Typical or 'isthmus-dependent' atrial flutter involves a macro re-entrant right atrial circuit around the tricuspid annulus. The wavefront circulates down the lateral wall of the right atrium, through the Eustachian ridge between the tricuspid annulus and the inferior vena cava, and up the interatrial septum, giving rise to the most frequent pattern, referred to as counterclockwise flutter. Re-entry can also occur in the opposite direction (clockwise or reverse flutter).

The ECG shows regular sawtooth-like atrial flutter waves (F waves) between QRS complexes (see **Fig. 17.42B**). In typical counterclockwise atrial flutter the F waves are negative in the inferior leads and positive in leads V_1 and V_2. In clockwise atrial flutter the deflection of the F waves is the opposite. If F waves are not clearly visible, for diagnostic clarification they can be revealed by slowing AV conduction by carotid sinus massage or by administering AV nodal-blocking drugs such as adenosine or verapamil.

Symptoms are largely related to the degree of AV block. Most often, every second flutter beat conducts, giving a ventricular rate of 150 b.p.m. When patients are receiving rate-controlling treatment, AV conduction block reduces the heart rate to approximately 75 b.p.m.

Management

The acute management of atrial flutter is discussed in Chapter 11 ('Atrial flutter' section). Recurrent paroxysms may be prevented by ***class III*** antiarrhythmic agents (sotalol, amiodarone). ***AV nodal blocking agents*** may be used to control the ventricular rate if the arrhythmia persists. However, the treatment of choice for patients with recurrent atrial flutter is catheter ***ablation*** (see 'Catheter ablation' section), which permanently interrupts re-entry by creating a line of conduction block within the isthmus between the inferior vena cava and the tricuspid valve ring. This technique offers patients whose only arrhythmia is typical atrial flutter an almost certain chance of a cure, although the later occurrence of atrial fibrillation is not uncommon.

Atrial tachycardia

This is an uncommon arrhythmia. Its prevalence is believed to be less than 1% in patients with arrhythmias. It is usually associated with structural heart disease, but in many cases it is referred to as idiopathic. Atrial tachycardia with block can be seen with digitalis poisoning.

The mechanisms of atrial tachycardia are attributed to enhanced automaticity, triggered activity, or intra-atrial re-entry. Atrial re-entrant tachycardia is usually relatively slow (125–150 b.p.m.) and can be initiated and terminated by atrial premature beats. The P′P′ intervals are regular. The PR interval depends on the rate of tachycardia and is longer than in sinus rhythm at the same rate.

Automatic tachycardia usually presents with higher rates (125–250 b.p.m.) and is often characterized by a progressive increase in the atrial rate with onset of the tachycardia ('warm-up') and progressive decrease prior to termination ('cool-down'). Atrial tachycardia is typically caused by a focus that is frequently located along the crista terminalis in the right atrium, adjacent to a pulmonary vein in the left atrium, or around one of the atrial appendages. Automatic atrial tachycardia may also present as an incessant variety leading to tachycardia-induced cardiomyopathy. Short runs of atrial tachycardia may provoke more sustained episodes of atrial fibrillation.

Carotid sinus massage may increase AV block during tachycardia, thereby facilitating the diagnosis, but does not usually terminate the arrhythmia. Management options include ***cardioversion***, ***antiarrhythmic drug therapy*** to maintain sinus rhythm, ***AV nodal slowing agents*** to control rate, and, in selected cases, radiofrequency catheter ***ablation***. Anticoagulation for isolated atrial tachycardia is not required.

Atrial ectopic beats

These often cause no symptoms, although they may be sensed as an irregularity or heaviness of the heartbeat. On the ECG, they appear as early and abnormal P waves and are usually, but not always, followed by normal QRS complexes. Treatment is not normally required unless the ectopic beats provoke more significant arrhythmias, when ***beta-blockers*** may be effective.

Ventricular tachyarrhythmias

The acute management of sustained ventricular tachycardia is discussed in Chapter 11 ('Paroxysmal supraventricular tachycardia' section). Conditions that may predispose to or commonly present with ventricular tachycardia are discussed below.

Brugada's syndrome

This inheritable condition accounts for part of a group of patients with idiopathic ventricular fibrillation who have no evidence of causative structural cardiac disease. It is more common in young male adults and in South-east Asia. The diagnosis is made by identifying the classic ECG changes that may be present spontaneously or may be provoked by the administration of a class I antiarrhythmic (flecainide or ajmaline—principally used as a diagnostic agent in patients where the diagnosis is suspected): right bundle branch block with coved ST elevation in leads V_1–V_3 (**Fig. 17.44**). Atrial fibrillation may occur.

In 20% of cases this is a monogenic inheritable condition associated with loss of sodium-channel function due to a mutation in the *SCN5A* gene. It can present with sudden death during sleep, resuscitated cardiac arrest, and syncope, or the patient may be asymptomatic and diagnosed incidentally or during familial assessment.

Management

Patients should be instructed to treat fevers aggressively (non-pharmacological measures, as well as paracetamol ± NSAIDs); high

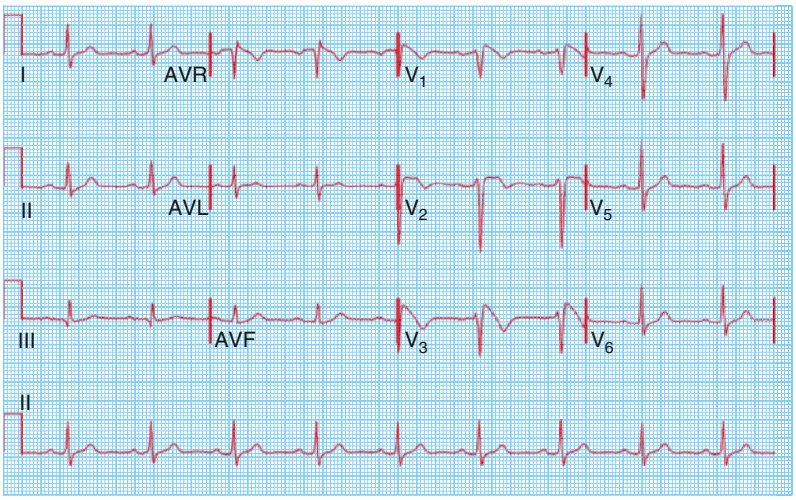


Fig. 17.44 The Brugada ECG. ST elevation (V_1, V_2, V_3) with right bundle branch block.

body temperatures are pro-arrhythmogenic. Certain prescribed medications are contraindicated (see https://www.brugadadrugs.org). An ICD should be considered in individuals with suspected arrhythmic syncope and a spontaneous type 1 Brugada ECG. There is an evolving role of quinidine to reduce the risk of ventricular fibrillation in this population. Beta-blockade is not helpful and may be harmful.

Long QT syndrome

This describes an ECG where the ventricular repolarization (QT interval) is greatly prolonged. The causes of long QT syndrome are listed in **Box 17.15**.

Congenital long QT syndrome

Two major syndromes have been described, one associated with congenital deafness (Jervell–Lange–Nielsen' syndrome) and one that is not (Romano–Ward' syndrome).

The molecular biology of the congenital long QT syndromes has been shown to be heterogeneous. It is usually a monogenic disorder and has been associated with mutations in cardiac potassium and sodium channel genes. The different genes involved appear to correlate with different phenotypes (**Fig. 17.45A**) that can exhibit such variable penetrance that carriers may have completely normal ECGs. Multiple long QT (LQT) subtypes have been identified but three major ones account for the majority of cases. These are:

- ***LQT1*** (*KCNQ1* gene mutation affecting the $I_{ks}\alpha$ subunit), in which the arrhythmia is usually provoked by exercise, particularly swimming
- ***LQT2*** (*KCNH2* gene mutation affecting the $I_{kr}\alpha$ subunit), in which arrhythmia provocation is associated with emotion and acoustic stimuli
- ***LQT3*** (*SCN5A* gene mutation affecting the $I_{Na}\alpha$ subunit), in which the arrhythmias occur during rest or when asleep.

It is likely that identification of the mutation involved will not only improve diagnostic accuracy, particularly with cascade screening in affected families, but also guide future therapy for the congenital long QT syndrome.

Acquired long QT syndrome

QT prolongation and torsades de pointes are usually provoked by bradycardia.

? Box 17.15 Causes of long QT syndrome

Congenital
- Jervell–Lange-Nielsen syndrome (autosomal recessive)
- Romano–Ward syndrome (autosomal dominant)

Acquired

Electrolyte abnormalities
- Hypokalaemia
- Hypomagnesaemia
- Hypocalcaemia

Drugs
- Quinidine, disopyramide
- Sotalol, amiodarone
- Tricyclic antidepressants, e.g., amitriptyline
- Phenothiazine drugs, e.g., chlorpromazine
- Antipsychotics, e.g., haloperidol, olanzapine
- Macrolides, e.g., erythromycin
- Quinolones, e.g., ciprofloxacin
- Methadone

Poisons
- Organophosphate insecticides

Miscellaneous
- Bradycardia
- Mitral valve prolapse
- Acute myocardial infarction
- Diabetes
- Prolonged fasting and liquid protein diets (long-term)
- Central nervous system diseases, e.g., dystrophia myotonica

Clinical features

Patients with a long QT are at risk of developing syncope and palpitations as a result of polymorphic ventricular tachycardia (torsades de pointes). They usually terminate spontaneously but may degenerate into ventricular fibrillation, resulting in sudden death.

Torsades de pointes is characterized on the ECG by rapid, irregular, sharp complexes that continuously change from an upright to an inverted position (**Fig. 17.45B**).

Between spells of tachycardia or immediately preceding the onset of tachycardia, the ECG shows a prolonged QT interval; the corrected QT (see **Box 17.9**) is usually over 0.50 seconds.

Management

Acute (acquired) long QT syndrome is treated as follows:
- Any electrolyte disturbance is corrected.
- Causative drugs are stopped.
- Magnesium sulphate 8 mmol (Mg^{2+}) is given over 10 to 15 min for acquired long QT.
- Intravenous isoprenaline may be effective when QT prolongation is acquired (isoprenaline is contraindicated for congenital long QT syndrome).

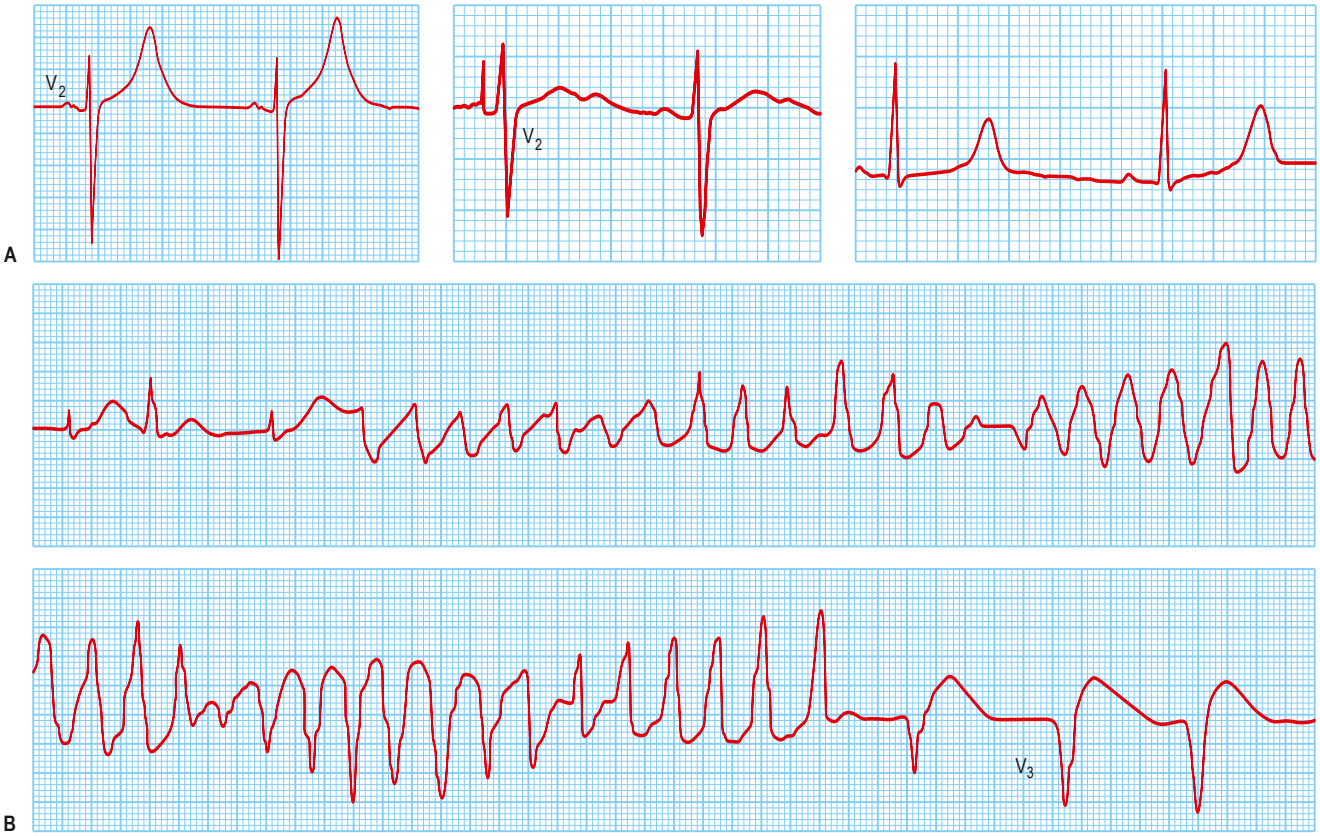


Fig. 17.45 Prolonged QT syndrome. **(A)** Three examples of a prolonged QT interval corresponding to three different types of long QT syndrome. **(B)** An ECG demonstrating a supraventricular rhythm with a long QT interval giving way to atypical ventricular tachycardia (torsades de pointes). The tachycardia is short lived and is followed by a brief period of idioventricular rhythm.

- A higher heart rate may need to be maintained with atrial or ventricular pacing.

Long-term, all patients with the syndrome should avoid drugs known to prolong the QT interval (see www.crediblemeds.org). Non-selective beta-blockade (nadolol or propranolol) reduces the risk of arrhythmic events. LQT3 patients respond additionally to sodium-channel blockers, such as mexiletine. Patients who have survived a cardiac arrest, syncope despite beta-blockers, or those with marked QT prolongation with a genetic mutation usually benefit from ICD therapy and are best evaluated in inherited cardiomyopathy clinics. Left cardiac sympathetic denervation can be considered in patients receiving ICD therapy despite optimal medical treatment or in those in whom ICD implant cannot be performed.

Short QT syndrome

This rare channelopathy is usually diagnosed in the presence of a QTC less than 320 ms or a QTC less than 360 ms with a family history of the syndrome or pathogenic mutation. Three definite genetic abnormalities predominately affecting potassium channels have been identified. Ventricular arrhythmias and sudden death may occur. ICD implantation is recommended for survivors of cardiac arrest and patients with either arrhythmic syncope or sustained ventricular tachycardia. Quinidine may be useful to reduce the burden of ventricular arrhythmia.

Normal heart ventricular tachycardia

Monomorphic ventricular tachycardia in patients with structurally normal hearts (idiopathic ventricular tachycardia) is usually a benign condition with an excellent long-term prognosis. Occasionally, it is incessant (so-called Gallavardin's tachycardia) and, if untreated, may lead to cardiomyopathy. An echocardiogram and CMR are frequently used to define 'structurally normal'.

Normal heart ventricular tachycardia arises from a focus in either the right ventricular outflow tract or the left ventricular septum. Treatment of symptoms is usually with beta-blockers. There is a special form of verapamil-sensitive tachycardia that responds well to non-dihydropyridine calcium antagonists. In symptomatic patients radiofrequency catheter ablation is highly effective, resulting in a cure in over 90% of cases. It is important to distinguish the more benign right ventricular outflow tract ventricular tachycardia from the more pathogenic ACM (see 'Hypertrophic cardiomyopathy' section) due to some similarities on ECG.

Non-sustained ventricular tachycardia

Non-sustained ventricular tachycardia (NSVT) is defined as ventricular tachycardia that is 5 consecutive beats or more but lasts for less than 30 seconds (Fig. 17.46A). NSVT can be found in 6% of patients with normal hearts and usually does not require treatment. It is documented in up to 60% to 80% of patients with heart disease. Antiarrhythmic suppression of NSVT is not usually advocated but beta-blockers may improve quality of life in symptomatic individuals.

Ventricular premature beats (ectopics)

These may be uncomfortable, especially when frequent. The patient complains of extra beats, missed beats, or heavy beats because it may be the premature beat, the post-ectopic pause, or the next

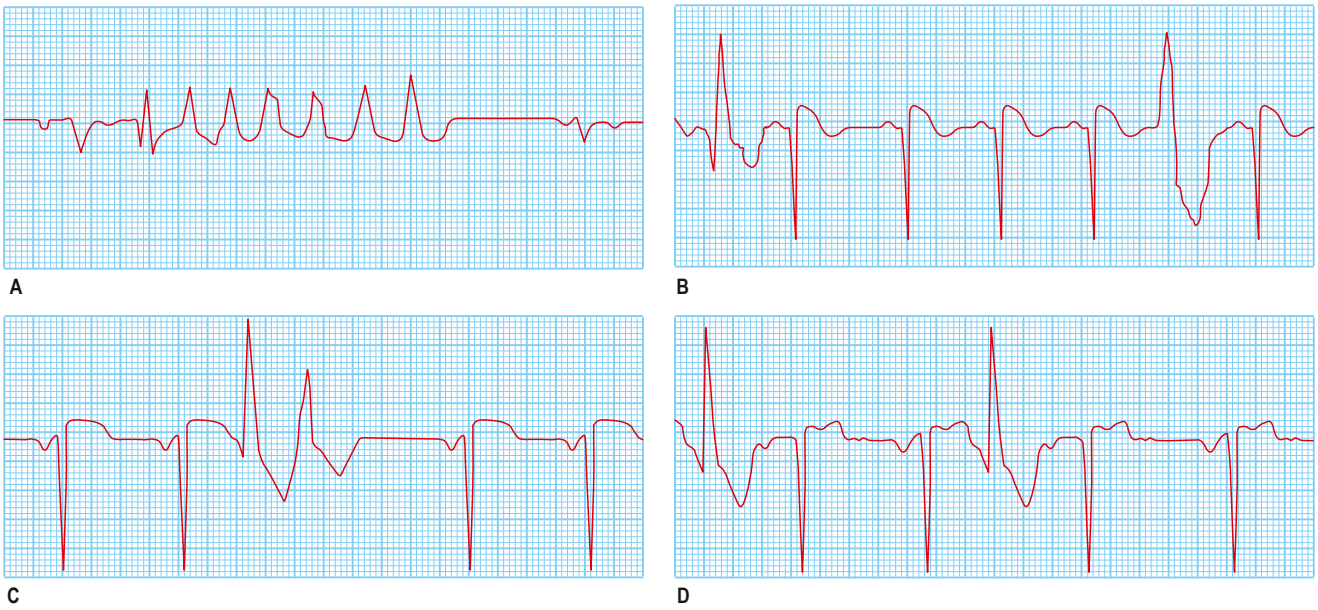


Fig. 17.46 Varieties of ventricular ectopic activity. **(A)** A brief run of ventricular tachycardia (*non*-sustained ventricular tachycardia) that follows previous ectopic activity. **(B)** Two ventricular ectopic beats of different morphology (multimorphological). **(C)** Two ventricular premature beats (VPBs) occurring one after the other (a pair or couplet of VPBs). **(D)** Frequently repetitive ventricular ectopic activity of a single morphology.

sinus beat that is noticed. The pulse is irregular owing to the premature beats. Some early beats may not be felt at the wrist. When a premature beat occurs regularly after every normal beat, 'pulsus bigeminus' may occur.

These premature beats (Fig. 17.46B–D) have a broad (>0.12 sec) and bizarre QRS complex because they arise from an abnormal (ectopic) site in the ventricular myocardium. Following a premature beat, there is usually a complete compensatory pause because the AV node or ventricle is refractory to the next sinus impulse. Early 'R-on-T' ventricular premature beats (occurring simultaneously with the upstroke or peak of the T wave of the previous beat) may induce ventricular fibrillation in patients with heart disease, particularly following MI.

Ventricular premature beats are usually treated only if symptomatic. Simple measures, such as reassurance and beta-blocker therapy, are normally all that is required. In patients with a 24-hour burden of ventricular ectopics of more than 10% and heart failure symptoms, an ablation could be considered to prevent further deterioration in symptoms and cardiac function.

Long-term management of cardiac tachyarrhythmias

Options for the long-term management of all cardiac tachyarrhythmias include:

- antiarrhythmic drug therapy
- ablation therapy
- device therapy.

Antiarrhythmic drugs

Drugs that modify the rhythm and conduction of the heart are used to treat cardiac arrhythmias. Antiarrhythmic drugs may aggravate or produce arrhythmias (proarrhythmia), and they may also depress ventricular contractility and must therefore be used with caution. They are classified according to their effect on the action potential (Vaughan Williams classification; Box 17.16 and Fig. 17.47).

Box 17.16 Vaughan Williams classification of antiarrhythmic drugs

Class/action	Drugs
Class I	Membrane-depressant drugs (sodium-channel blockers)
Ia Lengthen action potential	Disopyramide
Ib Shorten action potential	Lidocaine, mexiletine
Ic No effect on action potential	Flecainide, propafenone
Class II Block β-adrenoceptors	Atenolol, acebutolol, bisoprolol, propranolol, esmolol
Class III Lengthen action potential	Amiodarone, dronedarone, sotalol, dofetilide, ibutilide, vernakalant
Class IV Reduce plateau phase of action potential	Calcium-channel blockers, e.g., verapamil, diltiazem
Other	Adenosine, digoxin

Class I drugs

These are membrane-depressant drugs that reduce the rate of entry of sodium into the cell (sodium channel blockers). They may slow conduction, delay recovery, or reduce the spontaneous discharge rate of myocardial cells. Class I agents have been found to increase mortality compared to placebo in post-MI patients with ventricular ectopy (Cardiac Arrhythmia Suppression Trial [CAST] trials—class Ic agents) and in patients treated for atrial fibrillation (class Ia agent, quinidine). In view of this class Ic agents, such as flecainide, and all other class I drugs should be reserved for patients who do not have symptomatic CAD, left ventricular dysfunction, or other forms of significant structural heart disease.

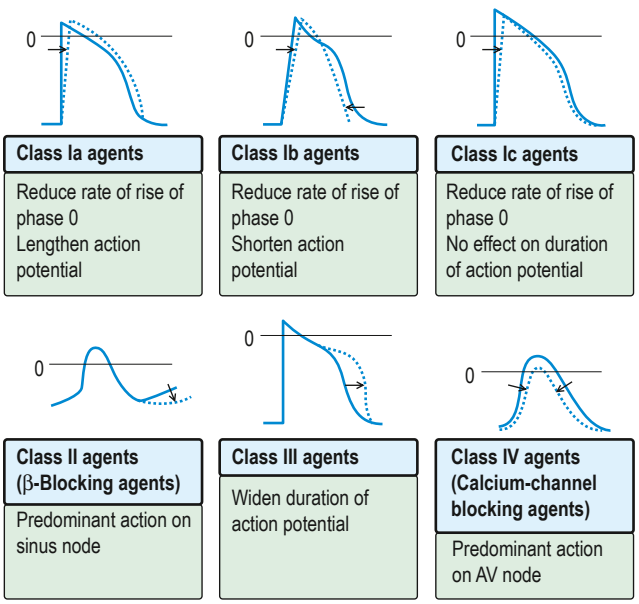


Fig. 17.47 Vaughan Williams classification of antiarrhythmic drugs based on their effect on cardiac action potentials (see also Fig. 17.13). (0 = 0 mV. The *dotted curves* indicate the effects of the drugs.) *AV,* Atrioventricular.

Class II drugs

These drugs prevent the effects of catecholamines on the action potential. Most are β-adrenoceptor antagonists. Cardioselective beta-blockers (β_1) include metoprolol, bisoprolol, atenolol, and acebutolol. Beta-blockers suppress AV node conduction, which may be effective in preventing attacks of junctional tachycardia and may help to control the ventricular rate during paroxysms of other forms of SVT (e.g., atrial fibrillation). In general, beta-blockers are anti-ischaemic and antiadrenergic and have proven beneficial effects in patients post MI (by preventing ventricular fibrillation) and in those with congestive heart failure. It is therefore advisable to use beta-blocker therapy either alone or in combination with other antiarrhythmic drugs in patients with symptomatic tachyarrhythmias, particularly those with CAD.

Class III drugs

These prolong the action potential, usually by blocking the rapid component of the delayed rectifier potassium current (I_{Kr}), and do not affect sodium transport through the membrane. The drugs in this class are amiodarone, dronedarone, and sotalol. Sotalol is also a beta-blocker.

Sotalol may result in acquired long QT syndrome and torsades de pointes, particularly in women. The risk of torsades is increased in the setting of renal dysfunction and hypokalaemia; particular care should be taken in patients taking diuretic therapy. Amiodarone therapy, in contrast to most other antiarrhythmic drugs, carries a low risk of proarrhythmia in patients with significant structural heart disease, but its use may be limited due to toxic and potentially serious side-effects. Dronedarone is a multichannel-blocking drug that suppresses the recurrence of atrial fibrillation and reduces hospital admissions in patients with cardiovascular risk. However, it has proven harmful in patients with left ventricular dysfunction and is contraindicated in heart failure.

Vernakalant is a multichannel blocker that is approved for the rapid intravenous medical cardioversion of new-onset atrial fibrillation.

Class IV drugs

The non-dihydropyridine calcium-channel blockers are particularly effective at slowing conduction in nodal tissue. These drugs can prevent attacks of junctional tachycardia (AVNRT and AVRT) and may help to control ventricular rates during paroxysms of other forms of SVT (e.g., atrial fibrillation).

Clinical use of antiarrhythmic drugs

Antiarrhythmic drugs have not been shown to prolong life. Patient safety is the main factor in determining the choice of antiarrhythmic therapy, and proarrhythmic risks need to be carefully assessed prior to initiating therapy. As a generalization, class Ic agents are employed in patients with structurally normal hearts, and class III agents are used in those with structural heart disease, although exceptions exist.

Patients with structurally normal hearts and normal QT intervals, or those with implantable defibrillators, either are at very low risk of proarrhythmia or are protected from any life-threatening consequences; in these individuals it is possible to persevere with drug therapy.

Catheter ablation

Catheter ablation (radiofrequency, cryoablation, and, more recently, pulse field) is frequently employed in the management of symptomatic tachyarrhythmias. Ablations are performed percutaneously by placing electrode catheters into the heart chambers, usually via femoral vessels. Successful ablation depends on accurate identification of either the site of origin of a focal tachycardia or a critical component of a macro re-entry tachycardia. Catheter ablation has been found to be highly effective in the following tachyarrhythmias:

- AVNRT
- AVRT with an accessory pathway, including WPW syndrome
- atrial fibrillation
- atrial flutter
- normal heart ventricular tachycardia
- atrial tachycardia.

Symptomatic patients with a ***pre-excited*** ECG because of accessory pathway conduction (WPW syndrome) are advised to undergo catheter ablation as first-line therapy, owing to the risk of sudden death should sustained atrial arrhythmia occur. Patients with accessory pathways that only conduct retrogradely from the ventricles to the atrium are not at increased risk of sudden death but experience symptoms due to AVRT. These individuals are commonly offered an ablation procedure if simple measures, such as AV nodal slowing agents, fail to suppress tachycardia. Asymptomatic patients with the WPW ECG pattern are now frequently offered an ablation procedure for prophylactic reasons. The main risk associated with accessory pathway ablation is thromboembolism in patients with left-sided accessory pathways.

Patients with ***normal hearts and documented ventricular tachycardia*** should be referred for specialist evaluation. Unlike ventricular tachycardia in patients with structural heart disease, normal heart ventricular tachycardia is not associated with increased risk of sudden death, and long-term success can be achieved by catheter ablation.

Catheter ablation is recommended in patients with ***atrial flutter*** that is not easily managed medically. Ablation of typical flutter is effective in 90% to 95% of cases. In the direct comparison of catheter ablation and antiarrhythmic therapy the rate of recurrence was significantly lower following ablation. ***Atrial tachycardia***, especially in patients with structurally normal hearts, may also be treated by catheter ablation. In atrial fibrillation adequate control of ventricular rates is sometimes not possible, despite optimal medical therapy. These patients experience

a marked symptomatic improvement following AV node ablation (which leads to complete heart block) and pacemaker implantation.

In ***younger patients*** with structurally normal hearts, ***atrial ectopic beats***, which commonly arise from a focus situated in the pulmonary veins, may trigger atrial fibrillation. Catheter ablation of this ectopic focus includes the application of radiofrequency energy around the pulmonary veins in order to abolish the connection between the sleeves of arrhythmogenic atrial myocardium surrounding or extending into the veins from the atrium (pulmonary vein isolation). The trigger is therefore eliminated, and the arrhythmia does not recur. These techniques appear to be highly effective, especially in young patients with paroxysmal atrial fibrillation, normal atrial size, and no underlying heart disease (70%–80% long-term success) but are presently time-consuming procedures (3 h or more) and carry a risk of serious complications such as stroke, tamponade, pulmonary vein stenosis, and atrio-oesophageal fistula in a small minority of patients (in experienced centres <2% altogether).

Implantable cardioverter–defibrillator

Up to 40% of patients diagnosed with life-threatening ventricular arrhythmias (ventricular fibrillation or rapid ventricular tachycardia with hypotension) will die within 1 year of diagnosis. Large multicentre prospective trials, such as the Antiarrhythmics Versus Implantable Defibrillator (AVID) trial (which compared an implantable defibrillator to amiodarone therapy), have proven that implantable defibrillators improve overall survival in patients who have experienced an episode of life-threatening ventricular tachyarrhythmia.

The ICD recognizes ventricular tachycardia or fibrillation and automatically delivers pacing or a shock to the heart to cause cardioversion to sinus rhythm. Modern ICDs are larger than a pacemaker and the generator is usually implanted in a pre- or subpectoral position (Fig. 17.48). The device may have leads to sense and pace both the right atrium and ventricle, and the lithium batteries employed are able to provide energy for over 100 shocks, each of around 40 J.

As an alternative to the traditional ICD with transvenous leads, there is also a subcutaneous device (S-ICD) where the generator is sited intramuscularly in the axilla, and a subcutaneous lead is tunnelled adjacent to the sternum. This option is increasingly favoured in the younger patient cohort as it minimizes the long-term complication of a defibrillator lead in the heart (infection, vascular obstruction, lead failure needing extraction, tricuspid valve dysfunction), but it does have limitations, requiring an individualized approach to patient selection.

ICD discharges are painful if the patient is conscious. However, ventricular tachycardia may often be terminated by overdrive pacing the heart, which is painless.

The ICD is superior to all other treatment options for preventing sudden cardiac death, and their use has cut the sudden death rate in patients with a history of serious ventricular arrhythmias to approximately 2% per year. However, the majority of these people have significant structural heart disease and overall cardiac mortality due to progressive heart failure remains high. As a result, the ICD is now first-line therapy in the secondary prevention of sudden death.

ICDs are also employed in the primary prevention of sudden cardiac death. The chances of surviving an out-of-hospital cardiac

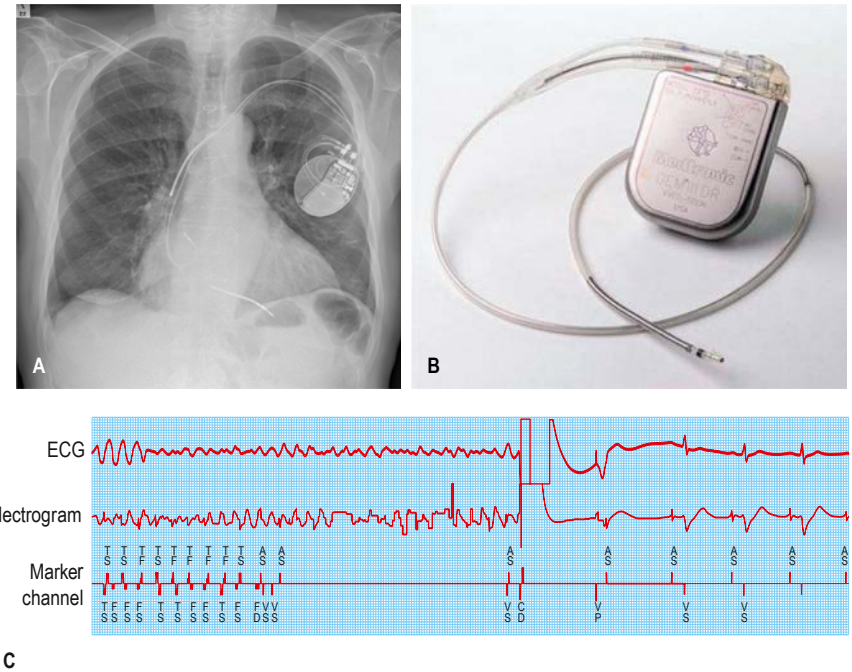


Fig. 17.48 Implantable cardioverter–defibrillator. **(A)** X-ray of a 'dual-chamber' defibrillator in a left pectoral position with atrial and ventricular leads. **(B)** Medtronic GEM® III DR defibrillator. **(C)** ECG showing the termination of ventricular fibrillation by a direct current shock at 20 J. The electrogram, recorded internally from the ventricular lead of an ICD, reveals chaotic ventricular activity consistent with ventricular fibrillation. This is confirmed by an electrocardiogram (lead V_2). The marker channel demonstrates that the device detects ventricular fibrillation correctly (FS, fibrillation sensing, lower line) and delivers an appropriate shock (CD) that terminates the arrhythmia and restores normal sinus rhythm (VS, ventricular sensing, lower line). Incidentally, atrial fibrillation is also detected before shock delivery (upper line on the marker channel).

arrest are as low as 10%. Therefore selected patients who have never experienced a spontaneous episode of life-threatening ventricular tachyarrhythmia but who are assessed to be at high risk of sudden death are advised to undergo ICD implantation. In two large primary prevention ICD trials Multicenter Automated Defibrillator Implantation Trial (MADIT II) and Sudden Cardiac Death in Heart Failure Trial (SCD-HeFT), therapy with an ICD reduced mortality by 23% to 31% on top of conventional treatment, which included revascularization, beta-blockers, and ACE inhibitors. ICD combined with cardiac resynchronization therapy may improve both symptoms and life expectancy of patients with any degree of heart failure (the COMPANION, CARE-HF, and MADIT-CRT trials).

The following groups of patients may merit prophylactic ICD placement for primary prevention:

- Patients with heart failure, with a left ventricular EF of 35% or less and NYHA functional class ≤3 (see **Box 17.20**). In such patients who also have LBBB (QRS > 120 ms) there is additional benefit in combining ICD with cardiac resynchronization therapy (CRT-D; see 'Cardiac resynchronization therapy or implantable cardioverter–defibrillator' section). In patients with DCM and low EF, scar on cardiac MRI (late gadolinium enhancement) identifies a subgroup of patients who are more likely to benefit from a primary prevention ICD.
- Those with a familial condition and high risk of sudden death, such as DCM and HCM, long QT syndrome, Brugada's syndrome, or other channelopathies, along with other high-risk factors.

Further reading

Glikson M, Nielsen J, Kronborg M, et al. 2021 ESC Guidelines on cardiac pacing and cardiac resynchronization therapy: *Eur Heart J* 2021; 42:3427–3520.
Hindricks G, Potpara T, Dagres N, et al. 2020 ESC Guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS). *Eur Heart J* 2021; 42:373–498.
National Institute for Health and Care Excellence. *NICE Interventional Procedures Guidance 603: Subcutaneous Implantable Cardioverter Defibrillator Insertion for Preventing Sudden Cardiac Death.* NICE 2017; https://www.nice.org.uk/guidance/ipg603.
Suravicz B, Childers R, Deal B, Gettes L. AHA/ACCG/HRS Recommendations for the standardization and interpretation of the electrocardiogram. part III: intraventricular conduction disturbances. *Circulation* 2009; 119:e235–e240.
Zeppenfeld K, Tfelt-Hansen J, dr Riva M, et al. 2022 ESC Guidelines for the management of patients with ventricular arrhythmias and prevention of sudden cardiac death: *Eur Heart J* 2022;43:3997–4126.

HEART FAILURE

Heart failure is a complex syndrome that can result from any structural or functional cardiac disorder that impairs the ability of the heart to function as a pump to support a physiological circulation. Worldwide, the incidence of heart failure is variable but increases with advancing age, with approximately 26 million people affected. In the UK overall incidence is about 2 in 1000. The prognosis of heart failure has improved over the past 10 years with evidence-based therapy, but the mortality rate remains high and approximately 50% of patients are dead at 5 years. Heart failure is the leading cause of medical hospital admissions in over 65-year-olds and accounts for 5% of admissions to hospital medical wards. The cost of managing heart failure in the UK exceeds £1 billion per year. CAD is the most common cause of heart failure in Western countries.

The causes of heart failure are shown in **Box 17.17**.

Pathophysiology

The physiological changes in heart failure are initially compensatory and maintain cardiac output and peripheral perfusion. However, as heart failure progresses, these mechanisms are overwhelmed and become pathological (**Box 17.18**). Peripheral vasoconstriction and sodium retention in heart failure by activation of the renin–angiotensin–aldosterone system entails a loss of beneficial compensatory mechanisms and contributes to the spiral of cardiac decompensation. Evaluation of cardiac function requires assessment of preload, afterload, and contractility.

? Box 17.17 Causes of heart failure

Main causes

- Ischaemic heart disease (35%–40%)
- Cardiomyopathy (dilated) (30%–34%)
- Hypertension (15%–20%)

Other causes

- Inherited: e.g., hypertrophic
- Valvular heart disease (mitral, aortic, tricuspid)
- Infiltrative: amyloidosis, sarcoidosis, haemochromatosis, Fabry's disease
- Congenital heart disease (atrial septal defect, ventricular septal defect)
- Alcohol and recreational drugs
- Prescribed medication related (trastuzumab, anthracyclines, tyrosine kinase inhibitors, checkpoint inhibitors, clozapine)
- Hyperdynamic circulation (anaemia, thyrotoxicosis, Paget's disease)
- Right heart failure (right ventricular infarct, pulmonary hypertension, pulmonary embolism, chronic obstructive pulmonary disease)
- Arrhythmias (atrial fibrillation, bradycardia [complete heart block, sick sinus syndrome])
- Pericardial disease (constrictive pericarditis, pericardial effusion)
- Infections: e.g., Chagas' disease
- Myocarditis (lymphocytic, eosinophilic, giant cell)

i Box 17.18 Pathophysiological changes in heart failure

- Ventricular dilation
- Myocyte hypertrophy
- Increased collagen synthesis
- Altered myosin gene expression
- Altered sarcoplasmic Ca^{2+}-adenosine triphosphatase density
- Increased atrial natriuretic peptide secretion
- Salt and water retention
- Sympathetic stimulation
- Peripheral vasoconstriction

Venous return (preload)

In the intact heart myocardial failure leads to a reduction of the volume of blood ejected with each heartbeat and an increase in the volume of blood remaining after systole. This increased diastolic volume stretches the myocardial fibres and, as Starling's law of the heart (see 'Starling's law of the heart' section) would suggest, myocardial contraction is restored. However, the failing myocardium results in depression of the ventricular function curve (cardiac output plotted against the ventricular diastolic volume) (see **Fig. 13.8**).

Mild myocardial depression is not associated with a reduction in cardiac output because it is maintained by an increase in venous pressure (and hence diastolic volume). However, the proportion of blood ejected with each heartbeat (EF) is reduced early in heart failure. Sinus tachycardia also ensures that any reduction of stroke volume is compensated for by the increase in heart rate; cardiac output (stroke volume × heart rate) is therefore maintained.

When there is more severe myocardial dysfunction, cardiac output can be maintained only by a large increase in venous pressure and/or marked sinus tachycardia. The increased venous pressure contributes to the development of dyspnoea, owing to the accumulation of interstitial and alveolar fluid, ascites with hepatic enlargement, and dependent oedema from increased systemic venous

pressure. However, the cardiac output at rest may not be much depressed, but myocardial and haemodynamic reserve is so compromised that a normal increase in cardiac output cannot be produced by exercise. In very severe heart failure the cardiac output at rest is depressed, despite high venous pressures. The inadequate cardiac output is redistributed to maintain perfusion of vital organs, such as the heart, brain, and kidneys, at the expense of the skin and muscle.

Outflow resistance (afterload)

Outflow resistance (afterload) (see Fig. 17.5) is the load or resistance against which the ventricle contracts. It is formed by:

- pulmonary and systemic resistance
- physical characteristics of the vessel walls
- the volume of blood that is ejected.

An increase in afterload decreases the cardiac output, resulting in a further increase of end-diastolic volume and dilation of the ventricle, which further exacerbates the problem of afterload. This is expressed by Laplace's law: the tension of the myocardium (T) is proportional to the intraventricular pressure (P) multiplied by the radius of the ventricular chamber (R)—that is, T∝PR. Systemic blood pressure correlates with afterload and is used as a surrogate measure.

Myocardial contractility (inotropic state)

The state of the myocardium also influences performance. The sympathetic nervous system is activated in heart failure via baroreceptors as an early compensatory mechanism, which provides inotropic support and maintains cardiac output. Chronic sympathetic activation, however, has deleterious effects by further increasing neurohormonal activation and myocyte apoptosis. This is compensated by a downregulation of β-receptors. Increased contractility (positive inotropism) can result from increased sympathetic drive, and this is a normal part of the Frank–Starling relationship (see Fig. 17.5). Conversely, myocardial depressants (e.g., hypoxia) decrease myocardial contractility (negative inotropism).

Neurohormonal and sympathetic system activation: salt and water retention

The increase in venous pressure that occurs when the ventricles fail leads to retention of salt and water as diminished renal perfusion, which activates the renin–angiotensin system (see Fig. 27.5). Salt and water accumulates in the interstitium, producing many of the physical signs of heart failure and further increasing venous pressure (Fig. 17.49). The retention of sodium is, in part, compensated by the action of circulating atrial natriuretic peptides and antidiuretic hormone.

Myocardial remodelling in heart failure

Left ventricular remodelling is a process of progressive alteration of ventricular size, shape, and function owing to the influence of mechanical, neurohormonal, and possibly genetic factors in several clinical conditions, including MI, cardiomyopathy, hypertension, and valvular heart disease. Its hallmarks include hypertrophy, loss of myocytes, and increased interstitial fibrosis. Remodelling continues for months after the initial insult, and the eventual change in the shape of the ventricle becomes responsible for the impairment of overall function of the heart (Fig. 17.50A). In cardiomyopathy the process of progressive ventricular dilation or hypertrophy takes place without ischaemic myocardial injury or infarction (Fig. 17.50B).

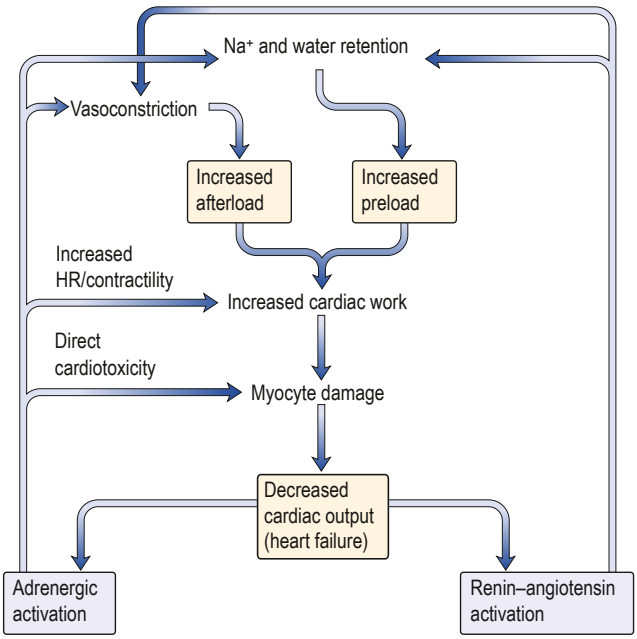


Fig. 17.49 The compensatory physiological response to heart failure. Chronic activation of the renin–angiotensin and adrenergic systems results in a 'vicious circle' of cardiac deterioration that further exacerbates the physiological response. *HR,* Heart rate.

Abnormal calcium homeostasis

Calcium ion flux within myocytes plays a pivotal role in the regulation of contractile function. Excitation of the myocyte cell membrane causes the rapid entry of calcium into myocytes from the extracellular space via calcium channels. This triggers the release of intracellular calcium from the SR and initiates contraction (see Fig. 17.3). Relaxation results from the uptake and storage of calcium by the SR (see Fig. 17.9), controlled by changes in nitric oxide. In heart failure there is a prolongation of the calcium current in association with prolongation of contraction and relaxation.

Natriuretic peptides (ANP, BNP, and CNP)

- ***Atrial natriuretic peptide (ANP)*** is released from atrial myocytes in response to stretch. ANP induces diuresis, natriuresis, vasodilation, and suppression of the renin–angiotensin system. Levels of circulating ANP are increased in congestive cardiac failure and correlate with functional class, prognosis, and haemodynamic state.
- ***B-type natriuretic peptide (BNP)*** is predominantly secreted by the ventricles in response to increased myocardial wall stress. N-terminal (NT)-proBNP is an inactive protein that is cleaved from proBNP to release BNP. Both BNP and NT-proBNP are increased in patients with heart failure, and levels correlate with ventricular wall stress and the severity of heart failure. BNP and NT-proBNP are good predictors of cardiovascular events and mortality, although monitoring levels are not routinely used to guide heart failure management.
- ***C-type natriuretic peptide (CNP)***, which is limited to vascular endothelium and the central nervous system, has similar effects to those of ANP and BNP.

Endothelial function in heart failure

The endothelium has a central role in the regulation of vasomotor tone. In patients with heart failure endothelium-dependent vasodilation in peripheral blood vessels is impaired and may be one mechanism of exercise limitation. The cause of abnormal endothelial

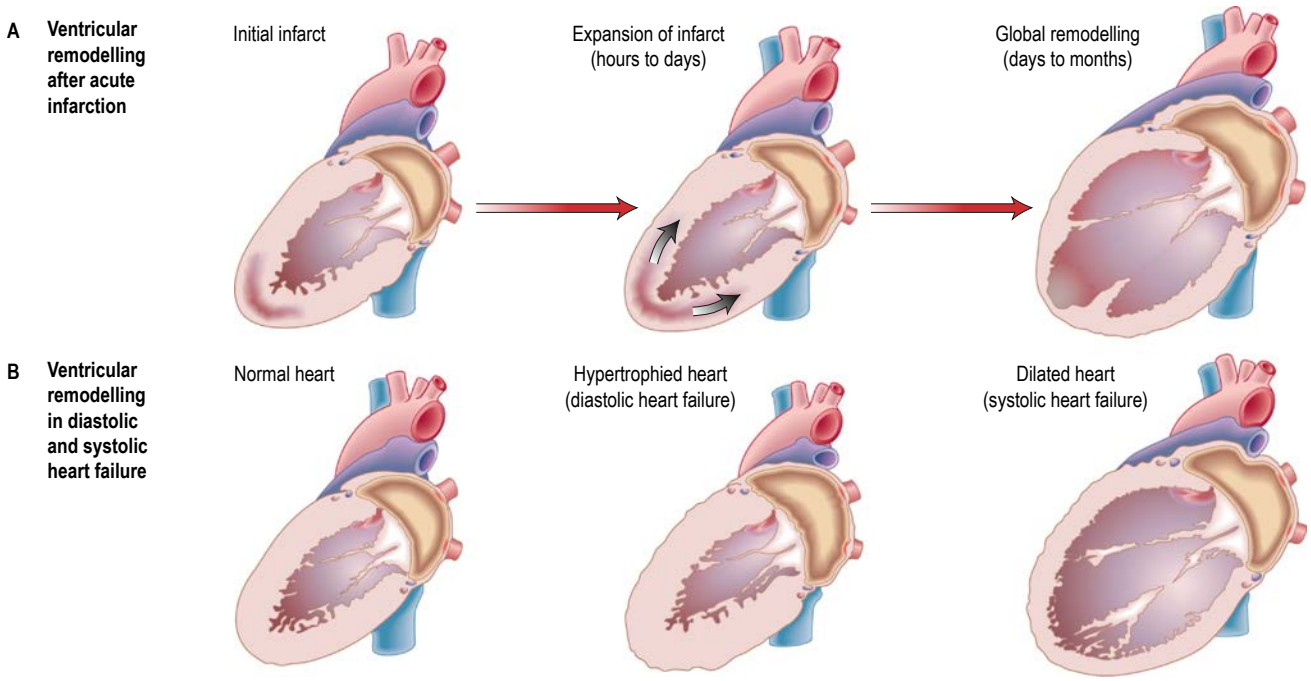


Fig. 17.50 Ventricular remodelling. **(A)** In ischaemic (myocardial infarction) heart failure. **(B)** In non-ischaemic heart failure (e.g., in cardiomyopathy).

responsiveness relates to abnormal release of both nitric oxide and vasoconstrictor substances, such as ET. The activity of nitric oxide, a potent vasodilator, is blunted in heart failure. ET secretion from a variety of tissues is stimulated by many factors, including hypoxia, catecholamines, and angiotensin II. The plasma concentration of ET is elevated in patients with heart failure, and levels correlate with the severity of haemodynamic disturbance.

ET has many actions that potentially contribute to the pathophysiology of heart failure: vasoconstriction, sympathetic stimulation, renin–angiotensin system activation, and left ventricular hypertrophy. Acute intravenous administration of ET antagonists improves haemodynamic abnormalities in patients with congestive cardiac failure, but oral ET antagonists have not shown long-term benefit in heart failure.

Arginine vasopressin (vasopressin)

Arginine vasopressin (AVP) is raised in severe chronic heart failure (CHF), particularly in patients on diuretic treatment. A high AVP concentration precipitates hyponatraemia, which is an ominous prognostic indicator. Oral short-term vasopressin antagonists can be considered in the management of patients with hospitalized decompensated heart failure.

Clinical syndromes of heart failure

There are many causes of heart failure (see **Box 17.17**) that can present suddenly, with AHF, or more insidiously, with CHF. One common classification of heart failure uses the left ventricular ejection fraction (LVEF):

- ***Heart failure with reduced ejection fraction (HFrEF)*** (EF ≤ 40%) is commonly caused by ischaemic heart disease but can also occur with valvular heart disease and hypertension.
- ***Heart failure with preserved ejection fraction (HFpEF)*** is a syndrome consisting of symptoms and signs of heart failure with an LVEF of 50% or greater. There is increased stiffness in the ventricular wall and decreased left ventricular compliance, leading to impairment of diastolic ventricular filling and hence decreased cardiac output. Echocardiography may demonstrate an increase in left ventricular wall thickness, increased left atrial size, and abnormal left ventricular relaxation with normal or near-normal left ventricular volume. HFPEF, which was historically described as ‘diastolic heart failure’, is more common in elderly hypertensive patients but may occur with primary cardiomyopathies (hypertrophic, restrictive, infiltrative).
- ***Heart failure with mildly reduced ejection fraction (HFmrEF)*** is a recently introduced classification for those patients in the ‘grey zone’ with an LVEF of 41% to 49%.
- ***Right ventricular systolic dysfunction (RVSD)*** may be secondary to chronic left-sided heart disease but can occur with primary and secondary pulmonary hypertension, right ventricular infarction, arrhythmogenic right ventricular cardiomyopathy, and adult congenital heart disease.

Clinical features of heart failure

The symptoms and signs of heart failure are shown in **Box 17.19**.

The NYHA classification of heart failure (**Box 17.20**) can be used to describe the symptoms of heart failure and limitation of exercise capacity and is useful for assessing response to therapy.

Diagnosis of heart failure

The diagnosis of heart failure should be based on a detailed history, clinical findings, natriuretic peptide levels, and objective evidence of cardiac dysfunction using measures of left ventricular structure and function (usually echocardiography). The underlying cause of heart failure should be established in all patients (**Box 17.21** and **Fig. 17.51**).

Investigations in heart failure

- ***Blood tests.*** Full blood count, serum creatinine and electrolytes, liver biochemistry, thyroid function, iron studies, fasting glucose, HbA_{1C}, and lipid profile should be considered. Cardiac enzymes (e.g., troponin) should be measured in AHF to assess for myocardial injury, and BNP or NT-proBNP may be used to diagnose CHF.
- ***Chest X-ray.*** Features of heart failure include cardiomegaly, pulmonary congestion with upper lobe diversion, fluid in fissures, Kerley B lines, and pulmonary oedema.

Box 17.19 Clinical features of heart failure

Symptoms
- Exertional dyspnoea
- Orthopnoea
- Paroxysmal nocturnal dyspnoea
- Fatigue

Signs
- Tachycardia
- Elevated jugular venous pressure
- Cardiomegaly
- Third and fourth heart sounds
- Bi-basal crackles
- Pleural effusion
- Peripheral ankle oedema
- Ascites
- Tender hepatomegaly

Box 17.20 New York Heart Association (NYHA) classification of heart failure

Class	Features
Class I	No limitation. Normal physical exercise does not cause fatigue, dyspnoea, or palpitations
Class II	Mild limitation. Comfortable at rest but normal physical activity produces fatigue, dyspnoea, or palpitations
Class III	Marked limitation. Comfortable at rest but gentle physical activity produces marked symptoms of heart failure
Class IV	Symptoms of heart failure occur at rest and are exacerbated by any physical activity

Box 17.21 Diagnosis of heart failure (European Society of Cardiology guidelines)

Diagnosis of HF-REF requires two conditions to be satisfied
1. Symptoms and/or signs typical of heart failure
2. Ejection fraction ≤ 40%

Diagnosis of HF-mREF requires two conditions to be satisfied
1. Symptoms and/or signs typical of heart failure
2. Ejection fraction 41%–49%

Diagnosis of HF-PEF requires three conditions to be satisfied
1. Symptoms and/or signs typical of heart failure
2. Ejection fraction ≥50%
3. Normal-sized LV with relevant structural heart disease (LV hypertrophy/left atrial enlargement) and/or diastolic dysfunction, including elevated natriuretic peptides

- ***ECG.*** Identify features of ischaemia, ventricular hypertrophy, or arrhythmia.
- ***Echocardiography.*** Assess cardiac chamber dimension, systolic and diastolic function, regional wall motion abnormalities, valvular disease, and cardiomyopathies.
- ***Stress echocardiography.*** Assess for ischaemia (and myocardial viability).
- ***Nuclear cardiology.*** Radionucleotide angiography (RNA) can quantify ventricular EF; SPECT or PET can demonstrate myocardial ischaemia and viability in dysfunctional myocardium. PET can be useful in diagnosing inflammatory cardiomyopathies.
- ***Cardiac MRI (CMR).*** Recommended for assessing cardiac structure, function, and viability, as well as myocardial characterization in suspected sarcoidosis, amyloidosis, haemochromatosis, Fabry's disease, and HCM.
- ***Coronary angiogram.*** For patients with angina and symptomatic ventricular arrhythmias.
- ***Right heart catheterization.*** This technique measures right atrial and pulmonary artery pressure (the 'wedge' pressure), a surrogate for left atrial pressure. It is valuable in assessment of pulmonary hypertension, prior to mitral or tricuspid valve intervention, in evaluation of patients with HFPEF, in heart transplant work-up, and in the management of cardiogenic shock.

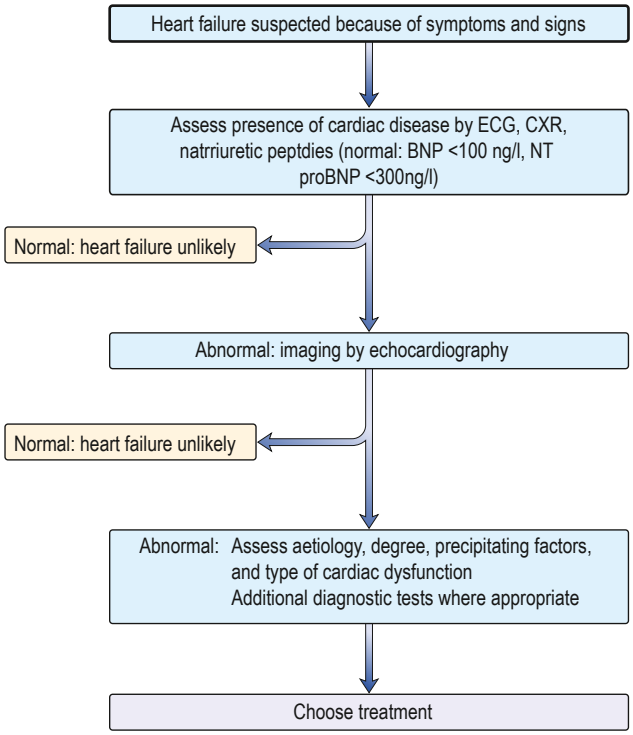


Fig. 17.51 Algorithm for the diagnosis of heart failure. *(Based on the European Society of Cardiology and NICE guidelines.)*

- ***Cardiac biopsy.*** This is used for diagnosis of cardiomyopathies, such as amyloid, myocarditis, and sarcoidosis, and for follow-up of transplanted patients to assess rejection.
- ***Cardiopulmonary exercise testing.*** Peak oxygen consumption (VO_2) is predictive of hospital admission and death in heart failure. A 6-minute exercise walk is an alternative. These assessments are clinically used in patients being assessed for heart transplant candidacy.
- ***Ambulatory 24-hour ECG monitoring (Holter).*** This is used in patients with suspected arrhythmia and may be employed in those with severe heart failure or inherited cardiomyopathy to determine whether a defibrillator is appropriate (non-sustained ventricular tachycardia).

Management of heart failure

The management of acute heart failure, presenting with pulmonary oedema, is discussed in Chapter 11 ('Acute heart failure' section). Management of CHF is aimed at relief of symptoms, prevention and control of disease leading to cardiac dysfunction, slowing disease progression, and improving quality and length of life.

Measures to prevent heart failure include cessation of smoking, alcohol, and illicit drugs, effective treatment of hypertension, diabetes and hypercholesterolaemia, and pharmacological therapy following MI.

The management of heart failure requires any factor involved in its aetiology to be identified and treated. Community nursing programs to help with drug compliance and detect early deterioration may prevent acute hospitalization.

General lifestyle advice

- ***Education.*** Effective counselling of patients and family, emphasizing weight monitoring and dose adjustment of diuretics, may prevent hospitalization. This is usually guided by primary care physicians and specialist community heart failure nurses.

- ***Dietary modification.*** Salt restriction is required, and foods rich in salt or salt added in cooking and at the table should be avoided. In severe heart failure fluid restriction is necessary and patients may need to weigh themselves daily. Alcohol has a negative inotropic effect and heart failure patients should moderate consumption.
- ***Smoking.*** Smoking should be stopped with help from anti-smoking clinics if necessary.
- ***Physical activity, exercise training, and rehabilitation.*** Low-level endurance exercise (e.g., 20–30 min walking 3–5 times per week or 20 min cycling at 70%–80% of peak heart rate 5 times per week) is actively encouraged in patients with compensated heart failure in order to reverse 'deconditioning' of peripheral muscle metabolism. Strenuous isometric activity should be avoided. For hospitalized patients with exacerbations of congestive cardiac failure, limiting activity reduces the demands on the heart; however, prolonged bed rest should be avoided, and consideration should be given to prophylaxis against venous thromboembolism in patients with reduced mobility.
- ***Vaccination.*** It is recommended that patients with heart failure be vaccinated against pneumococcal disease, influenza, and COVID-19.
- ***Sexual activity.*** Can be continued according to physical capabilities. Phosphodiesterase type 5 inhibitors (e.g., sildenafil) can be considered but are contraindicated in patients using nitrate medication as the combination may induce profound hypotension.
- ***Driving.*** Driving of cars and motorcycles may continue, provided that there are no symptoms that distract the driver's attention. Symptomatic heart failure or an LVEF of less than 40% disqualifies patients from holding a commercial driving license and driving large lorries and buses in the UK.

Monitoring

The clinical condition of a person with heart failure fluctuates; lengthy and repeated hospital admissions are common, with an average inpatient stay of between 5 and 10 days. Monitoring of clinical status is necessary and this responsibility should be shared between primary and secondary healthcare professionals.

Essential monitoring includes assessment of:

- functional capacity (e.g., NYHA functional class, exercise tolerance test, echocardiography)
- fluid status (body weight, clinical assessment, and serum creatinine and electrolytes)
- cardiac rhythm (ECG, Holter monitoring).

Multidisciplinary team approach

Heart failure care should be delivered by a multidisciplinary team with an integrated approach across the healthcare community. The team should involve specialist healthcare professionals: cardiologist or physician with a specialist interest in heart failure, heart failure nurse, dietician, pharmacist, occupational therapist, physiotherapist, and palliative care adviser. Understanding the information needs of patient and carers is vital. Good communication is essential for best clinical management and should include advice on anxiety, depression, and 'end-of-life' issues.

Drug management

Box 17.22 and **Fig. 17.52** show the treatment options. Four classes of medicine have been shown to reduce premature mortality and heart failure hospitalization in patients with heart failure with reduced EF (sometimes called the 'four pillars' of heart failure management):

- ***inhibitors of the renin–angiotensin system*** (ACE inhibitors, angiotensin receptor antagonists, or combined angiotensin receptor-neprilysin inhibitors)
- ***beta-blockers***
- ***aldosterone receptor antagonists*** (such as spironolactone or finerenone)
- ***sodium-glucose co-transporter 2 inhibitors***.

Systematic network meta-analysis has shown that patients with heart failure who are administered all four classes of agents benefit from a 73% reduction in all-cause mortality over 2 years. The guidelines also support the use of these agents in patients with heart failure with mildly reduced EF.

Prioritizing rapid simultaneous initiation of all four foundational therapies at low doses (Fig. 17.52) precedes individual drug dose optimization, as initiating a new drug class generally provides greater benefits than increasing the dose of existing medications. In the hospital setting, this can reduce absolute mortality risk by up to 25% over 1 year, compared to patients not receiving guideline directed therapy (STRONG-HF trial). This translates to a number needed to treat (NNT) of 4 to prevent one death.

Inhibitors of the renin–angiotensin system

Angiotensin-converting enzyme inhibitors

The use of ACE inhibitors in patients with heart failure has been demonstrated in multiple large randomized controlled trials (CONSENSUS, SOLVD) to improve symptoms and reduce mortality significantly. ACE inhibitors also benefit patients with asymptomatic heart failure following MI. Thus ACE inhibitors improve survival in patients in all functional classes (NYHA I–IV) and are recommended in all patients at risk of developing heart failure. The main adverse effects of ACE inhibitors are cough, hypotension, hyperkalaemia, and renal dysfunction. Contraindications to their use include bilateral renal artery stenosis, pregnancy, and previous angio-oedema. In patients with heart failure ACE inhibitors should be introduced at a low initial dose and gradually titrated, with regular monitoring of blood pressure and renal function.

Elevated creatinine is frequently cited as a reason not to start RAS receptor blockers in patients with heart failure due to concerns of these classes of drugs effecting a further deteriorating renal function. However, the STOP ACEi trial demonstrated that in chronic kidney disease patients with an estimated GFR of less than 30 mL/min/1.73 m^2, discontinuation of RAS did not impact the natural deterioration in estimated GFR over 3 years.

Angiotensin II receptor antagonists

The angiotensin II receptor antagonists (ARAs; candesartan, losartan, valsartan) are indicated as second-line therapy in patients intolerant of ACE inhibitors. Unlike ACE inhibitors, they do not affect bradykinin metabolism and do not produce a cough. The CHARM Alternative Trial showed that candesartan reduced the risk of heart failure hospitalization compared to placebo in patients intolerant of ACE inhibitors. Other trials (Val-HeFT and ELITE II) have assessed other ARAs. However, unlike ACE inhibitors ARAs have not been shown to reduce all-cause mortality in clinical trials.

Angiotensin receptor–neprilysin inhibitor

Angiotensin receptor–neprilysin inhibitors (ARNIs) are a new class of drug that produce dual inhibition of the angiotensin (AT_1) receptor and the natriuretic system. The PARADIGM-HF trial studied sacubitril/valsartan in patients with established symptomatic heart failure and reduced EF and demonstrated superiority over ACE inhibitor therapy in lowering morbidity and mortality. Symptomatic ambulatory patients who are on ACE inhibitors or ARA should be swapped to ARNIs. Patients on ACE inhibitors should have this drug discontinued at least 36 hours prior to commencement of ARNI to reduce the risk of angioedema. Further trial data (PIONEER-HF and TRANSITION) support the commencement of *de novo* ARNI in hospitalized patients with heart failure.

Box 17.22 Drugs used in heart failure

Drug	Dose (initial/maximum)	Precautions
Drugs blocking the renin-angiotensin system		
ACE inhibitors		
Captopril	6.25 mg ×3 daily/50 mg ×3 daily	
Enalapril	2.5 mg ×2 daily/10–20 mg ×2 daily	
Lisinopril	2.5–5 mg daily/20–40 mg daily	
Perindopril	2 mg daily/8–16 mg daily	
Ramipril	1.25–2.5 mg ×2 daily/5 mg ×2 daily	Monitor renal function and use with caution if baseline serum creatinine >250 μmol/L or baseline blood pressure <90 mmHg
ARAs		
Candesartan	4 mg daily/32 mg daily	
Losartan	50 mg daily/100 mg daily	
Valsartan	40 mg ×2 daily/160 mg ×2 daily	
ARNIs		
Angiotensin receptor neprilysin inhibitor Sacubitril/valsartan	24/26–49/51 mg ×2 daily/97/103 mg × daily	Monitor renal function, check for hyperkalaemia. Discontinue ACE inhibitor 3 days prior to starting, discontinue ARA the day of starting
Beta-adrenoceptor-blocking drugs		
Bisoprolol	1.25 mg daily/10 mg daily	Use with caution in obstructive airways disease, bradyarrhythmias
Carvedilol	3.125 mg ×2 daily/50 mg ×2 daily	Avoid in acute heart failure until patient is cardiovascularly stable
Metoprolol succinate (CR/XL)	12.5–25 mg daily/200 mg daily	
Nebivolol	1.25 mg daily/10 mg daily	
Aldosterone antagonists		
Eplerenone	25 mg daily/50 mg daily	
Spironolactone	12.5–25 mg daily/50 mg daily	Monitor renal function, check for hyperkalaemia, gynaecomastia with spironolactone
Sodium-glucose co-transporter 2 inhibitor (SGLT2i)		
Dapagliflozin	10 mg daily	Advise genital hygiene post urination to reduce risk of candida
Empagliflozin	10 mg daily	
Other drugs used in heart failure		
Diuretics		
Bendroflumethiazide	2.5 mg daily	Rarely need more than 2.5 mg daily. Reduced efficacy when eGFR<30 mL/min/1.73 m^2
Bumetanide	0.5–1.0 mg daily/5–10 mg daily	
Furosemide	20–40 mg daily/250–500 mg daily	Monitor renal function and check for hypokalaemia and hypomagnesaemia
Metolazone	2.5 mg daily/10 mg daily	Use in severe heart failure
Cardiac glycosides		
Digoxin	0.125–0.25 mg daily (reduce dose in elderly or in renal impairment)	Use with caution in renal impairment or conduction disease, and with amiodarone
Vasodilators		
Hydralazine	37.5–75 mg ×3 daily	
Isosorbide dinitrate	20–40 mg ×3 daily	
I_F *channel blocker*		
Ivabradine	5 mg daily/7.5 mg ×2 daily	Use with caution in sick sinus syndrome; AV block
Soluble guanylate cyclase inhibitor		
Vericiguat	2.5 mg/10 mg daily	

ACE, Angiotensin-converting enzyme; *ARA*, angiotensin II receptor antagonist; *AV*, atrioventricular.

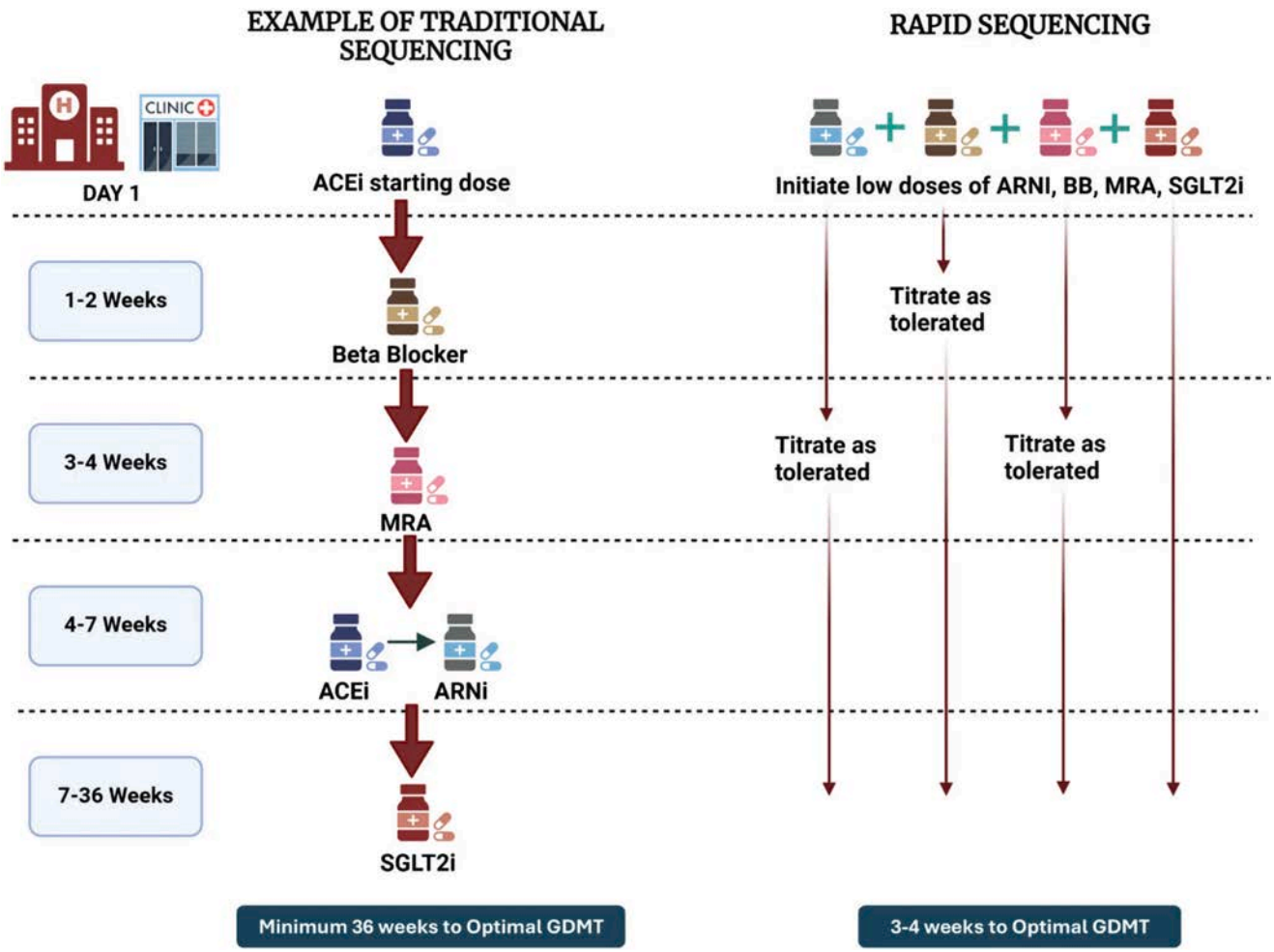


Fig. 17.52 Traditional versus rapid sequence initiation of guideline-directed medical therapy in heart failure with reduced ejection fraction. *(Shahid, I., Khan, M.S., Butler, J. et al. Initiation and sequencing of guideline-directed medical therapy for heart failure across the ejection fraction spectrum.* Heart Fail Rev *[2025]. https://doi.org/10.1007/s10741-025-10481-7.)*

Beta-blockers

Beta-blockers have been shown to improve functional status and reduce cardiovascular morbidity and mortality in patients with heart failure. Several trials (CIBIS II, MERIT-HF, COMET, SENIORS) have assessed the effects of beta-blockers in varying degrees of heart failure. Bisoprolol, carvedilol, and sustained-release metoprolol reduce mortality in heart failure. Nebivolol is used in the treatment of stable mild to moderate heart failure in patients over 70 years old. In practice, beta-blockers are started once a patient is euvolaemic at a low dose and gradually increased, with monitoring of heart rate and blood pressure.

Aldosterone antagonists

The aldosterone antagonists spironolactone and eplerenone have been shown to improve survival in patients with symptomatic heart failure. In the RALES study spironolactone reduced total mortality by 30% in severe heart failure. However, gynaecomastia or breast pain occurred in 1 in 10 men taking spironolactone. In EPHESUS eplerenone given to patients with an acute MI and heart failure reduced total mortality by 15% and sudden cardiac death by 21%, with no gynaecomastia.

Sodium-glucose co-transporter 2 inhibitors

Sodium-glucose co-transporter 2 inhibitors (SGLT2i) reduce the risk of heart failure admission and cardiovascular mortality (DAPA-HF, EMPEROR-reduced and SOLOIST-WHF). This class of drug, though initially used in the management of diabetes, is now widely accepted as a prognosis-altering heart failure medication that can be used in the absence of diabetes. SGLT2i are the only drugs that reduce the risk of heart failure hospitalization and cardiovascular mortality in patients with heart failure with preserved ejection fraction (HFPEF). Patients should be counselled about hygiene post urination to reduce the risk of genital candida. This drug should be withheld in situations of prolonged fasting to reduce the risk of euglycaemic ketoacidosis.

Other medication used in heart failure

The following agents lack the clear prognostic benefits for the four classes of agents above but may still have benefit in selected patients.

Diuretics

These act by promoting the renal excretion of salt and water by blocking tubular reabsorption of sodium and chloride (see Chapter 26). Loop diuretics (e.g., furosemide and bumetanide) should be given in patients with fluid overload. Although diuretics provide symptomatic relief of dyspnoea and improve exercise tolerance, there is limited evidence (due to lack of a clinical trial) that they affect survival. In patients with severe heart failure the combination of a loop and thiazide/thiazide-type diuretic (e.g., bendroflumethiazide, metolazone) may be required. There is emerging data for the use of acetazolamide (carbonic anhydrase inhibitor) or tolvaptan

(vasopressin receptor antagonist) as an add-on therapy to loop diuretics in acute decompensated heart failure. Serum electrolytes and renal function must be monitored regularly (risk of hypokalaemia and hypomagnesaemia).

Cardiac glycosides

Digoxin is a cardiac glycoside that is indicated in patients with atrial fibrillation with heart failure. Although the DIG study demonstrated that digoxin reduced hospital admissions in patients with heart failure, a subanalysis in the ROCKET AF trial suggested that mortality may, in fact, be increased. It is now infrequently used as add-on therapy in symptomatic heart failure patients already receiving ACE inhibitors and beta-blockers. If it is used, digoxin levels (6 hours post dose) should be measured, with the knowledge that a concentration of more than 1 ng/mL is associated with excess mortality.

Vasodilators and nitrates

The combination of hydralazine and nitrates reduces preload and afterload and is used in patients intolerant of ACE inhibitors or ARAs. The Veterans Administration Cooperative Study demonstrated that it improved survival in patients with CHF. The A-HeFT trial showed that the same combination reduced mortality and hospitalization for heart failure in black patients with heart failure.

Ivabradine

Ivabradine selectively decreases heart rate without affecting blood pressure by inhibiting the I_F channels in the SA node. An elevated heart rate in patients with heart failure is associated with worse cardiovascular outcomes; in the first instance it is important to exclude fluid overload. The SHIFT study reported a reduction in cardiovascular death and heart failure hospitalization in patients in sinus rhythm with CHF and left ventricular dysfunction (LVEF ≤ 35%). Ivabradine can be used in patients in sinus rhythm with a heart rate of more than 70 beats per minute despite beta-blocker treatment or in those who are unable to tolerate beta-blockers.

Vericiguat

Vericiguat is a stimulator of soluble guanylate cyclase and increases the functionality of nitric oxide, causing vasodilation, reduction in inflammation, and subsequent improvement in cardiac and vascular function. In the VICTORIA study, in patients with an EF of less than 45% and a recent decompensation requiring intravenous diuretics, vericiguat, as compared to placebo, reduced the frequency of heart failure hospitalizations but had no effect on mortality. The drug was also associated with a slightly higher incidence of anaemia and symptomatic hypotension.

Intravenous iron

Iron deficiency is common in heart failure and is multifactorial. Oral iron substitution is ineffective due to ineffective absorption in heart failure and hepcidin block (see Fig. 38.11). Multiple studies (IRONMAN, AFFIRM-AHF) have shown that intravenous iron therapy improves symptoms and quality of life irrespective of the presence of anaemia in patients with heart failure and reduced or mildly reduced EF and iron deficiency (transferrin saturation <20% or serum ferritin <100 μg/L).

Inotropic and vasopressor agents

Intravenous inotropes and vasopressor agents (see **Box 17.29**) are used in patients with AHF or cardiogenic shock (see Chapter 13, 'Cardiogenic shock' section). They have a limited role in the management of CHF and in this context are restricted to bridging therapy while awaiting heart transplantation.

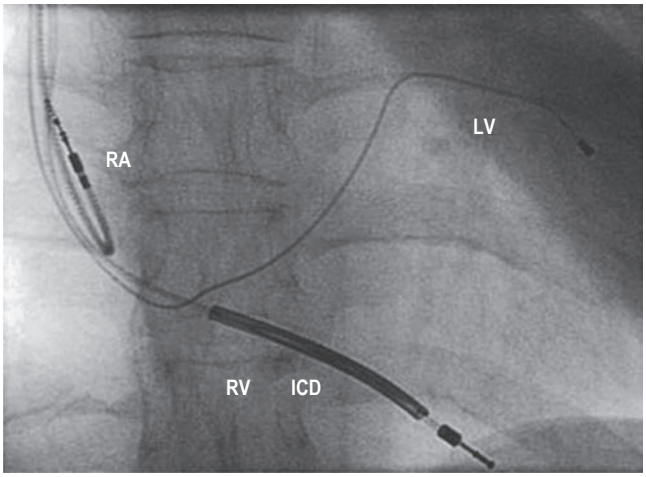


Fig. 17.53 Implanted biventricular pacemaker with implantable cardioverter–defibrillator device (CRT-D). *LV,* Left ventricle; *RA,* right atrium; *RV,* right ventricle.

Non-pharmacological treatment

Cardiac resynchronization therapy or implantable cardioverter–defibrillator

Cardiac resynchronization therapy (CRT) entails simultaneous pacing of both ventricles (biventricular pacing) using a lead placed in the right ventricle and another in the coronary sinus to pace the left ventricle (**Fig. 17.53**). It is an effective therapy in addition to optimal tolerated medical treatment in patients with significant left ventricular impairment and a prolonged QRS interval (data supporting benefit is predominately in patients with LBBB). Resynchronization may reverse the process of ventricular remodelling, reduce functional mitral regurgitation, and improve left ventricular function. The CARE-HF and COMPANION trials reported symptomatic benefit and a reduction in heart failure events and mortality following CRT implantation in patients with heart failure in NYHA classes III and IV. Similar findings have been noted in more recent trials in patients with mild heart failure or no symptoms (RAFT, MADIT-CRT). Resynchronization is now also attempted with conduction system pacing.

Most patients with heart failure who receive CRT also meet the criteria for use of an ICD and could receive a combined device (CRT-D, 'D' denoting 'defibrillator'). Patients with end-stage heart failure (NYHA IV) or other co-morbidities that may significantly reduce lifespan are generally not candidates for an ICD.

Ablation

There is increasing evidence that catheter ablation to achieve sinus rhythm in patients with atrial fibrillation/flutter is effective in reducing risk of death or worsening heart failure compared to placebo in patients with an EF of less than 35% (CASTLE-AF).

In patients with frequent ventricular ectopics (>10% burden) and impaired systolic burden, catheter ablation to reduce the burden of ectopics can improve cardiac function.

Revascularization

While CAD is the most common cause of heart failure, the role of revascularization in patients with heart failure is unclear. Patients with angina and left ventricular dysfunction have a higher mortality from surgery (10%–20%) but have the most to gain in terms of improved symptoms and prognosis. Factors that must be considered before recommending surgery include symptoms, age, co-morbidities, and evidence of reversible myocardial ischaemia.

Hibernating myocardium and myocardial stunning

'Hibernating' myocardium can be defined as reversible left ventricular dysfunction due to chronic CAD that responds positively to inotropic stress and indicates the presence of viable heart muscle that may recover after revascularization. It is caused by reduced myocardial perfusion, which is just sufficient to maintain viability of the heart muscle. Myocardial hibernation results from repetitive episodes of cardiac stunning that occur, for example, with repeated exercise in a patient with CAD.

Myocardial stunning is reversible ventricular dysfunction that persists following an episode of ischaemia when the blood flow has returned to normal: that is, there is a mismatch between flow and function.

The prevalence of hibernating myocardium in patients with CAD can be estimated from the frequency of improvement in regional abnormalities in wall motion after revascularization. Techniques to try to identify hibernating myocardium include stress echocardiography, nuclear imaging, CMR, and PET.

The clinical relevance of the hibernating and stunned myocardium is that ventricular dysfunction due to these mechanisms may be wrongly ascribed to myocardial necrosis and scarring, which seems untreatable, whereas reversible hibernating and stunned myocardium respond to coronary revascularization.

The evidence supporting revascularization as a treatment for patients with heart failure and reduced ejection in the absence of angina or ACS is very limited (REVIVED-BCIS2).

Advanced heart failure

Advanced heart failure therapies comprise both cardiac transplantation and durable mechanical circulatory supports (or ventricular assist devices).

Cardiac transplantation

Cardiac transplantation is the treatment of choice for younger patients with severe intractable heart failure. Identifying patients who might benefit from advanced heart failure treatments is challenging and the acronym '*I NEED HELP*' in **Box 17.23** may help. With careful recipient selection, the expected 1-year survival for patients following transplantation is over 90%, and median survival is 14 years. Irrespective of survival, quality of life is dramatically improved for the majority. The availability of organs for heart transplantation is limited.

Heart allografts do not function normally (but are still better than a severely diseased one), as cardiac denervation results in a high resting heart rate, loss of diurnal blood pressure variation, and impaired renin–angiotensin–aldosterone regulation. Some patients develop a 'stiff heart' syndrome, caused by rejection, denervation, and ischaemic injury during organ harvest and implantation.

The complications of heart transplantation are summarized in **Box 17.24**. Many (infection, malignancy, hypertension, and hyperlipidaemia) are related to immunosuppression. Allograft coronary atherosclerosis is the major cause of long-term graft failure and is present in 30% to 50% of patients at 5 years. It is due to a 'vascular' rejection process in conjunction with hypertension and hyperlipidaemia.

There are specific contraindications to cardiac transplantation (**Box 17.25**); notably, high pulmonary vascular resistance (PVR) and active malignancy are absolute contraindications.

Durable ventricular assist devices

Ventricular assist devices (VADs) are mechanical systems that replace or help the failing ventricles in delivering blood around the body and are powered by a driveline that exits the abdomen. The Heartmate 3™ (Abbott) is currently the main left ventricular assist device (LVAD) by market share (**Fig. 17.54**); it receives blood from the left ventricle and delivers it to the aorta in a continuous, non-pulsatile flow; these devices can also be implanted to receive blood from the right atrium and deliver it to the pulmonary artery, effectively a right ventricular assist device (RVAD). The Syncardia Total Artificial Heart™ has independently pumping right and left chambers that are surgically implanted in the place of the existing ventricles and deliver total cardiac support with pulsatile flow. Both devices are intracorporeal (suitable for long-term support as a bridge to transplantation or in some countries as destination therapy in patients with end-stage heart failure who are not candidates for transplantation). The main problems with VADs include thromboembolism, bleeding, infection, and device malfunction.

Box 17.23 Acronym for clinical indicators that should trigger possible referral to an advanced heart failure unit: I NEED HELP

I Intravenous inotropes (previous or on-going)
N New York Heart Association (NYHA) class IIIb to IV or persistently elevated natriuretic peptides
E End-organ dysfunction (worsening renal or liver dysfunction)
E EF ≤20%
D Defibrillator shocks (recent appropriate shocks)
H Hospitalizations (>1 in the last 12 months)
E Oedema despite escalating diuretics
L Low systolic BP (≤90 mmHg)
P Prognostic medication (progressive intolerance or down-titration of GDMT)

(Adapted from Baumwol J 'I Need Help'—A mnemonic to aid timely referral in advanced heart failure. J Heart Lung Transplant *2017;36:593-594.)*

Box 17.24 Complications of cardiac transplantation

- Allograft rejection:
  - 'Humoral'
  - 'Vascular'
  - 'Cell-mediated'
- Infections:
  - Early: hospital-acquired organisms—staphylococci; Gram-negative bacteria
  - Late (2–6 months): opportunistic (toxoplasmosis, cytomegalovirus, fungi, *Pneumocystis*)
- Allograft vascular disease
- Hypertension
- Hypercholesterolaemia
- Malignancy

Box 17.25 Contraindications to cardiac transplantation

- Age >60 years (some variations between centres)
- Alcohol/drug misuse
- Uncontrolled psychiatric illness
- Uncontrolled infection
- Severe renal/liver failure
- High pulmonary vascular resistance
- Systemic disease with multiorgan involvement
- Treated cancer in remission but with <5 years' follow-up
- Recent thromboembolism
- Other disease with a poor prognosis

Further reading

Felker GM, Ellison DH, Mullens W, Cox ZL, Testani JM. Diuretic therapy for patients with heart failure: JACC State-of-the-Art Review. *JACC* 2020; 75(10):1178–1195.

McDonagh TA, Metra M, Adamo M, Gardner RS et al. 2023 Focused update on the 2021 ESC guidelines for the diagnosis and treatment of acute and chronic heart failure. *Eur Heart J* 2023; 00:1–13.

CORONARY ARTERY DISEASE

Coronary artery disease (CAD) is the largest single cause of death in Europe, the USA, and many other parts of the world. It is estimated that 2.3 million people in the UK are living with coronary heart disease. Cardiovascular diseases were responsible for 168 030 deaths

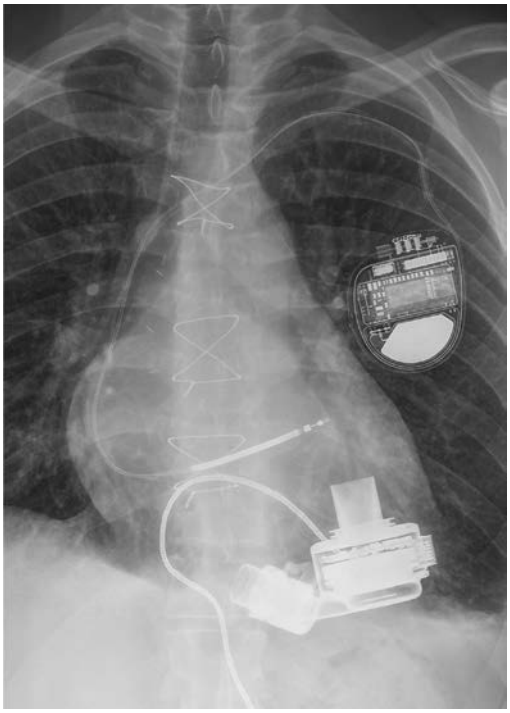


Fig 17.54 Radiograph of a HeartMate 3 left ventricular assist device.

in the UK, with 1 in 8 men and 1 in 14 women dying from coronary heart disease. This equates to one person dying from coronary heart disease every 8 minutes.

Myocardial ischaemia occurs when there is an imbalance between the supply of oxygen (and other essential myocardial nutrients) and the myocardial demand for these substances.

Coronary blood flow to a region of the myocardium may be reduced by a mechanical obstruction that is due to:

- atheroma
- thrombosis
- spasm
- embolus
- coronary ostial stenosis (seen after cardiac surgery)
- coronary arteritis (e.g., in systemic lupus erythematosus).

Myocardial ischaemia can also be caused or exacerbated by a decrease in the flow of oxygenated blood to the myocardium due to:

- anaemia
- carboxyhaemoglobulinaemia
- hypotension, causing decreased coronary perfusion pressure.

A higher demand for oxygen may cause myocardial ischaemia due to an increase in cardiac output (e.g., thyrotoxicosis) or myocardial hypertrophy (e.g., from aortic stenosis or hypertension).

Myocardial ischaemia most commonly arises as a result of obstructive CAD in the form of coronary atherosclerosis. In addition to this fixed obstruction, variations in the tone of smooth muscle in the wall of a coronary artery may add another element of dynamic or variable obstruction.

Pathophysiology of coronary atherosclerosis

Coronary atherosclerosis is a complex inflammatory process characterized by the accumulation of lipid, macrophages, and smooth muscle cells in intimal plaques in the large and medium-sized epicardial coronary arteries. The vascular endothelium plays a critical role in maintaining vascular integrity and homeostasis. Mechanical shear stresses (e.g., from hypertension), biochemical abnormalities (e.g., elevated low-density lipoprotein [LDL] or diabetes mellitus), immunological factors (e.g., free radicals from smoking), inflammation, and genetic alteration may contribute to the initial endothelial 'injury' or dysfunction, which is believed to trigger atherogenesis.

The ***development of atherosclerosis*** follows endothelial dysfunction, with increased permeability to and accumulation of oxidized lipoproteins, which are taken up by macrophages at focal sites within the endothelium to produce lipid-laden foam cells. Macroscopically, these lesions are seen as flat yellow dots or lines on the endothelium of the artery and are known as ***'fatty streaks'***. The 'fatty streak' progresses with the appearance of extracellular lipid within the endothelium ***('transitional plaque')***. Release of cytokines, such as platelet-derived growth factor and transforming growth factor beta (TGF-β), by monocytes, macrophages, or the damaged endothelium promotes further accumulation of macrophages, as well as smooth muscle cell migration and proliferation. The proliferation of smooth muscle with the formation of a layer of cells covering the extracellular lipid separates it from the adaptive smooth muscle thickening in the endothelium. Collagen is produced in larger and larger quantities by the smooth muscle and the whole sequence of events cumulates as an '***advanced*** or ***raised fibrolipid plaque***'. The 'advanced plaque' may grow slowly and encroach on the lumen or become unstable, undergo thrombosis, and produce an obstruction ***('complicated plaque')***.

Two different mechanisms are responsible for thrombosis on the plaques (Fig. 17.55):

- ***The first process*** is superficial endothelial injury, which involves denudation of the endothelial lining of the plaque. Subendocardial connective tissue matrix is then exposed and platelet adhesion occurs because of reaction with collagen. The thrombus is adherent to the surface of the plaque.
- ***The second process*** is deep endothelial fissuring, which involves an advanced plaque with a lipid core. The plaque cap tears (ulcerates, fissures or ruptures), allowing blood from the lumen to enter the inside of the plaque itself. The plaque core contains lamellar lipid surfaces, tissue factor (which triggers platelet adhesion and activation) produced by macrophages, and exposed collagen, which are highly thrombogenic. Thrombus forms within the plaque, expanding its volume and distorting its shape. Thrombosis may then extend into the lumen.

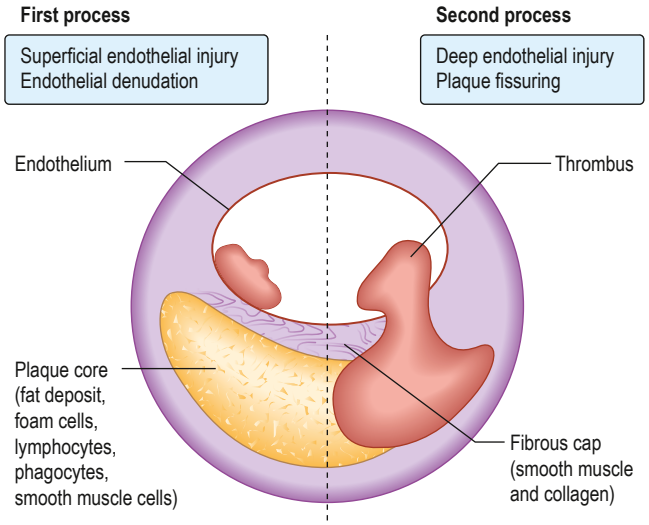


Fig. 17.55 Mechanisms underlying the development of thrombosis on plaques.

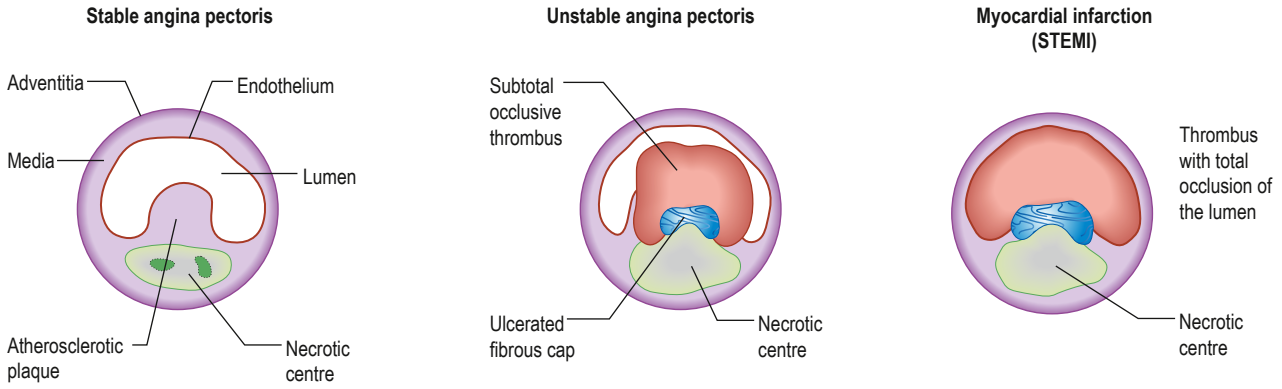


Fig. 17.56 Mechanisms underlying the development of thrombosis on plaques in various clinical syndromes. Relationship between the state of the coronary artery vessel wall and the various syndromes. *STEMI,* ST-elevation myocardial infarction.

A 50% reduction in luminal diameter (producing a reduction in luminal cross-sectional area of approximately 70%) causes a haemodynamically significant stenosis. At this point, coronary autoregulation through dilatation of the distal arterioles maintains adequate resting blood flow to the myocardium, but any further increase in myocardial oxygen demand would provoke ischaemia.

CAD gives rise to a wide variety of clinical presentations, ranging from relatively stable angina through to the ACSs of unstable angina (UA) and MI (Fig. 17.56). Fig. 17.57 shows a plaque rupture.

Risk factors for coronary artery disease

CAD is an atherosclerotic disease that is multifactorial in origin, giving rise to the risk factor concept. Certain living habits promote atherogenic traits in genetically susceptible persons, and more than 90% of coronary heart disease events will occur in individuals with at least one 'risk' factor (Box 17.26). Some of these, such as age, gender, race, and family history, cannot be modified, whereas other major risk factors, such as serum cholesterol, smoking habits, diabetes, and hypertension, can.

Atherosclerotic disease manifest in one vascular bed is often advanced in other territories. Patients with peripheral vascular disease (intermittent claudication) have a two- to fourfold increased risk of CAD, stroke, or heart failure. After a first MI, there is a three- to sixfold increase in the risk of heart failure and stroke. After stroke, the risk of heart failure and MI is increased twofold, and following a diagnosis of heart failure, risk of stroke is increased two- to threefold.

The disease can be asymptomatic even in its most severe form, with 1 in 3 MIs going unrecognized. Some 30% to 40% of individuals who present with an ACS have had no prior warning symptoms to suggest the presence of underlying disease.

Diagnosis

Cardiovascular risk assessment for primary and secondary prevention of cardiovascular disease

Primary prevention can be defined as prevention of the atherosclerotic disease process and ***secondary prevention*** as treatment of the atherosclerotic disease process (i.e., treatment of the disease or its complications). The objective of prevention is to reduce the incidence of first or recurrent clinical events due to CAD, ischaemic stroke, and peripheral artery disease.

In the UK, NICE guidelines recommend that primary care should use the QRISK®3 risk assessment tool (see Further reading) to identify people who are likely to be at high risk (10-year risk of cardiovascular disease ≥10%).

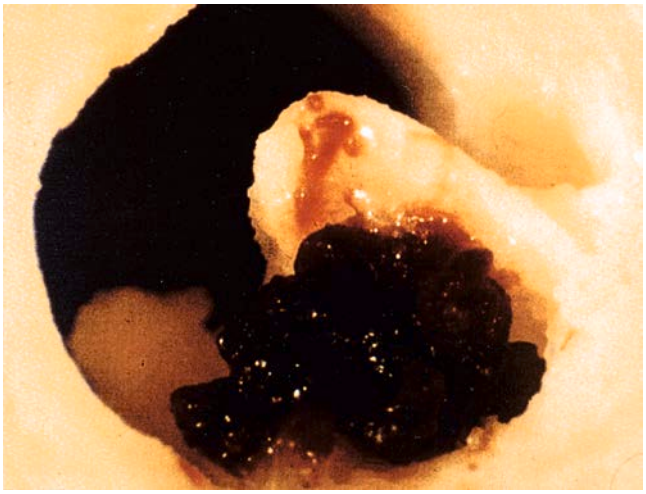


Fig. 17.57 Acute coronary thrombus. Cross-section (×30) of the epicardial coronary artery, demonstrating a rupture of the shoulder region of the plaque with a luminal thrombus.

Box 17.26 Risk factors for coronary disease

Fixed
- Age
- Male sex
- Positive family history
- Deletion polymorphism in the angiotensin-converting enzyme *(ACE)* gene (DD)

Potentially changeable
- Hyperlipidaemia
- Cigarette smoking
- Hypertension
- Diabetes mellitus
- Sedentary lifestyle
- Blood coagulation factors—high fibrinogen, factor VII
- Elevated C-reactive protein
- Homocysteinaemia
- Obesity
- Gout
- Soft water
- Drugs, e.g., contraceptive pill, nucleoside analogues, cyclo-oxygenase 2 (COX-2) inhibitors, rosiglitazone
- Heavy alcohol consumption

Lipids

A full lipid profile should be obtained, including total cholesterol, high-density lipoprotein (HDL) cholesterol, non-HDL cholesterol, and triglyceride concentrations. Total cholesterol concentration of more than 7.5 mmol/L and a family history of premature coronary heart disease may have ***familial hypercholesterolaemia***; in the UK those with a QRISK®3 score of more than 10% should be offered statin therapy regardless of their serum cholesterol (see Chapter 19, 'Risk prediction tables' section).

Lifestyle modifications

It is recommended that patients should eat a healthy, balanced diet with at least five portions of fruit and vegetables per day, ideally varying the type of fruit that is eaten, along with unsalted nuts, seeds, and legumes. Processed and ultra-processed foods should be minimized, as they contain fewer nutrients and fibre but more energy than whole foods. With each 10% increase in daily intake of ultra-processed food, there is a 6% increase in heart disease risk. Similarly, foods high in sugar or containing refined sugars such as fructose should be avoided, as should sugar-sweetened beverages, which contain a significant number of calories but are of limited nutritional benefit. Animal-based foods, particularly red meat and processed meats, are linked with a higher risk of cardiovascular disease, whereas fish intake is associated with a lower risk. At least two portions of fish weekly are recommended, one of which should be oily (e.g., sardines, salmon). Food containing trans fatty acids (TFAs) should be avoided and those with saturated fatty acids (SFAs) should preferably be replaced by polyunsaturated fatty acids (rapeseed and olive oils) to lower CVD risk.

Eating habits, however, are difficult to influence, and change via transformation of the whole food system, from production to consumption, is required with government legislation. For example, sugar taxes have now been introduced by various countries and have been shown to have a meaningful impact.

Regular exercise reduces blood pressure and has favourable effects on lipid profile. It has been shown to reduce cardiovascular events and death. The weekly exercise recommendations are 150 minutes or more of moderate-intensity aerobic activity or 75 minutes of vigorous-intensity aerobic activity. Weight/body mass index (BMI) should be less than 25 kg/m^2 (see Chapter 37).

In the UK the recommendation is not to drink more than 14 units of alcohol per week (men and women), ideally spread evenly over 3 or more days, and smoking should be discontinued. Blood pressure and diabetes should be managed according to guidelines (see Chapter 18).

Statin treatment

NICE guidance recommends atorvastatin 20 mg for the primary prevention of cardiovascular disease in people with a 10-year risk of cardiovascular disease of 10% or higher if lifestyle modification is ineffective or inappropriate. In patients with established cardiovascular disease atorvastatin 80 mg is recommended, unless there are potential drug interactions, a high risk of adverse effects, or a different patient preference.

ANGINA

Myocardial ischaemia in patients with stable CAD occurs when there is a mismatch between blood supply and metabolic demand. This results in cardiac ischaemic pain or angina and ST–T changes on the 12-lead ECG and may cause regional wall motion abnormalities on echocardiography. Ischaemic metabolites, including adenosine, stimulate nerve endings and produce pain.

Epidemiology

The prevalence of angina increases with age in both sexes: in women it is 5% to 7% at 45 to 64 years as opposed to 10% to 12% at 65 to 84 years, while in men it is 4% to 7% at 45 to 64 years as opposed to 12% to 14% at 65 to 84 years. The annual mortality rate ranges from 1.2% to 2.4% and there is an annual incidence of cardiac death of 0.6% to 1.4%.

Diagnosis

The diagnosis of angina is largely based on the clinical history.

- ***Classical angina*** or ***typical angina*** is characterized by chest pain, which is usually predictable and reproducible:
  - 'Heavy', 'tight', 'constricting', or 'gripping' central or retrosternal discomfort that may radiate to the jaw and/or arms.
  - Pain occurs with exercise or emotional stress or may be exacerbated by the cold.
  - Pain eases rapidly with rest or with GTN.
- ***Atypical angina*** is described as chest pain with 2 out of 3 of the features above.
- ***Non-angina chest pain*** is described as chest pain with 1 out of 3 of the features above. Contemporary data now suggest that only 15% of patients suspected of having CAD present with typical symptoms.
- ***Stable angina*** can be classified according to the Canadian Cardiovascular Society guidelines (**Box 17.27**).
- ***Unstable angina*** refers to angina of recent onset (<24 h) or deterioration in previous stable angina, with symptoms frequently occurring at rest. The utilization of high-sensitivity troponins has transformed the diagnosis of patients who previously would have been labelled as having UA to having an NSTEMI in most cases.
- ***Refractory angina*** refers to patients with severe coronary disease in whom revascularization is not possible and angina is not controlled by medical therapy.
- ***Vasospastic*** or ***variant (Prinzmetal's) angina*** refers to angina that occurs without provocation, usually at rest, as a result of coronary artery spasm. It occurs more frequently in women. Characteristically, there is ST-segment elevation on the ECG during the pain.
- ***Microvascular angina*** patients have exercise-induced angina but normal or unobstructed coronary arteries (on coronary angiography, CTCA). Intracoronary acetylcholine may cause coronary spasm. While they have a good prognosis, they are often highly symptomatic and can be difficult to treat. In women with this syndrome the myocardium shows an abnormal metabolic response to stress, consistent with the suggestion that the myocardial ischaemia results from abnormal dilator responses of the coronary microvasculature to stress.

Examination

There are usually no abnormal findings in angina. Signs to suggest anaemia, thyrotoxicosis, or hyperlipidaemia (e.g., lipid arcus,

Box 17.27 Canadian Cardiovascular Society functional classification of angina

Class	Description
Class I	No angina with ordinary activity; angina with strenuous activity
Class II	Angina during ordinary activity, e.g., walking up hills, walking rapidly upstairs, with mild limitation of activities
Class III	Angina with low levels of activity, e.g., walking 50–100 m on the flat, walking up one flight of stairs, with marked restriction of activities
Class IV	Angina at rest or with any level of exercise

xanthelasma, tendon xanthoma) should be sought. It is essential to exclude aortic stenosis (i.e., slow-rising carotid impulse and ejection systolic murmur radiating to the neck) as a possible cause of the angina. The blood pressure should be taken to identify coexistent hypertension. BMI or waist circumference should be measured.

Investigations

Patients with suspected angina should have initial investigations (Box 17.28). The diagnosis of stable angina can be made on clinical assessment alone *or* by clinical assessment combined with anatomical (cardiac catheterization or CTCA) or functional imaging (SPECT, stress echocardiography, stress MRI).

Patients *without known CAD* presenting with chest pain can be categorized as having typical angina, atypical angina, or non-anginal chest pain.

- Patients with typical or atypical angina or patients with non-anginal chest pain but with ST changes or Q waves should be referred for CTCA.
- If the results from CTCA are inconclusive, the patient should be referred for non-invasive functional tests (SPECT, stress echocardiography, stress MRI).
- Patients with ***non-anginal chest pain*** (e.g., if the pain is continuous, unrelated to exertion, exacerbated by respiration, or associated with dizziness, palpitations, or difficulty in swallowing), where the ECG and cardiac biomarkers are normal, should be considered for alternative diagnoses and investigated appropriately.
- It is important to note that up to 30% of ACS presentations can have very atypical or non-anginal symptoms, particularly in women.

Box 17.28 Initial investigations in stable angina

Laboratory tests
- Full blood count
- Thyroid function tests
- Fasting glucose
- HbA_{1c}
- Fasting lipid profile
- Glomerular filtration rate
- Troponin if unstable

12-Lead ECG
- Exclude coronary syndrome, pathological Q waves, left ventricular hypertrophy, left bundle branch block

Echocardiography
- Regional wall motion abnormalities
- Left ventricular ejection fraction
- Diastolic function
- Alternative causes of chest pain

Ambulatory ECG
- Paroxysmal arrhythmia
- Vasospastic angina

Chest X-ray
- Atypical presentation
- Pulmonary disease
- Heart failure

Management of stable angina

Patients should be informed of the nature of their condition and reassured that the prognosis is good (annual mortality <2%). Lifestyle management should be instigated, as for prevention of CAD. The stable angina algorithm in Fig. 17.58 should be used to guide initial patient management, while Box 17.29 outlines pharmacological therapy.

Revascularization

Revascularization with PCI or coronary artery bypass surgery is indicated for patients with on-going angina symptoms despite appropriate medical therapy and/or to improve prognosis.

Percutaneous coronary intervention

PCI is the process of dilating a coronary artery stenosis by introducing an inflatable balloon and metallic stent (Fig. 17.59) into the arterial circulation via the femoral, radial, or brachial artery (Fig. 17.60); the radial artery is the best (see later).

- Fractional flow reserve (FFR) is an invasive intracoronary measurement that can be used to assess the haemodynamic significance of a coronary stenosis prior to PCI using intracoronary adenosine. A normal FFR value is 1, and values less than 0.80 are associated with ischaemia, while if greater than 0.8, medical treatment (i.e., no stent) is associated with improved patient outcomes.
- PCI can also be guided by IVUS and optical coherence tomography (OCT) to assess coronary lesions both before and after angioplasty.

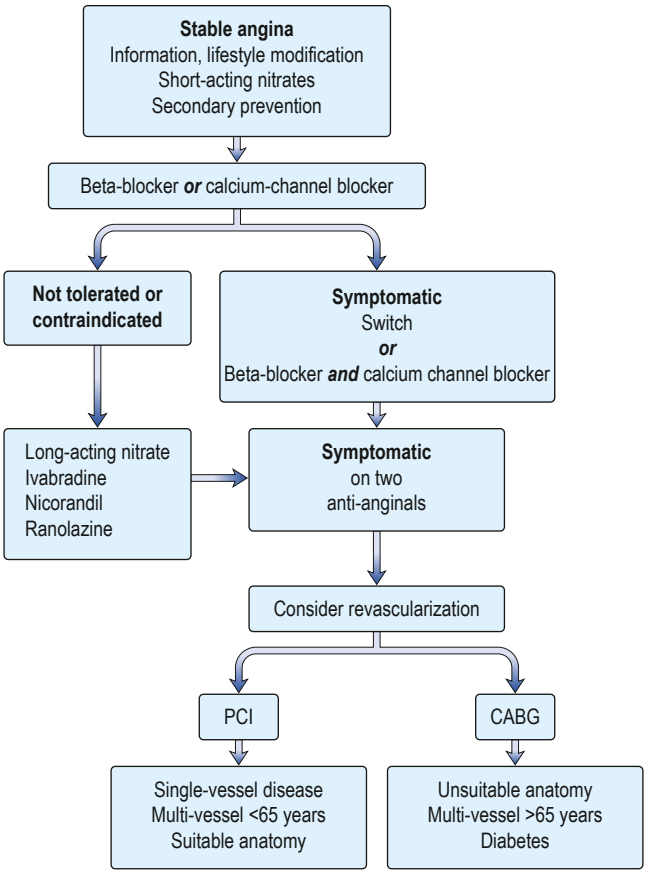


Fig. 17.58 Algorithm for treatment of patients with stable angina. *CABG,* Coronary artery bypass grafting; *PCI,* percutaneous coronary intervention. *(From National Institute for Health and Care Excellence. Clinical Guideline CG126L: Management of Stable Angina. NICE July 2011, with permission.)*

- Adjunct tools that can be used in PCI to treat highly calcified lesions include cutting balloon, laser atherectomy, rotational atherectomy, and intravascular lithotripsy.

Complications of PCI include bleeding, haematoma, dissection, and pseudo-aneurysm from the arterial puncture site, although use of the radial artery reduces these risks and is the standard technique in the UK. Serious complications include acute MI (2%), stroke (0.4%), and death (1%).

Bare metal stents (BMSs) are associated with a 20% to 30% risk of coronary artery restenosis within 6 to 9 months of

Box 17.29 Pharmacological therapy in stable angina

Drug	Dose	Indications/mechanism of action/cautions
Vasodilators		
Glyceryl trinitrate	0.3–1.0 mg sublingual 2.0–3.0 mg buccal	Prophylaxis and treatment of angina—rapid onset Repeat after 5 min if symptomatic Vasodilation (causes headache and flushing)
Isosorbide mononitrate	10–60 mg ×2 daily (slow-release preparations available)	Prophylaxis of angina Side-effects—headache and flushing Contraindicated with phosphodiesterase type 5 inhibitors
Beta-blocker		
Bisoprolol	2.5–10 mg daily	Inhibits β-adrenoceptors, reduces heart rate and BP, reduces myocardial oxygen consumption Caution—COPD, acute heart failure, AV conduction Side-effects—fatigue, peripheral vasoconstriction (cold peripheries), sexual dysfunction, bronchospasm
Calcium-channel blockers		
Verapamil (phenylalkylamines)	80–120 mg ×3 daily (or 240–480 mg daily slow-release)	Inhibit calcium channels in myocardium, cardiac conductive tissue, and vascular smooth muscle Diltiazem and verapamil—contraindicated in severe bradycardia, left ventricular failure with pulmonary congestion, second- or third-degree AV block Side-effects—constipation (verapamil), ankle oedema (amlodipine, diltiazem), reflex tachycardia (amlodipine)
Diltiazem (benzothiazepines)	60–120 mg ×3 daily (longer-acting preparations available)	
Amlodipine (dihydropyridines)	5–10 mg daily	
Other anti-anginal drugs		
Ivabradine	2.5–7.5 mg ×2 daily	Inhibits pacemaker I_f current in SA node Use in sinus rhythm ± beta-blocker Side-effects—bradycardia, phosphenes Contraindications—sick sinus syndrome, AV block
Nicorandil	5–30 mg ×2 daily	Activates ATP-sensitive potassium channels and has nitrate properties—peripheral and coronary vasodilation Use as adjunctive therapy Side-effects—headache, flushing, oral ulceration
Ranolazine	375–750 mg ×2 daily	Inhibits late sodium channels into cardiac cells Use as adjunctive therapy Metabolized by cytochrome P450 3A4 Side-effects—constipation, dizziness, lengthened QT
Event-reducing drugs		
Antiplatelet	Aspirin 75–100 mg daily	Reversible inhibition of platelet COX-1 and thromboxane production
	Clopidogrel 75 mg daily	Thienopyridine that antagonizes platelet ADP receptor P2Y12. Use in aspirin intolerance
ACE inhibitor or ARB	Enalapril 10 mg daily	Indicated if treating other condition, e.g., hypertension, heart failure, chronic kidney disease
Statins	Atorvastatin 80 mg daily	Use to reduce LDL cholesterol to <1.8 mmol/L

ACE, Angiotensin-converting enzyme; *ADP,* adenosine diphosphate; *ARB,* angiotensin receptor blocker; *ATP,* adenosine triphosphate; *AV,* atrioventricular; *BP,* blood pressure; *COPD,* chronic obstructive pulmonary disease; *COX-1,* cyclo-oxygenase-1; *LDL,* low-density lipoprotein; *SA,* sinoatrial.

implantation and are now infrequently used. First-generation ***drug-eluting stents*** (DESs) with ***sirolimus*** and ***paclitaxel*** reduced restenosis but suffered from late and very late stent thrombosis. Second-generation DESs (with thinner struts and biodegradable or biocompatible polymers) demonstrated better safety and efficacy and are recommended for most patients. BMSs may be preferred in patients requiring early non-cardiac surgery or who have active bleeding at the time of PCI or significant future bleeding risk on antiplatelet therapy. An alternative is the use of drug-eluting balloons (DEBs) which can be used to expand lesions while delivering antiproliferative agents locally, suppressing neointimal hyperplasia. These can be used in smaller coronary vessels which may not be amenable to stenting and also lessen the duration of dual antiplatelet therapy, which is particularly helpful in patients predisposed to bleeding. Dual antiplatelet therapy (aspirin and a P2Y12 inhibitor, e.g., clopidogrel) should continue for 6 to 12 months, and most patients will then continue on a single antiplatelet agent life-long.

Coronary artery bypass grafting

With CABG, autologous arteries or veins are anastomosed to the ascending aorta and to the native coronary arteries distal to the area of stenosis (Fig. 17.61). Arterial grafts (with *in situ* internal mammary or transposed radial arteries grafted onto the stenosed coronary arteries) demonstrate improved long-term patency compared to vein grafts. Operative mortality is ~1% in patients with normal left ventricular function but increases with age and co-morbidities.

PCI versus CABG

A multidisciplinary team (made up of interventional and non-interventional cardiologists, cardiac surgeons, cardiac anaesthetists, and intensivists) should review whether patients should

continue with medical therapy or be referred for PCI or CABG. The STS score and/or EuroSCORE II (see Further reading) can be utilized to determine in-hospital mortality and morbidity in patients referred for surgery. The SYNTAX score should be used for complex CAD to determine long-term mortality and morbidity in patients referred for PCI. In the current European Society of Cardiology (ESC) guidelines for revascularization (see Further reading), PCI is preferred to CABG in patients with single- or double-vessel disease not involving the proximal LAD or left main stem (LMS) vessel. CABG and PCI are both appropriate in patients with proximal LAD stenosis, LMS, or three-vessel disease (without diabetes mellitus) and a low SYNTAX score (0–22). CABG is preferred to PCI in patients with three-vessel disease and diabetes or elevated SYNTAX score

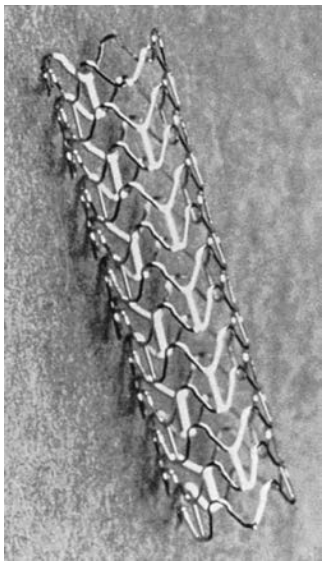


Fig. 17.59 **An intracoronary stent.**

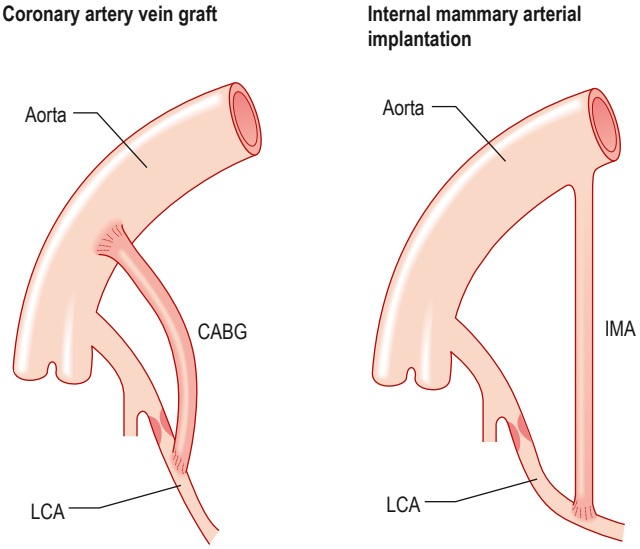


Fig. 17.61 **Relief of coronary obstruction by surgical techniques: coronary artery bypass grafting (CABG) and internal mammary arterial implantation (IMA).** In both of these examples the graft bypasses a coronary obstruction in the left coronary artery (LCA).

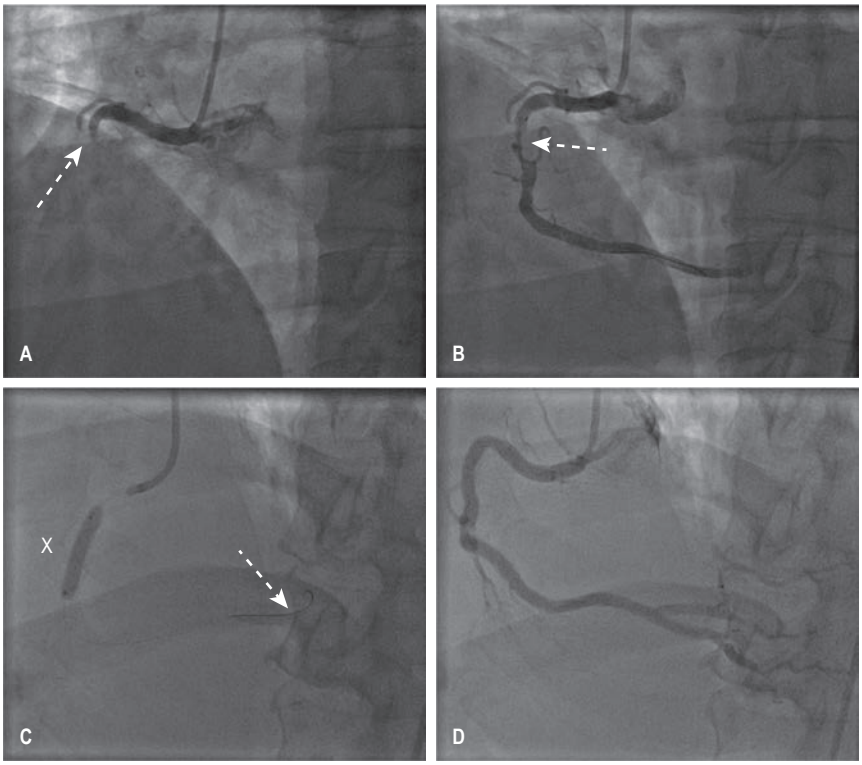


Fig. 17.60 **Percutaneous transluminal coronary angioplasty (PTCA).** **(A)** Coronary angiography demonstrates an occluded right coronary artery *(arrowed)*. **(B)** A soft wire passed through the guide catheter reopens the artery but a severe stenosis remains *(arrowed)*. **(C)** A balloon (X) is inflated to dilate the stenosis. The soft guidewire can be seen in the distal posterior descending artery *(arrowed)*. **(D)** The right coronary artery has now been successfully reopened with good antegrade flow.

(>22) and in patients with LMS and SYNTAX score of more than 22. Some patients may undergo a hybrid procedure with a combination of minimally invasive bypass surgery with an arterial graft and PCI. Patient clinical characteristics (severe co-morbidities, reduced life expectancy, 'porcelain' [heavily calcified] aorta, or the need for additional surgery to the aorta or valves) may alter the modality of revascularization.

Patients with refractory angina

Some patients remain symptomatic despite medication and are not suitable for (further) revascularization. These individuals need ongoing lifestyle advice and a pain management programme. Spinal cord stimulation and coronary sinus reducing devices show some promise.

Further reading

Knuuri J, Wijns W, Saraste A, et al. 2019 ESC Guidelines for the diagnosis and management of chronic coronary syndromes. *Eur Heart J* 2020; 41:407–477.
Neumann FJ, Sousa-Uva M, Ahlsson A et al. 2018 ESC/EACTS Guidelines on myocardial revascularization. *Eur Heart J* 2019; 40:87–165.
Xaplanteris P, Fournier S, Pijls, NHJ et al. Five-Year Outcomes with PCI Guided by Fractional Flow Reserve. *N Engl J Med* 2018; 379:250–259.
http://riskcalc.sts.org/stswebriskcalc/calculate. *Society of Thoracic Surgeons (STS) score*.
http://www.euroscore.org/calc.html. *EuroSCORE II*.
http://www.syntaxscore.com/calculator/start.htm. *SYNTAX score*.

ACUTE CORONARY SYNDROMES

Acute coronary syndromes (ACSs) comprise a spectrum of clinical presentations ranging from UA to non-ST-elevation myocardial infarction (NSTEMI) and ST-elevation myocardial infarction (STEMI). Each of these underlying conditions typically present with similar symptoms and signs (most commonly with sudden-onset 'heavy' or 'tight' feeling chest pain), and are managed similarly until subsequent tests (ECG, cardiac biomarkers and different imaging modalities) provide the ultimate diagnosis.

Pathophysiology

ACSs result from an imbalance between myocardial oxygen supply and demand, primarily due to a sudden reduction in coronary blood flow. This is commonly caused by the rupture or erosion of an atherosclerotic plaque, leading to thrombus formation and arterial occlusion. The extent of myocardial ischaemia and necrosis depends on the location and duration of the obstruction, the adequacy of collateral blood flow around the occluded vessel, and myocardial oxygen demand.

Partial occlusion or transient complete occlusion can result in NSTEMI or UA, where the ischaemia is typically subendocardial with less myocardial damage.

- UA represents myocardial ischaemia in the absence of myocardial damage. Patients may have prolonged (>20 mins) angina at rest, new onset severe angina, or rapidly worsening angina (increased frequency and/or longer duration).
- In patients with NSTEMI there is occluding thrombus, which leads to myocardial necrosis and a rise in serum troponin levels. The use of high-sensitivity troponins (see 'Biomarkers' section) has made the distinction between UA and NSTEMI more challenging, as minor troponin elevations can now be detected.
- STEMI (see section below) occurs in the setting of complete coronary artery occlusion and is characterized by transmural ischaemia and extensive myocardial injury.

The pathophysiological process of ACS is dynamic, and the clinical presentation can evolve, underscoring the importance of early recognition and intervention to restore coronary perfusion and minimize myocardial damage.

Diagnosis

A rise and/or fall in troponin levels above the 99th percentile of healthy individuals suggests a diagnosis of MI, based on the criteria in the fourth universal definition of MI developed jointly by European and American cardiological associations, provided that the clinical presentation is consistent with myocardial ischaemia. There are five different types, each with distinct pathological, clinical, and prognostic features:

- **Type 1 MI:** this is the traditional clinical picture, characterized by atherosclerotic plaque disruption complicated by thrombosis. It involves reduced blood supply to the myocardium due to coronary atherothrombotic obstruction to blood flow.
- **Type 2 MI:** there is reduced oxygen supply or increased demand secondary to other causes unrelated to acute coronary atherothrombosis, for example, coronary artery spasm, coronary embolism, anaemia, arrhythmias, hypertension, or hypotension.
- **Type 3 MI:** refers to a cardiac event resulting in death before measurement of biomarkers.
- **Type 4 MI:** increase in troponin in the setting of coronary angioplasty (PCI) (4a), stent thrombosis (4b), or in-stent restenosis (4c).
- **Type 4b MI:** increase in troponin in the setting of coronary angioplasty (PCI).
- **Type 5 MI:** increase in troponin related to coronary artery bypass grafting (CABG)

Clinical features

The primary symptom of ACS is substernal chest pain, often described as a crushing or pressure-like feeling, which may radiate to the jaw and/or left arm. This pain may come on quickly, come and go, or get worse with exercise. Unlike stable angina, it is not usually predictable.

Other symptoms can include dyspnoea, diaphoresis, nausea or vomiting, and unexplained fatigue. Some patients may also experience light-headedness, syncope, or a racing heartbeat. Atypical presentations are more common in women, elderly patients, and those with diabetes, who may experience more vague and subtle symptoms such as isolated jaw or left arm pain, epigastric pain, and weakness. These patients are often misdiagnosed and undertreated, and they have a higher mortality across the spectrum of ACS. Physical examination may note general distress and diaphoresis. Heart sounds are frequently normal. It is important to identify adverse clinical signs, such as hypotension, basal crackles, fourth heart sounds, and cardiac murmurs. A new pansystolic murmur may represent mitral regurgitation from papillary muscle dysfunction or rupture, or a VSD from septal rupture.

Investigations

Electrocardiogram

A resting 12-lead ECG is the initial diagnostic tool in assessing patients with suspected ACS, and it should be repeated, as required, if the patient has chest pain. In UA and NSTEMI there may

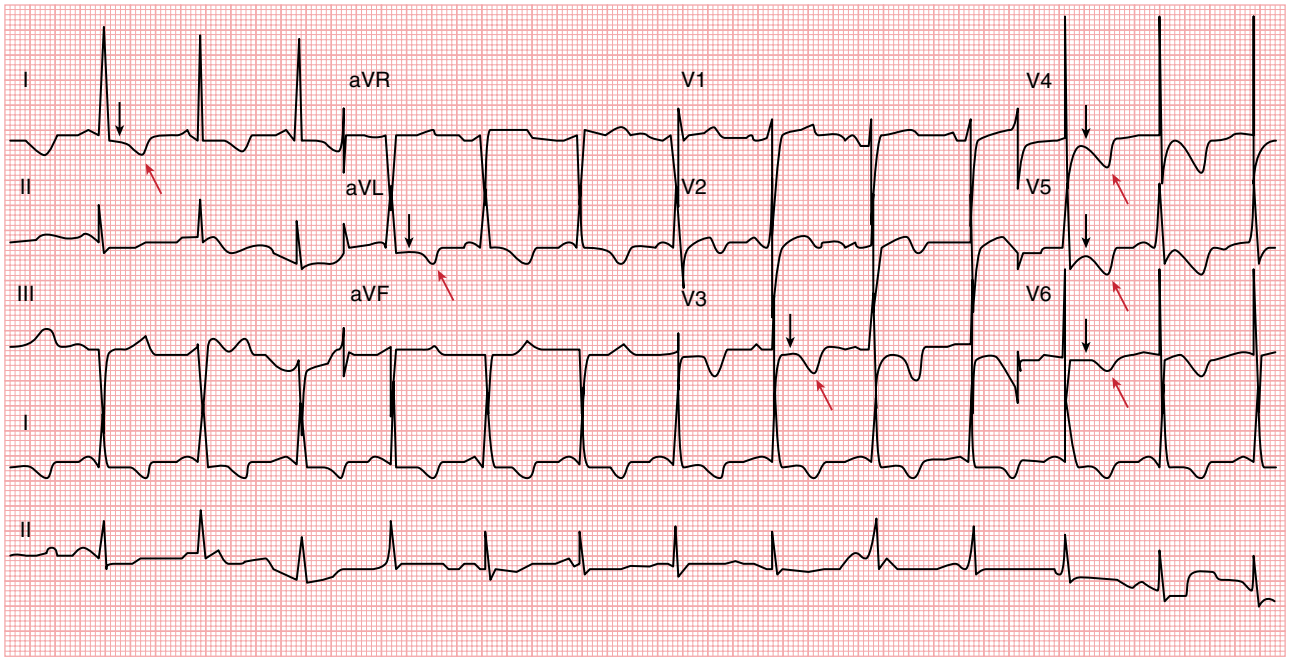


Fig. 17.62 NSTEMI with ST depression *(black arrows)* and T-wave inversion *(red arrows)* in anterior and lateral leads (V_3–V_6, aVL, I). *(Harding, M., et al. (2019). Coronary artery disease and acute coronary syndrome. In:* Lewis's Medical-Surgical Nursing: Assessment and Management of Clinical Problems, *11th ed. Elsevier.)*

be ST depression and/or T-wave inversion due to subendocardial ischaemia, occurring with a particular vascular territory (see Fig. 17.62), although about one-third of patients may have a normal ECG. Dynamic ECG changes, such as intermittent ST-segment elevations or depressions and T-wave inversions, are highly suggestive of ongoing ischaemia. In STEMI the hallmark ECG finding is persistent ST-segment elevation, typically in the leads facing the ischaemic zone, as a result of transmural ischaemia. Additionally, new onset of LBBB is considered a STEMI equivalent and may be indicative of acute coronary artery occlusion. Transient ST elevation can be seen in coronary vasospasm (Prinzmetal's angina), pericarditis, and aortic dissection involving the coronary arteries.

Biomarkers

Cardiac troponins are the cornerstone biomarkers for the diagnosis of ACS as they are highly specific and sensitive indicators of myocardial injury. Previous biomarkers (creatine kinase-MB [CK-MB], LDH, and aspartate transaminase [AST]) were once widely used, but due to their lower specificity and sensitivity for cardiac injury, they have been supplanted by troponins.

The cardiac troponin complex is composed of three subunits: troponin T (cTnT), troponin I (cTnI), and troponin C. Troponin T and I are cardiac specific, and their presence in the bloodstream is indicative of myocardial damage.

- Troponin T attaches the troponin complex to the actin filament.
- Troponin I inhibits the interaction with myosin heads in the absence of sufficient calcium ions.
- Troponin C acts as the calcium-binding site during excitation–contraction coupling.

High-sensitivity troponin assays have improved the early detection of myocardial injury, with elevations detectable within 60 minutes of an indexed cardiac event. The NPV for MI in those with normal troponin concentrations exceeds 99%. This high NPV indicates that a 'rule out' pathway is highly effective in correctly identifying patients who do not have an MI, thus allowing for safe discharge and reducing unnecessary hospital admissions and further testing. However, such an approach increases the likelihood of detecting minor troponin elevations that may not be related to acute ischaemic heart disease. If the initial troponin assay is negative, it should be repeated 1 to 3 hours later, with rising concentrations suggesting myocardial ischaemia.

The troponin assay provides prognostic information: that is, a high serum troponin level carries an increased mortality risk in ACS (Box 17.30). It also defines which patients may benefit from aggressive medical therapy and early coronary revascularization.

NSTEMI and unstable angina

Risk stratification

Both the Thrombolysis in Myocardial Infarction (TIMI) score (Box 17.31) and the Global Registry of Acute Coronary Events (GRACE) prediction score can be used in patients with ACS to define risk. The GRACE score is based on age, heart rate, systolic blood pressure, serum creatinine, and the Killip score (Boxes 17.32 and 17.33).

Investigations and management

All patients require immediate management of their chest pain, as outlined in Chapter 11 ('Chest pain' section) and in Box 17.34.

Antiplatelet drugs

Platelets play a crucial role in the thrombosis cascade associated with ACS. When an atheromatous plaque ruptures, circulating platelets are exposed to various substances such as ADP, thromboxane A2 (TxA2), adrenaline (epinephrine), thrombin, and collagen tissue factor. This exposure triggers platelet activation, with thrombin being a particularly potent stimulant. Activated platelets express glycoprotein (GP) IIb/IIIa receptors on their surface, which facilitate the bridging of fibrinogen between adjacent platelets, leading to the formation of platelet aggregates (see Fig. 38.33).

ACS patients should be treated with dual antiplatelet agents to inhibit different pathways of platelet activation and aggregation, to significantly reduce the risk of recurrent ischaemic events:

- ***aspirin 300 mg*** loading dose, then 75 to 100 mg daily *and*
- ADP-receptor $P2Y_{12}$ inhibitor (choice depends on bleeding risk and local guidelines):
  - **clopidogrel 300 to 600 mg** loading then 75 mg daily *or*
  - **prasugrel 60 mg** loading then 10 mg daily *or*
  - **ticagrelor 180 mg** loading then 90 mg twice daily.

Anticoagulants

Anticoagulant drugs inhibit thrombin, a key enzyme involved in both platelet and coagulation system activation, and should be prescribed with dual antiplatelets in ACS. Parenteral anticoagulants are essential to mitigate the risk of thrombus propagation and new thrombus formation. The choice of anticoagulant is influenced by the patient's characteristics and clinical setting:

- ***Low-molecular-weight heparin (LMWH),*** for example, enoxaparin—preferentially inhibits factor Xa while having a lesser effect on thrombin and has a predictable anticoagulant response, usually not requiring laboratory monitoring.

Box 17.30 Relationship between troponin I and risk of death in acute coronary syndrome

Serum troponin levels (ng/mL)	Mortality at 42 days (% of patients)
0 to <0.4	1.0
0.4 to <1.0	1.7
1.0 to <2.0	3.4
2.0 to <5.0	3.7
5.0 to <9.0	6.0
>9.0	7.5

(Adapted From Antman EM, Tanasijevic MJ, Thompson B et al. Cardiac-specific troponin I levels to predict the risk of mortality in patients with acute coronary syndromes. N Engl J Med *1996; 335:1342–1349.)*

Box 17.31 TIMI risk score in acute coronary syndrome (NSTEMI/UA)

Risk factor	Score
Age >65 years	1
>3 CAD risk factors—hypertension, hyperlipidaemia, family history, diabetes, smoking	1
Known CAD (coronary angiography stenosis >50%)	1
Aspirin use in the last 7 days	1
Severe angina (>2 episodes of rest pain in 24 h)	1
ST deviation on ECG (horizontal ST depression or transient ST elevation >1 mm)	1
Elevated cardiac markers (CK-MB or troponin)	1
Total score	
0–1: rate of death/MI in 14 days 3%; rate of death/MI/urgent revascularization 4.75% 2: 3%; 8.3% 3: 5%; 13.2% 4: 7%; 19.9% 5: 12%; 26.2% 6–7: 19%; 40.9%	

CAD, Coronary artery disease; *CK-MB,* creatine kinase-MB; *MI,* myocardial infarction; *NSTEMI,* non-ST-elevation myocardial infarction; *TIMI,* Thrombolysis in Myocardial Infarction; *UA,* unstable angina.

Box 17.33 Mortality in low-, intermediate-, and high-risk categories according to the GRACE risk score

Syndrome type/ timeframe	Risk category (tertiles)	GRACE risk score	Deaths (%)
NSTEMI-ACS			
In hospital	Low	≤108	<1
	Intermediate	109–140	1–3
	High	>140	>3
Post discharge to 6 months	Low	≤88	<3
	Intermediate	89–118	3–8
	High	>118	>8
STEMI-ACS			
In hospital	Low	49–125	<2
	Intermediate	126–154	2–5
	High	155–319	>5
Post discharge to 6 months	Low	27–99	<3
	Intermediate	100–127	3–8
	High	128–263	>8

ACS, Acute coronary syndrome; *GRACE,* Global Registry of Acute Coronary Events; *NSTEMI,* non-ST-elevation myocardial infarction; *STEMI,* ST-elevation myocardial infarction.
(From http://www.outcomes-umassmed.org/grace/grace_risk_table.aspx.*)*

Box 17.32 Risk calculator for 6-month post-discharge mortality after hospitalization for acute coronary syndrome (GRACE score)

Age		Findings at initial hospital presentation		Findings during hospitalization		History of CCF	History of MI	Systolic BP		Elevated cardiac enzymes	ST-segment depression	No in-hospital PCI
Age (years)	Points	Resting heart rate (b.p.m.)	Points	Initial serum creatinine (mg/dL)	Points	Points	Points	(mmHg)	Points	Points	Points	Points
≤29	0	≤49.9	0	0–0.39	1	24	12	≤79.9	24	15	11	14
30–39	0	50–69.9	3	0.4–0.79	3			80–99.9	22			
40–49	18	70–89.9	9	0.8–1.19	5			100–119.9	18			
50–59	36	90–109.9	14	1.2–1.59	7			120–139.9	14			
60–69	55	110–149.9	23	1.6–1.99	9			140–159.9	10			
70–79	73	150–199.9	35	2–3.99	15			160–199.9	4			
80–89	91	≥200	43	>4	20			≥200	0			
≥90	100											

BP, Blood pressure; *CCF,* congestive cardiac failure; *GRACE,* Global Registry of Acute Coronary Events; *MI,* myocardial infarction; *PCI,* percutaneous coronary intervention.

Box 17.34 Pharmacological therapy in acute coronary syndrome

Drug	Dose	Notes
Antiplatelet drugs		
Aspirin	150–300 mg chewable or soluble aspirin, then 75–100 mg orally daily	Caution if active peptic ulceration
Clopidogrel	300 mg orally loading dose, then 75 mg orally daily	Caution: increased risk of bleeding; avoid if CABG planned
Prasugrel	60 mg oral loading dose, then 10 mg orally daily (5 mg daily if <60 kg or >75 years old)	
Ticagrelor	Initially 180 mg, then 90 mg ×2 daily	
Antithrombin drugs		
Heparin	5000 units i.v. bolus, then 0.25 units/kg per hour	Measure anticoagulant effect with APTT at 6 h
Low-molecular-weight heparins, e.g., enoxaparin	1 mg/kg s.c. ×2 daily	Risk of bleeding
Bivalirudin	0.75 mg/kg i.v. bolus, then 1.75 mg/kg per hour for 4 h post PCI	
Fondaparinux	2.5 mg s.c. daily, for up to 8 days	
Rivaroxaban	Oral 2.5–10 mg daily	
Glycoprotein IIb/IIIa inhibitors[a]		
Abciximab	0.25 mg/kg i.v. bolus, then 0.125 μg/kg per min up to 10 μg/min i.v. ×12 h	Indicated if coronary intervention likely within 24 h
Eptifibatide	180 μg/kg i.v. bolus, then 2 μg/kg per min ×18 h	Indicated in high-risk patients managed without coronary intervention or during PCI
Tirofiban	25 μg/kg over 5 min, then 0.15 μg/kg per min for up to 18 h	Indicated in high-risk patients managed without coronary intervention or during PCI
Analgesia		
Diamorphine or morphine	2.5–5.0 mg i.v.	Prescribe with antiemetic, e.g., metoclopramide 10 mg i.v.
Myocardial energy consumption		
Atenolol	5 mg i.v. repeated after 15 min, then 25–50 mg orally daily	Avoid in asthma, heart failure, hypotension, bradyarrhythmias
Metoprolol	5 mg i.v. repeated to a maximum of 15 mg, then 25–50 mg orally ×2 daily	
Coronary vasodilation		
Glyceryl trinitrate	2–10 mg/h i.v./buccal/sublingual	Maintain systolic BP >90 mmHg
Plaque stabilization/ventricular remodelling		
HMG-CoA reductase inhibitors (statins)		
Rosuvastatin	20–40 mg orally	Combine with dietary advice and modification
Pravastatin	20–40 mg orally	
Atorvastatin	80 mg orally	
ACE inhibitors		
Ramipril	2.5–10 mg orally	Monitor renal function
Lisinopril	5–10 mg orally	

[a]Not now used in patients pre-treated with clopidogrel and aspirin prior to coronary intervention.
ABG, Arterial blood gases; *ACE*, angiotensin-converting enzyme; *APTT*, activated partial thromboplastin time; *BP*, blood pressure; *CABG*, coronary artery bypass graft; *COPD*, chronic obstructive pulmonary disease; *i.v.*, intravenous; *PCI*, percutaneous coronary intervention; *s.c.*, subcutaneous.

- ***Unfractionated heparin (UFH):*** a fast-acting anticoagulant with a short half-life, which can be advantageous for potential rapid reversal. It acts by binding to antithrombin and potentiating its effect, primarily inhibiting thrombin and factor Xa. UFH requires monitoring of the activated partial thromboplastin time (aPTT) to ensure therapeutic anticoagulation.
- ***Fondaparinux:*** fondaparinux is a synthetic pentasaccharide that selectively inhibits factor Xa. It has a longer half-life and does not inhibit thrombin directly. Fondaparinux is associated with a lower risk of heparin-induced thrombocytopenia (HIT) compared to heparin.
- Bivalirudin: as a direct thrombin inhibitor, bivalirudin directly inhibits both circulating and clot-bound thrombin. It has a rapid onset and offset of action and may be preferred in patients at high risk of bleeding or those with HIT.

Anti-ischaemia agents

In patients with no contraindications (asthma, AV block, acute pulmonary oedema) beta-blockers are administered orally, to reduce myocardial ischaemia by blocking circulating catecholamines. This will lower the heart rate and blood pressure, reducing myocardial

oxygen consumption. The dose can be titrated to produce a resting heart rate of ~60 b.p.m.

In patients with on-going angina, nitrates should be given either sublingually or intravenously. They effectively reduce preload and produce coronary vasodilation. However, tolerance can become a problem and patients should be weaned off intravenous administration if symptoms resolve.

Plaque stabilization/remodelling

HMG-CoA reductase inhibitor drugs (statins) and ACE inhibitors are routinely administered to patients with ACS. These agents may produce plaque stabilization, improve vascular and myocardial remodelling, and reduce future cardiovascular events. Starting the drugs while the patient is still in hospital increases the likelihood of these individuals receiving long-term therapy.

Coronary angiography and intervention

Very high-risk patients require urgent coronary angiography (<2 h). This includes those individuals with persistent or recurrent chest pain not responding to medical therapy, clinical signs of heart failure, haemodynamic instability or cardiogenic shock, or life-threatening arrhythmias (ventricular fibrillation, ventricular tachycardia).

High-risk patients with rising or falling cardiac troponin levels, dynamic ST- or T-wave changes, or elevated GRACE scores (>140) require coronary angiography within 24 hours.

Intermediate-risk patients with diabetes mellitus, renal impairment (estimated glomerular filtration rate <60), LVEF below 40%, congestive cardiac failure, early post-infarction angina, previous PCI or CABG, or a GRACE score of over 109 but less than 140 require coronary angiography within 72 hours.

Low-risk patients can be managed initially with medical therapy, but an invasive strategy with cardiac catheterization is preferred in most patients.

ST-elevation myocardial infarction

MI occurs when cardiac myocytes die due to prolonged myocardial ischaemia. In Europe approximately 1.8 million people die annually from ischaemic heart disease.

Pathophysiology

Rupture or erosion of a vulnerable coronary artery plaque can produce prolonged occlusion of a coronary artery, leading to myocardial necrosis within 15 to 30 minutes. The subendocardial myocardium is initially affected but, with continued ischaemia, the infarct zone extends through to the subepicardial myocardium, producing a transmural 'Q wave' MI. Early reperfusion may salvage regions of the myocardium, reducing future mortality and morbidity.

The in-hospital mortality rate is between 6% and 14%. Several risk factors can be identified that predict death rate at 30 days (GRACE ACS score, **Box 17.35**).

Clinical features

Any patient presenting with severe chest pain lasting more than 20 minutes may be suffering from MI. The pain may not respond to sublingual glyceryl trinitrate. The pain may radiate to the left arm, neck, or jaw. However, many patients present atypically (particularly females, the elderly, those with diabetes, or those from ethnic minorities), and symptoms in these groups may include dyspnoea, fatigue, pre-syncope, or syncope. Autonomic symptoms are common, and on examination, the patient may be pale and clammy, with marked sweating. In addition, the pulse may be thready with significant hypotension, bradycardia, or tachycardia.

Investigations

Electrocardiography

An ECG in patients with chest pain should be performed within 10 minutes of first contact with the emergency medical team (ambulance or emergency department). The baseline ECG is rarely normal, but if it is, it should be repeated every 15 minutes while the patient remains in pain. Continuous cardiac monitoring is required because of the high likelihood of significant cardiac arrhythmias. ECG changes (**Box 17.36**) are usually confined to the ECG leads that 'face' the infarction. The presence of new ST elevation (due to opening of the K^+ channels) of 0.2 mV or more at the J-point in leads V_1–V_3 and 0.1 mV or more in other

Box 17.35 TIMI risk score in ST-elevation myocardial infarction (STEMI)

Risk factor	Score
Age >65 years	2
Age >75 years	3
History of angina	1
History of hypertension	1
History of diabetes	1
Systolic BP < 100 mmHg	3
Heart rate >100 b.p.m.	2
Killip score II–IV	2
Weight >67 kg	1
Anterior MI or LBBB	1
Delay to treatment >4 h	1
Total score	
0: risk of death at 30 days 0.8% 1: 1.6% 2: 2.2% 3: 4.4% 4: 7.3% 5: 12.4% 6: 16.1% 7: 23.4% 8: 26.8% 9–16: 35.9%	

BP, Blood pressure; *LBBB,* left bundle branch block; *MI,* myocardial infarction; *TIMI,* thrombolysis in myocardial infarction.

Box 17.36 Typical ECG changes in myocardial infarction (STEMI)

Infarct site	Leads showing ST elevation
Anterior	
Small	V_3–V_4
Extensive	V_2–V_5
Anteroseptal	V_1–V_3
Anterolateral	V_4–V_6, I, AVL
Lateral	I, AVL
Inferior	II, III, AVF
Posterior	V_1, V_2 reciprocal changes, i.e., ST depression
Subendocardial	Any lead
Right ventricle	VR_4

leads suggests anterior MI (Fig. 17.63). An inferior wall MI is diagnosed when ST elevation is seen in leads II, III, and AVF (Fig. 17.64). Lateral MI produces changes in leads I, AVL, and V_5/V_6. In patients with a posterior MI there may be ST depression in leads V_1–V_3 with a dominant R wave and ST elevation in lead V_5/V_6. New (or presumed new) LBBB is compatible with coronary artery occlusion requiring urgent reperfusion therapy. It may not be possible to exclude MI in patients with a paced rhythm or right bundle branch block, and patients with symptoms compatible with MI should be triaged in the MI pathway. The evolution of the ECG during the course of STEMI is illustrated in Fig. 17.65.

Blood tests and other imaging

Blood samples should be taken for cardiac troponin I or T levels, although treatment should not be deferred until the results are available. Full blood count, serum electrolytes, glucose, and lipid profile should be obtained. Transthoracic echocardiography (TTE) may be helpful to confirm an MI, as wall-motion abnormalities are detectable early in STEMI. TTE may detect alternative diagnoses, such as aortic dissection, pericarditis, or pulmonary embolism.

Management

Initial assessment involves rapid triage for chest pain *('time is muscle')* and referral for reperfusion therapy (primary PCI or thrombolysis). Initial medical therapy includes aspirin (300 mg), nitrates, and, if refractory pain, intravenous opioids for relief. Routine oxygen administration is not recommended unless oxygen saturations are less than 90% (see Box 17.34).

Primary percutaneous coronary intervention

The preferred reperfusion therapy for STEMI in interventional cardiology centres is primary PCI with a target of 60 minutes for wire crossing of the culprit vessel. Recent updated ESC guidelines recommend performing complete revascularization during the index hospital admission. Data from the RIVAL and RIFLE STEACS studies

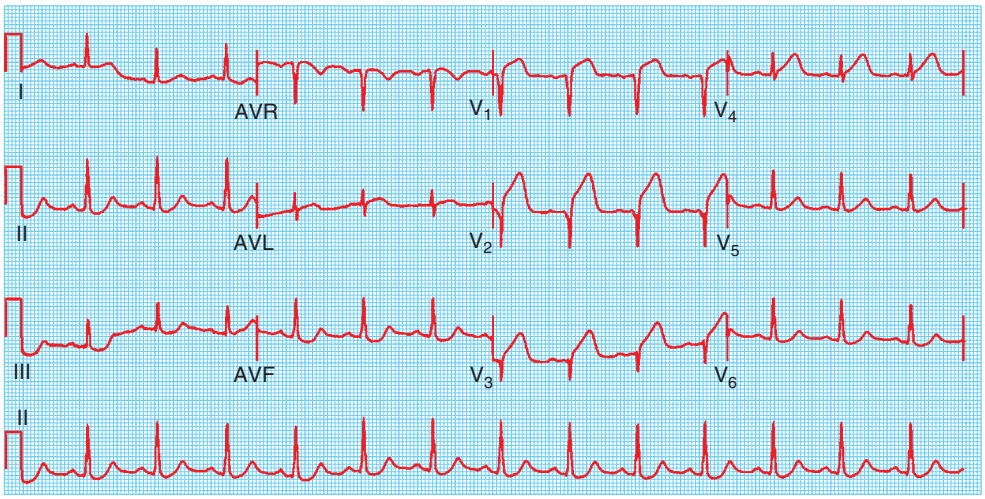


Fig. 17.63 An acute anteroseptal myocardial infarction, shown by a 12-lead ECG. Note the ST-segment elevation and pathological Q waves in leads V_1–V_3. Ischaemic changes with ST-segment depression are also seen in the inferior leads (II, III, and AVF).

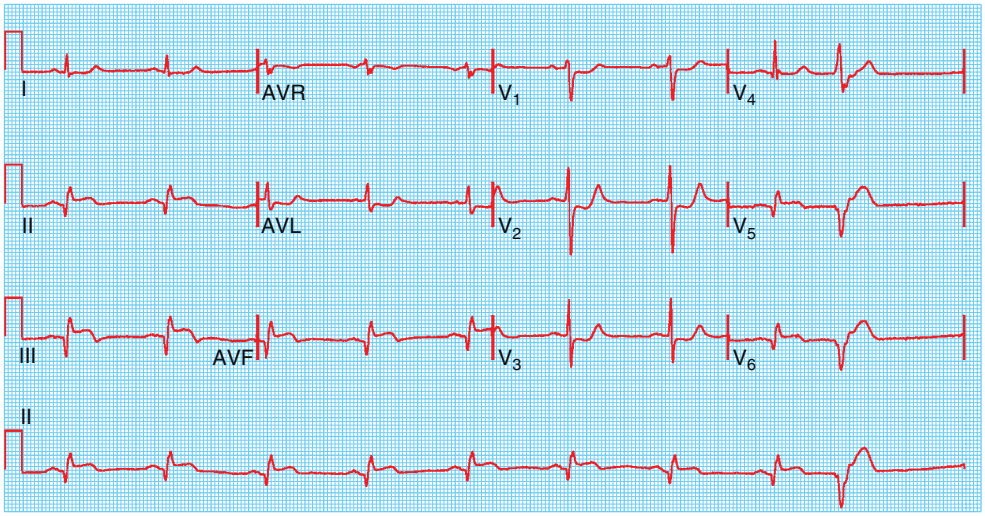


Fig. 17.64 An acute inferior wall myocardial infarction, shown by a 12-lead ECG. Note the raised ST segment and Q waves in the inferior leads (II, III, and AVF).

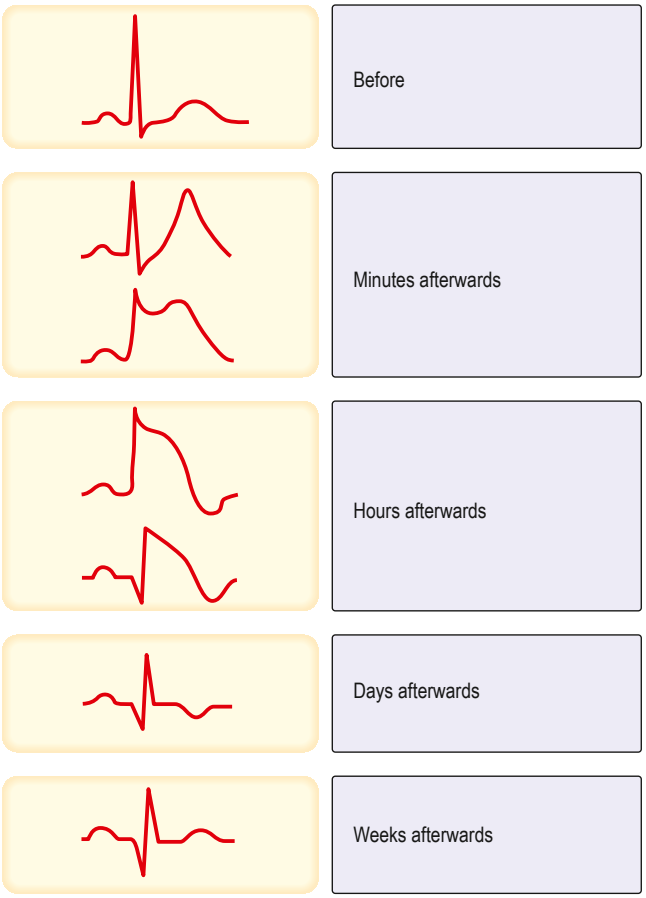


Fig. 17.65 Electrocardiographic evolution of myocardial infarction (STEMI). After the first few minutes, the T waves become tall, pointed, and upright, and there is ST-segment elevation. After the first few hours, the T waves invert, the R-wave voltage is decreased and Q waves develop. After a few days, the ST segment returns to normal. After weeks or months the T wave may return to upright but the Q wave remains.

demonstrate that radial access is the preferred route, as this reduces complications and may improve survival. Primary PCI performed with drug-eluting stents is preferred to BMSs. Dual antiplatelet therapy is with aspirin and an ADP-receptor blocker (prasugrel or ticagrelor). Anticoagulant options include UFH, enoxaparin, or bivalirudin. The routine use of GP IIb/IIIa inhibitors is no longer recommended.

Thrombolysis

Thrombolytic agents enhance the breakdown of occlusive thromboses by the activation of plasminogen to form plasmin. Thrombolysis is appropriate if transfer to a primary PCI-capable centre is not possible within 120 minutes of STEMI diagnosis. A meta-analysis showed that thrombolysis within 6 hours of STEMI or LBBB MI prevented 30 deaths in every 1000 patients treated. Between 7 and 12 hours, 20 in every 1000 deaths were prevented. After 12 hours the benefits are limited, and there is evidence to suggest less benefit for older patients, possibly because of the increased risk of strokes. For patients who fail to reperfuse 60 to 90 minutes following thrombolysis administration, as demonstrated by 50% resolution of the ST-segment elevation, re-thrombolysis or referral for rescue coronary angioplasty is recommended. The contraindications to thrombolysis are provided in **Box 17.37**. Aspirin and clopidogrel are recommended in patients undergoing thrombolysis. Anticoagulant options include UFH, enoxaparin, or fondaparinux.

Box 17.37 Contraindications to thrombolysis

Absolute contraindications
- Haemorrhagic stroke or stroke of unknown origin at any time
- Ischaemic stroke in the preceding 3 months
- Central nervous system damage/neoplasm/vascular malformation
- Recent major trauma/surgery/head injury (within the preceding 3 weeks)
- Gastrointestinal bleeding within the last month
- Known bleeding disorder
- Aortic dissection

Relative contraindications
- Ischaemic stroke in the preceding 3 months
- Oral anticoagulant therapy
- Pregnancy or within 1 week postpartum
- Non-compressible vascular punctures
- Traumatic resuscitation
- Refractory hypertension (systolic blood pressure >180 mmHg)
- Advanced liver disease
- Internal bleeding, e.g., active peptic ulcer
- Dementia

Coronary artery bypass surgery

Cardiac surgery is reserved for a small proportion of patients with unsuitable anatomy for PCI or for those with complications of MI, such as VSD or mitral regurgitation. Operative mortality is highest in the first 72 hours after STEMI.

Complications of myocardial infarction

Heart failure

Cardiac failure following STEMI is a poor prognostic feature that necessitates medical and invasive therapy to reduce the death rate. The Killip classification is used to assess patients with heart failure post MI:

- ***Killip I***—no crackles and no third heart sound
- ***Killip II***—crackles in less than 50% of the lung fields or a third heart sound
- ***Killip III***—crackles in more than 50% of the lung fields
- ***Killip IV***—cardiogenic shock.

Patients may develop AHF and should be managed with diuretics (e.g., intravenous furosemide 40–80 mg) and vasodilators. ACE inhibitors can be given within 24 to 48 hours if the blood pressure is satisfactory. Oxygen may be required, with regular oxygen monitoring. Intravenous inotropes, such as dobutamine or milrinone, are used in severe heart failure. If the patient is in cardiogenic shock, revascularization with or without temporary mechanical circulatory support may be required (see 'Temporary mechanical cardiac support' section). Patients with cardiogenic shock may require intracardiac pressure from right heart catheterization to guide treatment.

Myocardial rupture and aneurysmal dilation

Rupture of the free wall of the left ventricle is usually an early, catastrophic, and fatal event from haemopericardium. The patient will have a haemodynamic collapse, then an electromechanical cardiac arrest. A subacute rupture may allow for pericardiocentesis, followed by surgical repair of the rupture. Aneurysmal dilation of the infarcted myocardium (**Fig. 17.66**) is a late complication that may require surgical repair. Left ventricular thrombus may develop and as a result of blood stasis within the aneurysmal segment, necessitating anticoagulation.

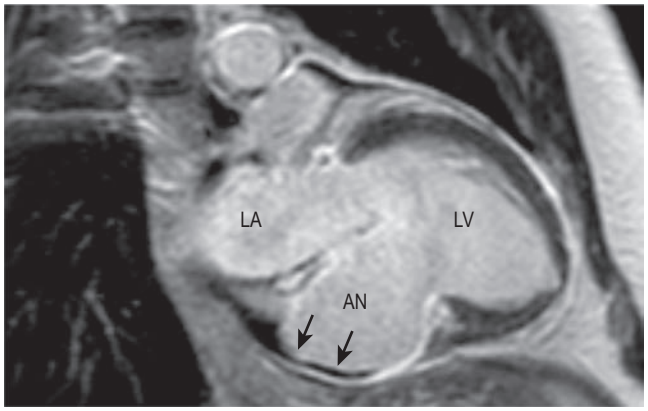


Fig. 17.66 Cardiac magnetic resonance image (vertical long axis) showing an inferior left ventricular aneurysm (*arrowed*, AN). The aneurysm was successfully resected surgically. *LA,* Left atrium; *LV,* left ventricle.

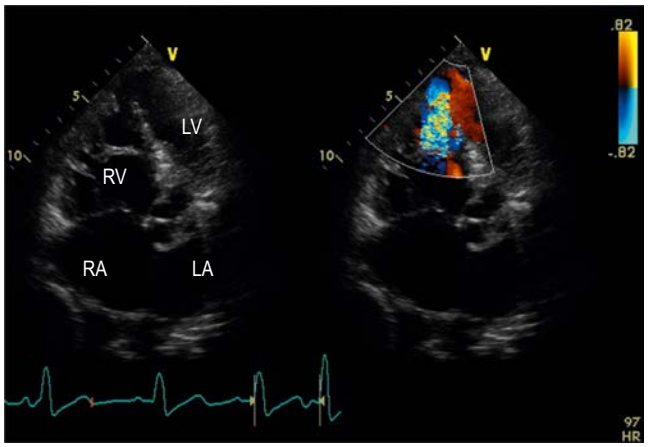


Fig. 17.67 Post-infarct ventricular septal defect. *LA,* Left atrium; *LV,* left ventricle; *RA,* right atrium; *RV,* right ventricle.

Ventricular septal defect

A ventricular septal defect (VSD) (Fig. 17.67) may occur in 1% to 2% of patients with STEMI and may be associated with delayed or failed revascularisation. However, mortality is very high and there is a 12-month non-operated mortality of 92%. Temporary mechanical cardiac support and coronary angiography may allow patient optimization prior to surgery (usually deferred till >7 days post infarction). In those not fit for surgery percutaneous VSD closure may be attempted.

Mitral regurgitation

Severe mitral regurgitation can occur early in the course of STEMI. Three mechanisms may be responsible, and a TOE may be necessary to confirm the aetiology:

- severe left ventricular dysfunction and dilation, causing annular dilation of the valve and subsequent functional regurgitation;
- MI of the inferior wall, producing dysfunction of the papillary muscle that may respond to coronary intervention; and
- MI of the papillary muscles (usually the posteromedial papillary muscle), producing sudden severe pulmonary oedema and cardiogenic shock.

Diagnosis is confirmed with echocardiography and patients may be managed with diuretics, nitrates, inotropic agents, and temporary mechanical support while they are stabilized for early mitral valve surgical repair or replacement.

Cardiac arrhythmias

Ventricular tachycardia and ventricular fibrillation are common in STEMI, particularly with reperfusion. Cardiac arrest requires defibrillation. Ventricular tachycardia that is haemodynamically tolerated should be treated with intravenous amiodarone 300 mg or lidocaine 50 to 100 mg. If the patient is hypotensive, synchronized cardioversion may be performed. Serum potassium should be maintained at >4.0 mmol/L and magnesium should be replenished. Refractory ventricular tachycardia or fibrillation may respond to magnesium 8 mmol/L over 15 minutes i.v.

Atrial fibrillation may occur in 1 in 6 patients and requires rate control with beta-blockers, anticoagulation, and consideration of rhythm control with antiarrhythmics such as amiodarone or synchronized DC cardioversion, especially for patients who are unstable.

Bradyarrhythmias can be treated initially with atropine 0.5 mg i.v. repeated up to six times in 4 hours. Temporary transcutaneous or transvenous pacemaker insertion may be necessary in patients with symptomatic heart block.

Heart block

Sinus bradycardia and varying levels of heart block including complete heart block may occur following acute MI, particularly with inferior MIs, as the right coronary artery supplies the SA and AV nodes. When associated with haemodynamic compromise, complete heart block can be managed with atropine and temporary transvenous pacing and may recover after a few days. Permanent pacing will be needed if complete heart block persists and is more common in anterior MI because of extensive infarction.

Post-MI pericarditis and Dressler's syndrome

See 'Pericardial disease' section.

Post-ACS lifestyle modification

After recovery from an ACS, patients should be encouraged to participate in a cardiac rehabilitation programme that provides education and information appropriate to their requirements. An exercise programme forms part of the rehabilitation.

- Patients should be physically active (30 min of moderate aerobic exercise five times per week).
- Dietary recommendations include calorie control of obesity, increased fruit and vegetables, reduced *trans* and saturated fats, and reduced salt intake in patients with hypertension.
- Alcohol consumption should be maintained within safe limits (≤14 units per week for both men and women), avoiding binge drinking.
- Patients should stop smoking (nicotine patches and buprenorphine are safe).
- A healthy weight (BMI < 25 kg/m^2) should be maintained.
- Blood pressure should be reduced to a systolic measurement of less than 130 mmHg.
- Patients with diabetes should be treated to maintain HbA_{1c} of less than 53 mmol/moL.

Post-ACS drug therapy and assessment

Extensive clinical trial evidence has been gathered in post-MI patients, demonstrating that a range of pharmaceuticals is advantageous in reducing mortality. Therefore patients who have suffered an MI should be taking most of the following medications:

- aspirin 75 mg daily (life-long)
- ADP-receptor blocker, for example, clopidogrel, prasugrel, and ticagrelor (at least for the first year)
- high-intensity statin (e.g., atorvastatin 80 mg) with target LDL cholesterol of less than 1.8 mmol/L
- beta-blocker (e.g., bisoprolol titrated to heart rate and blood pressure), particularly if there is heart failure with EF < 40%
- ACE inhibitor or angiotensin receptor blocker, particularly if LVEF is less than 40%
- aldosterone antagonist, if there is clinical evidence of heart failure and LVEF is less than 40%; the serum creatinine is less than 221 μmol/L (men) or less than 177 μmol/L (women); and the serum potassium is less than 5.0 mEq/L, for example, eplerenone 25 to 50 mg
- proton pump inhibitor for patients at high risk of bleeding while on dual antiplatelet therapy

The role of colchicine remains uncertain and requires careful patient selection. While earlier trials demonstrated benefits in reducing hospitalization for angina and revascularization, the largest trial to date (CLEAR SYNERGY) did not confirm these benefits.

Patients with permanent atrial fibrillation receiving medical therapy or CABG should be treated with dual therapy: aspirin or clopidogrel and anticoagulation.

Patients with permanent atrial fibrillation receiving stenting should be assessed for their bleeding risk using the HAS-BLED score (see **Box 17.14**):

- If this is 0 to 2, they can receive triple therapy for 6 months (aspirin and clopidogrel and warfarin) and then dual therapy (aspirin or clopidogrel and anticoagulation with warfarin or a DOAC).
- If it is over 2, they can receive triple therapy for 4 weeks (aspirin and clopidogrel and warfarin) and then dual therapy (aspirin or clopidogrel and anticoagulation with warfarin or a DOAC) for a further 11 months.

Further reading

Berry JD, Dyer A, Cal X et al. Lifetime risks of cardiovascular disease. *N Engl J Med* 2012; 366:321–329.
Byrne R, Rossello X, Coughlan J et al. 2023 ESC Guidelines for the management of acute coronary syndromes: developed by the task force on the management of acute coronary syndromes of the European Society of Cardiology (ESC). *Eur Heart J* 2023:ehad191.
Thygesen K, Alpert JS, Jaffe AS et al. Fourth universal definition of myocardial infarction (2018). *Eur Heart J* 2019; 40:237–269.
https://qrisk.org/three/. *QRISK®3-2018 Cardiovascular disease risk assessment tool*.

VALVULAR HEART DISEASE

MITRAL VALVE

The mitral valve consists of the fibrous annulus, anterior and posterior leaflets, chordae tendineae, and the papillary muscles (**Fig. 17.68**).

Mitral stenosis

Mitral stenosis is mostly due to rheumatic heart disease following previous rheumatic fever due to infection with a group A beta-haemolytic streptococcus; in low- to medium-income countries this affects nearly 20 million people. Genetically determined immune markers influence susceptibility to the initial infection and play a role in the risk of developing chronic rheumatic disease. The condition is more common in those with a history of multiple episodes of pharyngitis and also in women. Inflammation leads to commissural fusion and a reduction in mitral valve orifice area, causing the characteristic doming pattern on echocardiography. It is a progressive condition over many years with valve thickening, cusp fusion, and calcium deposition leading to increasing immobility of the valve cusps and a severely narrowed (stenotic) valve orifice.

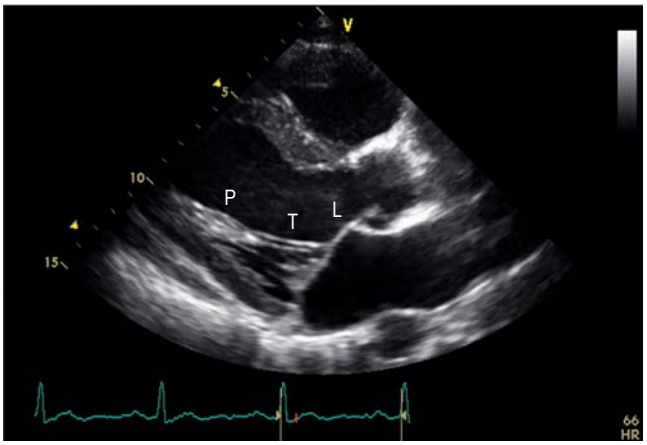


Fig. 17.68 Mitral valve apparatus—parasternal long axis. *L,* Leaflets; *P,* papillary muscle; *T,* chordae tendineae.

Other causes of mitral stenosis include:

- congenital mitral stenosis
- Lutembacher's syndrome (the combination of acquired mitral stenosis and an atrial septal defect)
- mitral annular calcification; rarely, this may lead to mitral stenosis if extensive, particularly in elderly patients and those with end-stage renal disease
- carcinoid tumours metastasizing to the lung, or primary bronchial carcinoid.

Pathophysiology

When the normal valve orifice area of 4 to 6 cm^2 is reduced to less than 1 cm^2, severe mitral stenosis is present. In order for sufficient cardiac output to be maintained, the left atrial pressure increases and left atrial hypertrophy and dilation occur, leading to increases in pulmonary venous, pulmonary arterial, and right heart pressures. Increases in the pulmonary capillary pressure can lead to the development of pulmonary oedema, particularly with the onset of atrial fibrillation with tachycardia and loss of coordinated atrial contraction. This is partially prevented by alveolar and capillary thickening and pulmonary arterial vasoconstriction (reactive pulmonary hypertension). Pulmonary hypertension leads to right ventricular hypertrophy, dilation, and failure, with subsequent functional tricuspid regurgitation.

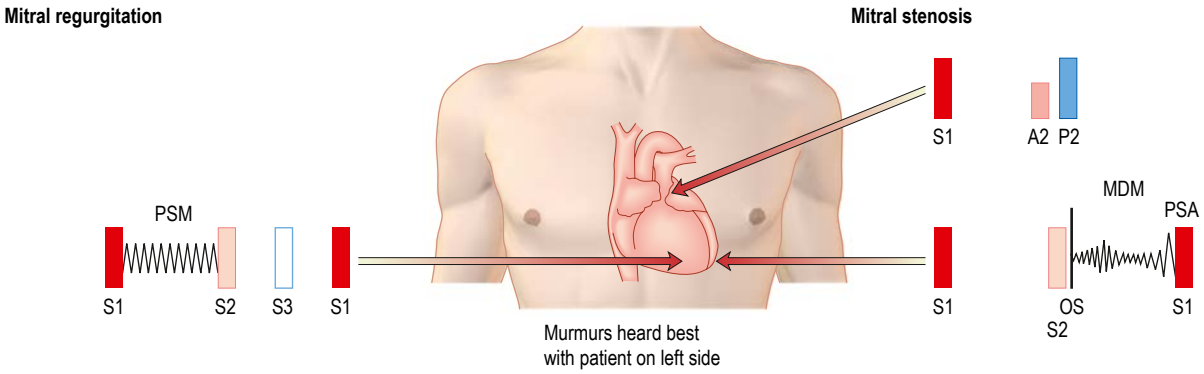
Clinical features

Symptoms

Usually, there is a long asymptomatic phase until the valve orifice is moderately stenosed (area <2 cm^2). This does not usually occur until several decades after the first attack of rheumatic fever, but in low- to medium-income countries, severe stenosis may occur at 10 to 20 years of age.

Progressive ***dyspnoea*** and ***reduced exercise tolerance*** develop from the elevation in left atrial pressure, vascular congestion, and interstitial pulmonary oedema. A cough productive of blood-tinged, frothy sputum or frank haemoptysis may occur. The development of pulmonary hypertension eventually leads to ***right heart failure*** and its symptoms of weakness, fatigue, and abdominal and lower limb swelling. The left ventricle, however, is safeguarded as it lies downstream of the mitral valve.

As the left atrium progressively enlarges, this predisposes to ***atrial fibrillation***, giving rise to symptoms such as palpitations.



Mitral regurgitation	
Pulse:	Sinus rhythm or atrial fibrillation
Apex:	Forceful, displaced, systolic thrill
Sounds:	Soft S1, S3
Murmurs:	Pansystolic

Mitral stenosis	
Face:	Mitral facies
Pulse:	Atrial fibrillation
RV:	Heaving, sustained
Apex:	Localized, tapping
Sounds:	Loud S1, loud P2 (if pulmonary hypertension), opening snap
Murmurs:	Mid-diastolic rumble at apex

Fig. 17.69 Features associated with mitral regurgitation and mitral stenosis. *A2,* Aortic component of the second heart sound; *MDM,* mid-diastolic murmur; *OS,* opening snap; *P2,* pulmonary component of the second heart sound (loud with pulmonary hypertension); *PSA,* pre-systolic accentuation; *PSM,* pansystolic murmur; *RV,* right ventricle; *S1,* first heart sound; *S2,* second heart sound; *S3,* third heart sound.

Atrial fibrillation may result in ***systemic emboli***, most commonly to the cerebral vessels, producing stroke, but mesenteric, renal, and peripheral emboli are also seen.

Signs

See the 'Clinical memo' in Fig. 17.69.

Face

Severe mitral stenosis with pulmonary hypertension is associated with 'mitral facies' or malar flush. This is a bilateral, cyanotic, or dusky pink discoloration over the upper cheeks arising from arteriovenous anastomoses and vascular stasis.

Pulse

A small-volume pulse may be present in severe mitral stenosis due to diminished left ventricular filling. The rhythm may be regular (sinus rhythm), but as the disease progresses, an 'irregularly irregular' pulse (atrial fibrillation) may occur and lead to symptomatic clinical deterioration.

Jugular veins

Right heart failure may develop with jugular venous distension. When pulmonary hypertension or tricuspid stenosis is present, the 'a'-wave will be prominent, provided that atrial fibrillation has not supervened.

Palpation

There is a tapping impulse felt parasternally on the left side which is the result of a palpable first heart sound combined with left ventricular backward displacement produced by an enlarging right ventricle. A sustained parasternal impulse due to right ventricular hypertrophy may also be felt.

Auscultation

Auscultation (see Fig. 17.69) reveals a loud first heart sound (S1) and an opening snap (OS) as the valve suddenly opens with the force of the increased left atrial pressure. With increasing severity, the OS moves closer to the second heart sound (S2), and with heavy calcification, the first heart sound softens and the OS may be absent. This is followed by a low-pitched, 'rumbling', mid-diastolic murmur, best heard with the bell of the stethoscope held lightly at the apex and the patient lying on the left side in expiration. If the patient is in sinus rhythm, the murmur becomes louder at the end of diastole as a result of atrial contraction (pre-systolic accentuation). Pulmonary hypertension may result in pulmonary valvular regurgitation, which causes an early diastolic murmur in the pulmonary area, known as a Graham Steell murmur.

Investigations

Chest X-ray

The chest X-ray may show left atrial enlargement with straightening of the left heart border and a 'double shadow' on the border of the right and left atria. Late in the course of the disease, a calcified mitral valve may be seen on a penetrated or lateral view. Pulmonary vascular congestion and enlargement of the main pulmonary arteries may also be present in severe disease.

ECG

In sinus rhythm the ECG may show a bifid P wave ('P mitrale') from delayed left atrial activation (Fig. 17.70). However, atrial fibrillation is frequently present. In the presence of pulmonary hypertension there may be features of right ventricular hypertrophy (right axis deviation and tall R waves in lead V_1) (Fig. 17.71).

Echocardiogram

Transthoracic echocardiography is able to determine the degree of thickening, calcification, and mobility of the mitral leaflets and the extent of commissural fusion and assess left atrial size (Fig. 17.72). The severity of the mitral stenosis (Box 17.38) can be defined by mitral valve area on two-dimensional echocardiography, with continuous wave (CW) Doppler to

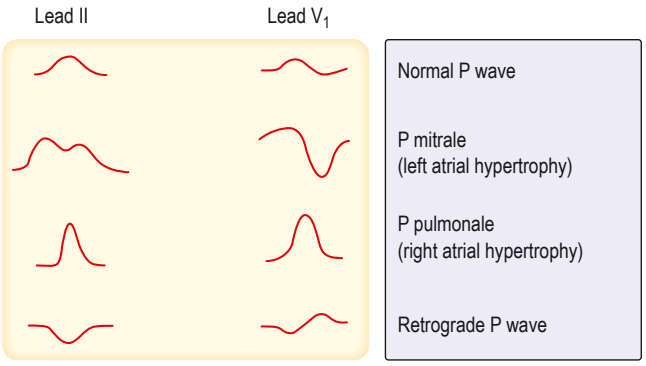


Fig. 17.70 A bifid P wave, as seen on the ECG in mitral stenosis (P mitrale). Other P wave abnormalities are also shown for comparison.

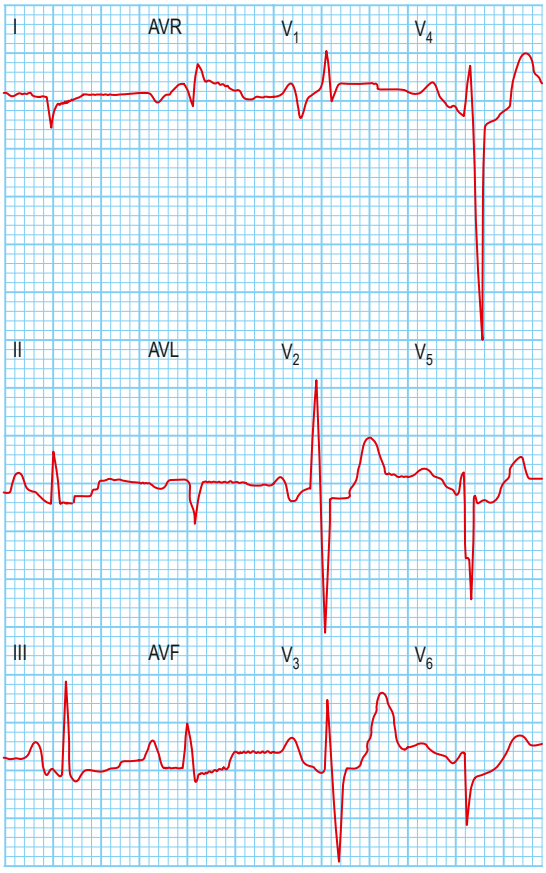


Fig. 17.71 Severe mitral stenosis, shown by a 12-lead ECG. Note the right axis deviation (frontal plane axis = +120 degrees), the left atrial conduction abnormality (large terminal negative component of the P wave in V_1), and the right ventricular hypertrophy (R wave in V_1 and right axis deviation).

measure the pressure half-time (the time taken for the pressure to halve from the peak value) and mean pressure drop across the valve. CW Doppler may also be used to estimate pulmonary artery pressure through measurement of the degree of tricuspid regurgitation.

TOE enables a detailed assessment of the mitral valve and other valves prior to consideration of surgical or percutaneous intervention and is able to detect thrombus in the left atrium or left atrial

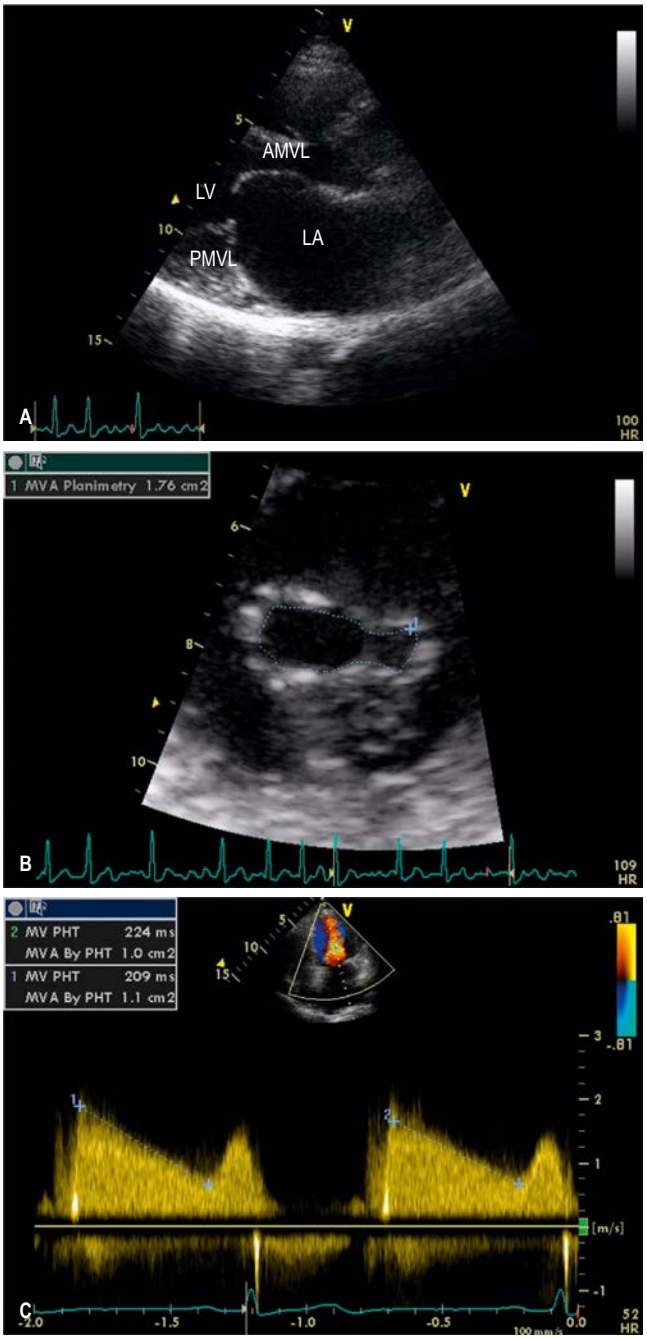


Fig. 17.72 Echocardiograms in rheumatic mitral valve disease. **(A)** A two-dimensional long-axis view, showing an enlarged left atrium and the 'hooked' appearance of the mitral valve leaflets, resulting from commissural fusion. **(B)** A magnified short-axis view, showing the mitral valve orifice as seen from the left atrium. The orifice area can be planimetered to assess the severity; in this case it is 1.5 cm^2, indicating moderately severe disease. **(C)** A continuous-wave (CW) Doppler recording, showing the slow rate of decay of flow velocity from the left atrium to the left ventricle during diastole. It is also possible to derive the valve orifice area from the velocity decay rate. *AMVL,* Anterior mitral valve leaflet; *LA,* left atrium; *LV,* left ventricle; *PMVL,* posterior mitral valve leaflet.

appendage. The Wilkins score is an echocardiographic assessment of the mitral valve (leaflet mobility, valve thickening, valve calcification, and subvalvular apparatus) and is helpful to determine suitability for percutaneous mitral valvuloplasty.

Cardiac catheterization

Left and right heart catheterization may be required in patients with severe mitral stenosis referred for intervention.

Management

Early symptoms of mitral stenosis, such as mild dyspnoea, can usually be treated with loop diuretics, although the presence of symptoms is an indication of surgical intervention. The onset of atrial fibrillation requires rate control (e.g., with beta-blockers or calcium-channel blockers) and consideration of DC cardioversion and anticoagulation with warfarin rather than a direct oral anticoagulant (DOAC), to prevent atrial thrombus and systemic embolization. Patients with severe mitral stenosis who develop symptoms, new-onset atrial fibrillation, or pulmonary hypertension are appropriate for intervention. Earlier prophylactic intervention may be indicated if the valve is amenable to balloon valvotomy and in pregnancy. There are four operative measures.

Trans-septal balloon valvotomy

This procedure is associated with low morbidity and mortality and involves introduction of a catheter into the right atrium via the femoral vein under local anaesthesia in the cardiac catheter laboratory. The interatrial septum is punctured and the catheter advanced into the left atrium and across the mitral valve. A balloon is passed over the catheter to lie across the valve and then inflated briefly to split the valve commissures. As with other valvotomy techniques, significant mitral regurgitation may result, necessitating valve replacement (see later). This procedure is ideal for patients with pliable valves in whom there is little involvement of the subvalvular apparatus and minimal mitral regurgitation (Wilkins score ≤8). TOE is required to exclude contraindications such as heavy calcification, more than mild mitral regurgitation, and left atrial thrombus.

Closed valvotomy

This operation is advised for patients with mobile, non-calcified, and non-regurgitant mitral valves. The fused cusps are forced apart by a dilator introduced through the apex of the left ventricle and guided into position by the surgeon's finger inserted via the left atrial appendage. Cardiopulmonary bypass is not required for this operation. Closed valvotomy may produce a good result for 10 years or more, but the valve cusps often re-fuse, and another operation may be necessary.

Open valvotomy

This operation is often preferred to closed valvotomy and requires cardiopulmonary bypass. The cusps are carefully dissected apart under direct vision, reducing the likelihood of traumatic mitral regurgitation,

Mitral valve replacement

Replacement with a mechanical or bioprosthetic mitral valve is necessary if:

- more than mild mitral regurgitation is present
- the valve is significantly calcified and stenotic and cannot be dissected open without producing significant mitral regurgitation
- there is severe mitral stenosis and thrombus in the left atrium despite anticoagulation
- there is a concomitant indication for surgery, for example, coronary artery bypass required.

Box 17.38 Echocardiographic severity of mitral stenosis

Parameter	Normal	Mild	Moderate	Severe
Pressure half-time (ms)	40–70	71–139	140–219	>219
Mean pressure drop (mmHg)		<5.0	5–10	>10
Valve area (cm^2)	4.0–6.0	1.5–2.0	1.0–1.5	<1.0

(Source: British Society of Echocardiography.)

Mitral regurgitation

Mitral regurgitation can occur due to a primary abnormality of the valvular apparatus involving the leaflets, mitral annulus, chordae tendineae, or papillary muscles or be secondary due to abnormalities of the left ventricle. The most frequent causes of mitral regurgitation are degenerative (mitral valve prolapse or myxomatous) disease, ischaemic heart disease, rheumatic heart disease, and infectious endocarditis. Mitral regurgitation may also be present in diseases of the myocardium (DCM and HCM), rheumatic autoimmune diseases (e.g., systemic lupus erythematosus), collagen diseases (e.g., Marfan's and Ehlers–Danlos syndromes), and disorders caused by drugs, including the centrally acting serotonin-releasing agent fenfluramine and the dopamine agonist cabergoline.

Pathophysiology

Regurgitation into the left atrium leads to left atrial dilation but little increase in left atrial pressure if the regurgitation is longstanding, as the regurgitant flow is accommodated by the enlarged left atrium. With acute mitral regurgitation (e.g., following MI, chordal rupture, or IE), the normal compliance of the left atrium does not allow dilation and the left atrial pressure rises. The left atrial *v*-wave is greatly increased and pulmonary venous pressure rises, leading to pulmonary oedema. Since a proportion of the stroke volume is regurgitated, the stroke volume increases to maintain the forward cardiac output and the left ventricle dilates with eccentric hypertrophy.

The Carpentier classification (Fig. 17.73) uses mitral leaflet motion to divide patients into different classes according to the mechanism of regurgitation, which can be useful when considering surgical intervention.

Clinical features

Symptoms

Mitral regurgitation can be present for many years and the cardiac dimensions greatly increased before any symptoms occur.

- ***Dyspnoea*** and ***orthopnoea*** develop because of pulmonary venous hypertension that arises as a direct result of mitral regurgitation and secondarily as a consequence of left ventricular failure.
- ***Fatigue*** and ***lethargy*** develop because of the reduced cardiac output.

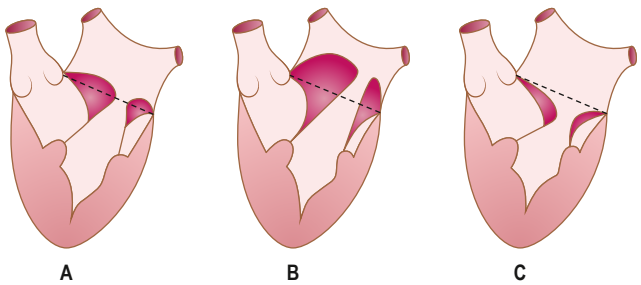


Fig. 17.73 Carpentier classification of mitral regurgitation and its causes. **(A)** Normal leaflet motion. **(B)** Increased leaflet motion/prolapse. **(C)** Restricted leaflet motion. *(Reproduced from Tuladhar SM, Punjabi PP. Surgical reconstruction of the mitral valve.* Heart *2006; 92:1373–1377, with permission.)*

- In the late stages of the disease the symptoms of ***right heart failure*** also occur and eventually lead to congestive cardiac failure.
- ***Cardiac cachexia*** may develop.
- **Palpitations** may occur from atrial fibrillation.

Signs

See Fig. 17.69.

The physical signs of uncomplicated mitral regurgitation are:

- ***Laterally displaced (forceful) diffuse apex beat*** and ***a systolic thrill*** (if severe)
- ***Soft first heart sound***, owing to the incomplete apposition of the valve cusps and their partial closure by the time ventricular systole begins
- ***Pansystolic murmur***, as regurgitation occurs throughout the whole of systole, being loudest at the apex but radiating widely over the precordium and into the axilla
- ***Mid-systolic click***, which may be present with a floppy mitral valve (see later); it is produced by the sudden prolapse of the valve and the tensing of the chordae tendineae that occurs during systole. This may be followed by a late systolic murmur owing to some regurgitation.
- ***Prominent third heart sound*** (S3), owing to the sudden rush of blood back into the dilated left ventricle in early diastole (sometimes a short mid-diastolic flow murmur may follow the third heart sound).

The signs related to atrial fibrillation, pulmonary hypertension, and left and right heart failure develop later in the disease. The onset of atrial fibrillation may have a much less dramatic effect on symptoms than in mitral stenosis.

Investigations

Chest X-ray

The chest X-ray may show left atrial and left ventricular enlargement. There is an increase in the cardiothoracic ratio and valve calcification may be present.

ECG

The ECG shows the features of left atrial delay (bifid P waves) and left ventricular hypertrophy (Fig. 17.74), as manifested by tall R waves in the left lateral leads (e.g., leads I and V_6) and deep S waves in the right-sided precordial leads (e.g., leads V_1 and V_2). Left ventricular hypertrophy occurs in about 50% of patients with mitral regurgitation. Atrial fibrillation may be present.

Echocardiogram

The echocardiogram (Fig. 17.75) shows a dilated left atrium and left ventricle. There may be specific features of chordal or papillary muscle rupture. The severity of regurgitation can be assessed with the use of colour Doppler, looking at the narrowest jet width (vena contracta) and area and calculating the regurgitant fraction, volume, or orifice area. Useful information regarding the severity of the condition can be obtained indirectly by observing the dynamics of ventricular function. ***TOE*** is used to identify the mechanism of regurgitation, the structural valve abnormalities, and the likelihood of successful surgical valve repair (see Fig. 17.75). Echocardiography during exercise is sometimes required in symptomatic individuals with discordant regurgitation severity at rest.

Cardiac catheterization

Left and right heart catheterization is appropriate in patients with suspected CAD and for assessment of pulmonary hypertension prior to surgery.

Management

Mild to moderate mitral regurgitation can be managed conservatively by following the patient with serial echocardiograms. Prophylaxis against endocarditis is discussed in the 'Prevention' section.

Current ESC guidelines recommend surgical intervention in patients with primary symptomatic severe chronic mitral regurgitation. Mitral valve repair is the preferred surgical technique as this preserves the functional components of the mitral apparatus, but mitral valve replacement may be required in those with advanced disease. ***Asymptomatic individuals*** should undergo regular surveillance in a Heart Valve Clinic, as certain echocardiographic parameters (LVEF ≤ 60%, LV end-systolic dimension > 40 mm, left atrium > 55 mm or pulmonary hypertension) are associated with a worse prognosis and should prompt surgical evaluation. The advantages of surgical intervention are diminished in more advanced disease. Sudden torrential mitral regurgitation, as seen with chordal or papillary muscle rupture or IE, necessitates emergency mitral valve surgery. When patients are not suited to surgical intervention or when surgery will be performed at a later date, management involves treatment with diuretics, ACE inhibitors, and anticoagulation if atrial fibrillation is present. A percutaneous mitral valve repair (MitraClip) may be appropriate in selected patients who are unsuitable for mitral surgery.

Mitral valve prolapse

This is also known as Barlow's syndrome or floppy mitral valve and is a common cause of mitral regurgitation. It is characterized by excessively large mitral valve leaflets, an enlarged mitral annulus, abnormally long chordae, or disordered papillary muscle contraction and can be diagnosed with echocardiography. Histology may demonstrate myxomatous degeneration of the mitral valve leaflets. Prolapsing mitral valve is more commonly seen in women than in men, and although most cases are sporadic, some have a familial preponderance. Its cause is unknown, but it may be associated with connective tissue disorders (Marfan's syndrome, Ehlers–Danlos syndrome, and pseudoxanthoma elasticum). Symptoms are usually due to mitral regurgitation, and some may experience atypical chest pain and develop atrial fibrillation. On auscultation, there is a systolic click from sudden stretch of the mitral chordae as the leaflets prolapse into the left atrium, and if regurgitation is present, this is late-systolic in the early phase and may progress to a pansystolic murmur.

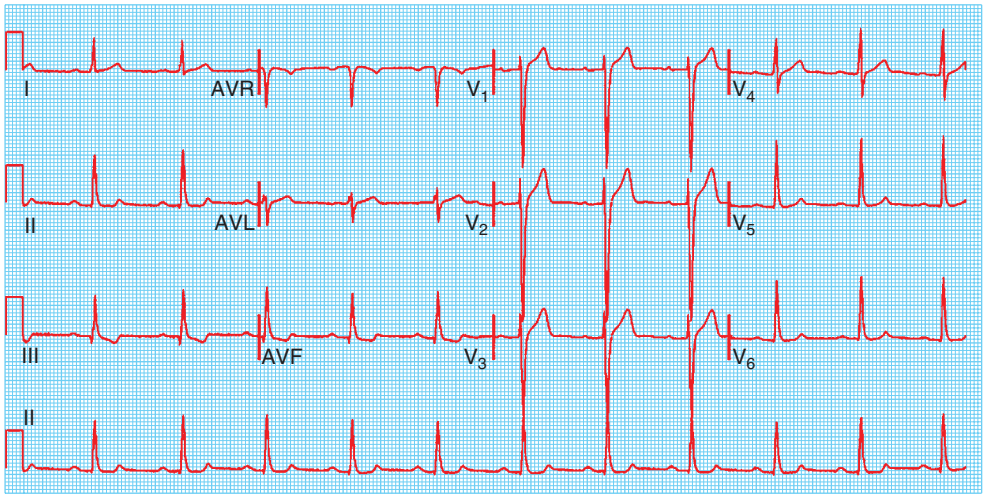
Secondary mitral regurgitation

Secondary mitral regurgitation refers to this abnormality in the presence of normal leaflet morphology and subvalvular apparatus. This is seen when there is either LV dysfunction and dilatation (due to cardiomyopathy) or atrial/annular dilatation (commonly seen with chronic AF). Unlike with primary mitral regurgitation, a clear benefit for surgical valve intervention has not been established. Treatment options for the individual patient in this instance are best decided by the Heart Valve team and may include medical treatment of the underlying cause, cardiac surgery, transcatheter intervention, heart transplant assessment, or palliation.

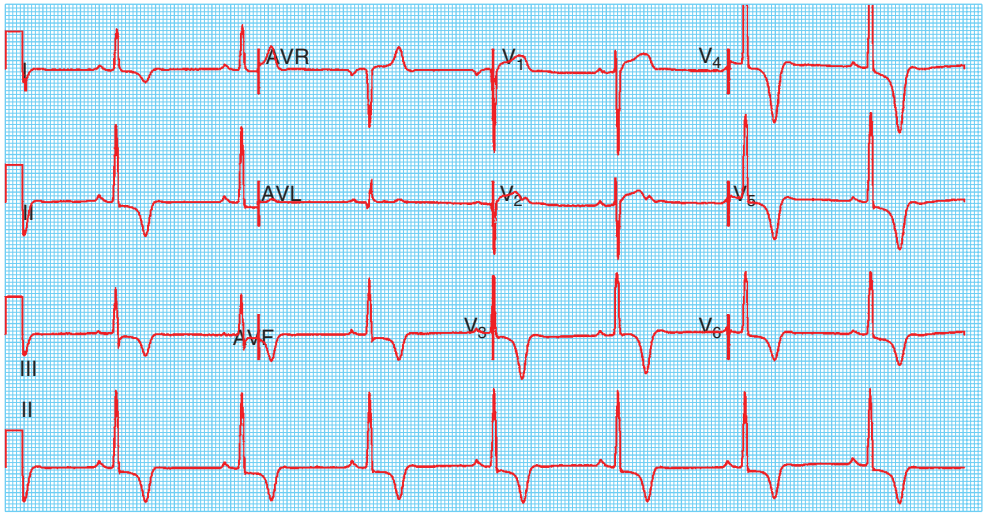
AORTIC VALVE

Aortic stenosis

Aortic stenosis is a chronic progressive disease that limits left ventricular outflow, leading to symptoms of chest pain, breathlessness, syncope, pre-syncope (exertional dizziness), and fatigue.

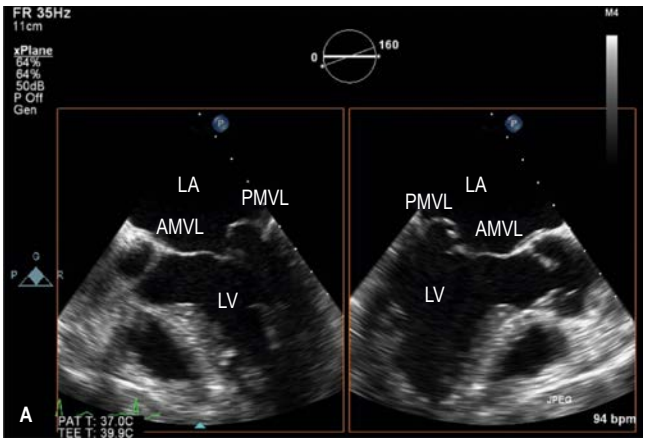


A



B

Fig. 17.74 Left ventricular hypertrophy, shown in 12-lead ECGs. **(A)** Note the size of the S wave seen in V_1 (26 mm); S in V_1 + R in V_6 = >35 mm. **(B)** Left ventricular hypertrophy in a patient with hypertrophic cardiomyopathy, with additional repolarization changes of deep T-wave inversion in leads V_3–V_6 and leads I, II, III, and AVF.



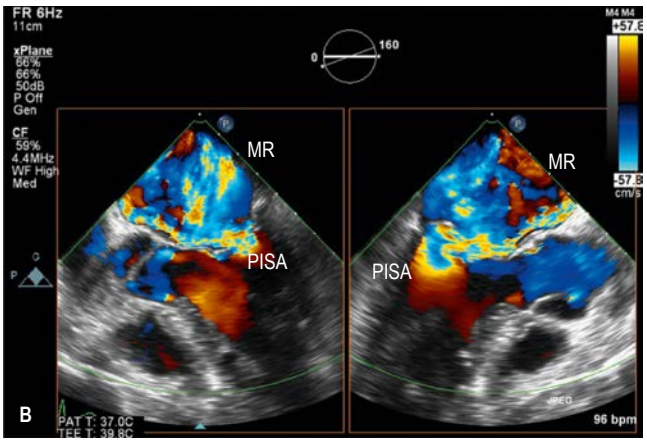


Fig. 17.75 Mitral regurgitation. **(A)** Transoesophageal echocardiography with marked prolapse of part of the posterior mitral valve leaflet (PMVL). **(B)** Image A with colour Doppler demonstrating severe mitral regurgitation (MR) into the left atrium (LA). *AMVL,* Aortic mitral valve leaflet; *LV,* left ventricle; *PISA,* proximal isovelocity surface area.

Aortic valve stenosis includes calcific stenosis of a trileaflet aortic valve, stenosis of a congenitally bicuspid valve, and rheumatic aortic stenosis. The prevalence of aortic stenosis increases with age and is primarily due to calcific disease in Europe, but worldwide, the commonest cause is rheumatic.

Calcific aortic valvular disease (CAVD) is the most common cause of aortic stenosis and occurs mainly in the elderly. This is an inflammatory process involving macrophages and T lymphocytes, initially with thickening of the subendothelium and adjacent fibrosis. The lesions contain lipoproteins, which calcify, increasing leaflet stiffness and reducing systolic opening. This can occur in a tri- or bileaflet aortic valve.

Bicuspid aortic valve (BAV) (Fig. 17.76) is the most common form of congenital heart disease, occurring in 1% to 2% of live births; in about 9% of cases it is familial. Patients with familial bicuspid valve tend to present at an earlier age. BAV is associated with aortic coarctation, root dilation, and, potentially, aortic dissection, and patients should have regular follow-up echocardiography.

Rheumatic fever can produce progressive fusion, thickening, and calcification of the aortic valve. In rheumatic heart disease the aortic valve is affected in about 30% to 40% of cases and there is usually associated mitral valve disease.

Other causes of valvular stenosis include chronic kidney disease, Paget's disease of bone, previous radiation exposure, and systemic lupus erythematosus. ***Valvular aortic stenosis*** should be distinguished from other causes of obstruction to left ventricular emptying (Fig. 17.77), which include:

- supravalvular obstruction—a congenital fibrous diaphragm above the aortic valve, often associated with learning difficulties and hypercalcaemia (Williams' syndrome)
- subvalvular aortic stenosis—a congenital condition in which a fibrous ridge or diaphragm is situated immediately below the aortic valve
- HCM—septal muscle hypertrophy obstructing left ventricular outflow.

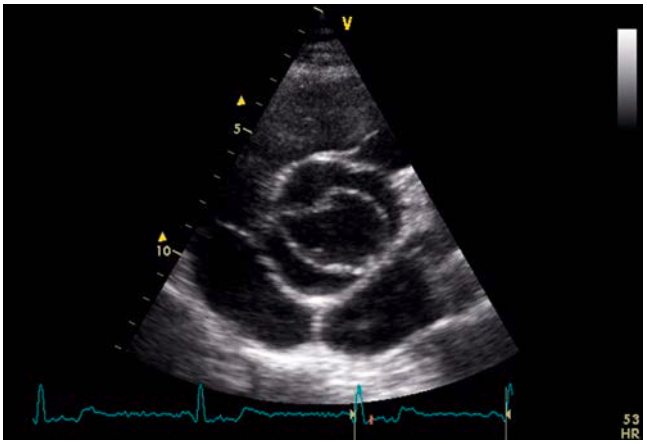


Fig. 17.76 **Bicuspid aortic valve.**

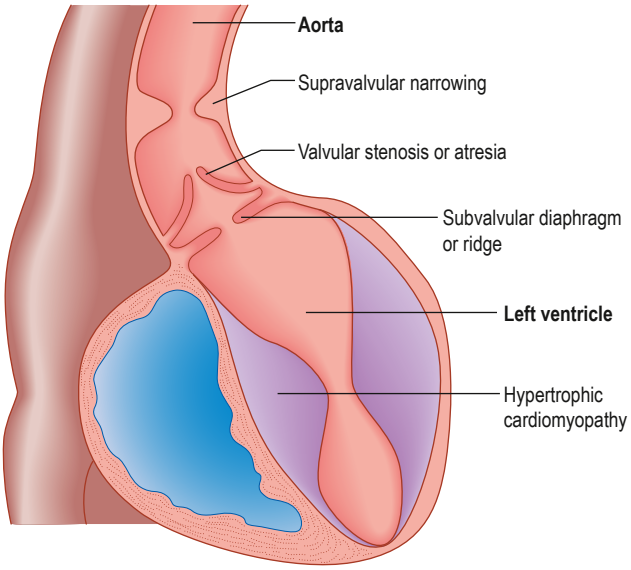


Fig. 17.77 **Several forms of left ventricular outflow tract obstruction.**

Pathophysiology

Obstructed left ventricular emptying leads to increased left ventricular pressure and compensatory left ventricular hypertrophy. In turn, this results in relative ischaemia of the left ventricular myocardium and consequent angina, arrhythmias, and left ventricular failure. The obstruction to left ventricular emptying is relatively more severe on exercise. Normally, exercise causes a many-fold increase in cardiac output, but when there is severe narrowing of the aortic valve orifice, the cardiac output can hardly increase. Thus, the blood pressure falls, coronary ischaemia worsens, the myocardium fails, and cardiac arrhythmias may develop. Left ventricular systolic function is typically preserved in patients with aortic stenosis.

Clinical features

Symptoms

There are usually no symptoms until aortic stenosis is moderately severe (when the aortic orifice is reduced to one-third of its normal size). At this stage, ***exercise-induced syncope***, ***angina***, and ***dyspnoea*** develop. When symptoms occur, the prognosis is poor: on average, death occurs within 2 to 3 years if there has been no valvular intervention.

Signs

See Fig. 17.78.

Pulse

The carotid pulse is small-volume and slow-rising or plateau in nature (see 'Character' section).

Precordial palpation

The apex beat is not usually displaced because hypertrophy (as opposed to dilation) does not produce noticeable cardiomegaly. However, the pulsation is sustained and obvious. A double impulse is sometimes felt because the fourth heart sound or atrial contraction ('kick') may be palpable. A systolic thrill may be felt in the aortic area.

Auscultation

There is an ejection systolic murmur that is usually crescendo–decrescendo ('diamond-shaped'), harsh in quality, and best heard in the second right intercostal space (aortic area). It may radiate into the carotid arteries and also the precordium. The murmur is usually longer when the disease is more severe, as a longer ejection time is required. Importantly, a soft or absent second heart sound reflects severe aortic stenosis. The intensity

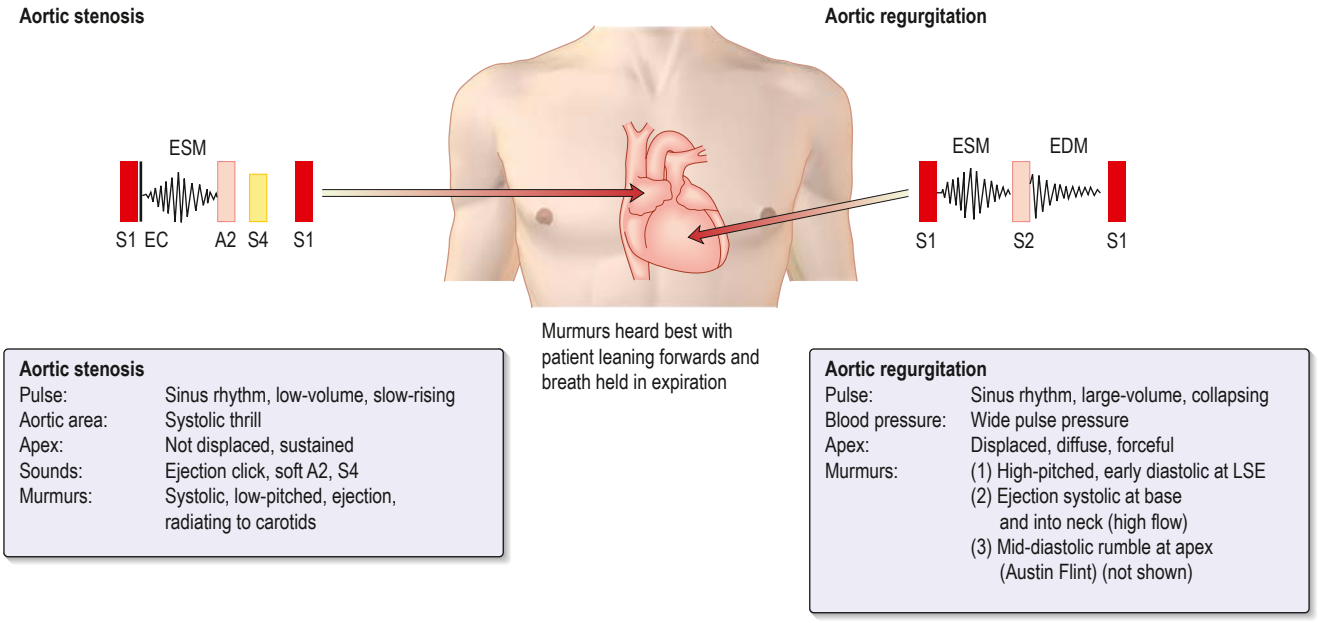


Fig. 17.78 Features of aortic stenosis and aortic regurgitation. *A2,* Aortic component of the second heart sound; *EC,* ejection click; *EDM,* early diastolic murmur; *ESM,* ejection systolic murmur; *LSE,* left sternal edge; *S1,* first heart sound; *S4,* fourth heart sound.

of the murmur is not a good guide to the severity of the condition because it is lessened by a reduced cardiac output. In severe cases the murmur may be inaudible.

Other findings include:

- ***systolic ejection click***, unless the valve has become immobile and calcified
- ***soft or inaudible aortic second heart sound*** when the aortic valve becomes immobile
- ***reversed splitting of the second heart sound*** (splitting on expiration) (see 'Second heart sound [S2]' section)
- ***prominent fourth heart sound***, caused by atrial contraction and heard unless coexisting mitral stenosis prevents this.

Investigations

Chest X-ray

The heart size is usually normal, although the cardiothoracic ratio increases as heart failure occurs. The aortic valve may be calcified. There may be a prominent, dilated, ascending aorta as a result of turbulent blood flow above the stenosed aortic valve leading to post-stenotic dilation.

ECG

The ECG shows left ventricular hypertrophy and left atrial delay. A left ventricular 'strain' pattern due to 'pressure overload' (depressed ST segments and T-wave inversion in leads orientated towards the left ventricle, i.e., leads I, AVL, V_5, and V_6) is common when disease is severe. Usually, sinus rhythm is present but atrial fibrillation and ventricular arrhythmias may be noted.

Echocardiogram

Echocardiography readily demonstrates the thickened, calcified, and immobile aortic valve cusps and can differentiate between a trileaflet and bicuspid valve. The severity of aortic stenosis is measured using the pressure gradient across the valve and the calculated valve area (Box 17.39 and Fig. 17.79). It can also assess for the presence of left ventricular hypertrophy, concomitant valve disease, and aortic coarctation.

Box 17.39 Echocardiographic severity of aortic stenosis

Parameter	Normal	Mild	Moderate	Severe
Peak velocity (m/s)	<1.7	1.7–2.9	3.0–4.0	>4.0
Peak pressure drop (mmHg)		<36	36–64	>64
Mean pressure drop (mmHg)		<25	25–40	>40
Valve area (cm^2)	>2.0	1.5–2.0	1.0–1.4	<1.0

(Source: British Society of Echocardiography.)

CMR and cardiac CT

These techniques are indicated for assessing the thoracic aorta for the presence of aneurysm, dissection, or coarctation. Cardiac CT has a key role in individuals being considered for transcatheter aortic valve replacement (TAVR, see below), but it can also measure the aortic valve calcium score to confirm stenosis severity if non-invasive testing is inconclusive.

Cardiac catheterization

Cardiac catheterization is mainly required to identify coexistent CAD prior to surgery but may be used to assess the pressure gradient across the aortic valve if further confirmation of severity is required.

Management

In patients with aortic stenosis symptoms are a good index of severity, and all symptomatic, appropriate patients should have ***aortic valve replacement***. Patients with a BAV and ascending aorta of over 50 mm or expanding at more than 5 mm per year should be considered for surgical intervention. Asymptomatic patients should be under regular review in a valve clinic for assessment of symptoms and serial echocardiography. Surgical intervention for asymptomatic people with severe aortic stenosis is recommended in those with:

- symptoms or fall in blood pressure during an exercise test
- very high peak velocity across aortic valve of more than 5.0 m/s

- rapid increase in velocity through the valve at more than 0.3 m/s/year
- the need for concurrent cardiac surgery, for example, CABG, ascending aorta, or other cardiac valve replacement.

Antibiotic prophylaxis against IE is discussed in the 'Prevention' section. Provided that the valve is not severely deformed or heavily calcified, critical aortic stenosis in childhood or adolescence can be treated by valvotomy (performed under direct vision by the surgeon or by balloon dilation using X-ray visualization). This produces temporary relief from the obstruction and aortic valve replacement will usually be required later. Balloon dilation (valvuloplasty) has been tried in adults, especially in the elderly, as an alternative to surgery, but results are poor and such treatment is reserved for patients unfit for surgery or as a 'bridge' to TAVR or surgery.

Transcatheter aortic valve replacement (TAVR or TAVI)

This is a minimally invasive procedure involving insertion of a new valve within the narrowed aortic valve, usually reserved for those at higher risk and not suitable for traditional surgical valve replacement due to frailty or underlying co-morbidities. A catheter is inserted into an artery in the groin and guided to the heart, and once in position, a balloon stent valve is expanded into place (Fig. 17.80). Compared with open heart surgery, procedural complications and mortality are lower, recovery time is quicker, and new-generation valves have been shown to be durable in follow-up studies; as such, the indications for this intervention are likely to expand to lower risk groups except in those who require concomitant cardiac surgery, for example, bypass grafting.

Aortic regurgitation

Aortic regurgitation can occur in diseases affecting the aortic valve, such as calcific valve disease or rheumatic valve disease, and diseases leading to aortic root dilatation, such as hypertension and Marfan's syndrome (MFS; Box 17.40).

Pathophysiology

Aortic regurgitation is reflux of blood from the aorta through the aortic valve back into the left ventricle during diastole. If net

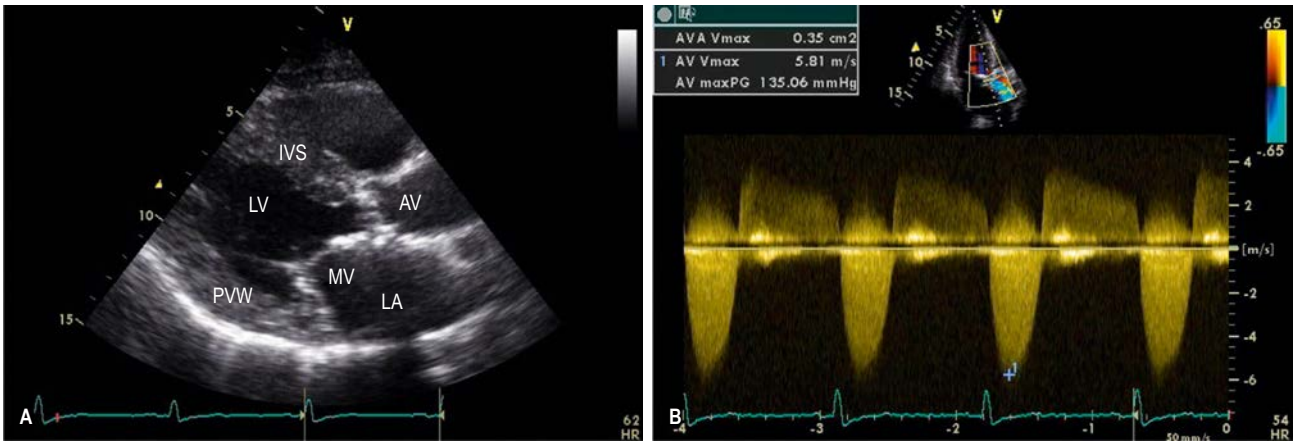


Fig. 17.79 Cardiac echograms. **(A)** Two-dimensional echocardiogram (long-axis view) in a patient with calcific aortic stenosis. The calcium in the valve generates abnormally intense echoes. There is some evidence of associated left ventricular hypertrophy. **(B)** Continuous-wave (CW) Doppler signals obtained from the right upper parasternal edge, where the high-velocity jet from the stenotic valve is coming towards the transducer. *AV,* Aortic valve; *IVS,* interventricular septum; *LA,* left atrium; *LV,* left ventricle; *MV,* mitral valve; *PVW,* posterior ventricular wall.

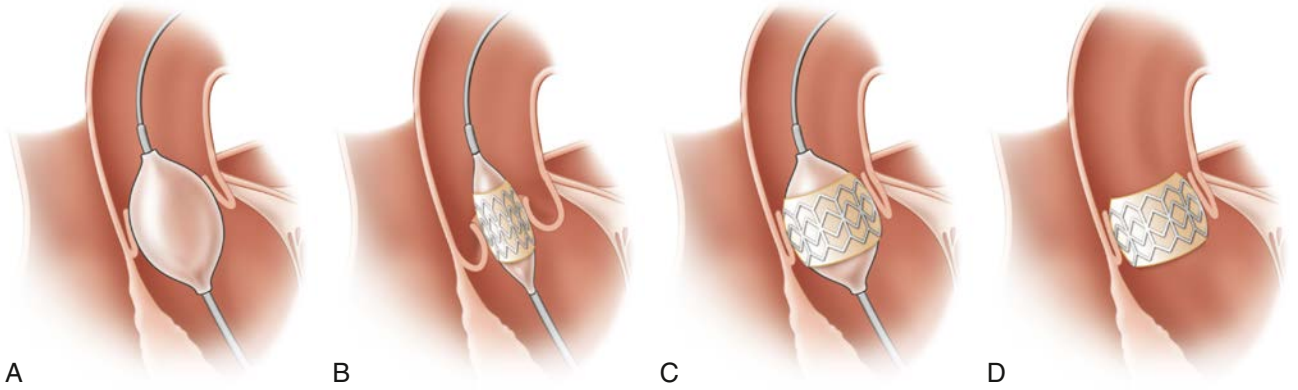


Fig. 17.80 Transcatheter aortic valve replacement (TAVR) procedure. **(A)** Balloon aortic valvuloplasty. **(B)** New valve placed around balloon. **(C)** Balloon inflated, deploying valve. **(D)** Catheter removed, valve in place. *(Concepts of Care for Patients With Cardiac Conditions. Medical-Surgical Nursing: Concepts for Clinical Judgment and Collaborative Care, Eleventh Edition. 2024. ISBN 978-0-323-87826-5.)*

cardiac output is to be maintained, the total volume of blood pumped into the aorta must increase and, consequently, left ventricular size must enlarge. Because of the aortic runoff during diastole, diastolic blood pressure falls, and coronary perfusion is decreased, whereas the increased stroke volume increases the systolic pressure, thereby also increasing the pulse pressure. In addition, the larger left ventricular size is mechanically less efficient, so the demand for oxygen is greater and cardiac ischaemia develops.

Clinical features

Symptoms

Patients may remain asymptomatic for many years until ***left ventricular failure*** develops with gradual exertional ***dyspnoea*** and lethargy, depending on the extent of left ventricular dilation and dysfunction. ***Angina pectoris*** may arise. ***Arrhythmias*** are relatively uncommon.

Signs

See Fig. 17.78. The signs of aortic regurgitation are many and are due to the high stroke volume, rapid return of blood back into the left ventricle, and increased left ventricular size.

The pulse is bounding or collapsing ('water hammer pulse') and there is a wide arterial pulse pressure (see 'Arterial pulse' section). The following signs, which are rare, also indicate a hyperdynamic circulation:

- ***Quincke's sign***—capillary pulsation in the nail beds
- ***de Musset's sign***—head nodding with each heartbeat
- ***Duroziez's sign***—a to-and-fro murmur heard when the femoral artery is auscultated with pressure applied distally (if found, it is a sign of severe aortic regurgitation)
- ***pistol shot femorals***—a sharp bang heard on auscultation over the femoral arteries in time with each heartbeat.

The apex beat is displaced laterally and downwards and is forceful in quality. On auscultation, there is a high-pitched early diastolic murmur best heard at the left sternal edge in the fourth intercostal space (Erb's point), with the patient leaning forwards and the breath held in expiration. Because of the volume overload there is commonly an ejection systolic flow murmur. The regurgitant jet can partially close the anterior mitral valve leaflet, causing a soft mid-diastolic murmur (Austin Flint rumble).

Investigations

Chest X-ray

The chest X-ray features are those of left ventricular enlargement and, possibly, dilation of the ascending aorta with calcification. The aortic valve may be calcified if valvular disease is responsible for the regurgitation.

ECG

The ECG reflects changes in the left ventricle with left ventricular hypertrophy due to 'volume overload': tall R waves and deeply inverted T waves in the left-sided chest leads, and deep S waves in the right-sided leads. Normally, sinus rhythm is present.

Echocardiogram

The severity of aortic regurgitation is assessed with a combination of colour Doppler (extent and width of the regurgitant jet; Fig. 17.81) and continuous-wave Doppler looking for diastolic flow reversal in the descending thoracic aorta, with characteristic pressure half-time measurements. Left ventricular volumes are increased and cardiac contraction is dynamic due to the increased stroke volume. The aortic root may also be enlarged. Diastolic fluttering of the mitral leaflets or septum occurs in severe aortic regurgitation (producing the Austin Flint rumble). TOE may provide additional information about the valves and aortic root.

Cardiac catheterization

Cardiac catheterization is appropriate for patients requiring valvular intervention, although CTCA or CT angiography may be an alternative in younger patients.

CMR and cardiac CT

These techniques may be indicated for assessing the thoracic aorta in cases of aortic dilation or dissection. CMR can be used to quantify left ventricular volumes and regurgitation severity.

Management

The underlying cause of aortic regurgitation (e.g., IE or syphilitic aortitis) may require specific treatment. Patients with acute aortic regurgitation may need treatment with vasodilators and inotropes. Patients should receive treatment for hypertension. ACE inhibitors are indicated in patients with left ventricular dysfunction and beta-blockers may slow aortic dilation in patients with MFS by reduction in shear stress. Patients should be followed by serial echocardiography as part of a valve clinic. As symptoms do not develop until the

? Box 17.40 Causes and associations of aortic regurgitation

Acute aortic regurgitation
- Acute rheumatic fever
- Infective endocarditis
- Dissection of the aorta
- Ruptured sinus of Valsalva aneurysm
- Failure of prosthetic valve

Chronic aortic regurgitation
- Rheumatic heart disease
- Syphilis
- Arthritides:
  - Reactive arthritis
  - Ankylosing spondylitis
  - Rheumatoid arthritis
- Hypertension (severe)
- Bicuspid aortic valve
- Aortic endocarditis
- Marfan's syndrome
- Osteogenesis imperfecta

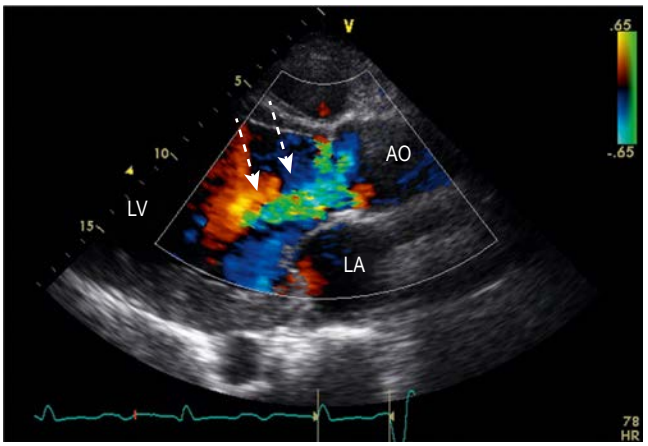


Fig. 17.81 Aortic regurgitation: colour Doppler. A regurgitant jet *(green)* can be seen passing retrogradely into the left ventricle from the aorta. *AO,* Aorta; *LA,* left atrium; *LV,* left ventricle.

myocardium fails, and because the myocardium may not recover fully after surgery, operative valve replacement may be performed before significant symptoms occur.

Aortic surgery is indicated in:

- acute severe aortic regurgitation, for example, as a result of endocarditis
- symptomatic patients (dyspnoea, NYHA class II–IV, angina) with chronic severe aortic regurgitation
- asymptomatic patients with an LVEF of less than or equal to 50%
- asymptomatic patients with an LVEF of more than 50% but with a dilated left ventricle (end-systolic dimension >50 mm)
- those undergoing CABG or surgery of the ascending aorta or other cardiac valve.

Both mechanical prostheses and tissue valves are used. Tissue valves are preferred in the elderly and in cases where anticoagulants must be avoided but are contraindicated in children and young adults because of the rapid calcification and degeneration of the valves.

Antibiotic prophylaxis against IE (see 'Prevention' section) is not recommended.

TRICUSPID VALVE

Tricuspid stenosis

This uncommon valve lesion, which is seen much more often in women than in men, is usually due to rheumatic heart disease and is frequently associated with mitral and/or aortic valve disease. Tricuspid stenosis is also seen in carcinoid syndrome and ergotamine treatment.

Pathophysiology

Tricuspid valve stenosis results in a reduced cardiac output, which is restored towards normal when the right atrial pressure increases. The resulting systemic venous congestion leads to hepatomegaly, ascites, and dependent oedema.

Clinical features

Symptoms

Patients with tricuspid stenosis are likely to have left-sided rheumatic valve disease that may be the main driver of symptoms which include abdominal pain (hepatomegaly), abdominal distension (ascites), and peripheral oedema.

Signs

See Fig. 17.82. The patient usually is in sinus rhythm and there is a prominent jugular venous *a*-wave. This pre-systolic pulsation may also be felt over the liver. There is usually a rumbling mid-diastolic murmur, which is heard best at the lower left sternal edge and is louder on inspiration. It may be missed because of the murmur of coexisting mitral stenosis. A tricuspid OS may occasionally be heard. Hepatomegaly, abdominal ascites, and dependent oedema may be present.

Investigations

Chest X-ray may show a prominent right atrial bulge. ECG demonstrates an enlarged right atrium as shown by peaked, tall P waves (>3 mm) in lead II. Echocardiography may show a thickened and immobile tricuspid valve and can quantify the severity of stenosis and other valvular lesions.

Management

Medical management consists of diuretic therapy and salt restriction. Tricuspid valvotomy is occasionally possible but tricuspid valve replacement is often necessary. Usually, other valves also need replacement because tricuspid valve stenosis is rarely an isolated problem.

Tricuspid regurgitation

Functional tricuspid regurgitation (see Fig. 17.23) may occur whenever the right ventricle dilates: for example, in cor pulmonale, MI, or pulmonary hypertension.

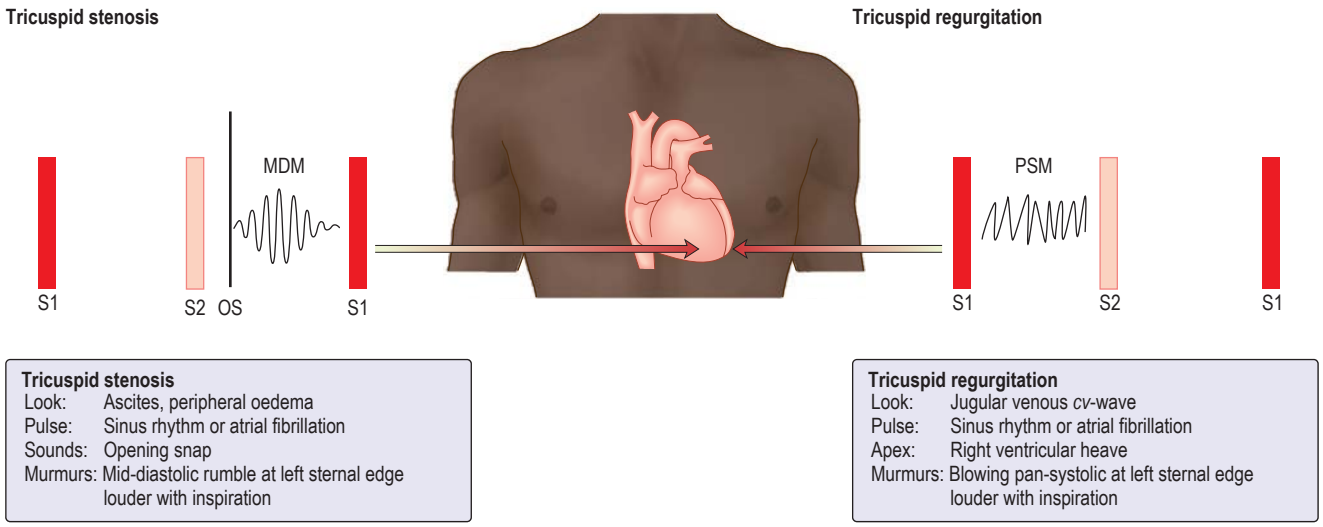


Fig. 17.82 Features of tricuspid stenosis and tricuspid regurgitation. *MDM,* Mid-diastolic murmur; *OS,* opening snap; *PSM,* pansystolic murmur; *S1,* first heart sound; *S2,* second heart sound.

Organic tricuspid regurgitation may occur with rheumatic heart disease, IE, carcinoid syndrome, Ebstein's anomaly (a congenitally malpositioned tricuspid valve), and other congenital abnormalities of the AV valves.

Clinical features

The valvular regurgitation gives rise to high right atrial and systemic venous pressures. Patients may experience the symptoms of ***right heart failure***.

Signs (see Fig. 17.82) include a large jugular venous '*cv*'-wave and a palpable liver that pulsates in systole. Usually, a right ventricular impulse may be felt at the left sternal edge, and there is a blowing pansystolic murmur, best heard on inspiration at the lower left sternal edge. Atrial fibrillation is common.

Investigations

An ***echocardiogram*** can assess the tricuspid leaflets (e.g., displaced in Ebstein anomaly, thickened in carcinoid, vegetations in endocarditis) and quantify dilation of the right ventricle.

Management

Functional tricuspid regurgitation usually improves with medical management. Severe organic tricuspid regurgitation may require operative repair of the tricuspid valve (annuloplasty or annuloplication) and very occasionally, tricuspid valve replacement may be necessary. In intravenous drug users with IE of the tricuspid valve, surgical removal of the valve is recommended to eradicate the infection. This is usually well tolerated in the short term. The insertion of a prosthetic valve for this condition is sometimes necessary.

PULMONARY VALVE

Pulmonary stenosis

This is usually a congenital lesion but may rarely result from rheumatic fever or from the carcinoid syndrome. Congenital pulmonary stenosis may be associated with Fallot's tetralogy, Noonan's syndrome, or congenital rubella syndrome. Pulmonary stenosis may be valvular, subvalvular, or supravalvular.

Clinical features

The obstruction to right ventricular emptying results in right ventricular hypertrophy, which, in turn, leads to right atrial hypertrophy. Severe ***pulmonary obstruction*** may be incompatible with life, but lesser degrees of obstruction give rise to ***fatigue***, ***syncope***, and the symptoms of ***right heart failure***. Mild pulmonary stenosis may be asymptomatic.

Physical signs (Fig. 17.83) are characterized by a harsh midsystolic ejection murmur, best heard on inspiration, to the left of the sternum in the second intercostal space. This murmur is often associated with a thrill. The pulmonary closure sound is usually delayed and soft. There may be a pulmonary ejection sound if the obstruction is valvular. A right ventricular fourth sound and a prominent jugular venous *a*-wave are present when the stenosis is moderately severe. A right ventricular heave (sustained impulse) may be felt.

Investigations

The chest X-ray usually shows a prominent pulmonary artery owing to post-stenotic dilation. The ECG demonstrates both right atrial and right ventricular hypertrophy, although it may sometimes be normal, even in severe pulmonary stenosis. Echocardiography can quantify the stenosis, assess the right heart, and look for other valve lesions or congenital cardiac abnormalities.

Management

Management of severe pulmonary stenosis requires pulmonary valvotomy (balloon valvotomy or direct surgery).

Pulmonary regurgitation

Physiological or mild pulmonary regurgitation is a common finding. Secondary pulmonary regurgitation is seen from dilation of the pulmonary valve ring, which occurs in pulmonary hypertension (Graham Steell murmur). It may also occur following tetralogy of Fallot repair. It is characterized by a decrescendo diastolic murmur, beginning

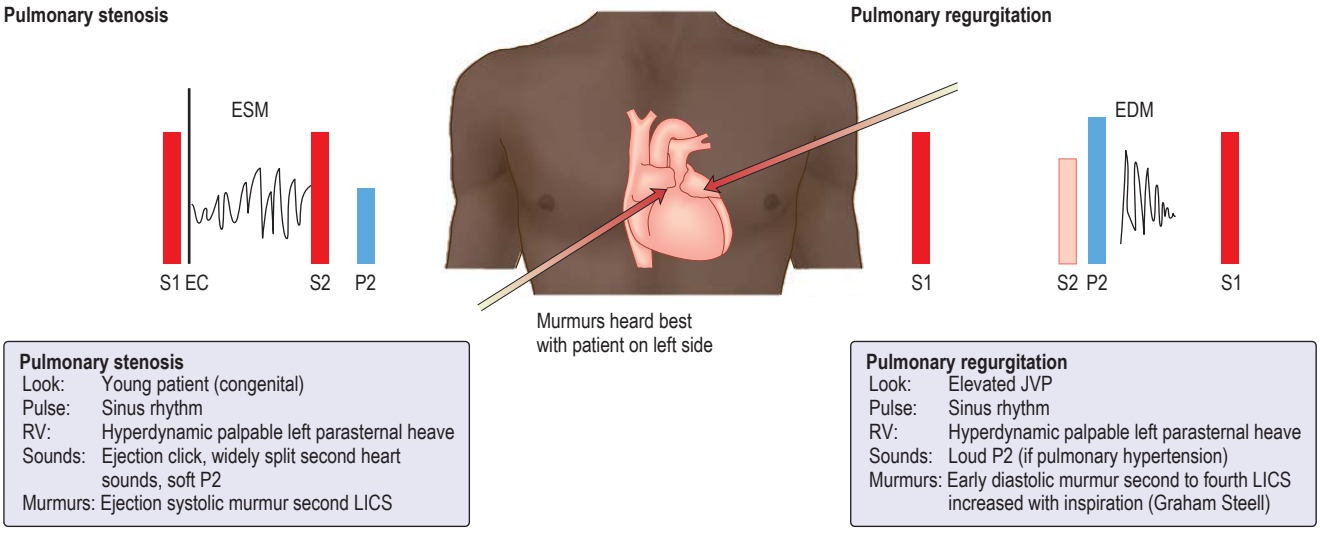


Fig. 17.83 Features of pulmonary stenosis and pulmonary regurgitation. *EC,* Ejection click; *EDM,* early diastolic murmur; *ESM,* end-systolic murmur; *ICS,* intercostal space; *LICS,* left intercostal space; *JVP,* jugular venous pulse; *P2,* pulmonary component of the second heart sound; *S1,* first heart sound; *S2,* second heart sound.

with the pulmonary component of the second heart sound that is difficult to distinguish from the murmur of aortic regurgitation ('Clinical memo', Fig. 17.83). Pulmonary regurgitation usually causes no symptoms and treatment is rarely necessary.

PROSTHETIC VALVES

There is no ideal replacement for our own normally functioning, native heart valves. There are two options for valve prostheses: mechanical (Fig. 17.84) or tissue (bioprosthetic).

The valves consist of two basic components: an opening to allow blood to flow through and an occluding mechanism to regulate the flow. Mechanical prostheses rely on artificial occluders: double tilting disc (St Jude), tilting disc (Björk–Shiley), or a ball and cage (Starr–Edwards). Tissue prostheses are derived from human (homograft) or porcine or bovine (xenograft) sources. A valve replacement from within the same patient (i.e., pulmonary to aortic valve position) is termed an autograft.

Mechanical versus tissue valves

Mechanical valves, being artificial structures, are more durable than their tissue counterparts, which tend to degenerate after 10 years. However, artificial structures are more thrombogenic. Mechanical valves require formal anticoagulation for the lifetime of the prosthesis. The target INR is determined by what type of valve is inserted, where it is positioned (risk of thrombosis higher for mitral than aortic valves), and whether the patient has additional risk factors for thromboembolism (mitral, tricuspid, and pulmonary disease is higher risk than aortic valve disease; previous thromboembolism; atrial fibrillation; left atrial diameter >50 mm; mitral stenosis; LVEF < 35%; hypercoagulable states, see Chapter 40).

All mechanical valves require anticoagulation, the precise strategy being dictated by the specifics of the prosthesis and other factors:

- ***low-thrombogenicity prosthesis*** (Carbomedics [aortic position], Medtronic Hall, St Jude Medical [without silzone]): INR 2.5 without and 3.0 with additional risk factors
- ***medium-thrombogenicity prosthesis*** (Björk–Shiley, other bileaflet valves): INR 3.0 without and 3.5 with additional risk factors
- ***high-thrombogenicity prosthesis*** (Lillehei–Kaster, Omniscience, Starr–Edwards): INR 3.5 without and 4.0 with additional risk factors.

Tissue valves require anticoagulation for a limited postoperative period only, while the suture lines endothelialize (the ESC recommends 3 months with a target INR of 2.5, although some centres use low-dose aspirin 75–100 mg daily). This can then be discontinued unless another risk factor for thromboembolism (e.g., atrial fibrillation) persists. There is currently no role for DOACs.

On auscultation, tissue valve heart sounds are comparable to those of a native valve. Mechanical valve heart sounds are generally louder, and both opening and closing sounds can be heard.

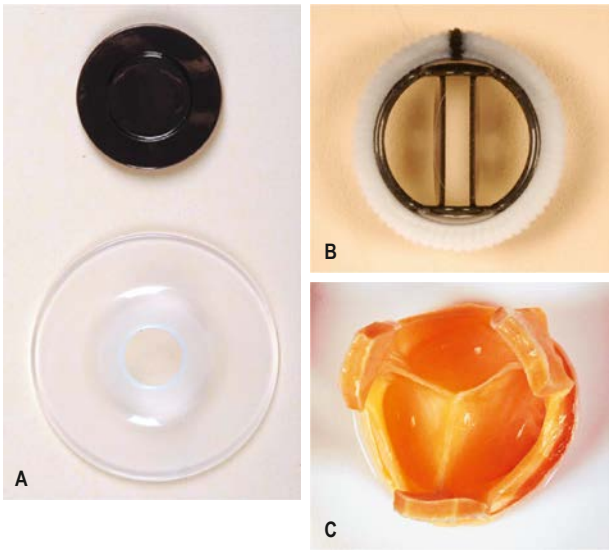


Fig. 17.84 Prosthetic valves. **(A)** Björk–Shiley mechanical prosthetic valve. **(B)** St Jude double tilting disc. **(C)** Aortic valve tissue prosthesis (aortic view).

Complications

All prostheses carry a risk of infection. Prosthetic valve endocarditis is associated with significant morbidity and mortality, and prevention is the cornerstone of management. Patient education regarding dental hygiene, avoidance of piercings and tattoos, and antibiotic prophylaxis is vital and should be reinforced at clinic visits. Details of antibiotic prophylaxis indicated in patients undergoing different types of interventional procedures are given in Box 12.8. Additional prophylaxis indicated in patients at higher risk of endocarditis is discussed in the 'Prevention' section. This must be borne in mind when managing a patient with a prosthetic heart valve, and steps should be taken to minimize the risk involved.

The prosthetic valve occluding mechanism can be interrupted by vegetations but also by thrombosis and calcification, resulting in either stenosis or regurgitation. The prosthesis can become detached from the valve ring, resulting in a ***para-prosthetic leak***. Haemolysis and anaemia from damage to red blood cells have become less common with current valves but may be seen in paravalvular regurgitation. Evidence of structural failure can be detected by simple auscultation, with echocardiography as the initial investigation of choice. TTE is non-invasive but scattering of echoes by mechanical valves makes assessment difficult. TOE provides alternative views and higher image resolution, making it the preferred investigation when prosthetic valve endocarditis is suspected.

Interruption of anticoagulant therapy

For minor surgical procedures, including dental extraction and diagnostic endoscopy, anticoagulation should not be interrupted, and the target INR may be reduced to 2.0 dependent on the valve. Percutaneous arterial puncture is safe with an INR below 2.0, although radial catheterization is safe at higher INR levels. For major surgical procedures, anticoagulation should be discontinued 5 days before the procedure and intravenous UFH or subcutaneous LMWH commenced when the INR is below 2.0.

Pregnancy and prosthetic heart valves

The types of valve prosthesis for use in women of childbearing age are as follows:

- ***Bioprosthetic valves*** are preferable during pregnancy, as they are less thrombogenic and do not require anticoagulation. However, valve degradation in women of childbearing age has been shown to be as high as 50% at 10 years and 90% at 15 years; women with a bioprosthetic valve are likely to require redo valve surgery.

- ***Mechanical heart valves*** have excellent durability but are thrombogenic and require life-long anticoagulation with warfarin. Pregnancy is a hypercoagulable state due to increased levels of fibrinogen and factors VII, VIII, and X; decreased levels of protein S activity; venous hypertension; and stasis.

Pregnancy in women with a mechanical heart valve is associated with increased maternal mortality (1%–4%) due to valve thrombosis. There is no anticoagulation regime that is ideal in these patients, so care should be managed by a pregnancy Heart Team who can balance the risks and benefits to the fetus and mother. Warfarin crosses the placenta and is associated with a 5%–12% risk of embryopathy during the first trimester. Warfarin also has an anticoagulant effect in the fetus, which may lead to spontaneous fetal intracranial haemorrhage. UFH or LMWH is preferable, particularly if doses of warfarin higher than 5 mg are required, as these drugs do not cross the placenta or cause fetal embryopathy. However, UFH may not provide consistent therapeutic anticoagulation during pregnancy, and there is a high incidence (25%) of valve thrombosis. LMWH provides a more consistent anticoagulant effect when given twice daily, with dose adjustment depending on the type of valve prosthesis to maintain anti-Xa levels.

Further reading

Lancellotti P, Vannan MA. Timing of Intervention in aortic stenosis. *N Engl J Med* 2020; 382:191–193.
Nishimura RA, Holmes DR. Treatment after TAVR—discordance and clinical implications. *N Engl J Med* 2020; 382:193–194.
Vahanian A, Beyersdorf F, Praz F et al. 2021 ESC/EACTS Guidelines for the management of valvular heart disease: Developed by the Task Force for the management of valvular heart disease of the European Society of Cardiology (ESC) and the European Association for Cardio-Thoracic Surgery (EACTS). *Eur Heart J* 2022; 43(7):561–632.

INFECTIVE ENDOCARDITIS

Infective endocarditis is an endovascular infection of cardiovascular structures, including cardiac valves, the atrial and ventricular endocardium, the large intrathoracic vessels, and intracardiac foreign bodies, such as prosthetic valves, pacemaker leads, and surgical conduits. The annual incidence in the UK is 6 to 7/100 000 but higher in low- and middle-income countries. Without treatment, mortality approaches 100%; even with treatment, there is significant morbidity and mortality.

Aetiology

Endocarditis is usually the consequence of two factors: the presence of organisms in the bloodstream and abnormal cardiac endothelium that facilitates their adherence and growth.

Bacteraemia may arise for patient-specific reasons (poor dental hygiene, intravenous drug use, soft tissue infections) or may be associated with diagnostic or therapeutic procedures (dental treatment, intravascular cannulae, cardiac surgery, or permanent pacemakers).

Damaged endocardium promotes platelet and fibrin deposition, which allows organisms to adhere and grow, leading to an infected vegetation. Valvular lesions may create non-laminar flow, and jet lesions from septal defects or a patent ductus arteriosus (PDA) result in abnormal vascular endothelium. Aortic and mitral valves are most commonly involved in IE; intravenous drug users are the exception, as right-sided lesions are more common in these patients.

Organisms

Common organisms and the sources of infection are shown in **Fig. 17.85**.

Rare causes

These include the HACEK group of organisms, which tend to run a more insidious course (**Box 17.41**).

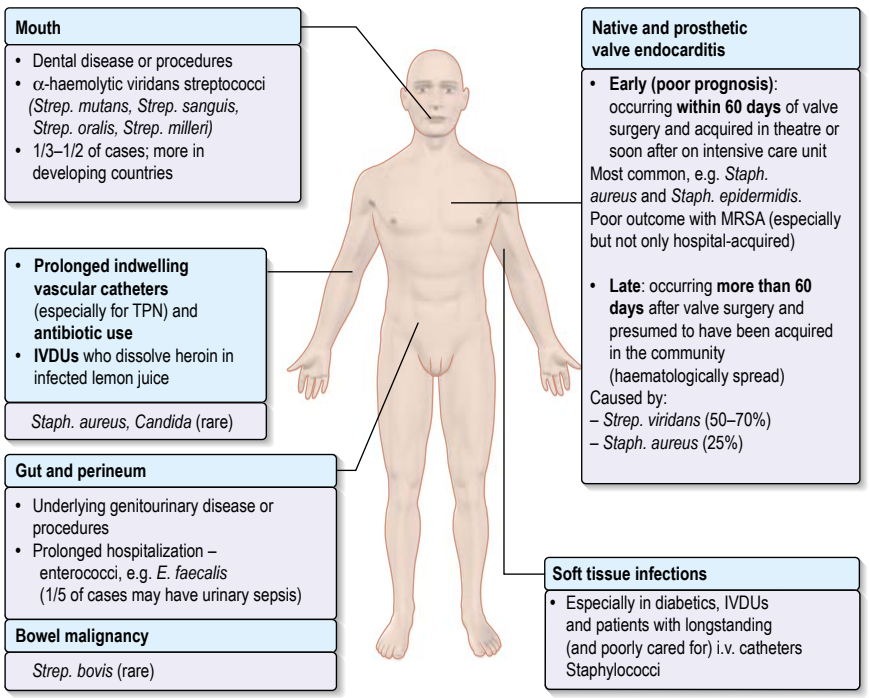


Fig. 17.85 Infective endocarditis: aetiology and sources of infection. *IVDU,* Intravenous drug user; *MRSA,* methicillin-resistant *Staphylococcus aureus*; *TPN,* total parenteral nutrition.

Box 17.41 Modified Duke criteria for endocarditis[a]

I. MAJOR CRITERIA
- **A. Microbiologic Major Criteria**
  - (1) Positive blood cultures
    - i. Microorganisms that commonly cause IE isolated from 2 or more separate blood culture sets (Typical)
    - *or*
    - ii. Microorganisms that occasionally or rarely cause IE isolated from 3 or more separate blood culture sets (Nontypical)
  - (2) Positive laboratory tests
    - i. Positive polymerase chain reaction (PCR) or other nucleic acid-based technique for *Coxiella burnetii*, *Bartonella* species, or *Tropheryma whipplei* from blood
    - *or*
    - ii. *C. burnetiid* antiphase I immunoglobulin G (IgG) antibody titer >1:800, or isolated from a single blood culture
    - *or*
    - iii. Indirect immunofluorescence assays (IFAs) for detection of IgM and IgG antibodies to *Bartonella henselae* or *Bartonella quintana* with immunoglobulin G (IgG) titer ≥1:800
- **B. Imaging Major Criteria**
  - (1) Echocardiography and cardiac computed tomography (CT) imaging
    - i. Echocardiography and/or cardiac CT showing vegetation, valvular/leaflet perforation, valvular/leaflet aneurysm, abscess, pseudoaneurysm, or intracardiac fistula
    - or
    - ii. Significant new valvular regurgitation on echocardiography as compared with previous imaging. Worsening or changing of pre-existing regurgitation is not sufficient.
    - or
    - iii. New partial dehiscence of prosthetic valve as compared with previous imaging
  - (2) Positron emission computed tomography with ^{18}F-fluorodeoxyglucose ([^{18}F]FDG PET/CT imaging)
    Abnormal metabolic activity involving a native or prosthetic valve, ascending aortic graft (with concomitant evidence of valve involvement), intracardiac device leads, or other prosthetic material
- **C. Surgical Major Criteria**
  Evidence of IE documented by direct inspection during heart surgery neither Major Imaging Criteria nor subsequent histologic or microbiologic confirmation

II. **MINOR CRITERIA**
- A. Predisposition
  - Previous history of IE
  - Prosthetic valve
  - Previous valve repair
  - Congenital heart disease
  - More than mild regurgitation or stenosis of any aetiology
  - Endovascular intracardiac implantable electronic device (CIED)
  - Hypertrophic obstructive cardiomyopathy
  - Injection drug use
- B. Fever
  Documented temperature greater than 38.0°C (100.4°F)
- C. Vascular Phenomena
  Clinical or radiological evidence of arterial emboli, septic pulmonary infarcts, cerebral or splenic abscess, mycotic aneurysm, intracranial haemorrhage, conjunctival haemorrhages, Janeway lesions, purulent purpura
- D. Immunologic Phenomena
  Positive rheumatoid factor, Osler nodes, Roth spots, or immune complex-mediated glomerulonephritis
- E. Microbiologic Evidence, Falling Short of a Major Criterion
  1) Positive blood cultures for a microorganism consistent with IE but not meeting the requirements for Major Criterion
  or
  2) Positive culture, PCR, or other nucleic acid-based test (amplicon or shotgun sequencing, *in situ* hybridization) for an organism consistent with IE from a sterile body site other than cardiac tissue, cardiac prosthesis, or arterial embolus; or a single finding of a skin bacterium by PCR on a valve or wire without additional clinical or microbiological supporting evidence
- F. Imaging Criteria
  Abnormal metabolic activity as detected by [^{18}F]FDG PET/CT within 3 mo of implantation of prosthetic valve, ascending aortic graft (with concomitant evidence of valve involvement), intracardiac device leads, or other prosthetic material
- G. Physical Examination Criteria
  New valvular regurgitation identified on auscultation if echocardiography is not available. Worsening or changing of pre-existing murmur not sufficient

[a]The diagnosis of infective endocarditis is definite when: (1) a microorganism is demonstrated by culture of a specimen from a vegetation, an embolism or an intracardiac abscess; (2) active endocarditis is confirmed by histological examination of the vegetation or intracardiac abscess; (3) two major clinical criteria, one major and three minor criteria, or five minor criteria are met.
Table 2. 2023 Duke-International Society for Cardiovascular Infectious Diseases Infective Endocarditis (IE) Criteria for the Diagnosis of IE, With Proposed Changes in Bold Type.
(From https://academic.oup.com/cid/article/77/4/518/7151107?login=true#415413859.*)*

Culture-negative endocarditis

This accounts for 5% to 10% of endocarditis cases. The usual cause is prior antibiotic therapy (good history-taking is vital), but some cases are due to a variety of fastidious organisms that fail to grow in normal blood cultures. These include *Coxiella burnetii* (the cause of Q fever), *Chlamydia* species, *Bartonella* species (organisms that cause trench fever and cat scratch disease), and *Legionella*.

Clinical features

The clinical presentation of IE (**Box 17.42**) is dependent on the organism and the presence of predisposing cardiac conditions. IE may occur as an acute, fulminating infection but also as a chronic or subacute illness with low-grade fever and non-specific symptoms. A high index of clinical suspicion is required to identify patients with IE, and certain criteria should alert the physician.

Box 17.42 Clinical features of infective endocarditis

Clinical feature	Approximate %
General	
Malaise	95
Clubbing	10
Cardiac	
Murmurs	90
Cardiac failure	50
Arthralgia	25
Pyrexia	90
Skin lesions	
Osler nodes	15
Splinter haemorrhages	10
Janeway lesions	5
Petechiae	50
Eyes	
Roth spots	5
Conjunctival splinter haemorrhages	Rare
Splenomegaly	40
Neurological	
Cerebral emboli	20
Mycotic aneurysm	10
Renal	
Haematuria	70

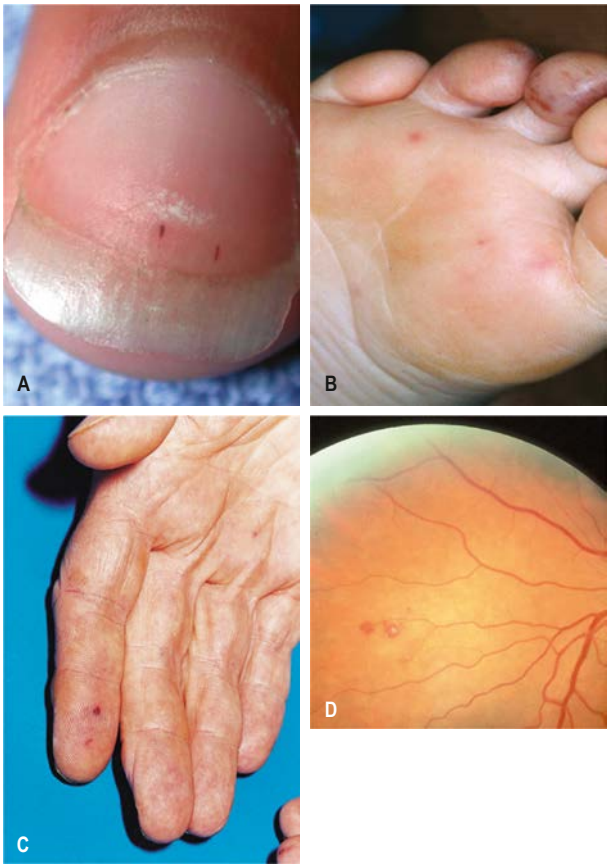


Fig. 17.86 Infective endocarditis. **(A)** Splinter haemorrhages. **(B)** Janeway lesions. **(C)** Osler nodes. **(D)** Roth spots. *(A, B From Moser DK, Riegel B.* Cardiac Nursing. *Philadelphia: Saunders; 2007, p. 1127, with permission from Elsevier. C From Forbes CD, Jackson WF.* Color Atlas and Text of Clinical Medicine, *3rd edn. St Louis: Mosby; 2003, p. 232; © Elsevier. D. Courtesy of Professor Ian Constable.)*

High clinical suspicion

- New valve lesion/(regurgitant) murmur
- Embolic event(s) of unknown origin
- Sepsis of unknown origin
- Haematuria, glomerulonephritis, and suspected renal infarction
- 'Fever' plus:
  - Prosthetic material inside the heart
  - Other high predisposition for IE, for example, intravenous drug use
  - Newly developed ventricular arrhythmias or conduction disturbances
  - First manifestation of congestive cardiac failure
  - Positive blood cultures (with typical organism)
  - Cutaneous (Osler, Janeway) or ophthalmic (Roth) lesions (Fig. 17.86)
  - Peripheral abscesses (renal, splenic, spine) of unknown origin
  - Predisposition and recent diagnostic/therapeutic interventions known to result in significant bacteraemia.

Diagnosis

The 2023 Duke-ISCVID criteria are used for the clinical diagnosis of endocarditis (see Box 17.41). These reflect the changing epidemiology of IE and the utility of the latest microbiologic testing and advanced imaging tools. Confirmation of 'Definite IE' requires pathological confirmation of active endocarditis from cardiac tissue, cardiac device, or an arterial embolus or clinical criteria from a combination of major and/or minor criteria. Major microbiological criteria include positive blood cultures or polymerase chain reaction (PCR) testing used alongside major imaging criteria with echocardiography, CT, or CT/PET.

Investigations

Blood cultures are the mainstay of diagnosis of IE to identify the organism and ensure appropriate targeted antimicrobial therapy (Box 17.43).

Echocardiography is the first-line imaging technique if used appropriately and can identify vegetations and paravalvular infection. It is not an appropriate screening test for patients with just a fever or an isolated positive blood culture when the pretest probability of endocarditis is low. A negative echocardiogram does not exclude a diagnosis of endocarditis.

Transoesophageal echocardiography (TOE) is invariably used to confirm the diagnosis and clarify the extent of infection with a higher specificity and sensitivity.

^{18}F-FDG PET/CT is particularly helpful in confirming infection in patients with prosthetic valves, cardiac devices, grafts, and individuals with congenital heart disease and can also assess distant infection, for example, discitis and septic emboli to the abdomen.

Other tests: The ECG may note a prolonged PR interval, or there may be heart block suggesting paravalvular infection, for example, aortic root abscess. Chest X-ray may demonstrate septic emboli to the lungs from right-sided lesions. CT may also confirm emboli to the lungs, kidneys, liver, and spleen or distant arterial emboli to the brain or peripheries.

Box 17.43 Investigations and findings in endocarditis

Investigation	Findings and notes
Blood cultures	Three sets from different venepuncture sites
Serological tests	Consider in culture-negative cases for *Coxiella*, *Bartonella*, *Legionella*, *Chlamydia*
Full blood count	Reduced haemoglobin, increased white cells, increased or reduced platelets
Serum creatinine and electrolytes	Increased urea and creatinine
Liver biochemistry	Increased serum alkaline phosphatase
Inflammatory markers	Increased erythrocyte sedimentation rate and C-reactive protein (CRP reduces in response to therapy and increases with relapse)
Complement proteins	C3 and C4 complement proteins commonly consumed due to immune complex formation.
Urine	Proteinuria and haematuria
Electrocardiogram	PR prolongation/heart block is associated with aortic root abscess
Chest X-ray	Pulmonary oedema in left-sided disease, pulmonary emboli/abscess in right-sided disease
Transthoracic echocardiography	First-line non-invasive imaging test with sensitivity of 60%–75%; demonstrates vegetations, valvular dysfunction, ventricular function, abscesses
Transoesophageal echocardiography	Second-line invasive imaging test with greater sensitivity (>90%) and specificity; useful in suspected aortic root abscess and essential in prosthetic valve endocarditis

Management

The location of the infection means that prolonged courses of antibiotics are usually required. The combination of antibiotics may be synergistic in eradicating microbial infection and minimizing resistance. Blood cultures should be taken prior to empirical antibiotic therapy (but this should not delay therapy in unstable patients). Antibiotic treatment should continue for 4 to 6 weeks and, in some cases, may be longer. Typical therapeutic regimens are shown in **Box 17.44**, but advice on specific therapy should be sought from the local microbiology department, according to the organism identified and current sensitivities. Serum levels of gentamicin and vancomycin need to be monitored to ensure adequate therapy and prevent toxicity. In patients with penicillin allergy one of the glycopeptide antibiotics, vancomycin or teicoplanin, can be used. Penicillins, however, are fundamental to the therapy of bacterial endocarditis; allergies therefore seriously compromise the choice of antibiotics. It is essential to confirm the nature of a patient's allergy to ensure that the appropriate treatment is not withheld needlessly. Anaphylaxis would be much more influential on antibiotic choice than a simple gastrointestinal disturbance.

Persistent fever

Most patients with IE should respond within 48 hours of initiation of appropriate antibiotic therapy, as evidenced by a resolution of fever, reduction in serum markers of infection, and relief of systemic symptoms of infection. Failure of these factors to occur needs to be taken very seriously. The following should be considered:

- perivalvular extension of infection and possible abscess formation
- drug reaction (the fever should resolve promptly after drug withdrawal)
- hospital-acquired infection (i.e., venous access site, urinary tract infection)
- pulmonary embolism (secondary right-sided endocarditis or prolonged hospitalization).

In such cases samples for culture should be taken from all possible sites and evidence sought of the above causes. A change of antibiotic dosage or regimen should be avoided unless there are positive cultures or a drug reaction is suspected. Emergence of bacterial resistance is uncommon. Close liaison with the microbiology department is recommended and a cardiothoracic surgical opinion should be sought.

Box 17.44 Recommended antibiotic regimens in infective endocarditis[a]

Clinical situation	Suggested antibiotic regimen to start[b,c,d]
Clinical endocarditis, culture results awaited, no suspicion of staphylococci (native valve or late prosthetic valve >12 months after surgery)	Amoxicillin 2 g 4-hourly Ceftriaxone 2 g 12-hourly Gentamicin 3 mg/kg 24-hourly
Suspected staphylococcal endocarditis (recent intravascular devices or cardiac surgery <12 months, intravenous drug user)	Amoxicillin 2 g 4-hourly Flucloxacillin 2 g 4-hourly Gentamicin 3 mg/kg 24 hourly
Streptococcal endocarditis (penicillin-sensitive) (4 weeks native; 6 weeks prosthetic)	Amoxicillin 2 g 4-hourly Gentamicin 3 mg/kg 24 hourly (2 weeks)
Enterococcal endocarditis (amoxicillin-susceptible Enterococci, usually *Enterococcus faecalis*) (at least 6 weeks for native and prosthetic)	Amoxicillin 2 g 4-hourly and ceftriaxone 2 g for 6 weeks OR Amoxicillin 2 g 4-hourly for 6 weeks and gentamicin 3 mg/kg 24 hourly for 2 weeks
Staphylococcal endocarditis (native valve) (4–6 weeks)	Flucloxacillin 2 g 4-hourly (first line) *or* Vancomycin (high dose 15–20 mg/L)
Staphylococcal endocarditis (prosthetic valve) (6 weeks)	Flucloxacillin 2 g 4-hourly and Rifampicin 600 mg orally 12-hourly and Gentamicin 3 mg/kg 24-hourly
Staphylococcal endocarditis (prosthetic valve; meticillin resistant organism) (6 weeks)	Vancomycin (high dose 15–20 mg/L) and Rifampicin 600 mg orally 12-hourly and Gentamicin 3 mg/kg 24-hourly

[a] Adapted from 2023 ESC Guidelines for the management of endocarditis.
[b] Optimum choice and duration of therapy needs close liaison with microbiology department. All antibiotics given intravenously except rifampicin. If penicillin allergic may require alternative. OPAT management possible in selected cases following discussion with microbiology.
[c] Monitor vancomycin and gentamicin levels and adjust if necessary. Gentamicin, trough (pre-dose) concentrations should be <1 mg/L and peak (1 hour post dose) concentrations should be 10–12 mg/L. Vancomycin trough (pre-dose) concentration should be 15–20 mg/L.
[d] Choice of antibiotic depends on sensitivities.

Surgery

Decisions about surgical intervention in patients with IE should be made by specialist endocarditis teams involving cardiologists, cardiothoracic surgeons, and microbiologists, taking into account patient-specific features (age, non-cardiac morbidities, presence of prosthetic material or cardiac failure) and IE features (infective organism, vegetation size, presence of perivalvular infection, systemic embolization).

Prevention

In 2023 the ESC produced guidelines for the management of IE. They identified patients who are at higher risk of developing IE, including:

- those with a prior history of previous endocarditis
- those with prosthetic valves (including transcatheter devices) or material used for valve repair
- those with congenital heart disease including cyanotic congenital heart disease, repair with prosthetic material, or those with palliative shunts
- those with VADs.

The ESC recommends that these groups should receive antibiotic prophylaxis during high-risk procedures. This includes dental procedures that involve manipulation of the gingival or periapical part of the teeth or perforation of the oral mucosa. (The American Heart Association also considers that cardiac transplant patients with valvular heart disease should be included as highest-risk patients.) This is not in line with NICE guidance, which states that 'antibiotic prophylaxis against IE is not *routinely* recommended for people undergoing dental procedures'.

The ESC also provided additional recommendations applicable to all patients with valvular heart disease (including the highest-risk patients). These include:

- regular dental check-ups (6 months for the highest-risk groups and 12 months for all others)
- disinfection of wounds and eradication of chronic bacterial carriage (skin, urine)
- curative antibiotics for any focus on bacterial infection
- no self-medication with antibiotics
- strict infection control during at-risk procedures
- avoidance of piercing or tattoos
- limitation of the usage of infusion catheters, preferring peripheral versus central catheters; peripheral catheters should be changed every 3 to 4 days.

Further reading

Delgado V, Ajmone Marsan N, de Waha S et al. 2023 ESC Guidelines for the management of endocarditis: Developed by the task force on the management of endocarditis of the European Society of Cardiology (ESC) Endorsed by the European Association for Cardio-Thoracic Surgery (EACTS) and the European Association of Nuclear Medicine (EANM). *Eur Heart J* 2023.

Fowler V, Durack D, Selton-Suty C et al. The 2023 Duke-international society for cardiovascular infectious diseases criteria for infective endocarditis: updating the modified duke criteria. *Clin Infect Dis* 2023; 77(4):518–526.

CONGENITAL HEART DISEASE

Fetal circulation

Oxygenated blood and nutrients are supplied to the developing fetus via the placenta and the umbilical vein (Fig. 17.87). Half of that blood is directed to the fetal ductus venosus and carried to the inferior vena cava and the remainder enters the liver. Blood moves from the inferior vena cava to the right atrium of the heart. In the fetus there is an opening between the right and left atrium (the foramen ovale), and most of the blood flows from the right into the left atrium, bypassing the pulmonary circulation. This blood goes into the left ventricle and is pumped through the aorta into the fetal body. Some of the blood flows from the aorta through the internal iliac arteries to the umbilical arteries and re-enters the placenta, where carbon dioxide and other waste products from the fetus are taken up and enter the woman's circulation.

Some of the blood from the right atrium does not enter the left atrium but rather enters the right ventricle and is pumped into the pulmonary artery. In the fetus there is a connection between the pulmonary artery and the aorta, the ductus arteriosus, which directs most of this blood away from the lungs. With the first breath after delivery, the lungs expand, vascular resistance in the pulmonary arteries falls, more blood moves from the right atrium to the right ventricle and pulmonary arteries, and oxygenated blood travels

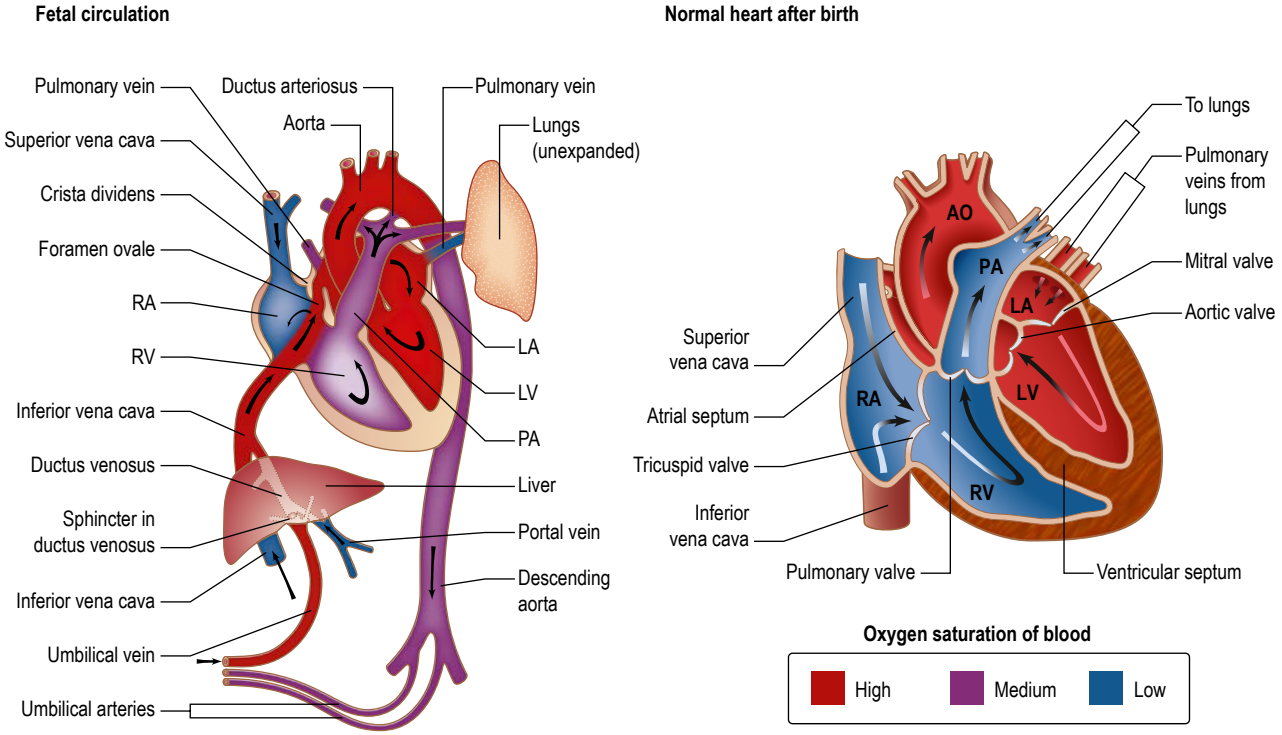


Fig. 17.87 Anatomy showing the circulation. *AO,* Aorta; *LA,* left atrium; *LV,* left ventricle; *PA,* pulmonary artery; *RA,* right atrium; *RV,* right ventricle.

back to the left atrium through the pulmonary veins. The decrease in right atrial pressure and relative increase in left atrial pressure result in closure of the foramen ovale.

The ductus arteriosus usually closes off within 1 or 2 days of birth, completely separating the left and right systems. The umbilical vein and the ductus venosus close off within 2 to 5 days of birth, leaving behind the ligamentum teres and the ligamentum venosus of the liver, respectively.

Adult congenital heart disease

Congenital heart defects (CHDs) are structural problems with the heart that are present at birth. The prevalence of CHD worldwide is estimated to be about 9 per 1000 newborns, with significant geographic variation. The prevalence of severe CHD is declining in many high income countries due to fetal screening and pregnancy termination (although migration from countries with higher rates of CHD to more economically high-income countries may offset this decline). However, the overall prevalence of CHD is increasing on a global scale, due to several factors, including:

- Increased survival of children with CHD due to medical, surgical, and technological advances; over 90% of individuals with CHD who are born now survive into adulthood.
- Ageing population: The global population is ageing, and CHD is more common in older adults who are surviving into late adulthood, having families, and leading active lifestyles.

Patients with congenital heart disease require life-long multidisciplinary specialist care and surveillance, as many need further intervention, surgery, and electrophysiological input in adulthood. In the UK the majority of these patients will be reviewed in designated specialist centres, at intervals depending on clinical need; however, non-specialists may also look after them in local hospitals, so all physicians should have a practical understanding of congenital heart disease.

Ventricular septal defect

VSD is the most common congenital cardiac malformation. They can occur just below the pulmonary and aortic valves, or in the perimembranous or muscular septum. The haemodynamic consequences of the VSD are dependent on the size and number of defects and resultant shunt size. In an otherwise structurally normal heart left ventricular pressure is higher than right and blood will shunt from left to right through a VSD (Fig. 17.88).

- In small defects the shunt velocity is high and the actual volume of blood shunting across the defect is small with no haemodynamic consequence.
- In larger defects the shunt, which is directed into the right ventricular outflow tract and the pulmonary vasculature and then back to the left heart, may lead to dilation of the ***left atrium and left ventricle***, and this is an indication for intervention or repair.
- In large, non-restrictive defects, a large volume of blood shunts into the pulmonary vasculature, leading to remodelling of the pulmonary vessels and pulmonary hypertension. Eventually, when the right ventricular pressure becomes suprasystemic (higher than left), blood starts to shunt from right to left, causing cyanosis. This is referred to as ***Eisenmenger's syndrome***.

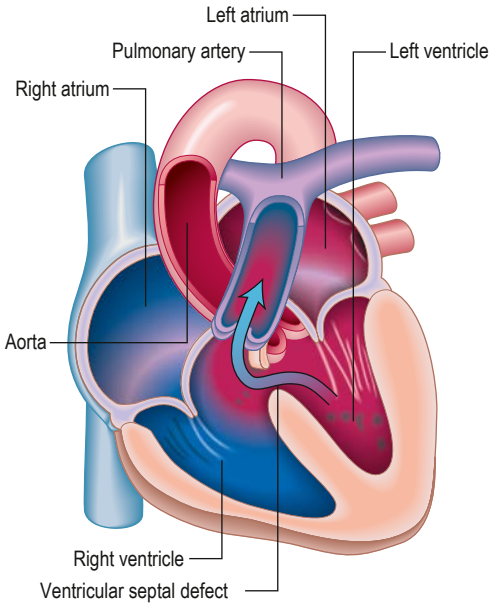
Clinical features

- ***Restrictive VSDs*** ('maladie de Roger') are often found incidentally, as patients are asymptomatic. These defects are associated with a loud pansystolic murmur best heard in the left third and fourth intercostal spaces. A systolic thrill may be present. The majority close spontaneously by the age of 10 years.
- ***Large (non-restrictive) VSDs*** usually present with heart failure symptoms in childhood and eventually lead to pulmonary hypertension and Eisenmenger's syndrome. As pressures equalize, the murmur becomes softer. With established Eisenmenger's syndrome, patients will be cyanosed and have evidence of clubbing.

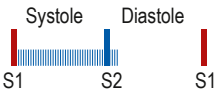
Investigations and treatment

A small VSD produces no abnormal X-ray or ECG findings. This is often a clinical diagnosis and echocardiography is used to confirm the VSD and assess for left heart dilation. If there is none, these lesions are managed conservatively, though there remains an increased risk of endarteritis or endocarditis.

In larger defects echocardiography is used to assess the size and location of the VSD and its haemodynamic consequences. Interventional options are either surgical repair or device closure. Closure (surgical or interventional) is offered for a significant shunt causing left atrial and ventricular enlargement with no evidence of irreversible pulmonary hypertension.



Heart sounds and murmurs



Harsh pansystolic murmur heard at the left sternal edge, accompanied by a systolic thrill ± left parasternal heave
Possible signs of pulmonary hypertension

Pathophysiology

- Left-to-right shunt
- When the left ventricle contracts, it ejects some blood into the aorta and some across the ventricular septal defect into the right ventricle and pulmonary artery
- **Small** VSDs ('maladie de Roger'): loud and sometimes long systolic murmur
- **Moderate** VSDs: loud, 'tearing' pansystolic murmur
- **Large** VSDs: cause pulmonary hypertension and soft murmur
  Eisenmenger's complex may result

Fig. 17.88 Ventricular septal defect (VSD). Pathophysiology and auscultatory findings.

Atrial septal defect

The diagnosis of an atrial septal defect (ASD) is often made in adulthood, as patients often remain asymptomatic through childhood and clinical signs can be subtle.

There are three main types of ASD (Fig. 17.89):

- ***Secundum defects*** (most common type): located in the mid-septum (fossa ovalis)
- ***Sinus venosus defects:*** located in the superior part of the septum near the superior vena cava (superior sinus venosus defect) or the inferior part of the septum near the inferior vena cava (inferior sinus venosus defect) entry point
- ***Ostium primum (atrioventricular septal) defects*** (15%): located in the lower part of atrial septum at the level of, and often involving, the AV valves.

An ASD with a significant left-to-right shunt causes right heart volume overload and right atrial and ventricular dilation. This may lead to arrhythmia, which is often the presenting symptom, along with breathlessness and exercise intolerance. There is raised pulmonary pressure due to increased blood flow through the pulmonary vessels, but significant pulmonary vascular disease develops in less than 5% of patients.

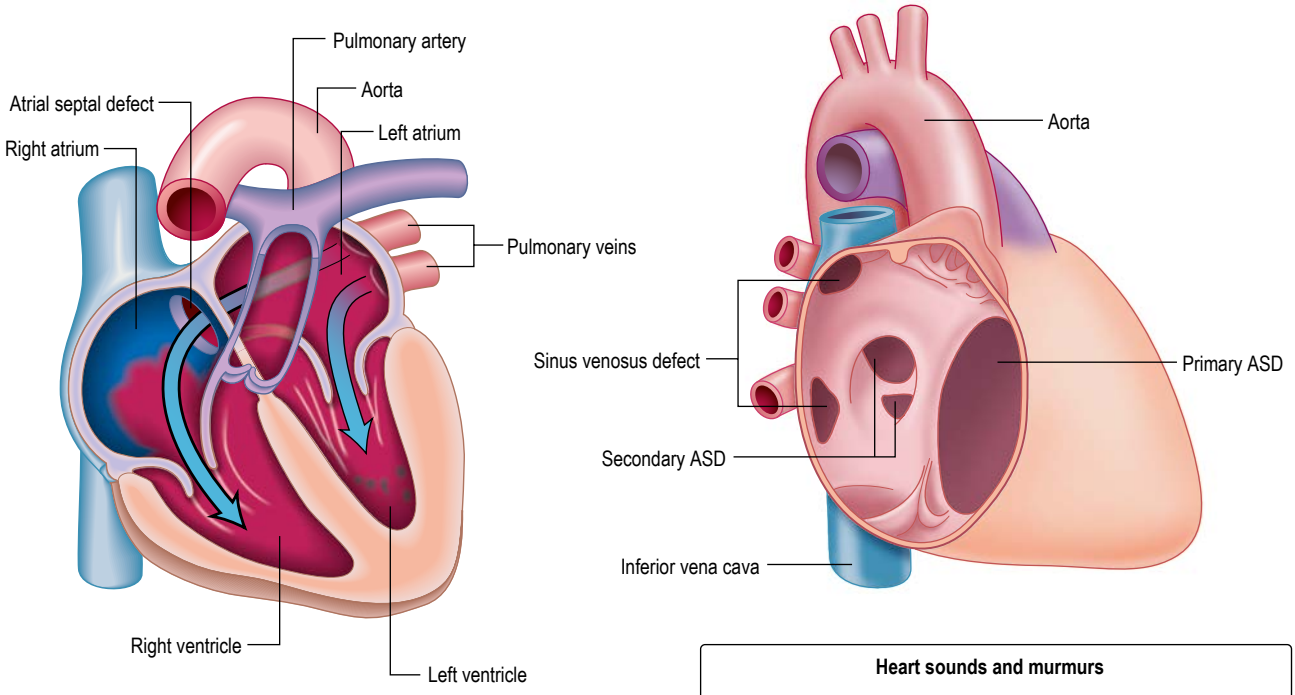
Clinical features

With larger ASDs, there may be a systolic flow murmur over the second and third left parasternal spaces from increased blood flow across the pulmonary valve and a wide, fixed, split-second heart sound (see Fig. 17.89).

Investigations and treatment

- ***Chest X-ray*** may demonstrate prominent pulmonary arteries and cardiomegaly due to right heart dilation.
- ***ECG*** usually shows complete or incomplete right bundle branch block and right axis deviation. Some patients with late diagnosis may have atrial arrhythmia.
- ***Echocardiography*** is used to demonstrate the defect in the septum and shunt direction is evaluated with colour Doppler. The right atrium and ventricle are assessed for dilation and the pulmonary pressure is measured. It is also important to assess the left heart, as if there is significant left ventricular or mitral pathology, this needs to be considered during closure of the defect (Fig. 17.90). TOE allows better imaging of the defect and guides device closure of the ASD.
- ***CMR and CT*** are helpful for assessing anomalous pulmonary venous drainage and confirming sinus venous defects which may not be seen on echocardiography.

Closure of an ASD is indicated if there is a significant left-to-right shunt, resulting in right atrial and ventricular enlargement with normal PVR and oxygen saturations. Secundum defects can often be closed with a device by an interventionalist if the defect is not too large and there are adequate rims of atrial septal tissue (Fig. 17.91).



Pathophysiology

- Left-to-right shunt through the defect in the interatrial septum
- Dilation of the pulmonary artery resulting from the shunt
- The murmur is produced by increased flow across the pulmonary valve and an increased stroke volume

Heart sounds and murmurs

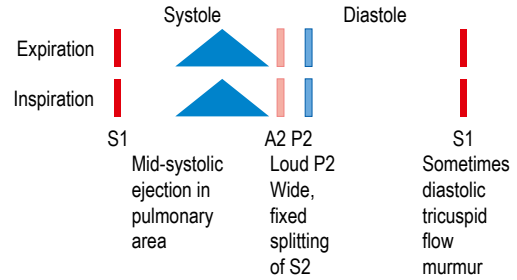


Fig. 17.89 Atrial septal defect (ASD). Pathophysiology and auscultatory findings.

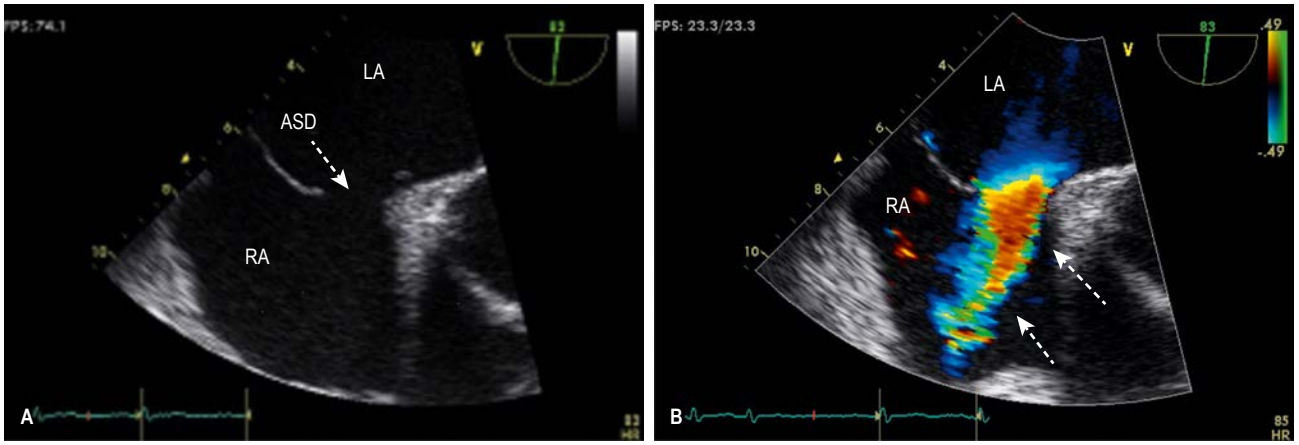


Fig. 17.90 Ostium secundum atrial septal defect (ASD, *arrowed*) in a young girl. **(A)** The defect is shown in a two-dimensional echocardiogram subcostal four-chamber view. **(B)** Colour Doppler can demonstrate the left-to-right shunt. *LA,* Left atrium; *RA,* right atrium.

Fig. 17.91 Angiographic appearance of a fully deployed atrial septal defect (ASD) closure device. The device bridges the ASD and wedges against the surfaces of the right and left atrial septa, occluding flow. The metal object in frame is the distal end of a transoesophageal echocardiography probe. *(Courtesy of Dr D Ward, St George's, University of London.)*

All other ASDs are closed surgically, and if there is anomalous pulmonary venous drainage, it is redirected to the left atrium at the same time.

Patent ductus arteriosus

A patent ductus arteriosus (PDA) is a persistent communication between the proximal left pulmonary artery and the descending aorta, resulting in a continuous left-to-right shunt (Fig. 17.92). Normally, the ductus arteriosus closes within a few hours of birth in response to decreased pulmonary resistance; however, in some cases (particularly in premature babies and in cases with maternal rubella) the ductus persists. Indomethacin (a prostaglandin inhibitor) is given to stimulate duct closure. The shunt size is dependent on the size of the duct and the pressure difference between the pulmonary arteries and aorta. If the left-to-right shunt is significant, it will result in left heart volume overload. Those with large unrestrictive defects may present with congestive heart failure in infancy and, if untreated, may develop pulmonary vascular disease; with time, Eisenmenger's syndrome will develop.

Clinical features

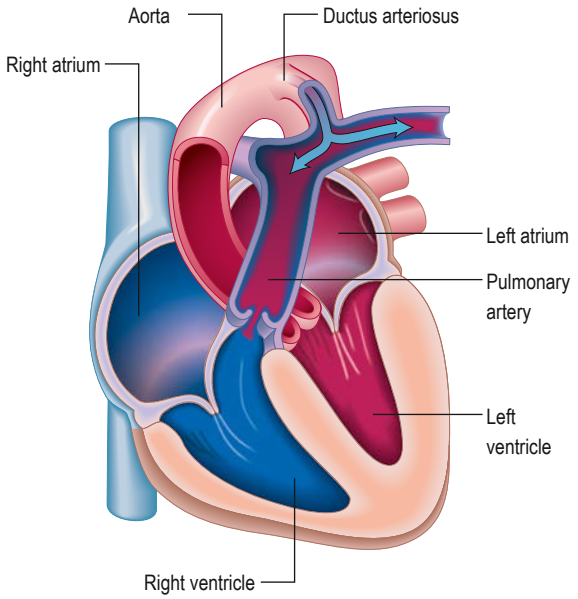
In adults, PDAs are graded as:

- ***Silent:*** tiny PDA, no murmur—incidental finding on echo.
- ***Small:*** no haemodynamic effect or significant shunt but audible systolic or continuous murmur.
- ***Moderate:*** bounding pulses (wide pulse pressure) and continuous murmur in the left infraclavicular region radiating to the back; displaced apex beat from left heart dilation and volume overload.
- ***Large:*** in adults these present with pulmonary hypertension; the murmur quietens and has often disappeared due to equalized pressures between the pulmonary artery and aorta. Patients with Eisenmenger's PDA may have more pronounced clubbing and lower saturations in their toes than in their fingers.

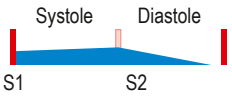
Investigations and treatment

- ***Chest X-ray*** in an adult with a significant left-to-right shunt will show enlarged pulmonary arteries and an increased cardiothoracic ratio (left heart dilation).
- ***ECG*** may demonstrate left atrial abnormality (broad P waves) and high-voltage QRS complexes related to left heart overload. If there is right axis deviation and a tall R in V_1, this may represent established Eisenmenger's syndrome.
- ***Echocardiography*** colour and CW Doppler imaging of the proximal pulmonary arteries may demonstrate the shunt. Echocardiography may show a dilated left atrium and left ventricle if there is a significant left-to-right shunt. Pulmonary pressure can be assessed. If there is pulmonary hypertension, the right ventricle will be hypertrophied.
- ***A right heart catheter study*** may be necessary in the adult who is diagnosed with a PDA and significantly raised pulmonary pressure to estimate the shunt and PVR.

Life expectancy is normal in silent PDAs and patients who have had appropriate and timely closure of a PDA. Indications for intervention (usually with percutaneous devices) include left ventricular dilation with significant shunt and no significant pulmonary vascular disease (assessed by PVR at right heart catheter). Small defects may predispose to endarteritis and should be considered for device closure unless clinically silent.



Heart sounds and murmurs

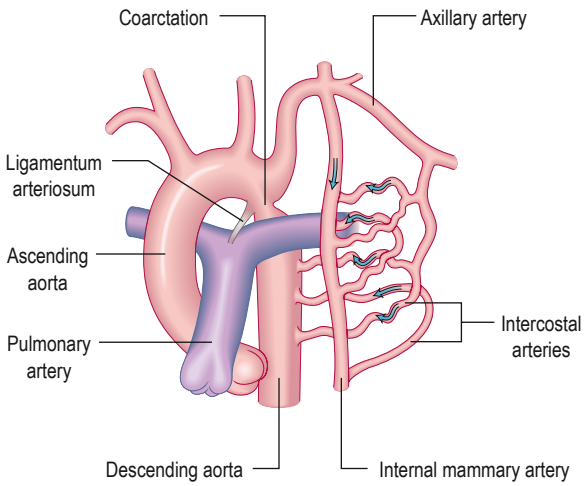


Continuous 'machinery' murmur best heard below the left clavicle in the first interspace or over the first rib
A thrill can often be felt

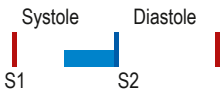
Pathophysiology

- Left-to-right shunt
- Some of the blood from the aorta crosses the ductus arteriosus and flows into the pulmonary artery
- The murmur is produced by the turbulent aortic-to-pulmonary artery shunting in both systole and diastole
- Dilation of the pulmonary artery, left atrium and left ventricle
- As pulmonary hypertension (Eisenmenger's reaction) develops, the murmur becomes quieter, may be confined to systole or may even disappear, causing central cyanosis

Fig. 17.92 Patent ductus arteriosus. Pathophysiology and auscultatory findings.



Heart sounds and murmurs



Mid–late systolic murmur may be heard over the upper precordium or the back
Vascular bruits from the collateral circulation may also be heard

Pathophysiology

Coarctation causes severe obstruction of blood flow in the descending thoracic aorta
The descending aorta and its branches are perfused by collateral channels from the axillary and internal thoracic arteries through the intercostal arteries

Fig. 17.93 Coarctation of the aorta. Pathophysiology and auscultatory findings.

Coarctation of the aorta

A coarctation of the aorta is a narrowing of the aorta at, or just distal to, the insertion of the ductus arteriosus (distal to the origin of the left subclavian artery; Fig. 17.93). Rarely, it can occur proximal to the left subclavian. In 80% of cases the aortic valve is bicuspid. There is an association with circle of Willis aneurysms (affecting approximately 10%) and Turner's syndrome. Severe narrowing of the aorta encourages the formation of a collateral arterial circulation involving the intercostal arteries.

Coarctation of the aorta may be asymptomatic for many years. Adults with undetected coarctation may present with hypertension, headaches, nosebleeds, and, in severe coarctation, ***claudication*** and cold legs may be present.

Clinical features

Physical examination of patients with significant coarctation reveals hypertension in the upper limbs and weak, delayed (radio-femoral delay) pulses in the legs. If coarctation is present in the aorta, proximal to the left subclavian artery, there will be asynchronous radial pulses in the right and left arms; often, the arch itself is hypoplastic and the blood pressure in the right arm is higher than in the left. For heart sounds and murmurs in coarctation of aorta, see Fig. 17.93. Those who have had surgery to correct coarctation may have a thoracotomy scar.

Investigations and treatment

- ***Chest X-ray*** may reveal a dilated aorta indented at the site of the coarctation. This kink in the aortic contour in the upper left

mediastinum is shaped like a number 3 and known as the '3 sign'. In adults tortuous and dilated collateral intercostal arteries may cause rib notching.

- ***ECG*** may demonstrate left ventricular hypertrophy.
- ***Echocardiography*** of suprasternal aortic arch windows can be difficult in adults but may demonstrate aortic luminal narrowing with turbulence in the descending aorta. Continuous Doppler of the descending aorta in significant coarctation will show high-velocity flow and slow diastolic runoff (diastolic tail) due to the obstruction.
- ***CT and MRI*** (Fig. 17.94) demonstrate the coarctation and the proximity to the aortic branch vessels, which need to be considered in intervention. MRI is preferable for surveillance, as it does not use radiation. If the aortic valve is bicuspid, it is also important to obtain images of the ascending aorta for associated aortopathy.

Fig. 17.94 Severe aortic coarctation in an adult. The coarctation is shown on cardiac magnetic resonance angiography. The *solid arrow* indicates the jump bypass graft; the *dotted arrow* indicates severe aortic coarctation.

Intervention/repair

Intervention is indicated if there is a peak–peak gradient across the coarctation of more than 20 mmHg in the cardiac catheter laboratory or greater than 50% luminal narrowing with associated hypertension or left ventricular hypertrophy. If there is significant collateralization, there may be minimal gradient, even in the setting of significant coarctation. In neonates coarctation is treated with surgical repair. In older children and adults the preferred option is balloon dilation and stenting. Balloon dilation is preferred for re-coarctation.

Tetralogy of Fallot

Tetralogy of Fallot (Fig. 17.95) consists of:

- a large, malaligned VSD
- an overriding aorta
- right ventricular outflow tract obstruction
- right ventricular hypertrophy.

The term tetralogy of Fallot represents a spectrum of disease: some patients will have only mild pulmonary stenosis, whereas the more severe extreme is pulmonary atresia, in which the pulmonary vasculature can also develop abnormally. Symptoms depend on the degree of pulmonary stenosis, but the majority of patients with tetralogy present in childhood. In infancy cyanosis usually develops due to increased right-sided pressures, resulting in a right-to-left shunt. Fallot's spells are episodes of severe cyanosis noted in children due to spasm of the subpulmonary muscle; these can be relieved by increasing systemic resistance using postural manoeuvres, such as squatting. In babies with severe pulmonary stenosis systemic-to-pulmonary artery shunts (i.e., a Blalock–Taussig subclavian-to-pulmonary artery shunt) may have been used initially to increase pulmonary blood flow until the pulmonary arteries are large enough for repair. Most adults with tetralogy of Fallot will have undergone complete repair but repair in childhood will often leave them with significant pulmonary regurgitation, which is generally well tolerated into early adulthood; the majority, however, will need pulmonary valve replacement with time.

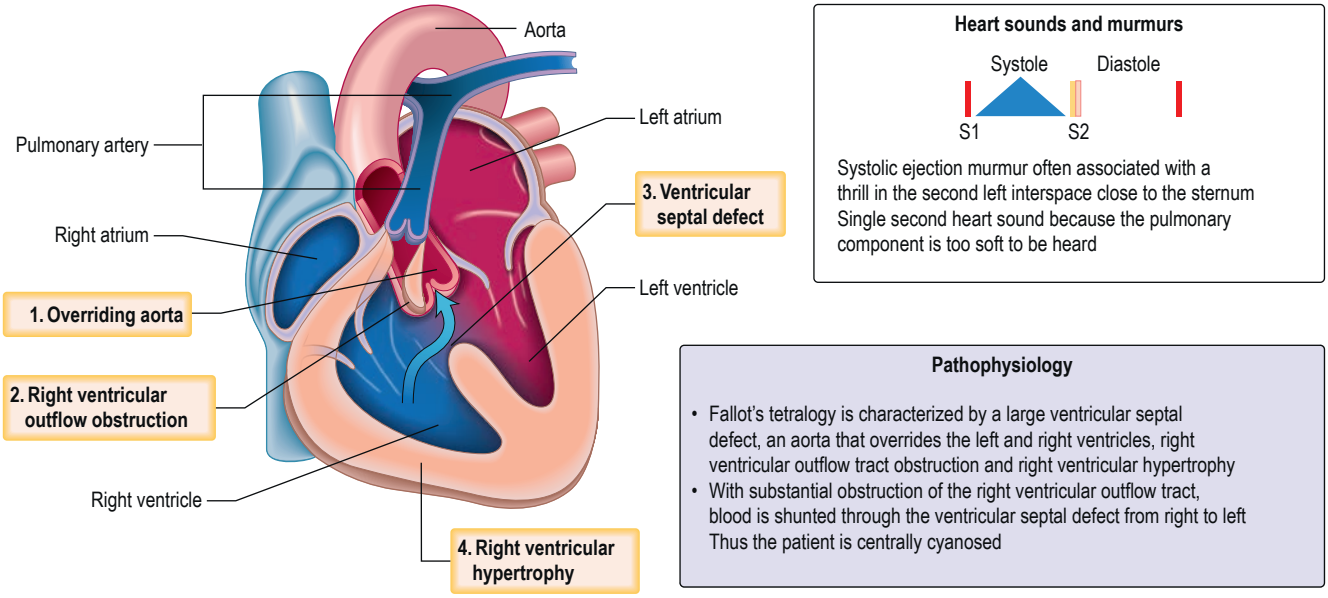


Fig. 17.95 Fallot's tetralogy. Pathophysiology and auscultatory findings.

The overall survival of those who have had operative repair is excellent. Many patients with repaired tetralogy of Fallot lead unlimited active lives. They do require on-going specialist follow-up as, in those with transannular patch repair (surgical or interventional), pulmonary valve replacements are needed during follow-up, and late complications include ventricular and atrial arrhythmias, heart failure, and sudden death.

Transposition of the great arteries

In transposition of the great arteries (TGA) the right atrium connects to the morphological right ventricle, which gives rise to the aorta, and the left atrium connects to the morphological left ventricle, which gives rise to the pulmonary artery (Fig. 17.96). This is incompatible with life, as blood circulates in two parallel circuits: that is, deoxygenated blood from the systemic venous return passes into the right heart and then, via the aorta, back to the systemic circulation. Babies with transposition are born cyanosed and rely on an ASD, VSD, or PDA, allowing oxygenated and deoxygenated blood to be mixed. In those without an adequate shunt an atrial septostomy is performed: a Rashkind's balloon is deployed to dilate the foramen ovale and is used to maintain saturations at 50% to 80% until a definitive procedure can be performed.

The majority of adult patients with TGA who were born between the 1960s and mid-1980s will have had an 'atrial switch' operation. In this procedure systemic venous blood was baffled across into the left atrium and then pumped via the left ventricle into the pulmonary artery; the pulmonary venous (oxygenated blood) returning to the left atrium was redirected via a baffle into the right atrium and right ventricle, and into the aorta. This palliation allowed patients a good quality of life into their fourth decade, but the right ventricle remains the systemic ventricle and with time starts to fail. Atrial arrhythmia is also common in this group.

The first arterial switch procedure was performed in 1975, and it is now routinely performed in babies with TGA in the first 2 weeks of life. The aorta is reconnected to the left ventricle and the pulmonary artery is connected to the right ventricle. The coronary arteries are re-implanted. Late complications include coronary complications, narrowing of the pulmonary arteries, which are stretched around the aorta, and neovalvular regurgitation.

Congenitally corrected transposition of the great arteries

In congenitally corrected transposition of the great arteries (ccTGA) systemic venous return to the right atrium enters a morphological left ventricle, which pumps deoxygenated blood into the pulmonary arteries. Pulmonary venous blood returns to the left atrium and then, via the morphological right ventricle, to the aorta. The circulation is physiologically corrected and these babies are not typically cyanosed, but the systemic circulation is supported by a morphological right ventricle. ccTGA is often associated with cardiac lesions:

- systemic (tricuspid) AV valve abnormalities with valve insufficiency
- VSD
- subpulmonary stenosis
- complete heart block
- WPW syndrome
- dextrocardia.

Some patients with ccTGA may live a normal life, while others develop progressive tricuspid regurgitation and systemic ventricular failure or require pacemaker insertion (the AV node is abnormal, leading to heart block).

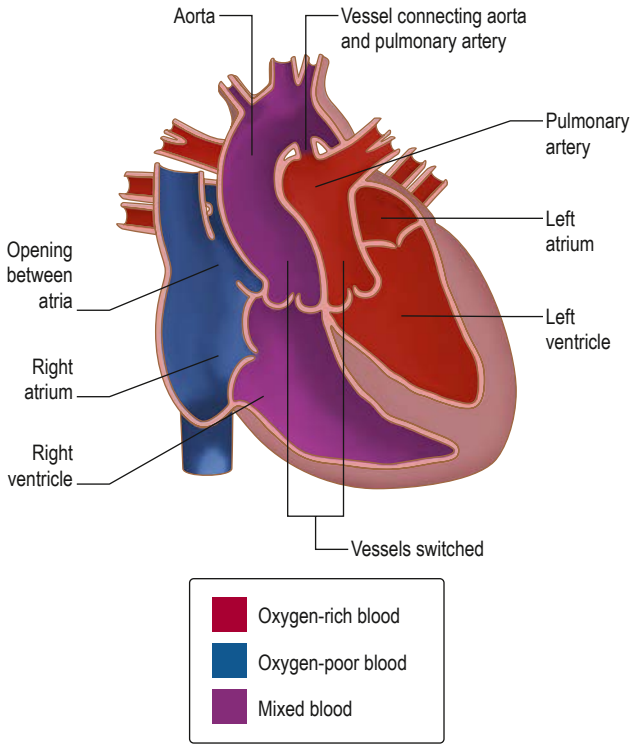


Fig. 17.96 **Transposition of the great arteries.**

Further reading

Baumgartner H, De Backer J, Babu-Narayan S et al. 2020 ESC Guidelines for the management of adult congenital heart disease: the task force for the management of adult congenital heart disease of the European Society of Cardiology (ESC). *Eur Heart J* 2021; 42(6):563–645.

Marfan's syndrome

Marfan's syndrome (MFS) is a multisystem connective tissue disorder with autosomal dominant inheritance pattern and a prevalence of 1:5000. One-quarter of cases are not inherited and are due to *de novo* mutations. MFS is caused by mutations in the *FBN1* gene on chromosome 15q21. The *FBN1* gene encodes the protein fibrillin-1, which is a component of structures called microfibrils that make up the extracellular matrix in elastic and non-elastic tissues.

Diagnosis

The diagnosis is currently made using the Modified Ghent Criteria (Box 17.45) and requires a complete medical and family history, examination, relevant investigations, and, increasingly, genetic analysis. The findings of aortic root aneurysm, lens dislocation (ectopia lentis), or a family history of MFS are major criteria in the diagnosis of MFS in these guidelines. The syndrome can affect the eyes, lungs, skeletal system, and cardiovascular system (Figs. 17.97 and 17.98).

Management

The management of patients with MFS requires multidisciplinary input from geneticists, cardiologists, cardiothoracic, vascular and orthopaedic surgeons, ophthalmologists, and physiotherapists, among others.

Box 17.45 Modified Ghent Criteria for Marfan's syndrome (MFS)

No family history of MFS

- Aortic root dilation Z-score ≥2 AND ectopia lentis = MFS
- Aortic root dilation Z-score ≥2 AND *FBN1* = MFS
- Aortic root dilation Z-score ≥2 AND systemic score ≥7 points* = MFS
- Ectopia lentis AND *FBN1* with known aortic root dilation = MFS

Family history of MFS

- Ectopia lentis = MFS
- Systemic score ≥7 points* = MFS
- Aortic root dilation Z-score ≥2 above 20 years old, ≥3 below 20 years old = MFS

*Systemic features	Score
Wrist AND thumb sign	3
Wrist OR thumb sign	1
Pectus carinatum (protrusion of chest wall)	2
Pectus excavatum (indentation of chest wall) or chest asymmetry	1
Hindfoot deformity	2
Pes planus (flat foot)	1
Spontaneous pneumothorax	2
Dural ectasia	2
Protrusio acetabuli	2
Scoliosis or thoracolumbar kyphosis	1
Reduced elbow extension	1
3/5 facial features (dolichocephaly, enophthalmos, down-slanting palpebral fissure, malar hypoplasia, retrognathia)	1
Skin striae	1
Severe myopia	1
Mitral valve prolapse	1
Reduced upper segment/lower segment (<0.85 whites, <0.78 blacks AND increased arm span/height >1.05).	1
	*Maximum total 20 points; score ≥7 indicates systemic involvement

Cardiovascular surveillance and management

The primary cause of death in MFS is cardiovascular. In the 1970s the mean age of death was 32 years, primarily from type A aortic dissection, but survival has significantly improved with better means of assessing for this condition and with prophylactic aortic root replacement for those with dilated ascending aortas. Patients with MFS are followed up with serial cardiovascular imaging, tailored to the specific cardiovascular findings of the individual. Yearly follow-up of aortic dimensions is recommended; however, patients with a cross-sectional diameter of more than 4.5 cm on the initial scan or a growth rate of more than 0.3 cm per year require more frequent (6-monthly) surveillance as they approach surgical thresholds. Echocardiography is readily available and useful for assessing the aortic sinuses, sinotubular junction, and part of the ascending aorta, as well as valvular and left ventricular function, which may be impaired in MFS. Cross-sectional imaging (CT or MRI) is necessary to assess the arch and descending aorta, which can also be involved. These modalities give more readily reproducible images to monitor interval change and may be preferred for this purpose in patients approaching the need for surgical intervention.

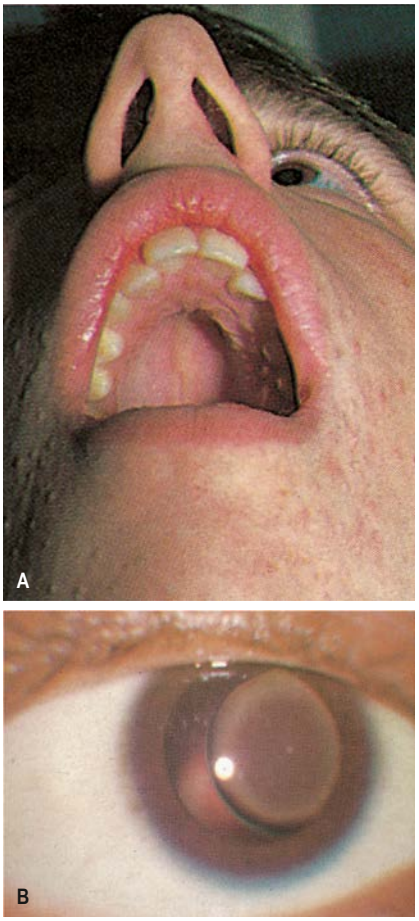


Fig. 17.97 Marfan's syndrome. **(A)** High-arched palate. **(B)** Eye lens dislocation. *(From Forbes CD, Jackson WF. Color Atlas and Text of Clinical Medicine, 3rd edn. St Louis: Mosby; 2003, p. 134, ©Elsevier.)*

Lifestyle

Contact sports and strenuous and isometric exertion should be avoided but moderate cardiovascular exercise is encouraged. Patients with Marfan's are excluded from joining the military.

Medical therapy

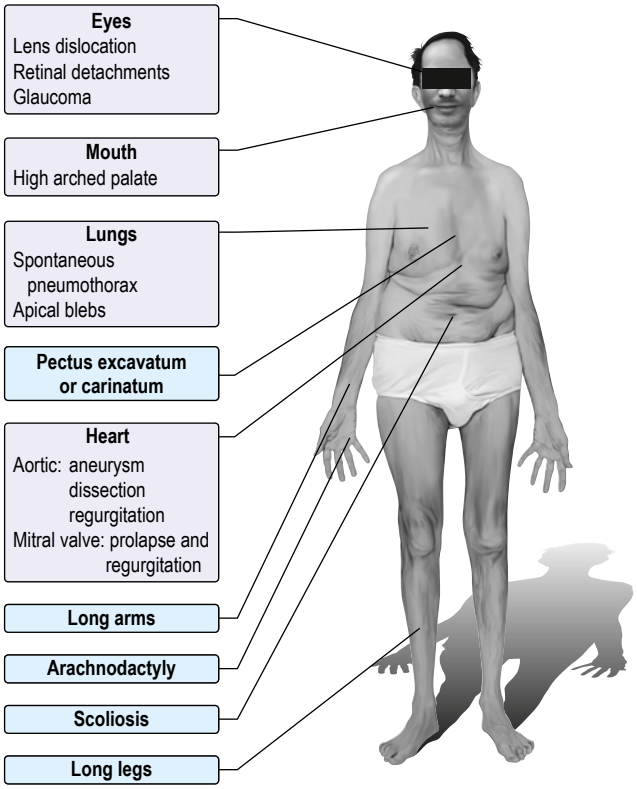
The current recommendation is to treat patients with MFS and aortic root dilatation with either a ***beta-blocker or angiotensin receptor blocker (ARB)***, or both if tolerated to reduce the rate of aortic dilatation.

Surgery

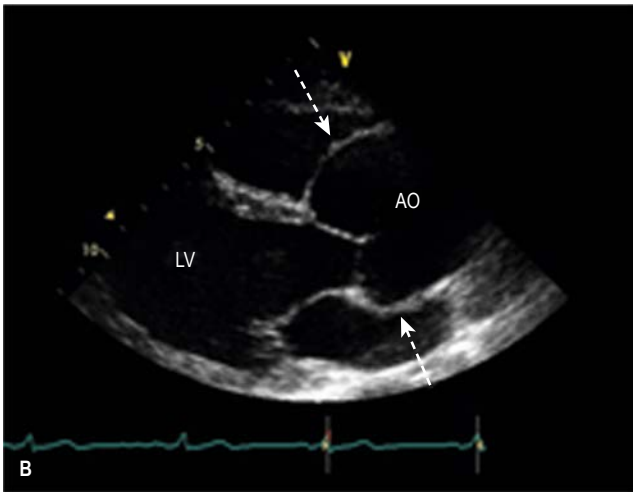
The current indication (European and American guidelines) for prophylactic aortic root replacement in patients with MFS is an aortic root measuring 5 cm or more (or ≥4.5 cm in those with a family history of dissection). This has very low surgical risk and prevents type A aortic dissection, thereby improving survival.

Pregnancy

All women with MFS should have preconceptual assessment and counselling. There is a 50% chance of their child being affected and there are options for pre-implantation genetic diagnosis (*in vitro* fertilization pregnancy where embryos can be assessed for the *FBN*



A



B

	Aortic size			
Yearly risk	***> 3.5cm***	***> 4cm***	***> 5cm***	***> 6cm***
Rupture	0.0%	0.3%	1.7%	3.6%
Dissection	2.2%	1.5%	2.5%	3.7%
Death	5.9%	4.6%	4.8%	10.8%
Any of the above	7.2%	5.3%	6.5%	14.1%

C

Fig. 17.98 Marfan's syndrome. **(A)** Photograph of a 63-year-old man. **(B)** Two-dimensional longitudinal echocardiogram revealing dilatation of the aortic root *(arrows)*. **(C)** Yearly risk of complications based on aortic size. *AO,* Aorta; *LV,* left ventricle.

mutation prior to implantation) and prenatal diagnosis (amniocentesis and chorionic villus sampling). Women having MFS with or without aortic dilation are at increased risk of aortic dissection during pregnancy. This is estimated at 1% if the aortic root measures less than 4 cm and 10% if it is over 4 cm. Pregnancy should be discouraged in women with aortic dimensions of more than 4.5 cm and they should be offered prophylactic aortic root replacement prior to pregnancy. This threshold is lowered to 4.0 cm if there has been rapid aortic growth of ≥0.3 cm/year. Current recommendations are that echocardiography should be performed every 6 to 8 weeks throughout pregnancy and the postpartum period. Blood pressure should be regularly monitored, and hypertension treated aggressively. As ARBs are teratogenic, women planning pregnancy should be given beta-blockers.

Further reading

Loeys BL, Dietz HC, Braverman AC et al. The revised Ghent nosology for the Marfan's syndrome. *J Med Genet* 2010; 47:476–485.

PULMONARY HEART DISEASE

The normal values for mean pulmonary artery pressure (mPAP), mean capillary wedge pressure (mPCWP), and cardiac output are 12 ± 2 mmHg, 6 ± 2 mmHg, and 5 L/minute, respectively, with an upper limit of normal for mPAP of 20 mmHg. The fall in pressure across the lung circulation is known as the transpulmonary gradient and reflects the difference between mPAP and mPCWP. The normal transpulmonary gradient is 6 ± 2 mmHg.

The ***pulmonary vascular resistance*** (PVR) is calculated by the formula:

$$\frac{\text{mPAP} - \text{mPCWP}}{\text{CO}}$$

It is normally about 1.5 mmHg/L per minute (1.5 Wood units). Approximately 60% of the body's endothelial surface is in the lungs and the lungs normally offer a low resistance to blood flow. This is because the media of the pre-capillary pulmonary arterioles is thin, as compared with their more muscular systemic counterparts that must respond constantly to postural changes under the influence of gravity. The fact that lung circulation normally offers a low resistance to flow explains the preferential passage of blood through the lungs in specific forms of congenital heart disease, which may eventually lead to remodelling of the lung circulation and pulmonary hypertension.

Pulmonary hypertension

Pulmonary hypertension is defined as an mPAP of more than 25 mmHg at rest, as measured on right heart catheterization. The clinical classification of pulmonary hypertension is provided in **Box 17.46**.

Box 17.46 Clinical classification of pulmonary hypertension[a]

Group	Aetiological classification	Subtypes
1	Pulmonary arterial hypertension (PAH)	1.1—Idiopathic 1.2—Heritable 1.2.1—*BMPR2* 1.2.2—*ALK1*, *ENG*, *SMAD9*, *CAV1*, *KCNK3* 1.2.3—Unknown 1.3—Drugs and toxins 1.4—Associated with: 1.4.1—Connective tissue disease[b] 1.4.2—HIV infection 1.4.3—Portal hypertension 1.4.4—Congenital heart disease 1.4.5—Schistosomiasis
1′	Pulmonary veno-occlusive disease (PVOD) and/or pulmonary capillary haemangiomatosis	
1″	Persistent pulmonary hypertension of the newborn (PPHN)	
2	Pulmonary hypertension due to left heart disease	2.1—Systolic dysfunction 2.2—Diastolic dysfunction 2.3—Valvular disease 2.4—Congenital/acquired left heart inflow/outflow tract obstruction and congenital cardiomyopathies
3	Pulmonary hypertension due to lung diseases and/or hypoxia	3.1—Chronic obstructive pulmonary disease 3.2—Interstitial lung disease 3.3—Other pulmonary disease with mixed restrictive and obstructive pattern 3.4—Sleep-disordered breathing 3.5—Alveolar hypoventilation disorders 3.6—Chronic exposure to high altitude 3.7—Developmental abnormalities
4	Chronic thromboembolic pulmonary hypertension (CTEPH)	
5	Pulmonary hypertension with unclear and/or multifactorial mechanisms	5.1—Haematological disorders: chronic haemolytic anaemia, myeloproliferative disorders, splenectomy 5.2—Systemic disorders: sarcoidosis, pulmonary histiocytosis, lymphangioleiomyomatosis 5.3—Metabolic disorders: glycogen storage disease, Gaucher's disease, thyroid disorders 5.4—Others: tumoral obstruction, fibrosing mediastinitis, chronic renal failure on dialysis, segmental pulmonary hypertension

[a]5th World Symposium on Pulmonary Hypertension, Nice, 2013.
[b]Now called 'autoimmune rheumatic disease'.
ALK1, Activin-like receptor kinase-1; *BMPR,* bone morphogenetic protein receptor type II; *CAV1,* caveolin-1; *ENG,* endoglin; *HIV,* human immunodeficiency virus; *KCNK3,* potassium channel superfamily K member-3; *SMAD9,* mothers against decapentaplegic 9.
(From Simonneau G, Gatzoulis MA, Adatia I et al. Updated clinical classification of pulmonary hypertension. J Am Coll Cardiol *2013; 62:D34–41.)*

Pathophysiology

The different groups are characterized by variable amounts of hypertrophy, proliferation, and fibrotic changes in distal pulmonary arteries (pulmonary arterial hypertension, PAH; pulmonary veno-occlusive disease, PVOD; pulmonary hypertension due to left heart disease; pulmonary hypertension due to lung disease and/or hypoxia). Pulmonary venous changes are seen in PVOD and pulmonary hypertension groups due to left heart disease, and the vascular bed may be destroyed in emphysematous or fibrotic areas seen in lung disease. In chronic thromboembolic pulmonary hypertension (CTEPH) organized thrombi are seen in the elastic pulmonary arteries. Patients with pulmonary hypertension with unclear and/or multifactorial mechanisms have variable pathological findings.

Patients with progressive pulmonary hypertension develop right ventricular hypertrophy, dilation, heart failure, and death.

Pulmonary artery hypertension

Epidemiology

A French registry of 674 patients with PAH identified 39.2% with idiopathic pulmonary artery hypertension (IPAH), 3.9% with familial (or heritable) disease, 9.5% caused by drugs and toxins (most prominently the now little used anorexigen fenfluramine), 15.3% with autoimmune rheumatic disease, 11.3% with congenital heart disease, 10.4% with portal hypertension, and 6.2% with HIV-associated disease. In familial or heritable PAH mutations in the bone morphogenetic protein receptor type 2 *(BMPR2)* gene are detected in over 70% of cases; other mutations are seen in patients with hereditary haemorrhagic telangiectasia (Osler–Weber–Rendu syndrome).

Clinical features

Patients with PAH may present with symptoms of dyspnoea, fatigue, weakness, angina, syncope, or abdominal distension. Clinical signs of PAH and right heart hypertrophy include a left parasternal heave, a loud P2 heart sound, a soft pansystolic murmur from tricuspid regurgitation, or an early diastolic murmur from pulmonary regurgitation. Right heart failure leads to jugular venous distension, hepatomegaly, ascites, and peripheral oedema. Clinical signs of associated diseases, such as systemic sclerosis or chronic liver disease, should be sought.

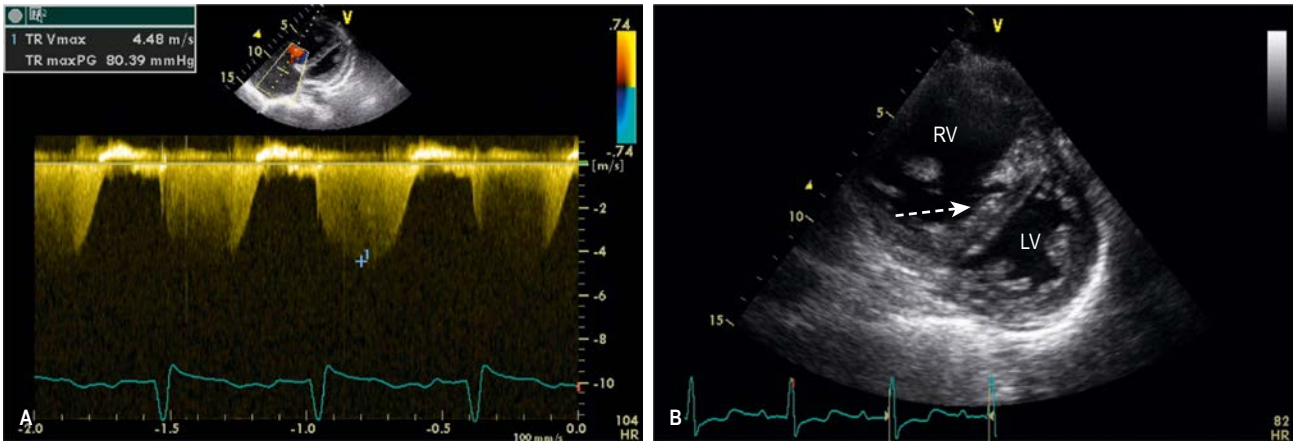


Fig. 17.99 Pulmonary hypertension. (A) Continuous wave Doppler echocardiography demonstrating markedly elevated tricuspid regurgitation velocity consistent with pulmonary hypertension. **(B)** Parasternal short-axis echocardiogram with a dilated right ventricle (RV) and septal flattening *(arrowed)* in a patient with pulmonary hypertension. *LV,* Left ventricle.

Investigations

- ***Routine blood tests*** include full blood count, renal and liver biochemistry, thyroid function tests, and serological assays for underlying autoimmune rheumatic diseases, HIV, and hepatitis.
- ***Chest X-ray*** shows enlargement of the pulmonary arteries and the major branches, with marked tapering (pruning) of peripheral arteries. The lung fields are usually lucent and there may be right atrial and right ventricular enlargement. There may be signs of chronic lung disease.
- ***ECG*** shows right ventricular hypertrophy and right atrial enlargement (P pulmonale).
- ***Echocardiography*** (Fig. 17.99) with tricuspid regurgitation can be used for determination of PAP using the simplified Bernoulli equation (PAP = 4 × (tricuspid regurgitation velocity)^2 + estimated right atrial pressure). Right atrial pressure can be assumed at 5 to 10 mmHg unless there is significant dilation of the inferior vena cava with reduced respiratory variation. mPAP = 0.61 × pulmonary artery systolic pressure +2 mmHg (although the Bernoulli equation may not be accurate in cases of severe tricuspid regurgitation).
- ***CMR*** may be useful in adult congenital heart disease and in assessing right ventricular function and volumes by serial assessment.
- ***Abdominal liver ultrasound*** is useful to exclude liver cirrhosis and portal hypertension.
- ***Right heart catheterization*** may be indicated as part of the clinical assessment to confirm the diagnosis (elevated PAP), determine the pulmonary wedge pressure (PWP), calculate the cardiac output, and assess PVR and reactivity. In PAH a vasodilator challenge (inhaled nitric oxide, intravenous adenosine, or epoprostenol) should be performed to identify patients who might benefit from these therapies. A responder is defined as a reduction in mean PAP of 10 mmHg or more to reach an absolute mPAP of 40 mmHg or less with increased or unchanged cardiac output. These vasodilator challenges are not recommended in patients with other types of pulmonary hypertension (types 2–5).

Management

- ***Physical activity.*** Patients should be encouraged to remain physically active but avoid exertion that precipitates severe dyspnoea, chest pain, or pre-syncope.
- ***Pregnancy.*** Patients with PAH have a very high mortality rate during pregnancy (30%–50%) and should be counselled against conception and contraceptive advice given.
- ***Travel.*** During plane travel, supplementary oxygen at 2 L/min may be appropriate for patients with reduced functional class and with resting hypoxia of less than 8 kPa.
- ***Vaccination.*** Vaccination should be given for influenza and pneumococcal pneumonia.
- ***Elective surgery.*** Epidural anaesthesia may be preferable to a general anaesthetic.
- ***Oral anticoagulation.*** There is evidence to support the use of oral anticoagulation in patients with IPAH, heritable PAH, and PAH due to anorexigens. The European target INR range is 2.0 to 3.0.
- ***Diuretics.*** These are used in patients with right heart failure and fluid overload.
- ***Digoxin.*** This may be helpful in patients with tachyarrhythmias.
- ***Calcium-channel blockers.*** These can be effective in high doses in selected patients with IPAH who demonstrate a response to a vasodilator challenge. Right heart catheterization should be repeated in 3 to 4 months to assess response to therapy.
- ***Prostanoids.*** Prostacyclin is a potent vasodilator that also inhibits platelet aggregation and cell proliferation. Synthetic prostacyclins are generally short-acting compounds requiring continuous intravenous or subcutaneous infusion or regular aerosol inhalation. They provide symptomatic relief and can improve exercise capacity; ***epoprostenol*** can improve survival in patients with both IPAH and associated pulmonary arterial hypertension. Oral ***selexipag*** has been approved by the US Food and Drug Administration for PAH treatment.
- ***Endothelin receptor antagonists.*** ET-1 is a potent vasoconstrictor and mitogen that binds to ET A and B receptors in the pulmonary vasculature. Both dual antagonists ***(bosentan)*** and selective A receptor antagonists ***(sitaxentan, ambrisentan)*** can improve symptoms, exercise capacity, and haemodynamics in patients with IPAH.
- ***Phosphodiesterase type 5 inhibitors.*** These produce vasodilation in the pulmonary vasculature and reduce cellular proliferation. ***Sildenafil*** and ***tadalafil*** have been demonstrated to provide symptomatic relief and improve exercise capacity in patients with IPAH.
- ***Balloon atrial septostomy.*** This technique may be considered palliative therapy in severe cases of pulmonary hypertension.

- ***Lung transplantation.*** This is used in patients with an adverse prognosis, although the 5-year survival following transplantation may be only 40% to 50%.

Other pulmonary hypertension groups

Left-sided heart disease (systolic and diastolic heart failure) and valvular heart disease are frequently associated with pulmonary hypertension, as is advanced chronic obstructive pulmonary disease (see Chapter 20, 'Chronic obstructive pulmonary disease' section), pulmonary fibrosis, and emphysema. Following acute pulmonary embolism, 0.5% to 2.0% of patients will develop CTEPH.

MYOCARDIAL AND ENDOCARDIAL DISEASE

Cardiac tumours

Primary cardiac tumours are rare and three-quarters of these are benign.

Atrial myxomas are the most common. Most are sporadic, but some are familial or part of a multisystem syndrome (Carney complex—hyperpigmented skin lesions with extracardiac tumours). Histologically, they are benign, polypoid, gelatinous structures. Most are solitary and occur in the left atrium (75%), attached by a pedicle to the atrial septum. The tumour may obstruct the mitral valve or may be a site of thrombi, which can embolize. The mean age of presentation is 50 years and there is a female predominance. Patients may be asymptomatic or present with dyspnoea, syncope, or a mild fever. Physical signs include a loud first heart sound, a tumour 'plop' (a loud third heart sound produced as the pedunculated tumour comes to an abrupt halt), a mid-diastolic murmur, and/or signs of embolization. Echocardiography, CT, and CMR are appropriate to confirm the diagnosis (Fig. 17.100) prior to cardiac surgery. Because of local recurrence, long-term surveillance is recommended.

Other benign cardiac tumours include lipoma, papillary fibroelastoma, fibroma, haemangioma, and rhabdomyoma.

Malignant primary cardiac tumours include angiosarcoma, undifferentiated sarcoma, rhabdomyosarcoma (children), osteosarcoma, and leiomyosarcoma. Primary cardiac lymphoma typically occurs in the right heart.

Secondary cardiac tumours from metastatic spread are 20 to 40 times more common than primary tumours. Direct invasion can occur with bronchogenic, breast, and oesophageal carcinoma. There may be haematogenous or lymphatic spread with lymphoma, leukaemia, transvenous renal cell carcinoma, hepatoma, and adrenal tumours. Pericardial effusions are common, and pericardiocentesis may be required for diagnosis or for cases with cardiac tamponade.

Myocardial disease

Myocardial disease that is not due to ischaemic, valvular, or hypertensive heart disease, or a known infiltrative, metabolic/toxic or neuromuscular disorder, may be caused by:

- an acute or chronic inflammatory pathology (myocarditis)
- idiopathic myocardial disease (cardiomyopathy).

Myocarditis

Acute inflammation of the myocardium has many causes (Box 17.47). In North America and Europe the most common causes of inflammation are viruses (enteroviruses, adenoviruses, human herpes virus-6, Epstein–Barr virus, cytomegalovirus, hepatitis C, and parvovirus B19). Chagas' disease, due to *Trypanosoma cruzi*, which is endemic in South America, is one of the most common causes of myocarditis worldwide. Additionally, toxins (including prescribed drugs), physical agents, hypersensitivity reactions, and autoimmune conditions may also cause myocardial inflammation.

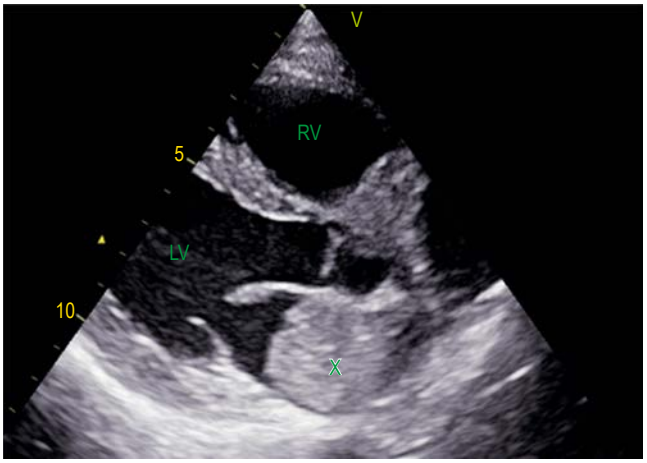


Fig. 17.100 Atrial myxoma shown by a two-dimensional echocardiogram (long-axis view). The myxoma (X) is an echo-dense mass obstructing the mitral valve orifice. It was removed surgically. *LV,* Left ventricle; *LA,* left atrium.

? Box 17.47 Causes of myocarditis

- Idiopathic
- Infective:
  - Viral: coxsackievirus, adenovirus, cytomegalovirus, echovirus, influenza, polio, hepatitis, HIV
  - Parasitic: *Trypanosoma cruzi*, *Toxoplasma gondii* (a cause of myocarditis in the newborn or immunocompromised)
  - Bacterial: streptococcus (most commonly, rheumatic carditis), diphtheria (toxin-mediated heart block common)
  - Spirochaetal: Lyme disease (heart block common), leptospirosis
  - Fungal
  - Rickettsial
- Toxic:
  - Drugs: those causing hypersensitivity reactions, e.g., methyldopa, penicillin, sulphonamides, anti-tuberculous, modafinil
  - Radiation: may cause myocarditis but pericarditis is more common
- Autoimmune:
  - An autoimmune form with autoactivated T cells and organ-specific antibodies may occur
  - Giant cell myocarditis
- Alcohol
- Hydrocarbons

Pathology

In the acute phase myocarditic hearts are flabby with focal haemorrhages; in chronic cases they are enlarged and hypertrophied. Histologically (Fig. 17.101), an inflammatory infiltrate is present—lymphocytes predominating with viral causes, polymorphonuclear cells with bacterial causes, and eosinophils with allergic and hypersensitivity causes.

Clinical features

Myocarditis typically presents in young adult patients. Presentation may vary from subclinical disease to mild fatigue, palpitations, chest pain, and dyspnoea through to fulminant congestive cardiac

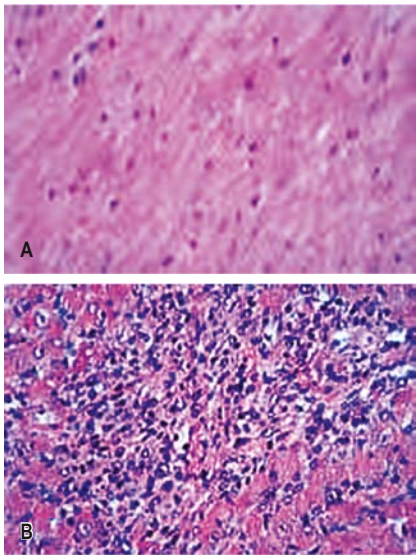


Fig. 17.101 Myocardial biopsies. (A) Normal myocardium. (B) Myocarditis, with increased interstitial inflammatory cells.

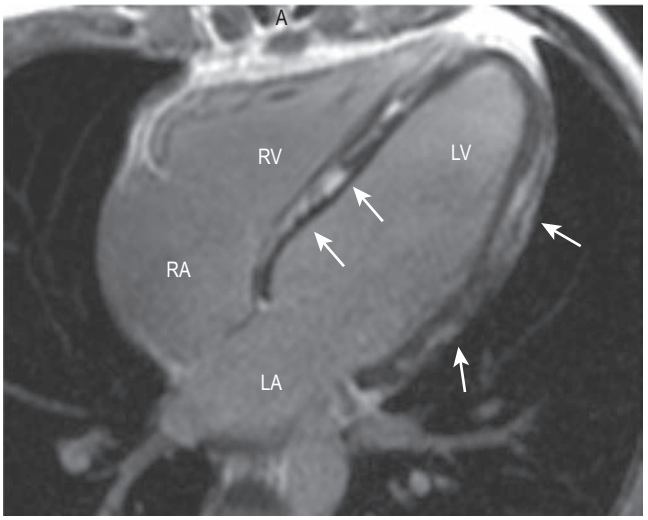


Fig. 17.102 Acute myocarditis. There is patchy mid-wall and epicardial myocardial enhancement *(arrowed)* with gadolinium. *A,* Anterior chest wall; *LA,* left atrium; *LV,* left ventricle; *RA,* right atrium; *RV,* right ventricle.

failure. Physical examination may reveal soft heart sounds, a prominent third sound, and a tachycardia. A pericardial friction rub may be heard.

Investigations

- ***ECG*** may demonstrate diffuse ST- and T-wave abnormalities (concave ST elevation) and arrhythmias. AV block may be seen with Lyme disease, sarcoid, giant-cell myocarditis, and Chagas' disease (see later).
- ***Echocardiography*** is required and may reveal normal or reduced biventricular function with resting regional wall motion abnormalities.
- ***Blood tests*** include ESR and CRP, which are frequently elevated in the acute presentation. Cardiac troponin and creatine kinase levels will also be elevated. Viral serology is not usually helpful.
- ***CMR*** is recommended to confirm the diagnosis. Patients with acute myocarditis may demonstrate myocardial oedema on T2-weighted images and patchy myocardial enhancement with gadolinium (Fig. 17.102).
- ***Endomyocardial biopsy*** is considered the gold standard for the diagnosis of myocarditis, although its use is reserved for selected patients (fulminant presentation or clinical deterioration) as this is an invasive test performed in specialized units. It may be helpful in giant cell myocarditis, sarcoidosis, and other autoimmune conditions as diagnosis will enable specific treatment.
- ***Coronary angiography*** is frequently performed to exclude acute ischaemia as the cause of the patient's presentation.

Management

Myocarditis resolves within a few weeks in the majority of patients, although a small percentage of patients deteriorate and may require urgent transfer to a cardiac centre with access to ECMO and/or cardiac transplantation. Bed rest is recommended in the acute phase of the illness and athletic activities should be avoided for 6 months. Heart failure should be treated with conventional medical therapy (see 'Management of heart failure' section).

Giant cell myocarditis

This severe form of myocarditis is characterized by the presence of multinucleated giant cells within the myocardium. The cause is unknown but it may be associated with sarcoidosis, thymomas, and autoimmune disease. The disease has a rapidly progressive course and a poor prognosis. Immunosuppression is recommended.

Chagas' disease

Chagas' disease (see Chapter 42, 'South American trypanosomiasis [Chagas' disease]' section) is caused by the protozoon *T. cruzi* and is endemic in South America, where upwards of 20 million people are infected. Acutely, features of myocarditis are present with fever and congestive heart failure. Chronically, there is progression to a DCM with a propensity towards heart block and ventricular arrhythmias. Treatment is discussed in Chapter 42 ('South American trypanosomiasis [Chagas' disease]' section).

Cardiomyopathy

Cardiomyopathies are a group of diseases of the myocardium leading to structural and functional abnormalities of the heart in the absence of other causes such as CAD, hypertension, valvular disease, and congenital heart disease. The ESC provides a classification system according to a phenotype-based integrated aetiological diagnosis using morphological and functional differences of the myocardium (Fig. 17.103):

- HCM
- ACM
- DCM
- RCMs
- unclassified.

Hypertrophic cardiomyopathy

HCM includes a group of inherited conditions that produce hypertrophy of the left ventricular myocardium. Left ventricular wall thickness is 15 mm or more (or if 13–14 mm requires evaluation of other features such as family history or genetics) in the absence of abnormal loading conditions such as aortic stenosis or hypertension. It is the most common cause of sudden cardiac death in young people and affects 1 in 500 of the population. The majority of cases are familial autosomal dominant and are due to mutations in the genes encoding sarcomeric proteins. The most common causes of HCM

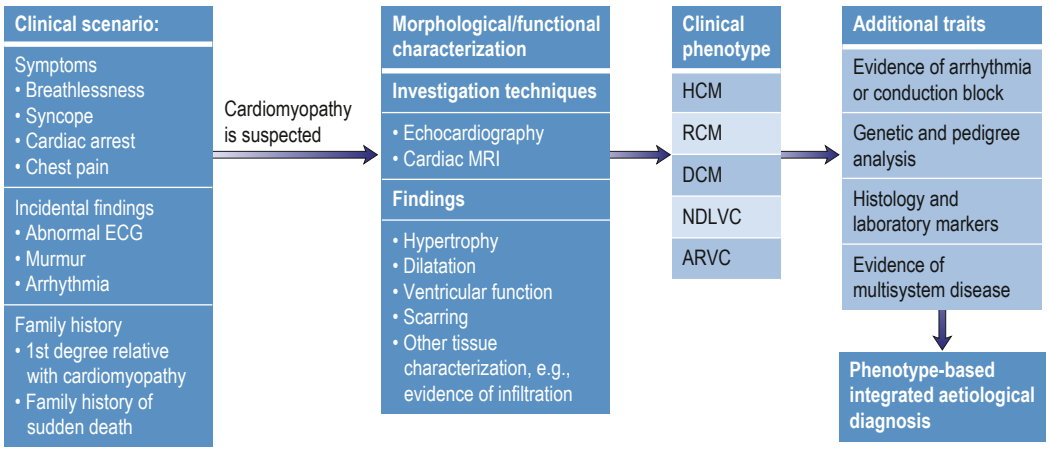


Fig. 17.103 Clinical diagnostic workflow of cardiomyopathy. *ARVC,* Arrhythmogenic right ventricular cardiomyopathy; *CMP,* cardiomyopathy; *CMR,* cardiac magnetic resonance; *DCM,* dilated cardiomyopathy; *ECG,* electrocardiogram; *HCM,* hypertrophic cardiomyopathy; *NDLVC,* non-dilated left ventricular cardiomyopathy; *RCM,* restrictive cardiomyopathy.

are mutations of the β-myosin heavy chain *MYH7* and myosin-binding protein C *MYBPC3*.

There are non-sarcomeric protein mutations in genes that control cardiac metabolism that result in glycogen storage diseases (Danon's, Pompe's, and Fabry's diseases) and cause cardiac hypertrophy that may be indistinguishable from inherited HCM on echocardiography.

Clinical features

Myocardial hypertrophy frequently involves the interventricular septum but can be in any distribution and histologically is characterized by disorganization ('disarray') of cardiac myocytes and myofibrils. Some 25% of patients have dynamic LVOT obstruction due to the combined effects of hypertrophy, systolic anterior motion of the anterior mitral valve leaflet, and rapid ventricular ejection. The salient clinical and morphological features of the disease vary according to the underlying genetic mutation. For example, marked hypertrophy is common with β-myosin heavy chain mutations, whereas mutations in troponin T may be associated with mild hypertrophy but a high risk of sudden death. The hypertrophy may not manifest before completion of the adolescent growth spurt, making the diagnosis difficult in children. HCM due to myosin-binding protein may not manifest until the sixth decade of life or later.

Symptoms

- Many cases are asymptomatic and are detected by family screening of an affected individual or by a routine ECG examination.
- Chest pain, dyspnoea, syncope or pre-syncope (typically with exertion), cardiac arrhythmias, and sudden death may occur.
- Sudden death can occur at any age but the highest rates (up to 6% per annum) occur in adolescents or young adults. Risk factors for sudden death are discussed later.
- Dyspnoea occurs due to impaired relaxation of the heart muscle or the LVOT obstruction that occurs in some patients. The systolic cavity remains small until the late stages of disease, when progressive dilation may occur. If a patient develops atrial fibrillation, there is often a rapid deterioration in clinical status due to the loss of atrial contraction and tachycardia, resulting in elevated left atrial pressure and acute pulmonary oedema.

Signs

- Double apical pulsation (forceful atrial contraction producing a fourth heart sound)
- Jerky carotid pulse because of rapid ejection and sudden obstruction to left ventricular outflow during systole
- Ejection systolic murmur due to left ventricular outflow obstruction late in systole; it can be increased by manoeuvres that decrease afterload, such as standing or Valsalva, and decreased by manoeuvres that increase afterload and venous return, such as squatting
- Pansystolic murmur due to mitral regurgitation (secondary to systolic anterior motion)
- Fourth heart sound (if not in atrial fibrillation).

Investigations

- ***ECG*** abnormalities in HCM include left ventricular hypertrophy (see Fig. 17.74), ST- and T-wave changes, and abnormal Q waves, especially in the inferolateral leads.
- ***Echocardiography*** is usually diagnostic. In classical HCM there is asymmetric left ventricular hypertrophy (involving the septum more than the posterior wall), systolic anterior motion of the mitral valve, and a vigorously contracting ventricle (Fig. 17.104). However, any pattern of hypertrophy may be seen, including concentric and apical hypertrophy.
- ***CMR*** can detect both hypertrophy and abnormal myocardial fibrosis (Fig. 17.105) and can differentiate between HCM and infiltrative cardiomyopathies.
- ***Genetic analysis***, where available, may confirm the diagnosis and provide prognostic information for the patient and relatives.

Management

The management of HCM includes treatment of symptoms and prevention of sudden cardiac death in the patient and relatives.

Risk factors for sudden death are:

- massive left ventricular hypertrophy (>30 mm on echocardiography)
- family history of sudden cardiac death (<50 years old)
- non-sustained ventricular tachycardia on 24-hour Holter monitoring
- prior unexplained syncope

- abnormal blood pressure response on exercise (flat or hypotensive response).

The presence of these cardiac risk factors is associated with an increased risk of sudden death, and patients with two or more should be assessed for ICD insertion. When the risk is less, amiodarone may be an appropriate alternative.

Patients with left ventricular outflow tract obstruction (LVOTO) should avoid dehydration and vasodilators, such as nitrates and phosphodiesterase type 5 inhibitors, as these may worsen obstruction or cause refractory hypotension. Treatment includes non-vasodilating beta-blockers (e.g., metoprolol) or, if not tolerated, verapamil or diltiazem. Disopyramide may be added if patients are still symptomatic. Mavacamten is a first-in-class cardiac myosin ATPase inhibitor and may be used as second-line therapy as it reduces LVOT gradient and improves exercise capacity in patients with HCM and symptomatic LVOTO.

The threshold for invasive treatment of significant LVOTO (usually considered ≥50 mmHg) involves careful evaluation and may include alcohol (non-surgical) ablation of the septum or surgical ventricular septal myectomy. Both appear to offer favourable results in reduction of outflow tract obstruction and subsequent improvement in exercise capacity. However, there is a risk of AV block and MI with alcohol ablation and iatrogenic VSD and AV block with surgery.

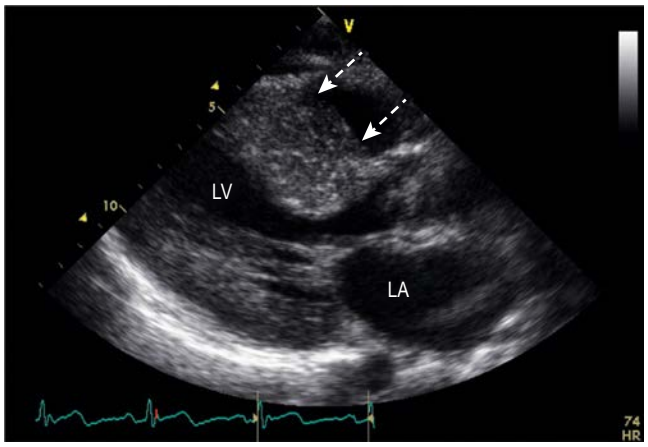


Fig. 17.104 Hypertrophic cardiomyopathy. A two-dimensional echocardiogram (short-axis view). The grossly thickened interventricular septum is shown, resulting in a small left ventricular cavity. This condition is associated with an abnormal anterior motion of the mitral valve during systole *(arrowed)*. *LA,* Left atrium; *LV,* left ventricle.

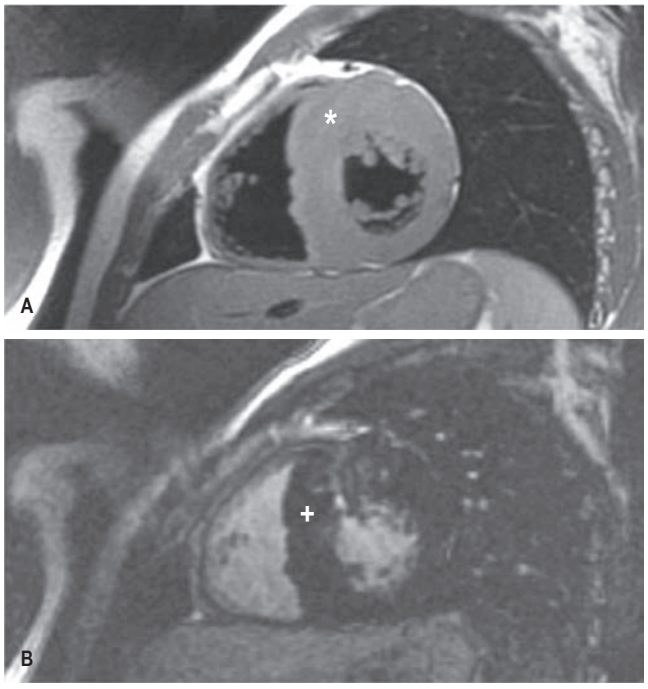


Fig. 17.105 Cardiac magnetic resonance imaging in a patient with hypertrophic cardiomyopathy. **(A)** Turbo spin-echo demonstrates marked thickening of the basal to mid-anterior wall and septum (*). **(B)** Inversion recovery image post gadolinium demonstrates regional fibrosis with enhancement (+).

Dilated cardiomyopathy

Dilated cardiomyopathy (DCM) has a prevalence of 7 to 12 in 10 000 and is characterized by dilation of the ventricular chambers and systolic dysfunction in the absence of abnormal loading conditions and CAD.

Familial DCM is predominantly autosomal dominant and can be associated with over 20 abnormal loci and genes (Fig. 17.106). Many of these are genes encoding cytoskeletal or associated myocyte proteins (dystrophin in X-linked cardiomyopathy; actin, desmin, troponin T, β-myosin heavy chain, sarcoglycans, vinculin, and lamin a/c in autosomal dominant DCM; Fig. 17.107). Many of these have prominent associated features, such as skeletal myopathy or

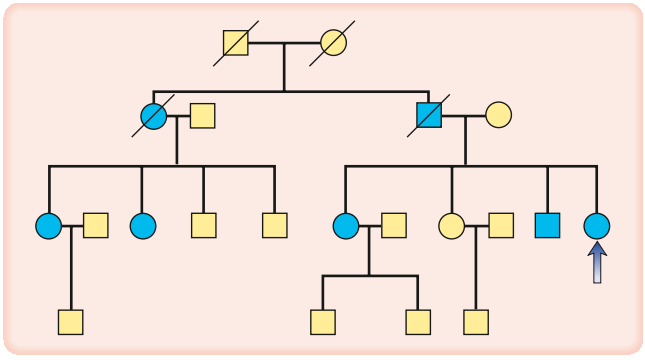


Fig. 17.106 Pedigree of a family with dilated cardiomyopathy. Blue symbols are affected family members. The *arrow* indicates the index case.

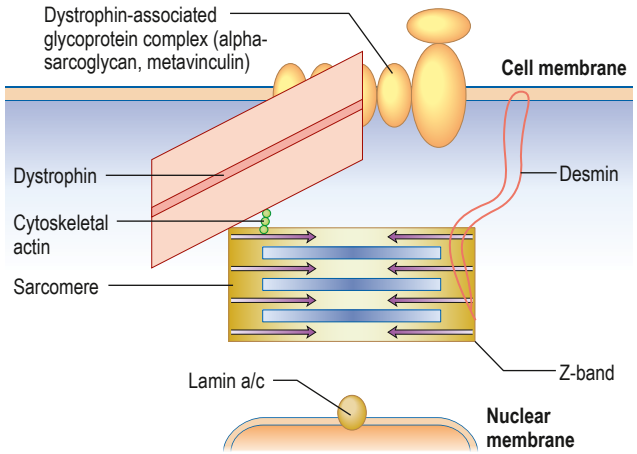


Fig. 17.107 Myocyte proteins implicated in dilated cardiomyopathy.

conduction system disease, and therefore differ from the majority of cases of DCM.

Sporadic DCM can be caused by multiple conditions:

- myocarditis: Coxsackievirus, adenoviruses, erythroviruses, HIV, bacteria, fungi, mycobacteria, parasitic (Chagas' disease)
- toxins: alcohol, chemotherapy, metals (cobalt, lead, mercury, arsenic)
- autoimmune disorders
- endocrine disorders
- neuromuscular disorders.

Clinical features

DCM can present with heart failure, cardiac arrhythmias, conduction defects, thromboembolism, or sudden death. Increasingly, evaluation of relatives of DCM patients is allowing identification of early asymptomatic disease, prior to the onset of these complications. Clinical evaluation should include a family history and construction of a pedigree where appropriate.

Investigations

- ***ECG*** may demonstrate diffuse non-specific ST-segment and T-wave changes. Sinus tachycardia, conduction abnormalities, and arrhythmias (i.e., atrial fibrillation, ventricular premature contractions, or ventricular tachycardia) are also seen.
- ***Echocardiogram*** reveals dilation of the left and/or right ventricle with reduced global function (Fig. 17.108).
- ***CMR*** may demonstrate other aetiologies of left ventricular dysfunction (e.g., previous MI) or show abnormal myocardial fibrosis (Fig. 17.109).
- ***Coronary angiography*** or CTCA should be performed to exclude CAD in all individuals at risk (generally, patients > 40 years, or younger if symptoms or risk factors are present).
- ***Biopsy*** is not normally indicated outside specialist care.
- ***Genetic testing:*** causative gene variants may be found in up to 40% and 10% to 15% of those with a presumed acquired cause such as peripartum, alcoholic, or chemotherapy-induced cardiomyopathy.

Management

Treatment is with conventional medical therapy for heart failure. Cardiac resynchronization therapy and/or ICDs may be indicated in those with an LVEF of 35% or less. ICDs may be considered in those with a high-risk gene even if LVEF is more than 35% (e.g., LMNA mutation encoding lamin A and C). Cardiac transplantation may be appropriate for certain patients.

Arrhythmogenic cardiomyopathy/ arrhythmogenic right ventricular cardiomyopathy

Arrhythmogenic cardiomyopathy (ACM) is an uncommon (1 in 1000–5000 population) inherited condition that predominantly affects the right ventricle with fatty or fibro-fatty replacement of myocytes, leading to segmental or global dilation (Fig. 17.110). Left ventricular involvement has been reported in up to 75% of cases, and some may have predominant LV disease. The fibro-fatty replacement leads to ventricular arrhythmia and risk of sudden death in its early stages and right ventricular or biventricular failure in its later stages.

Autosomal dominant ACM has been mapped to mutations in genes coding for desmosomal proteins. These are the cardiac ryanodine receptor RyR2 (also responsible for familial catecholaminergic polymorphic ventricular tachycardia, CPVT), desmoplakin, plakophilin-2, and mutations altering the regulatory sequences of the TGF-β gene.

There are two recessive forms: ***Naxos disease*** (associated with palmoplantar keratoderma and woolly hair), which is due to a

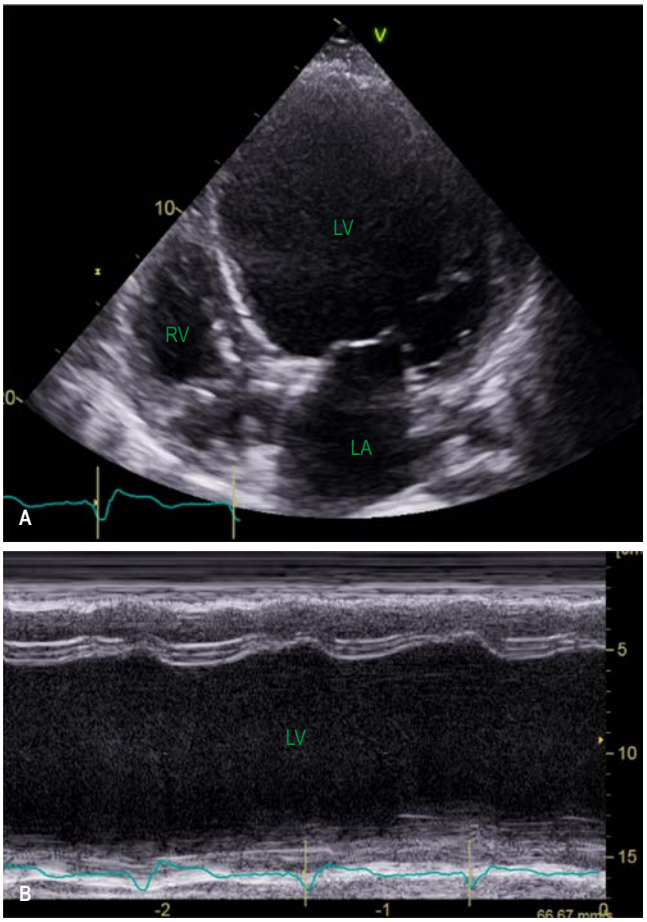


Fig. 17.108 Dilated cardiomyopathy. **(A)** Two-dimensional (apical four-chamber view) echocardiogram. The heart has a 'globular' appearance with all four chambers dilated. **(B)** M mode echocardiogram. The extremely impaired left ventricular function can be appreciated by comparing the systolic shortening fraction with that of Fig. 17.21. *LA,* Left atrium; *LV,* left ventricle; *RA,* right atrium.

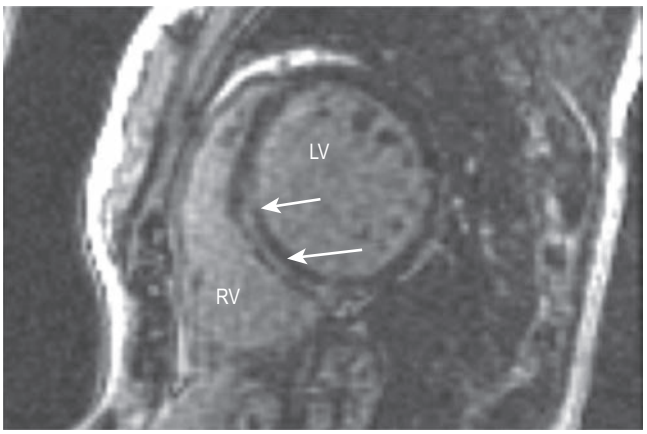


Fig. 17.109 Dilated cardiomyopathy. Inversion recovery magnetic resonance scan post gadolinium demonstrates septal mid-wall hyperenhancement *(arrowed)* consistent with fibrosis in a patient with dilated cardiomyopathy. *LV,* Left ventricle; *RV,* right ventricle.

Fig. 17.110 Gross pathological specimen demonstrating thinning and fibro-fatty replacement of the right ventricular free wall from a patient with arrhythmogenic cardiomyopathy. *(From Basso C, Thiene G, Corrado D et al. Arrhythmogenic right ventricular cardiomyopathy. Dysplasia, dystrophy, or myocarditis?* Circulation *1996; 94:983–991, with permission.)*

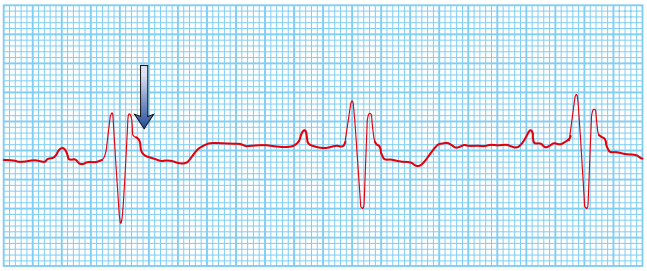


Fig. 17.111 Electrocardiogram from an adult with arrhythmogenic cardiomyopathy. The ECG demonstrates right bundle branch block and precordial T-wave insertion, with epsilon waves visible at the terminal of the QRS complex *(arrowed)*.

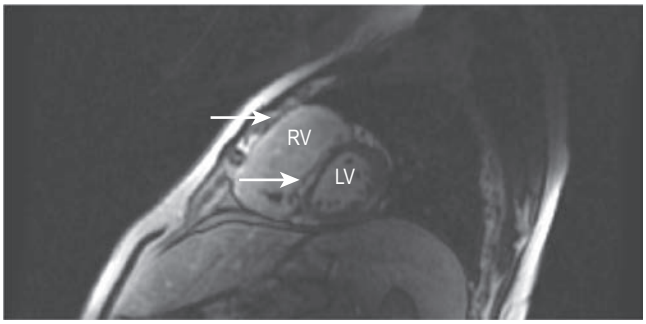


Fig. 17.112 Arrhythmogenic cardiomyopathy (ACM). Inversion recovery magnetic resonance image post gadolinium demonstrates marked hyper-enhancement *(arrowed)* in the right ventricular free wall and inferoseptum consistent with fibro-fatty replacement in right ventricle (RV) and left ventricle (LV).

mutation in junctional plakoglobin, and the rarer variant ***Carvajal's syndrome***, due to a mutation in desmoplakin.

Clinical features

Most patients are asymptomatic and men appear to be affected more. Symptomatic ventricular arrhythmia, syncope, or sudden death occurs. Occasionally, presentation is with symptoms and signs of right heart failure, although this is more common in the later stages of the disease. Some patients may be detected through family screening, although frequently, the morphological appearance of the right ventricle is normal, despite significant cardiac arrhythmias.

Investigations and diagnosis

- ***ECG*** is usually normal but may demonstrate T-wave inversion in the precordial leads related to the right ventricle (V_1–V_3). Small-amplitude potentials occurring at the end of the QRS complex (epsilon waves) may be present (Fig. 17.111), and incomplete or complete right bundle branch block is sometimes seen. Signal-averaged ECG may indicate the presence of late potentials and the delayed depolarization of individual muscle cells; 24-hour Holter monitoring may demonstrate frequent extrasystoles of right ventricular origin or runs of non-sustained ventricular tachycardia.
- ***Echocardiography*** is frequently normal but, with more advanced cases, may demonstrate right ventricular dilation and aneurysm formation; there may be left ventricular dilation.
- ***CMR*** can assess regional wall motions, size, and aneurysms of the right ventricle more accurately and in some cases can demonstrate fibro-fatty infiltration (Fig. 17.112).
- ***Genetic testing and family screening:*** the pattern of inheritance is autosomal dominant involving mutations of the desmosomal genes.

Clinical diagnosis is made using Task Force Criteria, which include:

- structural abnormalities of the right ventricle and right ventricular outflow tract (dilation and abnormal wall motion on echocardiography or MRI)
- fibro-fatty replacement of myocytes on tissue biopsy
- repolarization and conduction abnormalities on ECG or signal-averaged ECG
- ventricular tachycardia or frequent ventricular extrasystoles on Holter monitoring
- family history of ACM in a first- or second-degree relative
- premature sudden death (<35 years) due to ACM.

Management

Beta-blockers are first-line treatment for patients with non-life-threatening arrhythmias. Amiodarone is used for symptomatic arrhythmias when beta-blocker therapy fails. For refractory or life-threatening arrhythmias and those at highest risk of sudden cardiac death, catheter ablation and an ICD is required. Occasionally, cardiac transplantation is indicated for either intractable arrhythmia or cardiac failure.

Left ventricular non-compaction

Left ventricular non-compaction (LVNC) is associated with a sponge-like appearance of the left ventricle. The condition predominantly affects the apical portion of the left ventricle and may be associated with congenital heart abnormalities. LVNC is diagnosed by echocardiography, CMR, or left ventricular angiography. The natural history is unresolved but includes congestive cardiac failure, thromboembolism, cardiac arrhythmias, and sudden death. Familial and spontaneous cases have been described. It is important to differentiate LVNC from phenotypically similar conditions which may arise during athletic adaptation and pregnancy, which resolve with deconditioning and following pregnancy, respectively.

Restrictive cardiomyopathy

This is a rare condition with restrictive left and/or right ventricular pathophysiology in the setting of normal or reduced diastolic volumes of either one or both ventricles with normal or reduced systolic volumes and normal ventricular wall thickness.

The restrictive physiology produces symptoms and signs of heart failure. Conditions associated with this form of cardiomyopathy include amyloidosis (most common), sarcoidosis, glycogen storage disorders, Loeffler's endocarditis, and endomyocardial fibrosis; in the latter two conditions there is myocardial and endocardial fibrosis, associated with eosinophilia. The idiopathic form of RCM may be familial.

Clinical features

Patients with RCM may present with dyspnoea, fatigue, and embolic symptoms. On clinical examination there will be elevated JVP with diastolic collapse (Friedreich's sign) and elevation of venous pressure with inspiration (Kussmaul's sign), hepatic enlargement, ascites, and dependent oedema. Third and fourth heart sounds may be present.

Investigations

- ***ECG*** may demonstrate low-voltage QRS and ST-segment and T-wave abnormalities.
- ***Echocardiography*** shows symmetrical myocardial thickening and often a normal systolic EF but impaired ventricular filling. There is usually significant bi-atrial enlargement. In patients with amyloid the myocardium typically appears speckled with reduced longitudinal shortening with relative apical sparing (Fig. 17.113).
- ***CMR*** may show abnormal myocardial fibrosis in amyloidosis or sarcoidosis.
- ***Cardiac catheterization*** and haemodynamic studies may help distinguish between RCM and constrictive pericarditis, although volume loading may be required.
- ***Endomyocardial biopsy*** is often useful in this condition, in contrast to other cardiomyopathies, and may permit a specific diagnosis, such as amyloidosis, to be made.

Management

There is no specific treatment. Cardiac failure will require diuretics and embolic manifestations managed with anticoagulation. Cardiac transplantation is necessary in some severe cases, especially the idiopathic variety. Management of primary amyloidosis is discussed elsewhere but may include tafamidis, a transthyretin stabilizer. Liver transplantation may be effective in familial amyloidosis (due to production of mutant pre-albumin) and may lead to reversal of the cardiac abnormalities.

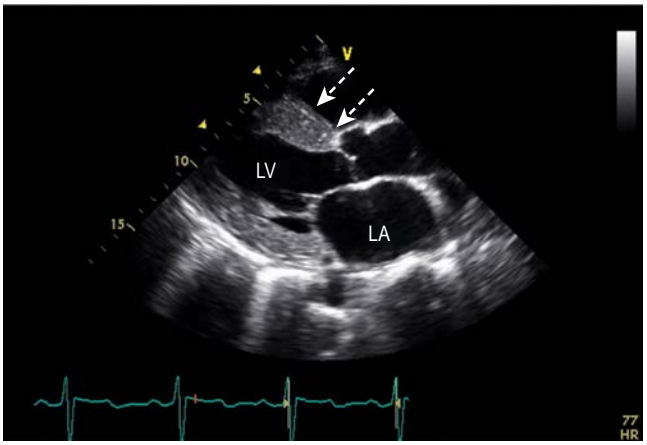


Fig. 17.113 Cardiac amyloidosis. Parasternal long-axis echocardiogram demonstrating left ventricular hypertrophy with reduced systolic function. The septum has a speckled appearance *(arrowed)*. *LA,* Left atrium; *LV,* left ventricle.

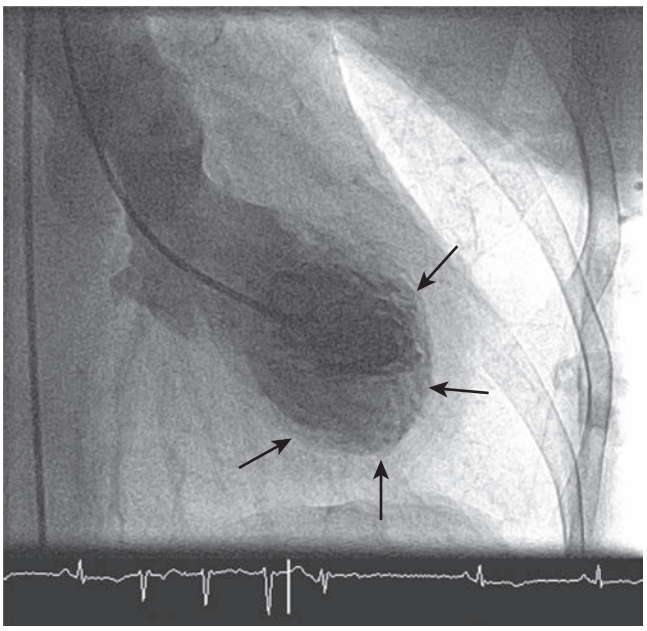


Fig. 17.114 Takotsubo cardiomyopathy. Left ventricular angiography in a patient admitted with chest pain following emotional stress demonstrates apical ballooning *(arrowed)* consistent with Takotsubo cardiomyopathy. Diagnostic coronary angiography showed normal coronary arteries.

Acquired cardiomyopathies

Takotsubo syndrome (stress cardiomyopathy)

Patients with this condition typically are postmenopausal women and present acutely with chest pain and breathlessness associated with ECG changes and elevated cardiac biomarkers consistent with acute MI. Diagnostic coronary angiography typically demonstrates unobstructed coronary arteries, with characteristic akinesia of the mid-apical segments of the left ventricle on ventriculography or echocardiography, and preserved basal function (Fig. 17.114). The pathophysiology is uncertain but disease may be due to transient catecholamine excess, coronary vasospasm, abnormalities of the coronary microcirculation, and hypertrophy of the basal septum. Severe cases may have cardiogenic shock and pulmonary oedema.

Patients with a significant left ventricular gradient may respond to cautious beta-blockade. Complete recovery of function is usual within 4 to 6 weeks, and therefore technically this is not classified as a cardiomyopathy. Recurrence is uncommon and may be reduced by long-term beta-blocker therapy.

Peripartum cardiomyopathy

This rare condition affects women in the last trimester of pregnancy or within 5 months of delivery. It presents as a DCM; is more common in obese, multiparous women over 30 years old and with Black ethnicity; and is associated with pre-eclampsia. Nearly half of patients will recover to normal function within 6 months, but in some it can cause progressive heart failure and sudden death. Genetic testing should be considered for an underlying cardiomyopathy, and counselling given regarding future pregnancies.

Arrhythmia-induced cardiomyopathy ('tachycardia cardiomyopathy')

Patients with prolonged periods of supraventricular arrhythmias (e.g., atrial fibrillation, atrial flutter, atrial tachycardia), and particularly with tachycardia and frequent ventricular ectopy (>15% of all heartbeats), are at risk of deterioration of left ventricular function. Patients present with heart failure symptoms and may have palpitations. Rate control with beta-blockers and conventional heart failure therapy is required, but cardioversion and ablation may be necessary to allow for recovery of cardiac function.

Further reading

Arbelo E, Protonotarios A, Gimeno J et al. 2023 ESC Guidelines for the management of cardiomyopathies: developed by the task force on the management of cardiomyopathies of the European Society of Cardiology. *Eur Heart J* 2023; ehad194.

PERICARDIAL DISEASE

The pericardium acts as a protective covering for the heart. It consists of an outer fibrous pericardial sac and an inner serous pericardium; the latter is made up of the inner visceral epicardium that lines the heart and great vessels, and its reflection, the outer parietal pericardium that lines the fibrous sac. The normal amount of pericardial fluid is 20 to 49 mL, and this fluid lubricates the surface of the heart. Presentations of pericardial disease include:

- acute, incessant (usually 4–6 weeks up to 3 months) and chronic (>3 months) pericarditis
- pericardial effusion and cardiac tamponade
- constrictive pericarditis
- pericardial masses.

Acute pericarditis

This term refers to inflammation of the pericardium, which may account for 5% of accident and emergency presentations with chest pain. The condition is more common in young adults, particularly men. Acute pericarditis has numerous aetiologies (**Box 17.48**), although, in most cases, a cause is not identified (idiopathic).

Clinical features and investigations

There are four common clinical features of acute pericarditis:

- ***chest pain*** that may be pleuritic, exacerbated by movement, and relieved by sitting forwards
- ***pericardial friction rub***

? Box 17.48 Aetiology of pericarditis

Infectious pericarditis
- Viral (Coxsackievirus, echovirus, mumps, herpes, HIV)
- Bacterial (staphylococcal, streptococcal, pneumococcal and meningococcal infections, *Haemophilus influenzae*, mycoplasmosis, borreliosis, chlamydia)
- Tuberculous
- Fungal (histoplasmosis, coccidioidomycosis, candidiasis)

Post-myocardial infarction pericarditis
- Acute myocardial infarction (early)
- Dressler's syndrome (late)

Malignant pericarditis
- Primary tumours of the heart (mesothelioma)
- Metastatic pericarditis (breast and lung carcinoma, lymphoma, leukaemia, melanoma)

Uraemic pericarditis

Myxoedematous pericarditis

Chylopericardium

Autoimmune pericarditis
- Collagen-vascular (rheumatoid arthritis, rheumatic fever, systemic lupus erythematosus, scleroderma)
- Drug-induced (procainamide, hydralazine, isoniazid, doxorubicin, cyclophosphamides)

Post-radiation pericarditis

Post-surgical pericarditis
- Post-pericardiotomy syndrome

Post-traumatic pericarditis

Familial and idiopathic pericarditis

- ***ECG abnormalities*** with widespread concave-upwards (saddle-shaped) ST elevation (**Fig. 17.115**), reciprocal ST depression in leads AVR and V_1, and PR segment depression
- ***pericardial effusion***.

A clinical diagnosis of acute pericarditis can be made with two out of four of these features.

Inflammatory blood tests will be elevated ***(CRP, ESR, white cell count)*** and can be used to monitor disease resolution. Cardiac troponin and creatine kinase may be elevated in cases with myopericarditis. Chest X-ray may demonstrate cardiomegaly in patients with a large pericardial effusion.

Management and treatment

In 70% to 90% of cases, acute pericarditis is self-limiting. Most patients can be treated with aspirin 750 to 1000 mg three times daily or ibuprofen 600 to 800 mg three times daily for 1 to 2 weeks. Colchicine for 3 months is also recommended to improve response and importantly reduce recurrence. Corticosteroids should be reserved for patients with a known immune cause, as their use is associated with an increased rate of recurrence. High-risk patients (fever >38°C, subacute onset, large effusion, tamponade, lack of response to therapy) should be admitted for treatment. Activity should be restricted until the inflammatory markers are normalized, and competitive athletes should avoid exercise for at least 3 months.

Incessant or chronic pericarditis

About 20% of cases of acute pericarditis go on to develop idiopathic relapsing pericarditis, which may be incessant (recurring within 6 weeks) or chronic (lasting >3 months). The first-line treatment is non-steroidal anti-inflammatory drugs (NSAIDs) or aspirin with colchicine for up to 6 months. In resistant cases low-dose oral corticosteroids may be effective, particularly in autoimmune disease

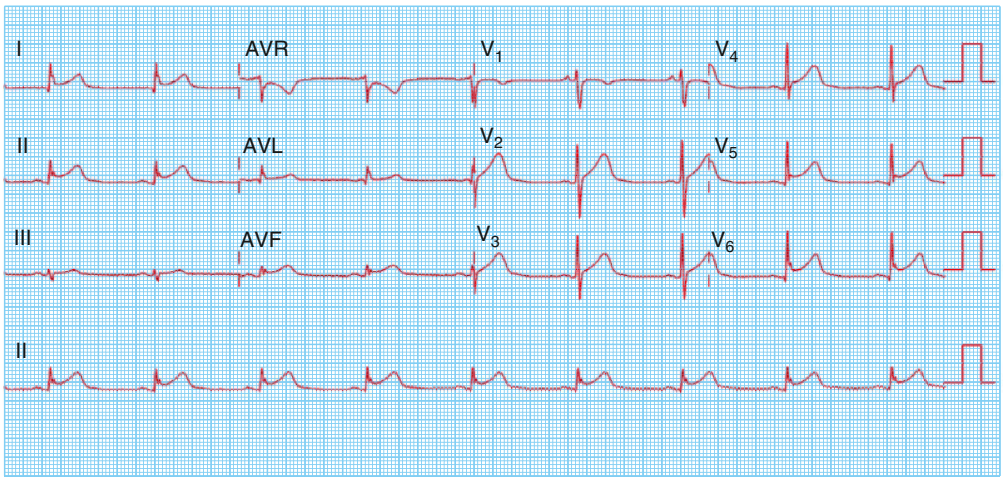


Fig. 17.115 Electrocardiogram associated with acute pericarditis. Note the diffuse, concave, raised ST segments seen in most leads.

and if an infectious cause has been excluded. CT or CMR may be helpful in demonstrating thickened (>4 mm) or inflamed pericardium.

Tuberculous pericarditis

This is uncommon (<5%) in the developed world but may account for over 50% of cases in low- and middle-income countries and over 90% in patients with HIV. Tuberculous pericarditis usually presents with chronic low-grade fever, particularly in the evening, associated with features of acute pericarditis, dyspnoea, malaise, night sweats, and weight loss. Pericardial aspiration is often required to make the diagnosis. Constrictive pericarditis is a frequent outcome. Treatment is as for pulmonary tuberculosis (see Chapter 20, 'Tuberculosis' section) with high-dose corticosteroids daily for 2 to 6 weeks.

Post-cardiac injury syndromes

This includes pericarditis following MI (Dressler's syndrome), cardiac surgery, or trauma. An autoimmune reaction is triggered by myocardial or pericardial damage, leading to pericarditis. Recurrences are common. Colchicine and anti-inflammatory agents are the treatment of choice.

Malignant pericarditis

Carcinoma of the bronchus, carcinoma of the breast, and Hodgkin's lymphoma are the most common causes of malignant pericarditis. Leukaemia and malignant melanoma are also associated with pericarditis. A substantial pericardial effusion is typical and is due to obstruction of lymphatic drainage from the heart. The effusion is often haemorrhagic. Radiation and therapy for thoracic tumours may cause radiation injury to the pericardium, resulting in serous or haemorrhagic pericardial effusion and pericardial fibrosis.

Pericardial effusion and cardiac tamponade

A pericardial effusion is a collection of fluid within the potential space of the serous pericardial sac (Fig. 17.116); it commonly accompanies an episode of acute pericarditis. When a large volume collects in this space, ventricular filling is compromised, leading to haemodynamic compromise. This is known as cardiac tamponade.

Clinical features

Symptoms of a pericardial effusion commonly reflect the underlying pericarditis. On examination:

- Heart sounds are soft and distant.
- Apex beat is commonly obscured.
- A friction rub may be evident due to pericarditis in the early stages, but this becomes quieter as fluid accumulates and pushes the layers of the pericardium apart.
- Rarely, the effusion may compress the base of the left lung, producing an area of dullness to percussion below the angle of the left scapula (Ewart's sign).
- As the effusion worsens, signs of cardiac tamponade may become evident:
  - raised JVP with sharp × descent
  - Kussmaul's sign (rise in JVP/increased neck vein distension during inspiration)
  - pulsus paradoxus (an exaggeration in the normal variation in pulse pressure seen with inspiration, such that there is a drop in systolic blood pressure of ≥10 mmHg)
  - reduced cardiac output.

Investigations

- ***ECG*** reveals low-voltage QRS complexes (<0.5 mV in limb leads) with sinus tachycardia and there may be electric alternans (alteration of QRS amplitude or axis between beats).
- ***Chest X-ray*** (see **Fig. 17.116A**) shows a large, globular heart with clear lung fields. Typically, the pulmonary veins are not distended.
- ***Echocardiography*** (see Fig. 17.116) is the most useful technique for demonstrating the effusion and looking for evidence of tamponade: late diastolic collapse of the right atrium, early diastolic collapse of the right ventricle, ventricular septum displacement into the left ventricle during inspiration, diastolic flow reversal in the hepatic veins during expiration, and dilated inferior vena cava with less than 50% reduction during inspiration.
- ***Cardiac CT or MRI*** is helpful if loculated pericardial effusions are suspected (post cardiac surgery).
- ***Pericardiocentesis*** is the removal of pericardial fluid with an aseptic technique under echocardiographic guidance (see **Fig. 17.32**). It is indicated to improve symptoms and also can aid with diagnosis, particularly when an infectious or malignant effusion is suspected.
- ***Pericardial biopsy*** may be needed if tuberculosis is suspected and pericardiocentesis is not diagnostic.

Other tests may be required to identify underlying causes: for example, blood cultures, autoantibody screen, and cross-sectional imaging usually with CT.

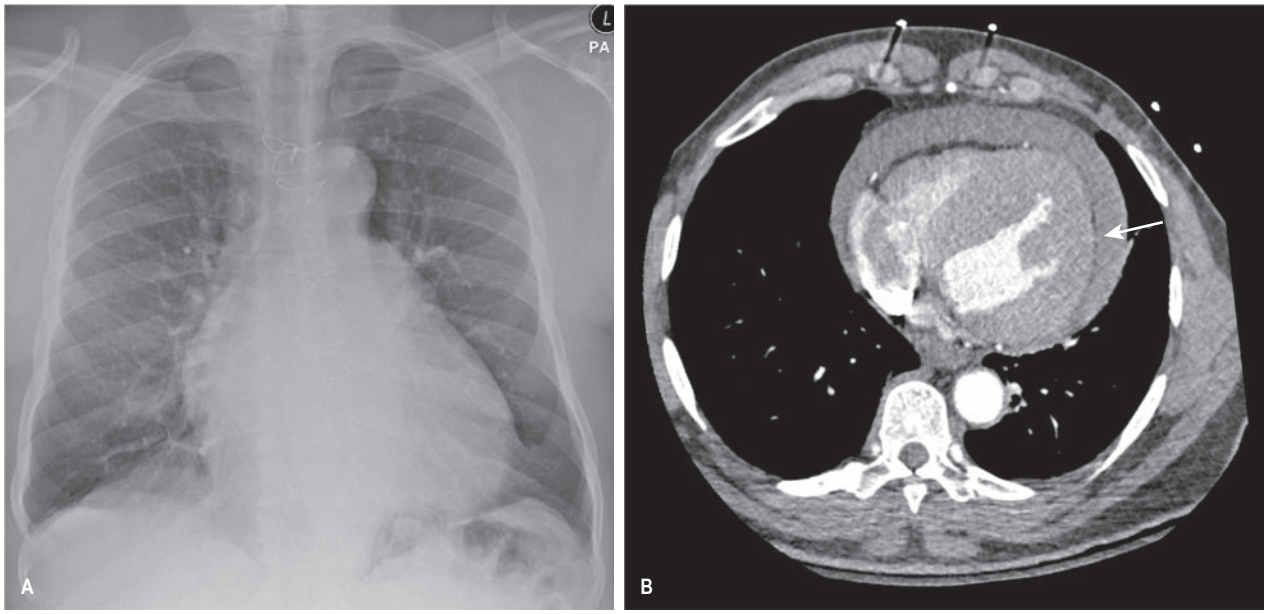


Fig. 17.116 Pericardial effusion. **(A)** Note the globular appearance to the cardiac silhouette on chest X-ray. **(B)** CT scan confirms the circumferential pericardial effusion *(arrowed)*. Note the absence of pericardial calcification that is a feature of chronic pericardial disease.

Management

An underlying cause should be sought and treated if possible. Most pericardial effusions resolve spontaneously. However, when the effusion collects rapidly, tamponade may result. Pericardiocentesis is then indicated to relieve the pressure; a drain may be left in place temporarily to allow sufficient release of fluid. Pericardial effusions may re-accumulate, most commonly due to malignancy (in the UK). This may require surgical creation of a pericardial window, or pericardiotomy, to allow the slow release of fluid into the surrounding tissues, usually under video-assisted thoracoscopy or open surgery.

Constrictive pericarditis

Certain causes of pericarditis, such as tuberculosis, haemopericardium, bacterial infection, and rheumatic heart disease, result in the pericardium becoming thick, fibrous, and calcified. This may also develop late after open heart surgery, and fibrosis also occurs with the use of dopamine agonists, such as cabergoline and pergolide. In many cases these pericardial changes do not cause any symptoms. If, however, the pericardium becomes so inelastic as to interfere with diastolic filling of the heart, constrictive pericarditis is said to have developed. As these changes are chronic, allowing the body time to compensate, this condition is not as immediately life-threatening as cardiac tamponade, in which the circulation is more acutely compromised.

Constrictive pericarditis should be distinguished from RCM. The two conditions are very similar in their presentation, but the former is fully treatable, whereas most cases of the latter are not. In the late stages of constrictive pericarditis the subepicardial layers of myocardium may undergo fibrosis, atrophy, and calcification.

Clinical features

The symptoms and signs of constrictive pericarditis occur due to:

- reduced ventricular filling (raised JVP with deep y descent, Kussmaul's sign, pulsus paradoxus)
- systemic venous congestion (ascites, dependent oedema, hepatomegaly and raised JVP)
- pulmonary venous congestion (dyspnoea, cough, orthopnoea, PND).

Less commonly, it may cause:

- reduced cardiac output (fatigue, hypotension, reflex tachycardia)
- rapid ventricular filling (a 'pericardial knock' is heard in early diastole at the lower left sternal border)
- atrial dilation (30% of cases have atrial fibrillation).

Investigations

- ***Chest X-ray*** shows a relatively small heart, disproportionate to the symptoms of heart failure. Pericardial calcification may be present in up to 50%. A lateral chest film may be useful for detecting calcification that is missed on a postero-anterior film. However, a calcified pericardium is not necessarily a constricted one.
- ***ECG*** reveals low-voltage QRS complexes with generalized non-specific ST- and T-wave changes.
- ***Echocardiography*** shows small ventricular cavities with normal wall thickness, moderate atrial enlargement, and a dilated inferior vena cava. A thickened, calcified pericardium may be seen. Doppler studies may be useful.
- ***CT and CMR*** are used to assess pericardial anatomy, thickness (≥4 mm), and inflammation (see Fig. 17.26).
- ***Endomyocardial biopsy*** may be helpful in distinguishing constrictive pericarditis from RCM in difficult cases.
- ***Cardiac catheterization*** will usually reveal equal end-diastolic pressures in the left and right ventricles, owing to pericardial constriction.

RCM is a close mimic of constrictive pericarditis, and all the above tests may help to distinguish the two conditions.

Management

The treatment for chronic constrictive pericarditis is complete resection of the pericardium. This is a risky procedure with a high

complication rate due to the presence of myocardial atrophy in many cases at the time of surgery. Thus early pericardiectomy is suggested in non-tuberculous cases, before severe constriction and myocardial atrophy have developed.

In cases of tuberculous constriction the presence of pericardial calcification implies chronic disease. Early pericardiectomy with anti-tuberculous drug cover is used in these cases. If there is no calcification, a course of anti-tuberculous therapy should be attempted first. If the patient's haemodynamic state remains static or deteriorates after 4 to 6 weeks of therapy, pericardiectomy is recommended.

Further reading

Adler Y, Charron P, Imazio M et al. 2015 ESC Guidelines for the diagnosis and management of pericardial diseases: The Task Force for the Diagnosis and Management of Pericardial Diseases of the European Society of Cardiology (ESC). *Eur Heart J* 2015; 36:2921–2964.
Bogaert J, Francone M. Cardiovascular magnetic resonance in pericardial diseases. *J Cardiovasc Magn Reson* 2009; 11:14.
Welch TD. Constrictive pericarditis: diagnosis, management and clinical outcomes. *Heart* 2018; 104:725–731.

PERIPHERAL VASCULAR DISEASE

Peripheral arterial disease

Peripheral vascular disease (PVD) is commonly caused by atherosclerosis and usually affects the aorto-iliac or infra-inguinal arteries. It is present in 7% of middle-aged men and 4.5% of middle-aged women and increases sharply with age. The mortality associated with coexisting cardiovascular and cerebrovascular disease means that only a proportion of patients with PVD progress to losing their leg.

Limb ischaemia

Limb ischaemia may be classified as chronic or acute.

Chronic lower limb ischaemia

Common risk factors are:

- smoking
- diabetes
- hypercholesterolaemia
- hypertension.

Other risk factors which may lead to premature atherosclerosis include autoimmune and chronic inflammatory conditions (e.g., SLE, rheumatoid arthritis), thrombophilia, and hyperhomocysteinaemia.

Clinical features

Symptoms

Peripheral arterial disease can be described using the Fontaine classification:

- Stage I: asymptomatic
- Stage II: intermittent claudication
- Stage III: rest pain/nocturnal pain
- Stage IV: necrosis/gangrene.

Patients with ***intermittent claudication*** complain of exertional discomfort, most commonly in the calf, which is relieved by rest. Patients with aorto-iliac disease may experience pain in the buttock, hip, or thigh and may notice erectile dysfunction. The 'claudication distance' may be reproducible.

Patients with ***rest pain*** experience severe, unremitting pain in the foot, which stops a patient from sleeping. It is partially relieved by dangling the foot over the edge of the bed or standing on a cold floor.

Patients with severe PVD or critical lower limb ischaemia may have ulceration or ***necrosis*** of the tissue ***(gangrene)***.

Signs

The lower limbs are cold with dry skin and lack of hair. Pulses may be diminished or absent. Ulceration may occur in association with dark discoloration of the toes or gangrene. The abdomen should be examined for a possible aortic aneurysm.

Differential diagnosis

Symptoms may be confused with those of:

- spinal canal claudication (when all pulses are present)
- osteoarthritis of the hip/knee (knee pain often at rest)
- peripheral neuropathy (associated with numbness and tingling)
- popliteal artery entrapment (young patients who may have normal pulses)
- venous claudication (bursting pain on walking with a previous history of a DVT)
- fibromuscular dysplasia
- Buerger's disease (young males, heavy smokers).

Investigations

An estimation of the anatomical level of disease may be possible with the examination of pulses. First-line non-invasive assessment of the severity of disease is by measurement of the ***ankle/brachial pressure index*** (ABPI). This is a measurement of the cuff pressure at which blood flow is detectable by Doppler in the posterior tibial or anterior tibial artery compared to the pressure in the brachial artery. Intermittent claudication is associated with an ABPI of 0.5 to 0.9, while values less than 0.5 are associated with critical limb ischaemia. The sensitivity of the test may be improved by a fall in ABPI after exercise. If the arteries are heavily calcified and incompressible, commonly where there is coexisting renal or diabetic disease, the ABPI will be falsely elevated. In these patients toe pressure values are more sensitive. Diagnostic imaging includes the following options:

- ***Duplex ultrasound (DUS)*** can provide an accurate anatomical map of the lower limbs with sensitivity of 87% and specificity of more than 95% to detect stenosis of greater than 50% compared to formal angiography.
- ***Contrast-enhanced MR angiography*** provides excellent imaging of both legs with a single contrast injection without exposure to ionizing radiation and is commonly used to assess the extent of disease prior to planned intervention. It can sometimes overestimate the degree of stenosis.
- ***CT angiography*** is an effective alternative to MR angiography, where MR is contraindicated, although extensive calcification may obscure stenoses. CT angiography requires ionizing radiation and iodinated contrast media.
- ***Digital subtraction angiography (DSA)*** provides an arterial map (Fig. 17.117) and may be required for below-the-knee arteries but needs peripheral arterial cannulation and exposes the individual to iodinated contrast. It should be reserved for use in patients immediately prior to intervention.

Management

Medical

All patients need aggressive risk factor management, as PVD also puts them at significant risk of ischaemic heart disease and cerebrovascular disease. Preventative strategies include antiplatelet treatment, lipid-lowering therapies, smoking cessation support, and optimal management of hypertension and diabetes. Regular review by a chiropodist may be needed.

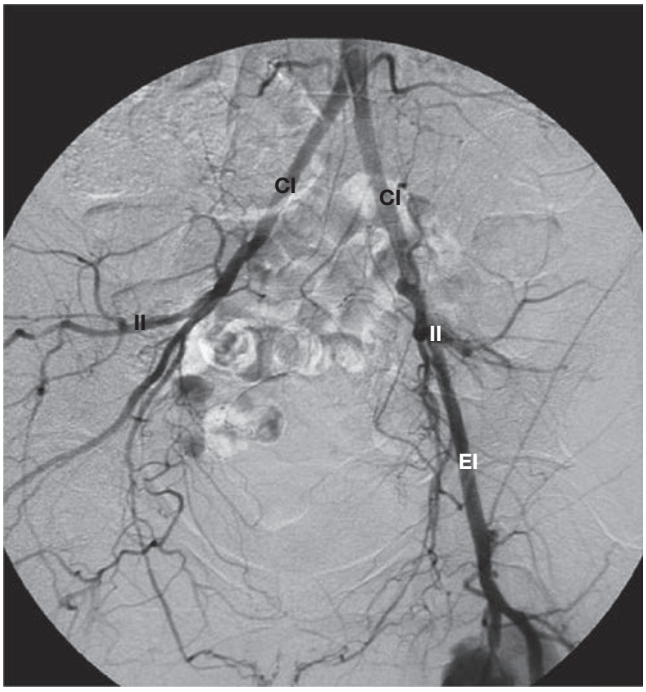


Fig. 17.117 Angiogram showing occlusion of the right external iliac artery. *CI,* Common iliac; *EI,* external iliac; *II,* internal iliac.

A supervised exercise programme should be offered to all patients with intermittent claudication, where they are encouraged to undergo aerobic exercise for at least 2 hours per week, up to the point of maximal pain.

Naftidrofuryl oxalate is a vasodilator agent that inhibits vascular and platelet 5-hydroxytryptamine$_2$ (5-HT$_2$) receptors and can reduce lactic acid levels. It may increase walking distance and improve quality of life and may be considered for a trial period of 3 to 6 months when patients are reluctant to undergo interventional treatment and supervised exercise has proved insufficient to control symptoms.

Surgical and radiological

Vascular intervention for stable claudication is not generally recommended, except when symptoms are severe or disabling. ***Percutaneous transluminal angioplasty*** is the first option and is carried out via a catheter inserted into the femoral artery. The long-term patency rates decrease as the angioplasty becomes more distal. Overall results of angioplasty appear to be similar to those achieved with a continued exercise programme. Arterial stents may be deployed in recurrent iliac disease, and DESs allowing long-term patency are also used. ***Bypass procedures*** may be performed using Dacron, polytetrafluoroethylene (PTFE), or autologous veins. Bypasses to distal vessels have poorer long-term patencies. Prosthetic grafts have equal patencies in above-knee bypasses but are inferior to veins in bypasses below the knee. In severe ischaemia with un-reconstructable arterial disease, an ***amputation*** may be necessary. An amputation commonly leads to loss of independence, with only 70% of below-knee and 30% of above-knee amputees achieving full mobility.

Acute lower limb ischaemia

Clinical features

Symptoms

Patients complain of the 'five Ps': ***pain***, the fact that the leg looks white ***(pallor)***, ***paraesthesia***, ***paralysis***, and the sensation that it is ***perishingly*** cold. The pain is unbearable and normally requires opioids for relief.

Signs

The limb is cold, with mottling or marbling of the skin. Pulses are diminished or absent. Sensation in and movement of the leg are reduced in severe ischaemia. Patients may develop a compartment syndrome with pain in the calf on compression.

Aetiology

Acute limb ischaemia (ALI) may occur because of embolic or thrombotic disease. ***Embolic disease*** is commonly due to cardiac thrombus and cardiac arrhythmias such as atrial fibrillation. Emboli may also occur secondary to aneurysm thrombus or thrombus on atherosclerotic plaques. Emboli from atrial myxomas are rare.

ALI is more often due to ***thrombotic disease***. Acute thrombus usually forms on a chronic atherosclerotic stenosis in a patient who has previously reported symptoms of claudication. Thrombus may also form in normal vessels in individuals who are hypercoagulable because of malignancy or thrombophilia defects. Prosthetic or venous grafts may also thrombose either *de novo* or secondary to a developing stenosis in the graft or the native vessels. Popliteal aneurysms may thrombose or embolize distally.

Acute upper limb ischaemia may be caused by similar processes or occur secondary to external compression with a cervical rib/band.

Investigations

Investigations are similar to those described for chronic lower limb disease and must be carried out urgently.

Management

Medical

Management is dependent on the degree of ischaemia. Patients showing improvement may be treatable with heparin, analgesia, and appropriate treatment of the underlying cause. Those with emboli following MI or atrial fibrillation need long-term anticoagulation. Wounds, ulceration, and infection require careful management.

Surgical and radiological

Patients with mild to moderate ischaemic symptoms who have an occluded graft may need graft thrombolysis. Intra-arterial thrombolysis may reveal an underlying stenosis within a graft or native vessel that could be treated with angioplasty.

Patients with severe ALI require emergency surgical or radiological attempts at revascularization. An embolus may benefit from its surgical removal (embolectomy). A bypass graft may be required after occlusion of a popliteal aneurysm or acute-on-chronic lower limb arterial disease. When an ischaemic limb is revascularized, the sudden improvement in blood flow can cause reperfusion injury, with release of toxic metabolites into the circulation. In muscle compartments the consequent oedema may lead to a 'compartment syndrome', which requires fasciotomies (release of the fascia to prevent muscle damage). An amputation may be warranted in un-reconstructable disease or severe ischaemia. In patients dying from other causes ALI may occur as a 'terminal event' and intervention may then be inappropriate.

Aneurysmal disease

Aneurysms are defined as a permanent dilation of the artery 50% greater than the normal diameter and may be classified as true or false. In true aneurysms the arterial wall forms the wall of the aneurysm. The arteries most frequently involved are the abdominal aorta, the iliac, popliteal and femoral arteries, and the thoracic aorta (in decreasing frequency). In false aneurysms (pseudoaneurysms) the surrounding tissues form the wall of the aneurysm. False aneurysms can occur following femoral artery puncture. A haematoma is formed because of inadequate compression of the entry site and continued bleeding into the surrounding compressed soft tissue forms the wall of this aneurysm.

Abdominal aortic aneurysm

Abdominal aortic aneurysms (AAAs) occur most commonly below the renal arteries (infra-renal). The incidence increases with age, with AAAs being present in 5% of the population over 60 years of age. They arise five times more frequently in men, and in 1 in 4 male children of an affected individual. Aneurysms may occur secondary to atherosclerosis, infection (syphilis, *Escherichia coli*, *Salmonella*), and trauma or may be genetic (Marfan's or Ehlers–Danlos syndrome).

Screening

In England and Wales approximately 6000 deaths occur each year from a ruptured aortic aneurysm. The mortality rate for elective surgery is 3.5% or lower, compared to 50% for an emergency open repair. As with all cardiovascular disease, there are fixed risk factors—age, male sex, strong family history—and modifiable risk factors—smoking, hypertension, hypercholesterolaemia.

The UK recommendation is that screening should be offered to men aged 65 to 74 years, who receive an abdominal ultrasound.

- ***Normal*** or ***less than 3 cm*** aortas do not require treatment or further scans.
- ***Small (3–4.4 cm)*** aortas require annual ultrasound surveillance and GP review to optimize lifestyle.
- ***Medium (4.5–5.4 cm)*** aortas require quarterly ultrasound surveillance and cardiovascular secondary prevention therapy.
- ***Large (≥5.5 cm)*** aortas are referred for assessment and possible elective repair.

Screening of female patients is not recommended, as the prevalence is much lower.

Clinical features

Symptoms

Most aneurysms are asymptomatic and are found on routine abdominal examination or plain X-ray or during investigations for other problems. Rapid expansion or rupture of an AAA may cause ***severe pain*** (epigastric pain radiating to the back). A ruptured AAA causes ***hypotension***, ***tachycardia***, ***profound anaemia***, and ***sudden death***. The symptoms of rupture may mimic renal colic, diverticulitis, and severe lower abdominal or testicular pain. Gradual erosion of the vertebral bodies may cause non-specific back pain. The aneurysm may embolize distally, causing ALI. Inflammatory aneurysms can obstruct adjacent structures, such as the ureter, duodenum, and vena cava. Rarely, aneurysms can present with severe haematemesis secondary to an aorto-duodenal fistula.

Signs

The aorta is retroperitoneal, and in overweight patients there may be no overt signs. An aneurysm is suspected if a pulsatile, expansile abdominal mass is felt. The presence of an AAA should alert a clinician to the possibility of popliteal aneurysms. Patients may present with 'trash feet', dusky discoloration of the digits secondary to emboli from the aortic thrombus.

Management

As with any operation, treatment of an asymptomatic aneurysm depends on the balance of operative risk and conservative management. The UK Small Aneurysm Trial showed that patients with infra-renal AAAs did best with an operation if the aneurysm was:

- 5.5 cm or more in diameter
- expanding at a rate of more than 1 cm/year
- symptomatic.

Medical

Aneurysmal disease needs careful control of hypertension, smoking cessation, and lipid-lowering medication. Patients with AAAs of less than 5.5 cm are followed up by regular ultrasound surveillance.

Surgical and endovascular intervention

Intervention to treat AAA can be performed by open aneurysm repair, hand-assisted or total laparoscopic repair, or endovascular stent insertion (via the femoral or iliac arteries). The Endovascular Aneurysm Repair (EVAR) studies (stent versus open surgical repair) and EVAR 2 (stent versus medical therapy in patients unsuitable for open repair) investigated the role of endovascular stents in patients with an AAA of 5.5 cm or more on CT. 3D-printed fenestrated grafts which are individualized to patients' anatomy may be employed. In EVAR the 30-day mortality rate was 1.7% with stenting versus 4.7% with surgery (p = 0.009), but the long-term mortality rate was similar in both groups at 4 years. In EVAR 2 the 30-day mortality rate with stenting was 9%. Long-term mortality rate was similar in both stent and medical therapy groups. Current UK guidance recommends surgical repair rather than EVAR for unruptured aneurysms in patients who are fit for open surgery, and EVAR or conservative (medical) management in patients at high surgical or anaesthetic risk. EVAR has a more prominent role in treating ruptured aneurysms, especially in older and frailer patients.

Prognosis

After repair, patients with an AAA should return to normal activity within a few months.

Thoraco-abdominal aneurysm

The ascending, arch, or descending thoracic aorta may become aneurysmal. Ascending thoraco-abdominal aneurysms (TAAs) occur most commonly in patients with MFS or hypertension. Descending or arch TAAs occur secondary to atherosclerosis and are now rarely due to syphilis.

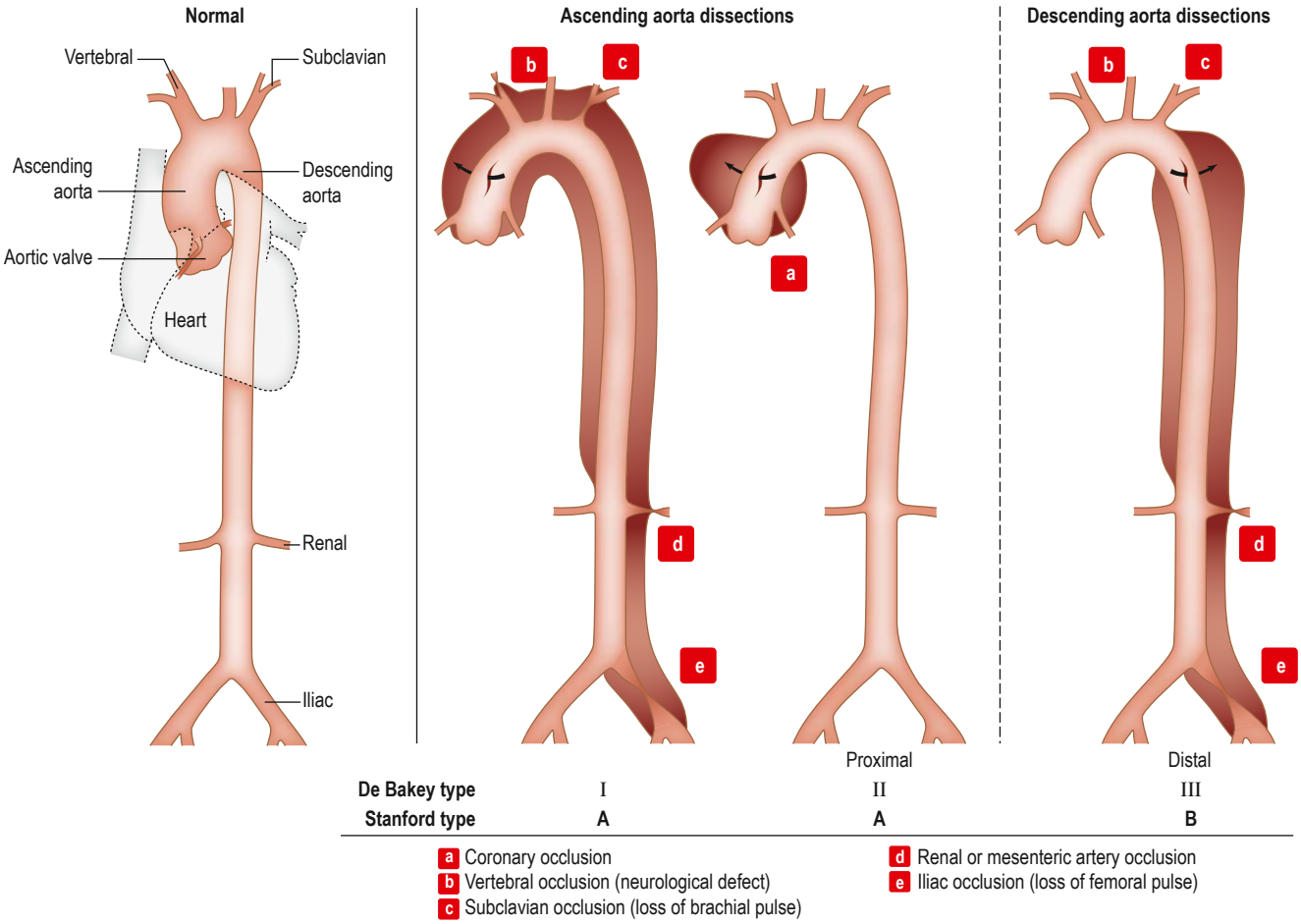


Fig. 17.118 **Classification of aortic aneurysms.**

Clinical features

Most aneurysms are asymptomatic and are found on routine chest X-ray or cardiological investigation. Rapid expansion may cause severe pain (chest pain radiating to the upper back) and rupture is associated with hypotension, tachycardia, and death. Chest symptoms from expansion may include stridor (compressed bronchial tree), haemoptysis (aortobronchial fistula), and hoarseness (compression of the recurrent laryngeal nerve). Aorto-oesophageal fistula is an uncommon cause of haematemesis.

Investigations

- ***CT or MRI scan*** is used for assessment of a TAA.
- ***Aortography*** may be helpful for assessing the position of the key branches in relation to the aneurysm.
- ***TOE*** can be useful for identifying an aortic dissection.

Management

If the aneurysm is bigger than 6 cm, then open repair or stenting may be appropriate, but these procedures can be technically difficult and carry a high risk of mortality and paraplegia. 3D-printed fenestrated aortic stents may be employed. EVAR is, at present, the procedure of choice for isolated descending thoracic aneurysms.

Acute aortic syndromes

Acute aortic syndromes include aortic dissection, intramural haematoma (IMH), and penetrating aortic ulcers. Aortic dissection usually begins with a tear in the intima. Blood penetrates the diseased medial layer and then cleaves the intimal laminal plain, leading to dissection. IMH is considered a precursor of dissection, in which there is rupture of the vasa vasorum in the aortic media with aortic wall infarction. IMH is typically in the descending thoracic aorta. Deep penetrating aortic plaques may lead to IMH, dissection, or ulceration/perforation. There is a predisposition to aortic dissection in patients with autoimmune rheumatic disorders and Marfan's and Ehlers–Danlos syndromes.

Aortic dissection can be classified according to the timing of diagnosis from the origin of symptoms: acute, less than 2 weeks; subacute, 2 to 8 weeks; and chronic, more than 8 weeks, with mortality and extension decreasing with time. It can also be classified anatomically:

- ***Type A*** involves the aortic arch and aortic valve proximal to the left subclavian artery origin. This category includes De Bakey type I (extends to the abdominal aorta) and De Bakey type II (localized to the ascending aorta).

- ***Type B*** involves the descending thoracic aorta distal to the left subclavian artery origin. This category includes De Bakey type III (Fig. 17.118).

Clinical features

Symptoms

Most patients present with a sudden onset of severe and central ***chest pain*** that often radiates to the back and down the arms, mimicking MI. The pain is often described as tearing in nature and may be migratory.

Signs

Patients may be shocked and may have neurological symptoms secondary to loss of blood supply to the spinal cord. They may develop aortic regurgitation, coronary ischaemia, and cardiac tamponade. Distal extension may produce acute kidney failure, acute lower limb ischaemia, or visceral ischaemia. Peripheral pulses may be absent.

Investigations

The mediastinum may be widened on chest X-ray; urgent CT scan, TOE, or MRI will confirm the diagnosis (see Fig. 17.98).

Management

At least 50% of patients are hypertensive and may require urgent antihypertensive medication to reduce blood pressure to below 120 mmHg; intravenous beta-blockers (labetalol, metoprolol) and vasodilators (GTN) are used. Type A dissections should undergo surgery (arch replacement) if the patient is fit enough, as medical management carries a high mortality (50% within 2 weeks). Type B dissections carry a better prognosis and have a survival rate of 89% at 1 month; initially, these patients should be managed medically unless they develop complications. Endovascular intervention with stents may be indicated in individuals with rapidly expanding dissections (>1 cm/year), critical diameter (>5.5 cm), refractory pain or malperfusion syndrome, blunt chest trauma, penetrating aortic ulcers, or IMH. Patients will require long-term follow-up with CT or MRI.

Other types of peripheral arterial disease

Raynaud's phenomenon or Raynaud's disease

Raynaud's phenomenon consists of spasm of the digital arteries, usually precipitated by cold and relieved by heat. If there is no underlying cause, it is known as Raynaud's disease. This affects 5% of the population, mostly women. The disorder is usually bilateral and fingers are affected more commonly than toes.

Clinical features

Vasoconstriction causes skin ***pallor*** followed by ***cyanosis*** due to sluggish blood flow, then redness secondary to hyperaemia. The duration of the attacks is variable but they can sometimes last for hours. Numbness, a burning sensation, and severe pain occur as the fingers warm up. In chronic, severe disease tissue, ***infarction*** and ***digital loss*** can occur.

Diagnosis

Primary Raynaud's disease needs to be differentiated from secondary treatable causes leading to Raynaud's phenomenon. These are the rheumatic autoimmune disorders, such as systemic sclerosis. It can be associated with atherosclerosis or occupations that involve the use of vibrating tools. Ergot-containing drugs and beta-blockers, as well as smoking, can aggravate symptoms.

Management

Patients should avoid cold provocation by wearing gloves and warm clothes and stop smoking. Vasodilators can be prescribed but are often unacceptable, as cerebral vasodilation causes severe headaches. Sympathectomy or prostacyclin infusion can be helpful in severe disease.

Takayasu's disease

This is rare, except in Japan. It is known as the pulseless disease or aortic arch syndrome. It is of unknown aetiology and occurs in females. There is a vasculitis involving the aortic arch, as well as other major arteries. A systemic illness is also present, with pain and tenderness over the affected arteries. Absent peripheral pulses and hypertension are common. Corticosteroids help the constitutional symptoms. Eventually, heart failure and strokes may occur, but most patients survive for at least 5 years. Treatment may involve a surgical bypass to improve perfusion of the affected areas.

Thromboangiitis obliterans (Buerger's disease)

This disease, involving the small vessels of the lower limbs, occurs in young men who smoke. Although sometimes clinically indistinguishable from atheromatous disease, pathologically there is inflammation of the arteries and sometimes veins, which may indicate a separate disease entity. Clinically, it presents with severe claudication and rest pain, leading to gangrene. A thrombophlebitis is sometimes present. Treatment is as for all peripheral vascular disease but patients must stop smoking.

Cardiovascular syphilis

This gives rise to:

- uncomplicated aortitis
- aortic aneurysms, usually in the ascending part
- aortic valvulitis with regurgitation
- stenosis of the coronary ostia.

The diagnosis is confirmed by serology. Treatment is with penicillin. Aneurysms and valvular disease are treated as necessary by the usual methods.

PERIPHERAL VENOUS DISEASE

Venous thromboembolic disease is covered in Chapter 40.

Varicose veins

Varicose veins are a common problem and sometimes give rise to pain. They may be treated by surgery, injection with ultrasound-guided foam sclerotherapy, or thermal ablation with endovenous lasers.

Superficial thrombophlebitis

This commonly affects the saphenous vein and is often associated with varicosities. Occasionally, the axillary vein is involved, usually as a result of trauma. There is local superficial inflammation in the vein wall, with secondary thrombosis.

The clinical picture is of a painful, tender, cord-like structure with associated redness and swelling.

The condition usually responds to symptomatic treatment with rest, elevation of the limb, and analgesics (e.g., NSAIDs). The Comparison of Arixtra in Lower Limb Superficial Vein Thrombosis with Placebo (CALISTO) trial demonstrated that a 45-day course of subcutaneous fondaparinux (2.5 mg daily), compared with placebo, significantly decreased the rate of thromboembolic events (pulmonary embolism and DVT) from 1.3% to 0.2%, and limited the extension of superficial vein thrombosis to the saphenofemoral junction from 3.4% to 0.3% with no increased risk of bleeding.

Significant websites

http://www.achd-library.com *Nevil Thomas Adult Congenital Heart Library*.
http://www.americanheart.org *American Heart Association*.
http://www.dh.gov.uk *UK National Service Framework: Coronary Heart Disease (2000)*.
http://www.ecglibrary.com/ecghome.html *ECG tracings library*.
http://www.erc.edu *European Resuscitation Council*.
http://www.resus.org.uk *Resuscitation Council (UK)*.

Questions for this chapter can be found on Elsevier eBooks+.

18 Hypertension

Vikas Kapil

Background 496
Aetiology and pathophysiology 497
Primary hypertension 497
Secondary hypertension 498
Hypertension in pregnancy 498
Clinical approach to the patient with hypertension 498
History 498
Examination and investigations 500
Measurement of blood pressure 500
Phenotypes of hypertension 502
Hypertensive target organ damage 503
Thresholds and targets in hypertension 503
Treatment 505
Population interventions 505
Lifestyle changes 505
Drug treatment 505
Adherence to treatment 506
Choice of therapy 506
Referral to specialist care 506
Renal sympathetic denervation 507
Management of cardiovascular risk 508
Managing blood pressure in hospital 508

CORE SKILLS AND KNOWLEDGE

Hypertension is a major cardiovascular (CV) risk factor affecting one-third of the adult population. It should be managed as part of a comprehensive CV risk reduction strategy while also assessing for hypertensive end-organ damage and ruling out secondary causes. Treatment involves modifying lifestyle to improve diet and body weight and making use of a range of antihypertensive medications.

Most hypertension is managed in primary care. However, young patients with hypertension, or patients who have uncontrolled hypertension despite taking three or more antihypertensives, should be referred for expert evaluation. Doctors may subspecialize in the management of hypertension after completing training in clinical pharmacology, cardiology, renal medicine, or another medical specialty.

Key skills in hypertension include:

- Diagnosing appropriately, and placing this in the context of an individual's overall CV risk
- Screening for secondary causes of hypertension and seeking evidence of end-organ damage
- Understanding the evidence-based guidelines for treating hypertension and becoming familiar with the main classes of drug used.

Gaining experience in the practical management of hypertension is best done during primary care attachments, for instance, by observing patients undergoing annual blood pressure medication reviews with their general practitioner. Attending a specialist hypertension clinic in a secondary or tertiary care setting gives an interesting insight into secondary causes and more severe phenotypes of hypertension.

BACKGROUND

Elevated blood pressure (BP) is a leading risk factor for death worldwide, outstripping tobacco smoking, poor sanitation and infectious diseases. It is a leading risk factor for atrial fibrillation, stroke, myocardial infarction and end-stage kidney disease. It is such a powerful predictor of these CV events that modern drug therapies are licensed for CV protection simply based on evidence of sustained BP reduction.

Hypertension refers to the ***persistent elevation of BP in the systemic arterial circulation***. It is the most common chronic health condition for which patients receive medication from their primary care doctor and is associated with 1 in 10 of all primary care appointments. It affects one-third of adults and this number continues to rise in all populations. Annual drug costs related to hypertension are well in excess of £1 billion in the UK.

Hypertension is diagnosed against threshold levels of BP from the evidence gathered from large multicentre, multinational studies. However, given the continuous relationship between BP and CV disease (**Fig. 18.1**) and renal morbidity, there is a trend towards basing treatment on even lower thresholds, and towards using total ***CV risk*** (see Chapter 17) to guide decisions on whether lowering BP is of benefit in individual patients.

The majority of patients with hypertension are managed in primary care, with specialist hospital services usually reserved for evaluation of secondary causes, or for those patients with problematic BP phenotypes (see 'Phenotypes of hypertension' section) or medication intolerances. Despite the wide availability of diagnostics (i.e., BP monitors) and proven treatments (i.e., medications), reduction of BP in hypertensive patients to target levels is less than 50% in most populations and remains the focus of renewed efforts on an international, national, and local scale.

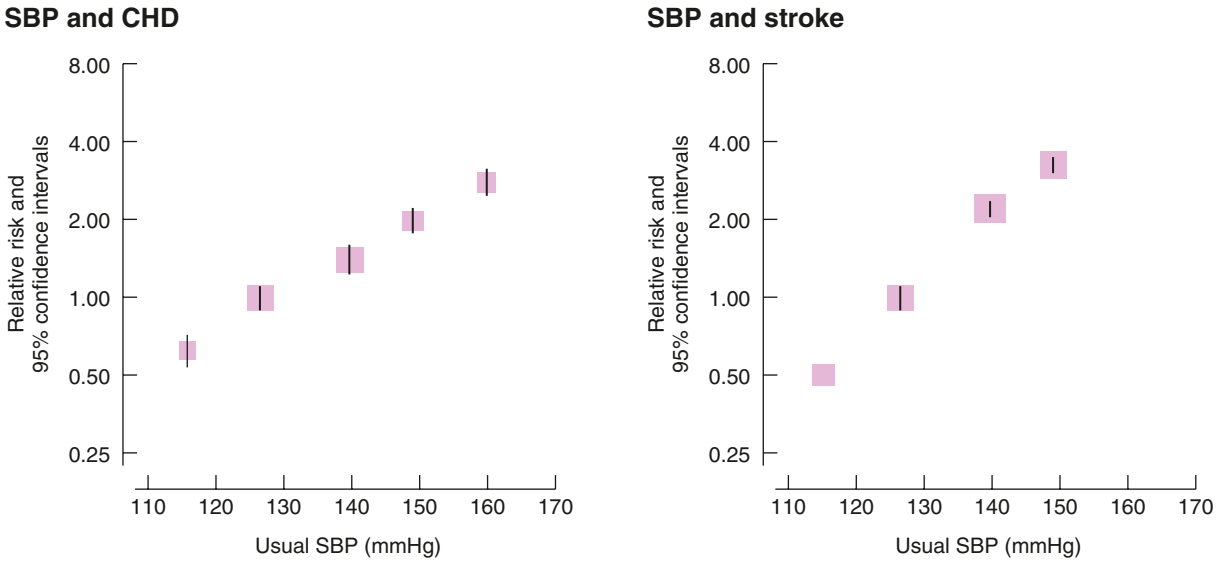


Fig. 18.1 The relationship between blood pressure (BP) and cardiovascular (CV) morbidity.
CHD, Coronary heart disease; *SBP,* systolic BP. (*Jackson et al.* Lancet *2005; 365: 434–441.*)

AETIOLOGY AND PATHOPHYSIOLOGY

Hypertension is a disease of ***ageing***. It is rare in children and adolescents, and a diagnosis in those under 30 years of age requires careful evaluation of possible underlying secondary causes (see 'Secondary hypertension' section). Due to age-related stiffening in large arteries (arteriosclerosis), BP tends to increase throughout middle and later life. Female sex hormones appear to protect against raised BP, as women have lower BP levels at all ages until menopause, when both prevalence of hypertension and achieved BP levels approximate to those of men. Hypertension is more common in certain ethnic groups (black African, African–Caribbean) and is also associated with urbanized and migrant populations, although an unequivocal explanation for these phenomena is not agreed.

Primary hypertension

Over 90% of patients have no singular identifiable cause of their elevated BP; this is called ***primary hypertension*** (now preferred to the historical term, 'essential hypertension', see later). Advances in genomic research have increased our understanding of polygenic effects on BP (see Chapter 5). We now know that several hundred genes are individually associated with prevalent BP levels but that each one contributes only small amounts (perhaps 0.5–1 mmHg). Combined, however, they may account for 60% of the BP level in any one person. As such, a family history of hypertension is common, though a family history of hypertension or strokes at unusually young ages should instigate careful evaluation for secondary causes. While there are a large number of distinct genetic influences on BP regulation, for the purposes of clinical assessment and management the main mechanisms affecting BP are ***vascular volume and tone***, and ***cardiac output*** (Fig. 18.2). The remaining effects on BP (up to 40%) are mediated by environmental (e.g., temperature, noise) and lifestyle stressors (e.g., diet).

Sudden development of severe hypertension or worsening of previously good control should alert healthcare professionals to the need for careful evaluation of ***adherence to treatment***, and for consideration of lifestyle and secondary causes of hypertension if adherence is confirmed. Adherence issues and causes of secondary hypertension should also be borne in mind in patients with ***resistant hypertension*** (usually defined as uncontrolled BP despite three separate, guideline-recommended antihypertensives).

Lifestyle-related issues

Diet

Relevant ***dietary*** information provided by patients is often inaccurate. Nevertheless, some clear dietary patterns are associated with raised BP:

- high salt intake
- low vegetable and fruit intake
- high saturated fat intake
- high simple carbohydrate intake
- excessive liquorice (inhibits an enzyme that normally prevents cortisol from activating the mineralocorticoid receptor).

Lack of exercise

Both CV and strength training forms of ***exercise*** are associated with lower BP values, along with other health benefits.

Drugs

Many patients with hypertension will have other co-morbidities that require additional pharmacological treatment, and consideration should therefore be given to the possible interference of these other drugs with BP control. Both over-the-counter medicines, such as non-steroidal anti-inflammatory drugs, and recreational drugs, such as alcohol and sympathomimetic stimulants, should be taken into account, as patients may not mention them when a history is being taken unless specific enquiry is made.

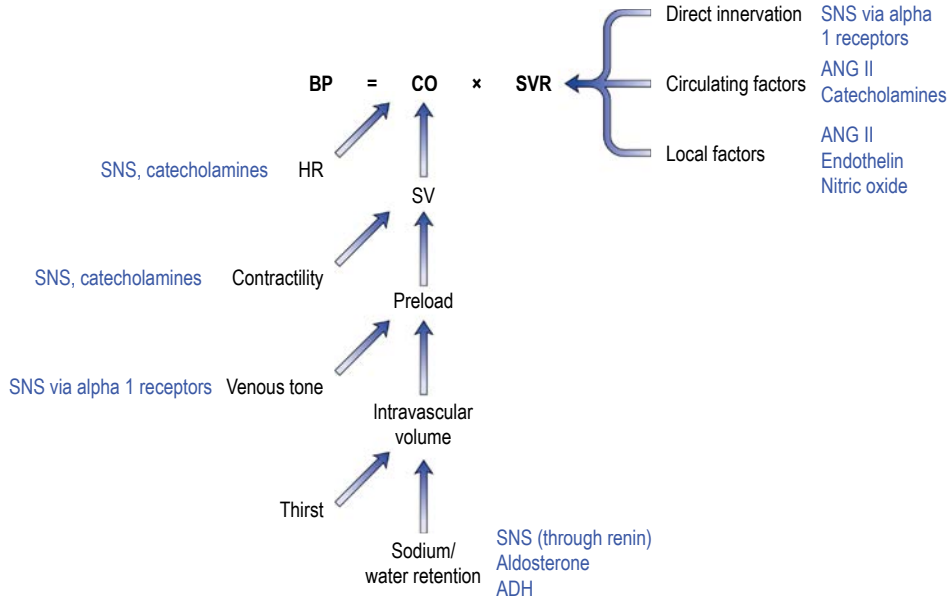


Fig. 18.2 The main mechanisms affecting BP. *ANGII,* Angiotensin II; *CO,* cardiac output; *HR,* heart rate; *SNS,* sympathetic nervous system; *SVR,* systemic vascular resistance.

Commonly implicated drugs that raise BP include:

- alcohol
- stimulant, recreational drugs
- oral contraceptive pill
- non-steroidal anti-inflammatory drugs
- corticosteroids
- calcineurin inhibitors
- vascular endothelial growth factor inhibitors
- some antidepressants (e.g., venlafaxine)
- mirabegron.

Secondary hypertension

Hypertension can have an identifiable singular cause, removal or reversal of which leads to normalization of BP. These are ***secondary causes*** of hypertension, and are broadly categorized into renal, vascular, endocrine, and neural problems (**Box 18.1**). Together, these are present in <10% of all hypertensive patients (though this proportion is higher in those under 30 years of age). The most common causes of secondary hypertension are largely thought to be ***primary hyperaldosteronism*** (increased salt/water retention), ***obstructive sleep apnoea*** and ***obesity*** (the last two causing sympathetic overdrive), though the precise proportions ascribed to these and other causes vary in different studies due to advances in evaluation and diagnostics.

Hypertension in pregnancy

Although it is not usually considered a secondary cause, ***pregnancy*** can be associated with elevated BP. The normal physiological response to pregnancy is an increase in circulating volume that is offset by significant vasodilation, leading to lower BP, particularly in the second trimester. However, in some pregnant women, defects in placental formation or vascularity lead to pregnancy-induced hypertension and/or ***pre-eclampsia*** (see Chapter 48). Hypertension associated with pregnancy often improves within a few months of the infant's delivery but women with hypertensive disorders of pregnancy are more likely to develop hypertension earlier in life than age-matched controls. The history should establish whether women of childbearing age are actively trying to become pregnant, as many of the first-line drugs are not recommended in this situation (see **Box 18.6**).

CLINICAL APPROACH TO THE PATIENT WITH HYPERTENSION

Hypertension is largely asymptomatic, though it is associated with increased prevalence of headaches, epistaxis and, less commonly, other neurological symptoms such as visual disturbance and dizziness. Furthermore, in the absence of a hypertensive emergency syndrome (see 'Hypertensive emergencies' section), symptoms related to raised BP are not an indication for treatment. Indeed, antihypertensive therapy is more likely to be the cause of adverse symptoms, such as postural intolerance. History and examination should seek out identifiable causes of elevated BP and asymptomatic organ damage.

There are four key considerations in the clinical assessment of patients with hypertension (**Fig. 18.3**) that can be determined through history, examination and investigations:

- What is the true BP level?
- Is there an identifiable cause for high BP?
- Does the BP level require lowering?
- Are there compelling reasons to use certain therapeutic approaches above others?

History

This should cover:

- symptoms of hypertension or secondary causes
- lifestyle issues: weight, alcohol excess, stimulant recreational drug use, tobacco smoking, exercise, stressors (work and personal)
- family history
- pregnancy history including episodes of pre-eclampsia
- adherence to antihypertensives.

Box 18.1 Drivers/secondary causes of hypertension

System	Causes	History	Examination	Investigations	Management
Lifestyle	Diet	Dietary history		24 h urinary sodium excretion (as a guide and to track improvement in salt intake)	Dietary advice
	Lack of exercise	<30 min/day moderate-intensity exercise			Exercise advice
Drugs	Alcohol	Alcohol history, features of alcohol dependence	May have signs of chronic liver disease	Only if indicated for underlying liver disease	Alcohol reduction <14 units/week
	BP-increasing medications	Medication history, including supplements, recreational and over-the-counter		Use of reference sources to identify BP-increasing medications	Cessation, switching where possible
	Poor adherence to antihypertensives	Validated adherence questionnaires Open, non-judgemental attitude		Urine or plasma qualitative drug monitoring	Shared decision-making, single-pill combinations
Metabolic	Obesity	Recent weight gain	Body mass index >30		Weight loss advice, bariatric surgery
Vascular	Coarctation of aorta		Unequal arm–arm or arm–leg pulses or BP Cardiac murmur or abdominal bruit	MR or CT angiography of whole aorta	Surgery, angioplasty or conservative
Renal	Renal artery stenosis	Sudden deterioration in renal function with ACE inhibitor/ARB Flash pulmonary oedema with normal cardiac function	Abdominal bruit	MR or CT renal angiogram	Balloon angioplasty ± stenting
	Chronic kidney disease	Often asymptomatic Childhood UTIs Features of vasculitis/GN	Often no physical signs Ballotable renal masses (PKD) Vasculitic rash	Urine dipstick and ACR eGFR Renal ultrasound Immunological tests (if GN likely) Renal biopsy (if GN likely)	Treatment of BP Treatment of underlying inflammatory renal disease
Endocrine	Primary hyperaldosteronism (Conn's syndrome)	Weakness, fatigue		Low K, raised aldosterone: renin MRI/CT adrenal ± adrenal vein sampling	Laparoscopic adrenalectomy if unilateral secretion
	Hypercortisolaemia (Cushing's syndrome)	Weight gain, fatigue	Centripetal weight distribution, hirsutism, pigmented striae	Low-dose dexamethasone suppression (cortisol should normally be low)	Laparoscopic adrenalectomy or pituitary surgery, depending on source
	Phaeochromocytoma	Palpitations, sweating, BP surges, dizziness, pallor, anxiety	Associated cutaneous neurofibromata (in neurofibromatosis-1)	Plasma metanephrines MIBG scan/MRI chest/abdomen/pelvis to localize	Surgery after sufficient alpha blockade
	Acromegaly	Sweating, skin tags, growth of soft tissues, visual disturbance	Large hands, feet, interdental separation, bitemporal hemianopia	Serum IGF-1	Pituitary surgery
	Thyroid (both hypo- and hyperthyroidism)	Weight gain, tiredness (hypothyroidism) Irritability, weight loss (hyperthyroidism)	Pretibial myxoedema, slow relaxing reflex (hypothyroidism) exophthalmos, acropachy (hyperthyroidism)	Serum TSH	Thyroxine replacement (hypothyroidism) or thyroid blockade/surgery (hyperthyroidism)

Box 18.1 Drivers/secondary causes of hypertension—cont'd

System	Causes	History	Examination	Investigations	Management
Respiratory	Obstructive sleep apnoea (Chapter 20)	Daytime somnolence (Epworth score), snoring, nocturnal sweating, morning headaches	Large neck, Mallampati score 3 or 4 (see Chapter 14)	Overnight oximetry (screening) Sleep study	Weight loss Mandibular advancement device CPAP
Neurological	Autonomic failure	Postural intolerance, thermoregulatory issues, parkinsonism	Variable BP Supine hypertension with normal or low BP sitting, standing	24 h BP: nocturnal hypertension Autonomic function testing	Symptom-based, nocturnal medication if supine hypertension
Obstetric	Pregnancy	Secondary amenorrhea Visual disturbance, oedema, headaches, seizures (for PET/eclampsia only)	Fundal palpation	Urine or serum β-hCG	Pregnancy-safe medicines Delivery of fetus

ACE, Angiotensin-converting enzyme; *ACR,* albumin creatinine ratio; *ARB,* angiotensin II receptor blocker; *BP,* blood pressure; *β-hCG,* beta-human chorionic gonadotrophin; *CPAP,* continuous positive airways pressure; *CT,* computed tomography; *eGFR,* estimated glomerular filtration rate; *GN,* glomerulonephritis; *IGF-1,* insulin-like growth factor 1; *MIBG,* meta-iodobenzylguanidine; *MRI,* magnetic resonance imaging; *PET,* positron emission tomography; *PKD,* polycystic kidney disease; *TSH,* thyroid stimulating hormone; *UTI,* urinary tract infection.

Examination and investigations

These should include:

- 'out-of-office' BP level
- asymptomatic organ damage: eyes, kidneys, heart
- exclusion of secondary causes where relevant
- estimation of total CV risk.

Measurement of blood pressure

Due to the non-continuous, pulsatile blood flow maintained by the cardiac cycle, BP is represented by two numbers: systolic (peak arterial pressure during cardiac contraction) and diastolic (lowest arterial pressure during cardiac relaxation). It is presented as ***systolic pressure/diastolic pressure in mmHg*** (millimetres of mercury). BP is dynamic and has patterns of variability over seconds (related to breathing, sympathetic activation), minutes (exertion), hours (wake–sleep, circadian hormonal patterns), and the longer term (seasonal). Therefore, the measurement of BP and its use to determine the treatment of hypertension are inherently imprecise and a single measure may not reliably represent the usual (real) BP and its contribution to overall CV risk. As with any biological variable, the true usual BP is better approximated with multiple measurements at each assessment, and sometimes over several different assessment episodes, before decisions are made regarding changes in treatment.

BP can be measured in several different locations on the body, though readings are traditionally taken in the non-dominant arm. Initially, BP should be checked in both arms; if there is a significant difference (usually considered to be >10 mmHg), then the arm with the higher BP is used for subsequent measurements.

It is important to measure BP correctly. Methodological variation can falsely and enormously over- or underestimate BP. For static measurements (office/surgery, home), BP should be measured in the seated position after 5 minutes' uninterrupted rest, with the back supported and the legs uncrossed. The BP bladder should encircle at least 80% of the upper arm circumference and should be placed at the level of the heart; the arm should be supported (**Fig. 18.4**). More than one reading should be taken (in a row with minimal waiting between) during a measurement episode and the average of all the readings used to determine BP. If there is a significant reduction (>10 mmHg) between subsequent readings, further readings should be taken until the difference is <10 mmHg and the initial higher readings should be discarded before averaging at least the last two readings where there is <10 mmHg variation.

Although the unit of BP is still 'millimetres of mercury' (mmHg), mercuric sphygmomanometers are obsolete. Modern monitors are usually electronic, oscillometric devices that are automated and have been validated against mercuric devices or intra-arterial BP catheters.

One situation where auscultatory (i.e., using a cuff and stethoscope to listen over the brachial artery) methods are still preferred is in ***atrial fibrillation*** (see Chapter 17), where variations in pulse volume and rapid variation in BP due to variable left ventricular filling time lead to inaccuracies when oscillometric BP machines are used, especially when ventricular rates are uncontrolled. In particular, diastolic BP tends to be poorly estimated in this situation.

Office/clinic measurement

This refers to measurement of the BP by a healthcare professional in a healthcare setting, such as a primary care clinic or hospital outpatient department. This is the most established method of measurement and all clinical trials of antihypertensive drugs have used it to assess for inclusion and for titration towards target.

Issues with this approach include:

- a requirement for healthcare professionals and setting (expense, time)
- a requirement for multiple visits to obtain sufficient data for decision-making (expense, time)
- a 'white coat effect': for most people, BP measured by healthcare professionals is approximately 5/5 mmHg higher than at home, though there is inter-individual variation in the magnitude

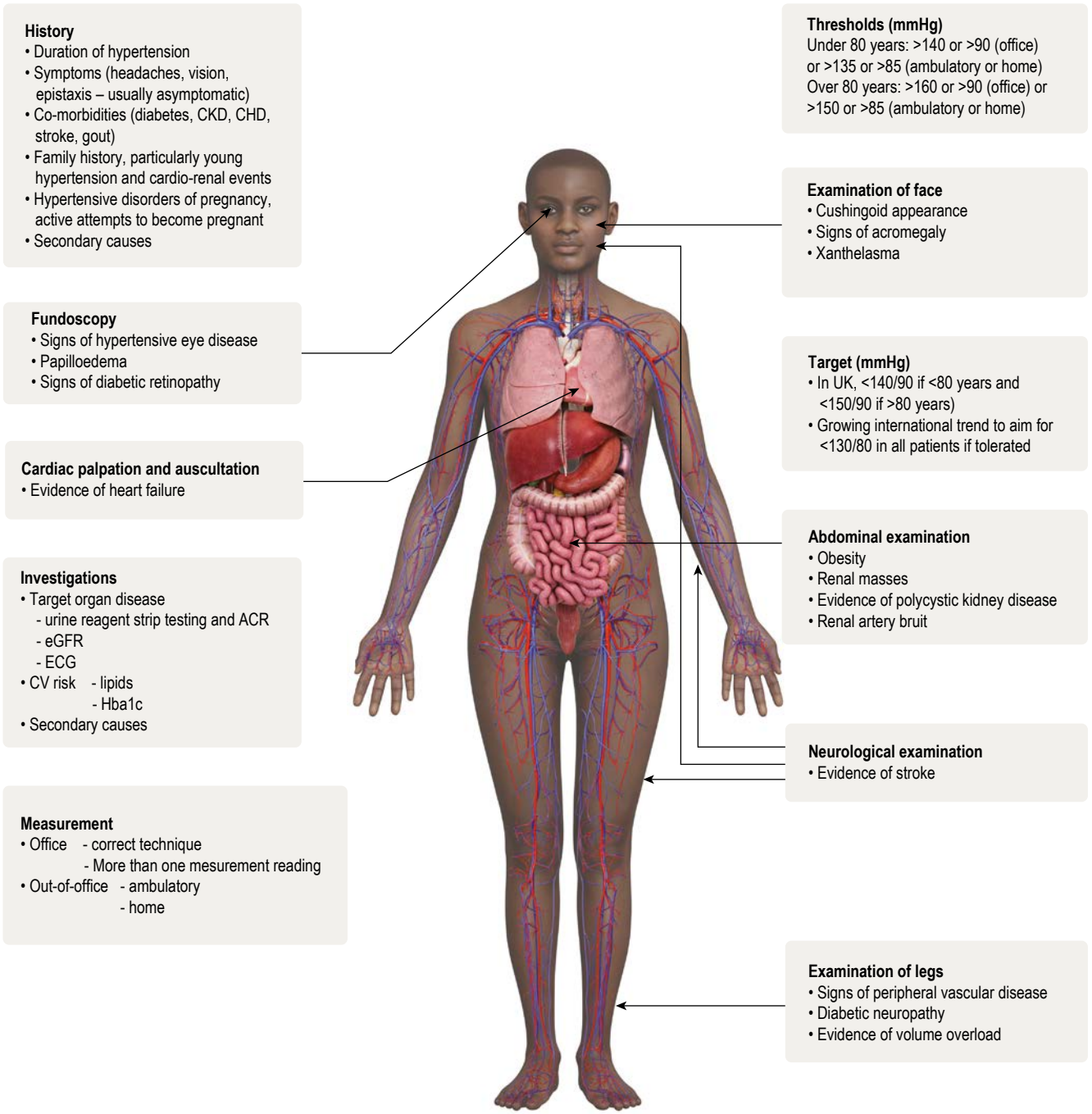


Fig. 18.3 Clinical assessment in hypertension. *ACR,* Albumin creatinine ratio; *CAD,* coronary artery disease; *CKD,* chronic kidney disease; *CV,* cardiovascular; *eGFR,* estimated glomerular filtration rate.

- home and ambulatory BP measurements are more predictive of CV events, as patients do not live in healthcare settings.

Although office BP is no longer recommended as the sole measure for diagnosis in many guidelines, it is still advocated as adjunctive for on-going monitoring and medication titration, as no CV outcome trials have used out-of-office readings to adjust treatment.

Home measurement

This is measurement of BP at home by the patients themselves, using the same type of machine as in the clinic. It is a type of 'out-of-office' BP measurement and is closely approximated to daytime ambulatory BP. It is particularly useful for:

- identification of different BP phenotypes
- engagement of patients in healthcare and shared decision-making
- long-term monitoring of BP between healthcare visits.

Current recommendations suggest that BP is measured on two or more occasions in sequence, in the morning prior to medicines and again in the evening for 4 to 7 days. The initial day's readings are discarded and the rest of the values are averaged to calculate home BP. This BP is highly predictive of CV events.

Issues with this approach include:

- a requirement for patients to obtain or be supplied with their own machine (expense)
- a requirement for training
- the lack of objective readings (i.e., the measurements are often patient-reported averages, rather than independent evaluations).

Ambulatory measurement

Ambulatory BP monitors are portable, oscillometric devices that measure BP discontinuously throughout a 24-hour period, most commonly every 20 to 30 minutes during waking hours and every 30 to 60 minutes during sleep.

The large number of readings that this provides gives an accurate estimate of true BP. The 24-hour mean of the sequence of daytime and nocturnal BP measurements is more predictive of CV events than both home and office BP measurements. Ambulatory BP monitoring is now recommended in the UK and some other high-income health economies instead of office BP for diagnosis, and is mostly recommended as a complementary strategy to office BP elsewhere.

Similar to home BP, ambulatory BP allows in a single measurement:

- improved diagnostic accuracy over office BP
- identification of different BP phenotypes.

Daytime ambulatory BP is 5 to 10/5 mmHg lower than office BP in the same patient. Issues with this approach include dislike of repeated BP measurement, especially at night.

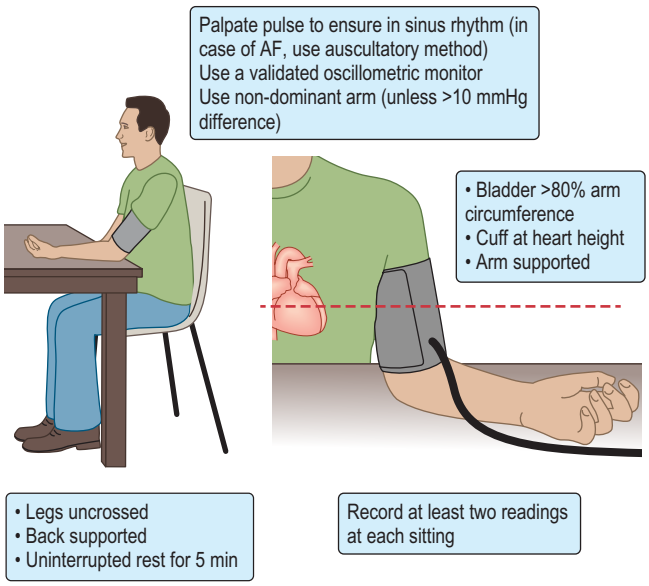


Fig. 18.4 The correct way to measure BP. *AF,* Atrial fibrillation.

Phenotypes of hypertension

The measurement techniques described can describe different phenotypes of hypertension (Box 18.2). Although both ***white coat hypertension*** and ***masked hypertension*** represent intermediate CV risk phenotypes, masked hypertension is treated using out-of-office values to guide treatment, while antihypertensives are not recommended for use in white coat hypertension, despite elevated CV risk, due to a current lack of evidence.

There are other common phenotypes of hypertension:

Isolated systolic hypertension

Due to age-related arterial stiffening, systolic BP continues to rise in patients above 50 years of age, with a corresponding reduction in diastolic BP. This widening ***pulse pressure*** (the difference between systolic and diastolic BP, usually <50 mmHg) is associated with increased vascular damage. Drug treatment in isolated systolic hypertension is the same as for combined hypertension (where both systolic and diastolic BP are elevated), though care is needed not to reduce diastolic BP below 60 mmHg, which theoretically could cause problems with coronary blood flow (which is dependent on diastolic pressure). Aortic valve incompetence can also cause an isolated systolic phenotype, though this is normally apparent on cardiac auscultation and/or echocardiography.

Orthostatic hypotension

This refers to a sustained fall in BP within 3 minutes of assuming an upright position of either more than 20 mmHg systolic or more than 10 mmHg diastolic BP. It is more common in older age (see Chapter 49), in conditions associated with ***autonomic neuropathies*** (such as diabetes mellitus and Parkinson's disease) and in those taking antihypertensive drugs. It is typically asymptomatic, although it can cause postural intolerance or instability, dizziness and falls. It should be actively screened for those with typical symptoms or risk factors. Most guidelines recommend using the standing BP as the true BP value for guiding treatment.

Variable blood pressure

Although BP is a highly dynamic variable, the degree of variability is exaggerated in some patients. This is often due to disorders of the autonomic nervous or endocrine system. Reliable assessment is achieved only through a combination of repeated short-term (ambulatory) and long-term (home) methodologies. Management of BP levels in such patients is often complex and best performed in specialist clinics.

? Box 18.2 Phenotypes of hypertension and corresponding BP measurements

Phenotype/method	Office BP	Out-of-office BP	Cardiovascular risk
Normal BP	Normal (<120/70)	Normal (<120/70)	Lowest
Elevated BP	Normal (<140/90)	Normal (<135/85)	Intermediate
White coat hypertension	High	Normal	Intermediate
Masked hypertension	Normal	High	Intermediate
Hypertension	High	High	Highest

BP, Blood pressure.

Hypertensive target organ damage

Hypertension can cause structural or functional changes in the heart, kidney and central nervous system that may be regarded as intermediate end-points with respect to reducing the absolute clinical end-points of myocardial infarction, heart failure, end-stage renal disease and strokes.

Eyes

Direct ophthalmoscopy (fundoscopy) of the dilated eye is recommended in most guidelines for evaluation of hypertensive vascular changes. The most commonly used grading classification is the Keith–Wagener–Barker system, comprising four distinct ophthalmic hypertensive phenotypes ***(retinopathy)***. More recent evaluation, however, has suggested combining the two milder phenotypes, leaving three grades to consider (**Fig. 18.5**):

- ***Mild:*** Generalized arteriolar narrowing, focal arteriolar narrowing, arteriovenous nicking (modest association with CV and cerebral events).
- ***Moderate:*** Haemorrhage (blot, dot or flame-shaped), microaneurysm, cotton-wool spot, hard exudate, or a combination of these signs (strong association with CV and cerebral events).
- ***Severe:*** Signs of moderate retinopathy plus papilloedema (swelling of optic head) (strong association with CV events, stroke and death).

Heart

The heart is a muscle and hypertrophy may occur in response to increased workload. Left ventricular remodelling is initially a compensatory phenomenon to reduce wall stress but eventually leads to a pathogenic increase in left ventricular mass.

Hypertensive left ventricular hypertrophy is usually asymptomatic but can be detected as a thrusting apex beat on precordial palpation. It is also sometimes detected on standard 12-lead electrocardiography (ECG), demonstrating various patterns of increased voltages in the chest leads and T-wave abnormalities (see Fig. 17.74). There are many different scoring systems for ECG criteria for diagnosing left ventricular hypertrophy; all are quite insensitive (20%–50%) but specificity is better (>90%), which means that if criteria are positive, true left ventricular hypertrophy is likely to be present.

Electrocardiograms are recommended in all hypertensive patients, as they are widely available and cheap, and do not require much in the way of additional training to interpret, especially as modern ECG machines have in-built reporting algorithms. However, cardiac imaging, using transthoracic echocardiography and cardiac magnetic resonance imaging, is more sensitive and specific for left ventricular hypertrophy than ECG (when compared to autopsy data), though access, cost and need for additional trained staff to conduct and report these tests limit their wide utility and acceptance in all hypertensive patients. These cardiac imaging modalities can estimate the actual mass of the left ventricle; this is indexed to height and weight, and has sex-specific normal values.

Kidney

Kidney disease is both a cause and a consequence of hypertension. Early hypertensive kidney damage may be detected through an increase in microalbuminuria on urine test strips or as an increased laboratory albumin : creatinine ratio, although in pure hypertensive nephropathy there may be little or only modest albuminuria. Estimates of glomerular filtration rate (GFR), based on serum creatinine and demographic criteria, are also obtained, especially as several antihypertensive medications are potentially nephrotoxic. However, there is usually an initial reduction in GFR associated with chronic BP-lowering from any cause, and this usually reflects intra-renal haemodynamic changes rather than intrinsic renal damage; it is therefore common practice to allow GFR to reduce by up to 10% on initiation of antihypertensive therapy. More regular follow-up is mandated to ensure that this does not indicate progressive renal decline.

Thresholds and targets in hypertension

Primary hypertension used to be termed ‘essential’ as, in the 19th century, raised BP was largely thought to be an appropriate physiological adaptation. However, with the advent of randomized clinical trials and modern therapeutics, there is now substantial evidence that lowering BP is beneficial, reducing both cardiorenal events and overall mortality.

The first Veterans Administration Cooperation study, commenced in 1964, randomized patients with a diastolic BP of 115 to 129 mmHg to active or placebo treatments. By modern standards, this would be considered to be severe hypertension. There were substantial differences between active and placebo-treated patients in a range of CV end-points, including mortality, and the trial was ended early because the degree of treatment benefit made it unethical to continue to give placebo.

Treatment threshold

Since around 1990, the threshold for diagnosis of hypertension has mostly been an office BP of more than 140 mmHg systolic or more than 90 mmHg diastolic (whichever is *worse*). Treatment has largely been reserved for those with ***grade 2 hypertension*** (**Box 18.3**), as there is a large evidence base for lowering BP with medications in these patients. For ***grade 1 hypertension***, it has been usual practice to persevere with lifestyle changes to lower BP and to recommend immediate treatment for those with an overall elevated CV risk (>1% per year, using validated CV risk equations), or evidence of organ injury due to hypertension (hypertensive target organ damage). Drug therapy is advocated, however, if BP fails to normalize after a 3-month period of best-tolerated lifestyle changes.

More recently, treating BP even in those not hypertensive but with elevated BP and at high CV risk has been proposed. For those at highest risk, either with existing CV disease or from coexisting risk factors, treatment initiation at lowers levels of BP than required for a hypertension diagnosis is associated with CV risk reduction.

Treatment target

For the most part, then, the target has been an office BP of less than 140/90 mmHg, with stricter targets for patients with diabetes and proteinuric chronic renal disease (<130/80 mmHg). However, recent trials such as SPRINT and meta-analysis of other trials have suggested that there is incremental benefit in targeting all patients to a BP 120–130/70–80 mmHg; this is reflected in most international guidelines, although not currently in those produced by the National Institute for Health and Care Excellence (NICE) in the UK.

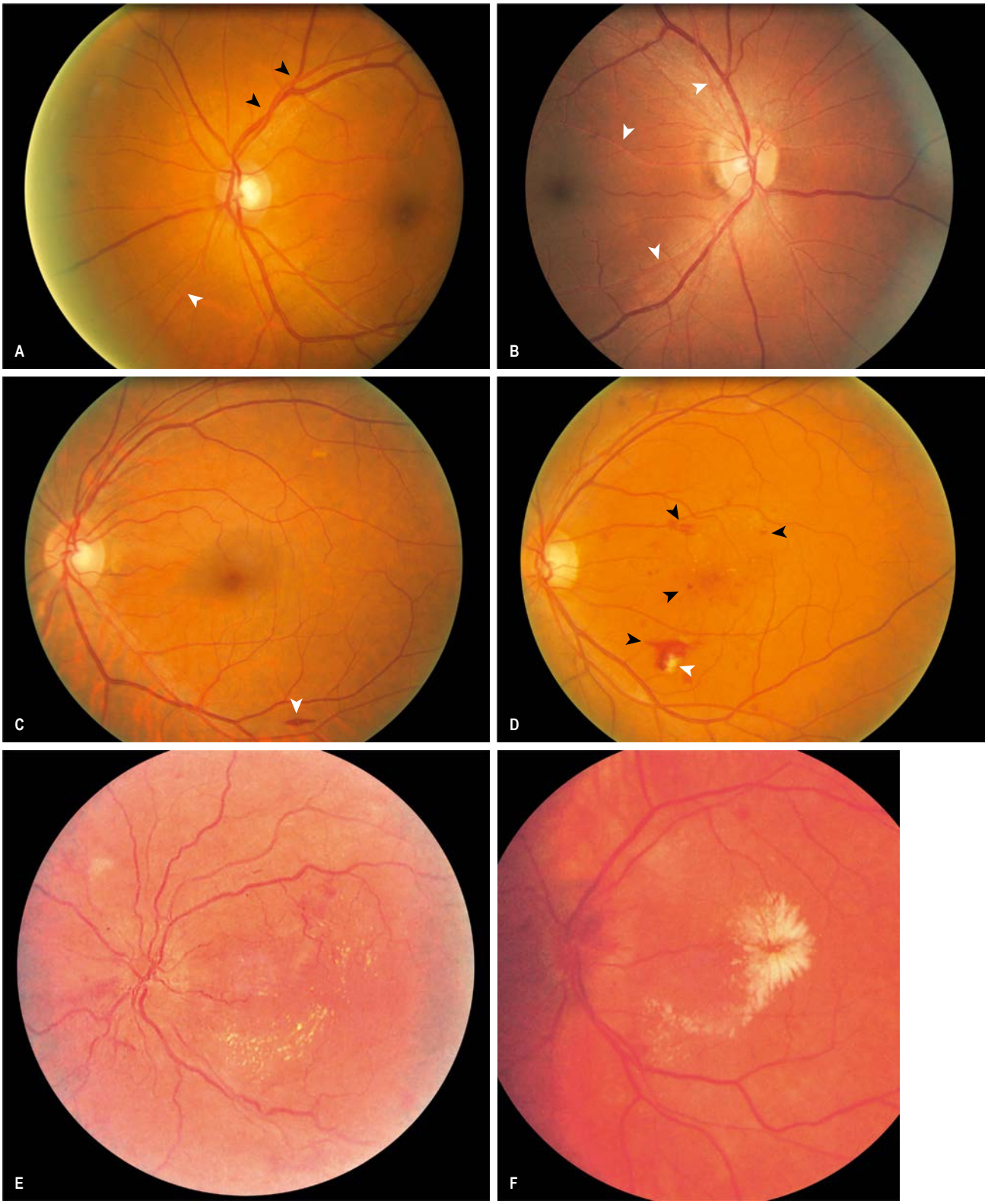


Fig. 18.5 Hypertensive retinopathy. **(A and B)** Examples of ***mild hypertensive retinopathy***. **(A)** Arteriovenous nicking *(black arrows)* and focal narrowing *(white arrow)*. **(B)** Opacification (silver or copper wiring) of arteriolar wall *(white arrows)*. **(C and D)** Examples of ***moderate hypertensive retinopathy***. **(C)** A flame-shaped retinal haemorrhage *(white arrow)*. **(D)** A cotton-wool spot *(white arrow)*, retinal haemorrhages and microaneurysms *(black arrows)*. **(E and F)** ***Severe hypertensive retinopathy***. **(E)** Exudates and flame haemorrhages in grade 3 retinopathy. **(F)** Signs of malignant hypertension in grade 4 disease, with a swollen optic disc and macular exudate. (*A–D From Schachat AP, Sadda SVR, Hinton DR et al. Ryan's Retina, 6th edn. Elsevier Inc., 2018, Figs. 52.1 and 52.2; E–F From Innes JA, Dover AR, Fairhurst K. Macleod's Clinical Examination, 14th edn. Elsevier Ltd, 2018, Fig. 8.18C and D.*)

Box 18.3 Thresholds for diagnosis of hypertension and related phenotypes

Description	Office BP (mmHg)		Daytime ambulatory or home BP (mmHg)	
	Systolic BP	Diastolic BP	Systolic BP	Diastolic BP
Normal BP	**<120**	**<70**	**120**	**70**
Elevated BP	120–139	70–89	120–135	70–85
Grade 1 hypertension	140–159	90–99	135–149	85–94
Grade 2 hypertension	160–179	100–109	150–169	95–104
Severe hypertension	>180	>110	>170	>105
Isolated systolic hypertension	>140	<90	>135	<85

BP, Blood pressure.

In older people (over 80 years), there is less evidence for treating mild hypertension, and most guidelines recommend treating only moderate (i.e., grade 2) hypertension. Some guidelines (e.g., NICE) suggest a lax target office BP of below 150/90 mmHg in patients diagnosed over 80 years of age. However, age is often used incorrectly as a surrogate for frailty, and although there is good reason to lower BP cautiously and judiciously in the frail elderly or in those at increased risk of falls or postural hypotension, there is a trend in updated international guidelines, to reduce BP to similar levels as for younger patients in biologically fit elderly patients, if tolerated.

There are additional considerations relating to these values. As mentioned, there has been a consistent shift to using out-of-office methods, preferably ambulatory, for diagnosis. In the hypertensive range, home and ambulatory BP is generally 5 to 10/5 mmHg lower than measurements taken contemporaneously in the office. At lower levels of BP, such as the 120/70 mmHg treatment target, there is less difference between measurements from different measurement techniques.

Further reading

Blood Pressure Lowering Treatment Trialists' Collaboration. Pharmacological BP lowering for primary and secondary prevention of CV disease across different levels of BP: an individual participant-level data meta-analysis. *Lancet* 2021; 397: 1625–1636.

The SPRINT Research Group. A randomized trial of intensive versus standard blood-pressure control. *N Engl J Med* 2015; 373:2103–2116.

TREATMENT

Meta-analyses of large-scale randomized controlled trials have shown that a 10/5 mmHg reduction in BP is associated with a 15% reduction in all-cause mortality, 35% reduction in stroke, 40% reduction in heart failure and 20% reduction in myocardial infarction. Hypertension is treated to reduce these major CV and renal events. For this reason, treatment of hypertension without managing other modifiable CV risk factors (raised cholesterol, diabetes mellitus, tobacco smoking, obesity) is suboptimal. CV risk scores, such as QRISK3®, are useful for integrating all of these risk factors, to judge when there is benefit in treating hypertensive patients. Ten-year risk estimates are less useful than lifetime risk estimates in young patients, given the powerful role of age in all CV risk scores.

Population interventions

In view of the potential effects of lifestyle issues, together with those associated with obesity (see Chapter 37), population approaches are increasingly being taken to reduce BP, including public health education drives to:

- reduce salt intake such as regulatory approaches to reducing salt in processed foods
- reduce simple carbohydrate intake through 'sugar taxes'
- reduce alcohol intake through minimum unit pricing
- reduce body weight
- increase exercise.

Lifestyle changes

These are recommended for all hypertensive patients, irrespective of grade or duration of disease. The BP reductions accompanying such lifestyle changes are complementary and similar in magnitude (compared to half-standard monotherapy) to those achieved by antihypertensive medications (**Box 18.4**).

Bariatric surgery (see Chapter 37) is indicated for those with hypertension and a body mass index of more than 35 kg/m^2, but it is not commonly performed without the additional co-morbidity of type 2 diabetes. Recent data confirm the profound BP-lowering effect of such surgery, which may be due to additional gut hormonal changes postoperatively, in addition to expected weight loss.

Drug treatment

A single antihypertensive medication at standard dose reduces BP by about 9/5 mmHg in mild hypertension. Although patients with mild hypertension may achieve control with monotherapy, those with moderate or worse hypertension generally require several drugs in combination. Greater effects are obtained by combining medications from different classes, targeting different pathophysiological mechanisms. Three antihypertensives at half-standard dose reduce BP by 20/11 mmHg. Most ***adverse effects*** with antihypertensives, in keeping with most pharmaceuticals, are type A adverse reactions (see Chapter 7) and are thus dose-dependent. Keeping doses to a minimum but combining different classes minimizes the chances of adverse effects and maximizes the likelihood of effectively lowering BP to target levels.

Physician inertia in the face of elevated BP should be avoided—leaving BP higher than target BP for more than a few months is

Box 18.4 Impact of lifestyle changes on BP

Lifestyle change	Expected mean BP reduction
Regular cardiovascular exercise (30 min daily)	5 mmHg
Weight reduction if overweight (body mass index >25 kg/m^2)	1 mmHg/kg
Increased intake of fruits and vegetables, reduced intake of saturated fat	10 mmHg
Dietary salt reduction <6 g day	5 mmHg
Alcohol <2 units daily	3 mmHg

BP, Blood pressure.

associated with excess CV morbidity. Achieving control early with combination medications is preferred.

Adherence to treatment

Adherence is increasingly recognized as a key barrier to achieving BP control. Although there are different validated measures to assess this, such as the Morisky-8 Medication Adherence Scale (Box 18.5), along with prescription refill rate (whether patients request a resupply at the right time to have therapy every day), modern drug analytical techniques on plasma or urine have demonstrated that question-/refill-based assessment correlates poorly to objective evidence of measurable drug in biological matrices. From these newer techniques, it appears that up to two-thirds of patients in specialist care do not take some or any of their antihypertensive medications. There are no easy solutions to such covert non-adherence, but keeping an open, non-judgemental approach can facilitate the volunteering of this information and then shared-decision-making can determine how best to proceed and what patients want for their own health.

In view of this, and with evidence that suggests there is an elevated CV risk when there is failure to achieve timely BP control, guidelines have recently tended to abandon monotherapy and suggest starting with a combination of medicines to achieve BP controleffectively and rapidly, including initiating patients with single-pill combination therapies.

Choice of therapy

There are numerous classes of antihypertensive medication available and their key features are described in Box 18.6.

Most guidelines recommend initial therapy with two of the following three classes of drugs:

- angiotensin-converting enzyme inhibitors/angiotensin II receptor blockers (ARBs),
- calcium channel blockers
- thiazide-like diuretics.

Monotherapy should be initiated only in the mildest of stage I hypertension, as most patients require two or medications to get to guideline-based targets (Fig. 18.6).

Current UK NICE guidelines use age and ethnicity as surrogates for plasma renin activity. Hypertension in older patients and those of black African ancestries is associated with low renin status and so patients have reduced responsiveness to angiotensin-converting enzyme inhibitors and ARBs as monotherapy: hence calcium-channel blockers are recommended in these groups. Younger white patients are recommended to start

Box 18.5 Morisky medication adherence scale[a]

Please answer each question, based on your personal experience with your medications. Note that there is no right or wrong answer.

1. Do you sometimes forget to take your medications?	No	Yes
2. People sometimes miss taking their medications for reasons other than forgetting. Thinking over the past 2 weeks, were there any days when you did not take your medications?	No	Yes
3. Have you ever cut back or stopped taking your medications without telling your doctor because you felt worse when you took it?	No	Yes
4. When you travel or leave home, do you sometimes forget to bring along your medications?	No	Yes
5. Did you take your medications yesterday?	No	Yes
6. When you feel that your health condition is under control, do you sometimes stop taking your medications?	No	Yes
7. Taking medications every day is a real inconvenience for some people. Do you ever feel hassled about sticking to your treatment plan?	No	Yes
8. How often do you have difficulty remembering to take all your medications?	Never/rarely Once in a while Sometimes Usually All the time	1 0 0 0 0

[a]'No' answers to questions 1–4, 6 and 7, a 'Yes' answer to question 5 and 'Never/rarely' to question 8 score 1. A score of 8 suggests good adherence. Score of 7 or less suggests poor adherence and is associated with poorer BP control.
(Adapted from Morisky DE, Ang A, Krousel-Wood M et al. Predictive validity of a medication adherence measure in an outpatient setting. J Clin Hypertens *2008;10:348–354.)*

on an ACEI or ARB, which are largely thought to be interchangeable, except that ARBs are better tolerated, with less cough and angio-oedema.

There is clear evidence from the PATHWAY-2 study that spironolactone is the best fourth-line drug, in preference to beta- or alpha-blockers, though it is unlicensed for this indication in most countries and patients should be appropriately counselled.

Referral to specialist care

The majority of hypertensive patients are managed in primary care. Once a patient is on three or four drugs and remains uncontrolled ***(resistant hypertension)***, they should be referred for specialist assessment. Other reasons to mandate referral to specialist secondary care include:

- Suspected secondary causes:
  - Young age (<30–40 years)
  - Suggestive comorbidities (i.e., obstructive sleep apnoea, phaeochromocytoma)
  - Examination findings (i.e., abdominal bruit suggesting renal artery stenosis)
  - Sudden change in BP
- Target organ damage detected with normal BP values
- Intolerance of medicines preventing guideline-based treatment

Box 18.6 Main classes of antihypertensive agent and their key features

Class	Mechanism	Adverse effects	Compelling indication	Compelling contraindication	Monitoring	Example
Angiotensin-converting enzyme (ACE) inhibitor	Reduce angiotensin II vasoconstriction	Cough Angio-oedema Hyperkalaemia	CKD with proteinuria, diabetes, heart failure with reduced ejection fraction	Bilateral RAS Angio-oedema	eGFR, K	Ramipril 2.5–10 mg daily
Angiotensin II receptor blockers (ARB)	Reduce angiotensin II vasoconstriction	Hyperkalaemia	CKD with proteinuria, diabetes, heart failure with reduced ejection fraction	Bilateral RAS	eGFR, K	Losartan 25–100 mg daily
DHP CCB	Peripheral vasodilation	Constipation, headache, oedema	Nifedipine/amlodipine recommended for use in pregnancy			Amlodipine 5–10 mg daily
Non-DHP CCB	Peripheral vasodilation	Constipation, bradycardia		High grade atrioventricular block Bradycardia	Heart rate	Verapamil 120–480 mg daily
Thiazide-like diuretic	Vasodilation, salt/water loss	Hyponatraemia, hypokalaemia, hyperuricaemia, hyperglycaemia	Heart failure	Gout	Serum electrolytes, HbA_{1c}, urate	Indapamide 1.5–2.5 mg daily
Potassium-sparing diuretic	Salt/water loss	Hyponatraemia, hyperkalaemia, gynaecomastia	Primary hyperaldosteronism Resistant hypertension	eGFR <30 mL/min K >5.0 mmol/L	Serum electrolytes	Spironolactone 25–50 mg daily
Beta-blocker	Vasodilation through renin inhibition, sympatholysis	Bradycardia, sleep disturbance, bronchospasm	Heart failure with reduced ejection fraction Ischaemic heart disease Labetalol licensed in pregnancy	Asthma High-grade atrioventricular block Bradycardia	Heart rate	Bisoprolol 2.5–10 mg daily
Alpha-blocker	Venodilation	Postural hypotension, urine incontinence, oedema	Suspected phaeochromocytoma Benign prostatic hypertrophy	Urinary incontinence Postural hypotension Heart failure		Doxazosin 1–8 mg twice daily
Central	Sympatholysis	Depression, drowsiness, dry mouth	Methyldopa licensed in pregnancy	Mood disorders Heart failure with reduced ejection fraction	Rebound hypertension on cessation	Methyldopa 250–500 mg 3 times daily

ACE, Angiotensin-converting enzyme; *ARB*, angiotensin receptor blocker; *CKD*, chronic kidney disease; *DHP CCB*, dihydropyridine calcium-channel blocker; *eGFR*, estimated glomerular filtration rate; *K*, potassium; *RAS*, renin–angiotensin system.

- Symptomatic hypertension or hypotension
- Labile or highly variable BP
- Elevated BP when it is not clear whether a patient would benefit from BP-lowering
- Presentation with a hypertensive emergency.

Renal sympathetic denervation

Neural sympathetic signalling has a key role in vascular homeostasis, with effects mediated via afferent renal and aortic/carotid baroreceptor functions. Novel interventional approaches to alter this signalling are currently under evaluation to improve BP in both medicine-naive and treated hypertensive patients. The most advanced, in terms of regulatory pathways and evidence base, is renal sympathetic denervation. This group of procedures uses modern endovascular catheter-based approaches to the renal sympathetic nerves, which lie on the outside of the renal artery. Catheters can then deliver energy (radiofrequency, ultrasound) or neurotoxins (such as alcohol) across the renal artery wall to reach the sympathetic nerves and abrogate afferent signalling to the vasomotor centres, thus reducing total sympathetic drive and BP. Although the technological advancements are impressive and there is evidence of robust BP-lowering using modern out-of-office measurements, the magnitude of these effects is similar to that of antihypertensive monotherapy and thus most patients will still require drug therapy to control BP in the long term. Furthermore, these are new technologies: the long-term safety of the procedures is not clear, as regards the renal artery or otherwise, and there are no outcome studies investigating cardiac, renal, stroke or death end-points using any technology apart from drugs. Most guidelines recommend considering renal denervation an add-on option in expert centres only for patients uncontrolled BP despite maximally or best tolerated therapies once shared-decision-making process to understand the uncertainty of benefit or risk is discussed and agreed.

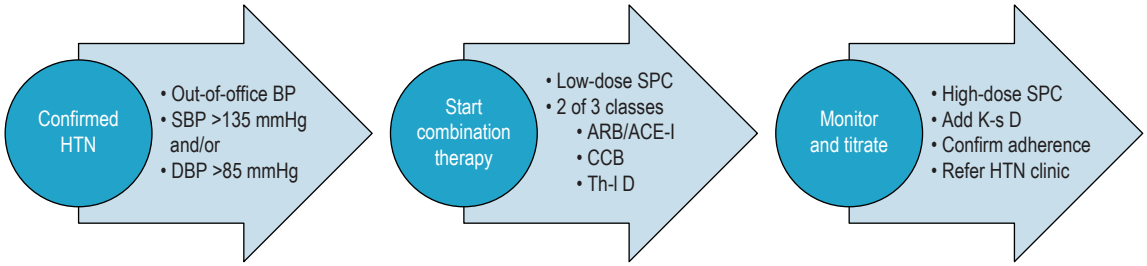


Fig 18.6 Summary of treatment recommendations. Elevated BP or hypertension should be confirmed with ambulatory or home BP (out-of-office monitoring) prior to initiation of therapy unless it is severe at presentation. Combination therapy, ideally in a single pill combination should be used in all patients unless there is clinical frailty or very mild elevated BP as most patients need 2 or more medications to control BP long term. BP should be monitored and medication adjusted proactively as leaving BP uncontrolled for more than a few months is associated with excess morbidity. Potassium sparing diuretics have the best evidence being added in as 4th line medication, though uncontrolled BP on 3 medications should prompt reviews for secondary causes and adherence and consideration of referral to a specialist clinic. *ACE-i,* Angiotensin converting enzyme inhibitor; *ARB,* angiotensin receptor blocker; *CCB,* calcium channel blocker; *DBP,* diastolic blood pressure; *HTN,* hypertension; *K-S D,* potassium sparing diuretic; *SBP,* systolic blood pressure; *SPC,* single pill combination; *th-l d,* thiazide like diuretic.

Management of cardiovascular risk

In keeping with the view that reduction in total CV risk is paramount, cholesterol-lowering therapy, most commonly with small doses of potent statins, is recommended in all patients with a total CV risk of more than 10% per year. This threshold may not be reached in younger patients due to the power of age in the risk calculators, and so a more long-term view of risk reduction should be taken in these patients. Although low-dose aspirin has been used in ***primary CV prevention*** for many decades, contemporary large-scale trials have not demonstrated a net benefit on mortality, even in patients with diabetes, with any gains from reduction in CV events offset by increases in bleeding events. Given these data, aspirin is no longer recommended in any patient group for primary prevention.

Further reading

Ettehad D, Emdin CA, Kiran A et al. Blood pressure lowering for prevention of cardiovascular disease and death: a systematic review and meta-analysis. *Lancet* 2016; 387:957–967.
Hunter PG, Chapman FA, Dhaun N. Hypertension. Current trends and future perspectives. *Br J Clin Pharmacol* 2021; 87:3721–3736
Morisky DE, Ang A, Krousel-Wood M et al. Predictive validity of a medication adherence measure in an outpatient setting. *J Clin Hypertens* 2008; 10:348–354.
Williams B, MacDonald TM, Morant M et al. Spironolactone versus placebo, bisoprolol, and doxazosin to determine the optimal treatment for drug-resistant hypertension (PATHWAY-2): a randomised, double-blind, crossover trial. *Lancet* 2015; 386:2059–2068.
https://www.qrisk.org/three/. *Cardiovascular risk score calculator.*

MANAGING BLOOD PRESSURE IN HOSPITAL

Blood pressure on the wards

Patients with hypertension are often admitted to hospital for other medical reasons. Various factors can contribute to acute elevations of BP, such as omission of regular antihypertensive medication, drug-induced hypertension, anxiety, pain, bladder distension, recreational drug withdrawal, and neurological injury. In the absence of a hypertensive emergency or other co-morbidity that requires acute BP management (e.g., to facilitate thrombolysis in acute stroke), uncontrolled BP does not require immediate management. However, as with most phenotypes, there is a threshold value that makes most healthcare professionals anxious about increased cerebral and CV risk, even in inpatients, and it is therefore common for BPs consistently above 180/110 mmHg to be treated, where there is no other clear precipitant. There is no role for short-acting drugs, such as sublingual glyceryl trinitrate or nifedipine, and titration of standard chronic therapies with on-going monitoring is preferred.

Hypertensive emergencies

Occasionally, hypertension is so severe with evidence of end-organ damage that urgent, specialist inpatient or closely-supervised outpatient care is required. The term ***hypertensive emergency*** is used to describe patients with significantly elevated BP (usually >180/120 mmHg) and signs or symptoms of acute target-organ damage (**Box 18.7**). The rate and magnitude of any increase in BP may be more important than the absolute level of BP in determining the severity of organ injury; this is key in obstetric medicine, where the usual BP in younger women may be verging on hypotensive (90–110/60–70 mmHg), and pre-eclampsia may develop with a BP over 140/90 mmHg.

In most hypertensive emergencies, intravenous antihypertensive medication therapy is indicated to cause rapid reduction of BP, as this is thought to minimize on-going organ damage and prevent or reduce the risk of morbidity and mortality. The exception to this is malignant hypertension with only eye changes and no other features (haemaotological, renal, etc.) where in specialist centres, treatment with oral therapy may be suitable with appropriate monitoring and follow up.

Box 18.7 Management of hypertensive emergencies

Presentation/syndrome	History	Examination	Investigation of choice	Target/timeframe	First line therapies	Notes
'Malignant' or 'accelerated' hypertension	Visual disturbance, headache	Moderate–severe retinopathy		25% MAP reduction within a few hours	i.v. labetalol, nicardipine (with retinopathy only, some consider oral treatment sufficient)	Often associated with acute kidney injury, due to fibrinoid small-artery necrosis
Hypertensive encephalopathy	Visual disturbance, headache, seizures, confusion, coma	Often features moderate–severe retinopathy	MRI brain: posterior fossa oedema (PRES)—see Chapter 28	Immediate MAP reduction by 25%	i.v. labetalol, nicardipine	
Acute aortic dissection	Tearing chest pain	Unequal arm pulses or BP Aortic valve incompetence	CT angiogram aorta, trans-oesophageal echo: dissection flap visible	Immediate SBP reduction to 100–120 mmHg *and* heart rate to 50–60 beats/min	i.v. labetalol Once rate control is adequate, vasodilators can be added (GTN, nicardipine)	Patients with Marfan's or proximal aorta syndromes more commonly proceed to early surgery
Acute pulmonary oedema	Shortness of breath	Bibasal crackles, elevated jugular venous pressure	Chest X-ray: interstitial oedema	Immediate SBP reduction to <140 mmHg	i.v. GTN, diamorphine	Loop diuretics are also vasodilating and lower BP
Acute coronary syndrome	Chest pain	Diaphoretic	ECG: ST/T-wave changes Serum troponin elevated	Immediate SBP reduction to <140 mmHg	i.v. GTN, labetalol	BP control should not delay procedural intervention
Pre-eclampsia	Oedema, visual disturbance, abdominal pain	Oedematous New proteinuria		Immediate SBP reduction to <160 and DBP to <105 mmHg	i.v. labetalol, nicardipine	i.v. magnesium used to prevent/treat eclamptic seizures

BP, Blood pressure; *CT,* computed tomography; *DBP,* diastolic blood pressure; *ECG,* electrocardiography; *GTN,* glyceryl trinitrate; *MAP,* mean arterial pressure; *MRI,* magnetic resonance imaging; *PRES,* posterior reversible encephalopathy syndrome; *SBP,* systolic blood pressure.

Blood pressure control in special populations

The management of hypertension in pregnancy is discussed in Chapter 48. The management of hypertension around surgical procedures is discussed in Chapter 14.

Questions for this chapter can be found on Elsevier eBooks+.

19 Lipid and metabolic disorders

Anthony S. Wierzbicki and Wiaam Al-Hasani

Background 510
Approach to the patient 510
History 510
Examination 510
Investigations 510
Physiology 510
Measurement of serum lipids 512
Plasma lipids and cardiovascular risk 512
Hyperlipidaemia 513
Disorders of VLDL and chylomicrons: moderate and severe hypertriglyceridaemia 513
Disorders of LDL: hypercholesterolaemia alone 513
Disorders of HDL 516
Combined hyperlipidaemia (hypercholesterolaemia and hypertriglyceridaemia) 516
Other lipid disorders 517
Management of hyperlipidaemia 517
Non-pharmacological approaches 518
Lipid-lowering drugs 518

CORE SKILLS AND KNOWLEDGE

Lipid disorders are common and of great public health importance. All doctors need to be comfortable with managing such patients in the context of their overall cardiovascular disease risk.

Key skills in this area include:

- History taking to quantify cardiometabolic risk
- Recognition of physical signs of hyperlipidaemia
- Familiarisation with the different drugs for lowering lipids and their role in reducing cardiovascular disease risk.

Opportunities for learning in this specialty include observing clinics in specialist metabolic medicine centres and shadowing a dietician giving advice to patients on reducing cardiovascular disease risk.

BACKGROUND

Lipid disorders are common, with 95% of the Western population having raised cholesterol levels compared with undeveloped societies. Cholesterol contributes to up to half of the population's cardiovascular disease (CVD) risk worldwide. A single genetic disorder of cholesterol metabolism—familial hypercholesterolaemia (FH)—is one of the commonest autosomal dominant disorders globally. Other disorders are less common, but defects of the lipoprotein lipase (LPL) pathway constitute a medical emergency as they can present as gross hypertriglyceridaemia associated with pancreatitis.

APPROACH TO THE PATIENT

History

Significant factors relevant to lipid disorders include:

- Demographics, including ethnicity
- A cardiovascular risk factor history, including current and previous smoking, diabetes or pre-diabetes, family history of CVD prior to age 60 (to second generation if available), and family history of hyperlipidaemia
- A history of unexplained pancreatitis.

Examination

Examination of a patient with a lipid disorder includes:

- Noting signs of soft tissue lipid accumulation, such as tendon lumps (xanthomata), eyelid xanthelasma, and eruptive or planar xanthomata
- Measuring waist circumference and calculating body mass index (BMI)
- Measuring blood pressure.

Investigations

Patients with lipid disorders require a random or fasting lipid profile and (at their first clinic visit) exclusion of secondary causes of hyperlipidaemia such as nephrotic syndrome (by measuring urine protein).

PHYSIOLOGY

Lipids are essential as fuel sources, structural components of membranes, and the basis for paracrine mediators.

Lipid components can be synthesized from both fatty acids and carbohydrates, via acetyl-CoA derived from the tricarboxylic acid (Krebs) cycle pathway. Acetyl-CoA either can then be made into linear chain fatty acids (to be used or stored as metabolic fuel) or can be made into the ring structures of cholesterol, which is an essential component of cell membranes (e.g., in lipid rafts).

The cholesterol synthesis pathway is regulated by its end-product through its rate-limiting initial enzyme, hydroxy-methyl-glutaryl coenzyme A reductase (HMG-CoAR), which is inhibited by 'statins', explaining their importance in the control of lipid disorders.

Lipids are insoluble in water and are transported in the bloodstream as macromolecular complexes. These complexes comprise a core of hydrophobic triglyceride (TG), cholesterol, and cholesterol esters surrounded by a phospholipid coat. Proteins (called apolipoproteins) on the surface of these 'lipoprotein' particles act as

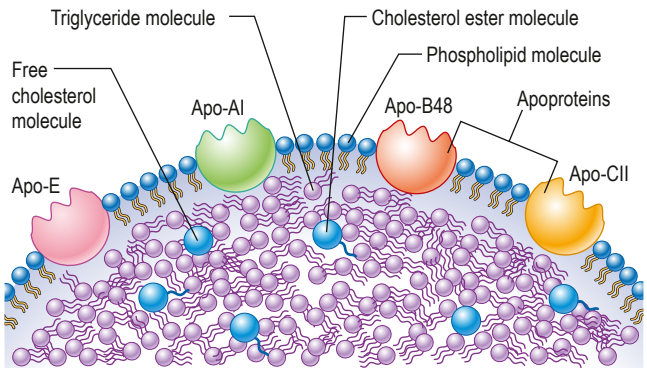


Fig. 19.1 A chylomicron particle (75–1200 nm) showing apoproteins lying in the surface membrane.

modulators of metabolism (e.g., enzyme cofactors) and ligands for receptors in the liver and the peripheral tissues. The structure of a chylomicron (one type of lipoprotein particle) is illustrated in Fig. 19.1.

Five principal types of lipoprotein particles are found in the blood: chylomicrons, the largest; very-low-density; intermediate-density; low-density; and high-density lipoproteins (VLDL, IDL, LDL, and HDL, respectively) (Fig. 19.2). They are structurally different and can be separated in the laboratory by their composition, density, and electrophoretic mobility. The larger particles give postprandial plasma its cloudy appearance (lipaemia). The major apolipoproteins and most receptors have been defined, but the functions of some remain unclear.

Plasma lipid levels are dependent on nutritional and environmental factors as well as genetic variants and contribute 55% of population-attributable risk for atherosclerotic CVD. Genetic disorders of lipid metabolism are relatively common and contribute to a highly increased risk of CVD in some individuals.

Chylomicrons

Chylomicrons (see Fig. 19.1) are synthesized in the small intestine postprandially in response to the presence of gut fats (TGs) and cholesterol. They are secreted into the intestinal lymphatic drainage and then pass along the thoracic duct into the bloodstream to be cleared eventually by the liver. They contain large quantities of TGs, some cholesterol, and cholesterol esters. They are the main mechanism for transporting dietary fat, cholesterol, and dietary-derived fat-soluble vitamins. Apolipoprotein C2 on the surface of chylomicrons binds to specific vascular surface (endothelial) receptors in adipose tissue, skeletal muscle, and the liver, where the endothelial enzyme LPL hydrolyzes TGs into fatty acids for local uptake. The chylomicron remnant particle, containing most of the cholesterol, is taken up by the liver by a specific receptor and also by the apo E/B-100 receptor.

Very-low-density lipoprotein particles

VLDL particles are synthesized continuously in the liver. They transport the body's endogenously synthesized TG and cholesterol. Apolipoprotein B-100 is the signature protein component of VLDL and its metabolic derivatives and is present at one per particle. VLDL particles become denser and turn into IDL as TGs are removed and apo C1-3 and E are incorporated from HDL particles.

Intermediate-density lipoprotein particles

IDL particles contain apo B-100 and apo C1-3 and apo E molecules on the particle surface. Most IDL particles bind to liver LDL (apo-E/B-100) receptors through the apo E molecule and are catabolized. Further TG removal leads to a denser remnant called LDL particles.

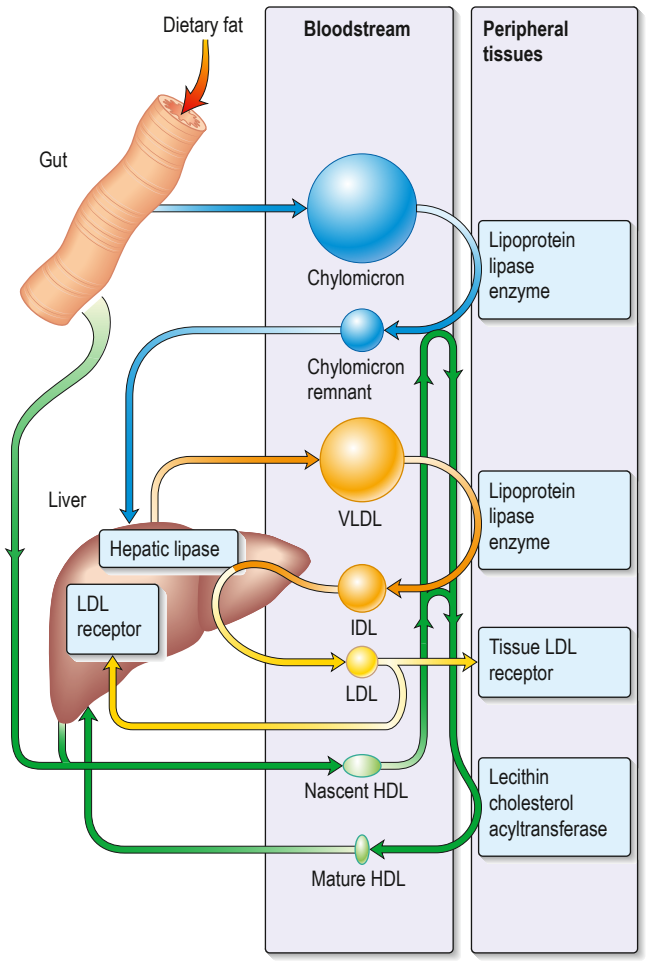


Fig. 19.2 The sites of origin of the major lipoprotein particles, showing their interaction and fate. *HDL,* High-density lipoprotein; *IDL,* intermediate-density lipoprotein; *LDL,* low-density lipoprotein; *VLDL,* very-low-density lipoprotein.

Low-density lipoprotein particles

LDL particles are the main carrier of cholesterol in fasting plasma and deliver it both to the liver and to peripheral cells. The surface of the LDL particle contains one apo B-100 molecule and varying quantities of apo C-3 and apo-E. Apolipoprotein B-100 is the principal ligand for the LDL (apo-E/B-100) receptor (LDLR). The LDLR binds LDL particles and forms coated pits on the surface of the hepatocyte, which invaginate and fuse with lysosomes and proteasomes. These destroy the LDL particle, releasing its components (Fig. 19.3). This process is, in part, regulated by a serine protease, proprotein convertase subtilisin/kexin type 9 (PCSK9). PCSK9 binds to LDL–LDLR complexes, increasing the rate of clearance of the complexes. If no PCSK9 is bound, then the LDLR can be recycled to the cell surface. The number of hepatic LDLR regulates the circulating LDL particle number and LDL cholesterol. Molecules that inhibit PCSK9 have been licensed in the UK for the control of dyslipidaemia, which cannot be controlled with statins and diet alone.

LDL particles are closely linked to CVD risk. They can deposit lipid into the walls of the peripheral vasculature through binding to elastin and other hyaline components, causing macrophage entry and atherosclerosis.

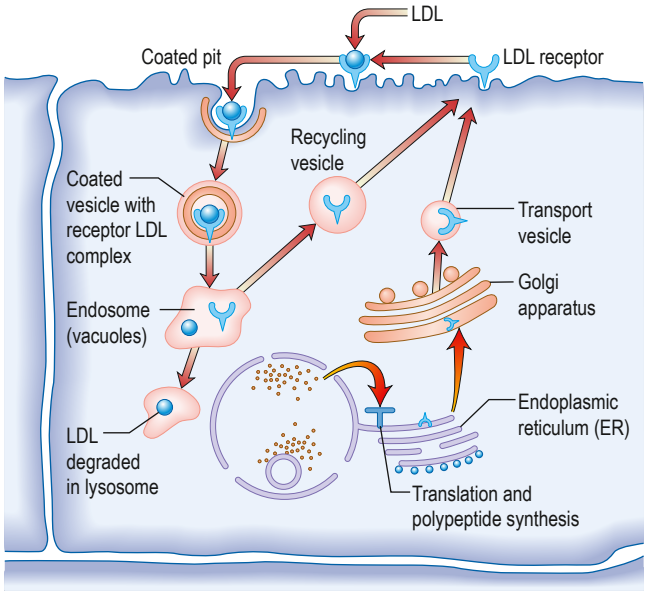


Fig. 19.3 Receptor-mediated endocytosis. Low-density lipoprotein (LDL) receptors are formed in the endoplasmic reticulum and transported via the Golgi apparatus to the surface of a hepatocyte. LDLs bind to these receptors and are internalized and taken up by the endosome. The receptor is recycled back to the surface, while the LDL is broken down by the lysosomes, freeing the cholesterol needed for membrane synthesis.

Lipoprotein(a)

Lipoprotein(a) (Lp(a)) is a subgroup of LDL particles associated with high rates of CVD. Lp(a) is derived from apo B-100 containing particles by addition and covalent binding of apo(a) at the liver cell surface. The apo(a) binds to apo B-100 in such a way as to impede clearance of the Lp(a) through the LDLR. High levels of Lp(a) are genetically co-dominantly inherited. Most Caucasians have little Lp(a), unlike West African populations. Lp(a) is 'sticky', binding avidly to arterial glycans and acting to promote atherosclerosis. Patients with high levels of Lp(a) show increased rates of CVD even if they have only moderate levels of LDL-C. The use of niacin, PCSK9 inhibitors, and other novel agents have all been tried or are being trialled to reduce CVD risk in patients with high Lp(a) concentrations.

High-density lipoprotein particles

Nascent HDL particles are produced in both the liver and the intestine. They are initially disc shaped and contain their signature protein apoprotein A1 (4 molecules), other apolipoprotein components, and some phospholipid. HDL particles take up cholesterol from cells through the ATP-binding cassette transporter-A1 (ABC-A1). A concentration gradient for export is maintained by esterifying this cholesterol through the action of lecithin-cholesterol acyltransferase (LCAT). The particles acquire further cholesterol through ABC-G1 and also phospholipids, apo C1-3, and apo-E by transfer from chylomicrons and VLDL particles.

The HDL particle transports excess cholesterol from the periphery and may transfer it to other particles, such as VLDL (through the action of cholesterol ester transfer protein, CETP), or may deliver cholesterol directly to the liver (reverse cholesterol transport) or to steroid-synthetic tissues (ovaries, testes, adrenal cortex). HDL particles are taken up by the liver scavenger receptor B1 (SRB1).

In contrast to apo B-100 particles such as LDL, the apo A1 core of HDL is recycled and re-secreted. HDL can be modified, and its handling changed by serum amyloid A substituting for apoA1 molecules, changing it to a more pro-atherogenic form. Highly depleted or modified HDL particles are eventually cleared by the kidney through megalin and cubilin receptors. Cholesterol in the liver can be oxidized to bile acids, which are secreted into the bile to promote cholesterol and TG uptake in the gut. Excess bile salts are resorbed in the terminal ileum and recirculate in the enterohepatic circulation (see Chapter 24). HDL particles also have multiple other actions including reducing inflammation, being anti-parasitic, modulating clotting factor activity, affecting bone marrow function, and promoting endothelial health.

Measurement of serum lipids

In a fasting serum lipid profile, the majority of the total cholesterol (TC) concentration is present in LDL particles, with about 20% to 30% in HDL particles. The TG concentration largely reflects VLDL particles, as chylomicrons are normally completely cleared by 8 to 12 hours of fasting. In non-fasting samples, the total TG concentration will be raised, due to the presence of TG-rich chylomicrons. The laboratory will usually supply the non-HDL cholesterol (difference between total and HDL cholesterol) concentration as this is the strongest index of CVD risk. In patients with TGs below 4.5 mmol/L, and ideally fasting, the laboratory typically calculates the LDL cholesterol (LDL-C) concentration using the Friedewald formula: LDL-C = TC – HDL-C – (TG/2.2) (in mmol/L). A newer formula to calculate LDL-C, which is less affected by fasting and accurate with TG below 8 mmol/L has recently been defined: LDL-C = (TC/0.948) – (HDL-C/0.971) – ((TG/3.74) + (TG × (TC-HDLC)/24.16) – (TG × TG/79.36)) – 0.244. Some laboratories measure concentrations of LDL-C directly, as well as those of Lp(a), apo B-100, and apo A1.

Plasma lipids and cardiovascular risk

Population studies have repeatedly demonstrated a strong association between total, non-HDL cholesterol (a surrogate for apo B-100-containing particles), and LDL-C with CVD risk; the strongest association is with coronary heart disease. Risk of CVD also rises progressively with TC: HDL-C ratio (**Fig. 19.4**). Higher HDL-C concentrations (up to 1.8 mmol/L), conversely, protect against CVD. There is a weak additional risk associated with TG-rich VLDL or IDL particles and CVD risk possibly driven by the cholesterol content of remnant particles. Chylomicrons do not confer an excess CVD risk but do raise the total plasma TG concentrations. Raised TG levels (>5 mmol/L), and especially those associated with continuing plasma persistence of chylomicrons, are exponentially associated with an increased risk of pancreatitis.

Screening

CVD secondary to atherosclerosis is common. Most patients with hyperlipidaemia are asymptomatic and have no clinical signs. Screening practices for CVD vary between countries ranging from opportunistic measurements to systematic programmes targeting individuals aged greater than 40 years. Opportunistic measurement of lipids is justified for:

- a family history of coronary heart disease (especially below 60 years of age);
- a family history of lipid disorders;
- the presence of physical signs, for example, xanthoma or xanthelasma at any age, or corneal arcus before the age of 40 years;
- diabetes mellitus;
- hypertension;
- obesity;
- acute pancreatitis; and

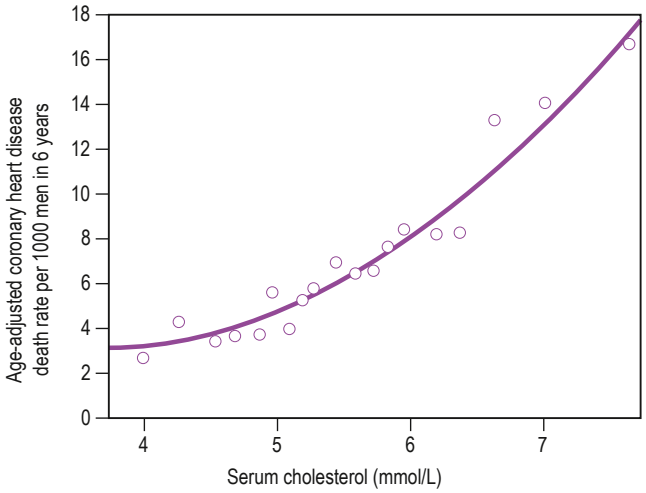


Fig. 19.4 The Multiple Risk Factor Intervention Trial.
Relationship between levels of serum cholesterol and risks of fatal coronary artery disease in a longitudinal study of more than 361 000 men screened for entry into the trial. *(Data from Stamler J, Wentworth D, Neaton JD. Is relationship between serum cholesterol and risk of death from coronary heart disease continuous and graded? Findings in 356222 primary screenees of the Multiple Risk Factor Intervention Trial [MRFIT].* JAMA *1986; 256:2823.)*

- patients undergoing renal replacement therapy.

Where one family member is known to have a monogenic disorder, such as FH, direct relatives, siblings, and children should have their plasma lipid concentrations measured.

Acute severe illnesses, such as myocardial infarction, reduce plasma lipid concentrations for up to 3 months. Lipid profiles should be measured either within 48 hours of an acute myocardial infarction (before derangement has had time to occur) or 3 months later. Cholesterol and HDL-C concentrations do not change significantly after a meal, so screening is performed using random blood samples for TC, TG, and HDL-C. Patients with moderate hypertriglyceridaemia should have a repeat fasting profile performed.

HYPERLIPIDAEMIA

Hyperlipidaemia results from genetic predispositions interacting with the environment, including nutritional factors.

Secondary hyperlipidaemia

Glucose and liver metabolism influence lipid levels. A clinical history, examination, and simple investigations need to be performed to detect the causes of secondary hyperlipidaemia (**Box 19.1**), which may need specific treatment in their own right. Hypothyroidism, untreated diabetes, renal disease (especially nephrotic syndrome), and abnormal liver function can all raise plasma lipid level.

Atherogenic dyslipidaemia is found in two-thirds of patients with metabolism-associated steatotic liver disease (MASLD, see Chapter 23), the most common chronic liver disease in the Western world. MASLD-associated dyslipidaemia is driven mainly by increased insulin resistance and is commonly characterized by excess TG-rich particles (VLDL), increased LDL-C, and low HDL-C. TG-rich apolipoprotein causes activation of the interleukin-1 (IL-1) family with subsequent induction of the inflammatory response through IL-6 and CRP pathway, promoting vascular inflammation, which plays a central role in the development of CVD. CVD is the most common cause of death among MASLD patients.

? Box 19.1 Causes of secondary hyperlipidaemia

- Hypothyroidism
- Diabetes mellitus (when poorly controlled)
- Obesity
- Chronic kidney disease
- Nephrotic syndrome
- Dysglobulinaemia
- Hepatic dysfunction and cholestasis
- Alcohol
- Anorexia nervosa
- Drugs:
  - Oral contraceptives in susceptible individuals
  - Retinoids, thiazide diuretics, corticosteroids, beta-blockers, sirolimus (and other immunosuppressive agents), anti-retroviral drugs

Primary hyperlipidaemia

Primary lipid disorders can be grouped as follows:

- Disorders of VLDL and chylomicrons—hypertriglyceridaemia
- Disorders of LDL—hypercholesterolaemia alone
- Disorders of HDL (rare)
- Combined hyperlipidaemia

Disorders of VLDL and chylomicrons: moderate and severe hypertriglyceridaemia

Most cases are caused by multiple gene variants acting together to produce a modest excess of circulating concentration of VLDL particles, with such cases being termed polygenic hypertriglyceridaemia.

Lipoprotein lipase pathway deficiencies

These are rare orphan diseases that produce greatly elevated TG concentrations (>20 mmol/L) owing to the persistence of chylomicrons (not VLDL particles), because their TG content cannot be properly metabolized. This may occur if the enzyme LPL is defective (95% of cases) or otherwise unable to function. Patients with homozygous or dual deficiency present in childhood with pancreatitis and hepatosplenomegaly, lipaemia retinalis (yellow lipid particles visible in retinal arteries), and sometimes eruptive xanthomas. Heterozygotes can present as adults with gross hypertriglyceridaemia often associated with pancreatitis. The presence of chylomicrons floating like cream on top of fasting lipaemic plasma suggests this diagnosis. The aetiology can be implied by measurement of the TG:apo B-100 ratio or ultracentrifugation, but the diagnosis is confirmed by DNA sequencing. Management is by strict reduction of dietary fat and use of lipid-lowering agents such as inhibitors of apo C3 (e.g., volanesorsen) and, in the case of heterozygotes, drugs such as fibrates to reduce TG levels.

Disorders of LDL: hypercholesterolaemia alone

Heterozygous familial hypercholesterolaemia

Heterozygous familial hypercholesterolaemia (HeFH) is an autosomal dominant monogenic disorder present in 1 in 350 of the population. Mutations affect the LDLR receptor (90%; more than 1500 mutations described), the apo B-100 LDLR ligand domain (5%; 1 principal mutation), and proteins such as PCSK9 that control LDLR expression (2%). The defect results in a reduction

Box 19.2 Genetic defects underlying some lipoprotein disorders

Disorder	Affected gene	Chromosome	Frequency and inheritance
Common disorders			
Heterozygous familial hypercholesterolaemia	*LDLR* (inactivating)	19	1:350 Autosomal dominant
Familial defective apolipoprotein B	*Apo B-100* (exon 26 mostly)	2	1:5000 Autosomal dominant
'Familial hypercholesterolaemia phenocopies'	*PCSK9* (activating) *apoE deletion Leu167*	1 19	1:30 000 Autosomal dominant
Autosomal recessive familial hypercholesterolaemia (ARFH)	*LDLRAP-1*	1	1:200 000 Autosomal recessive
Hypobetalipoproteinaemia	*Apo B-100* (various exons), MTP	2 4	1:10 000 Autosomal recessive
Familial combined hyperlipidaemia	Complex	Multiple	1:250 Complex
Familial hypertriglyceridaemia	Mostly polygenic	Multiple	1:1000 Complex
Some rarer disorders			
'Homozygous' familial hypercholesterolaemia	*LDLR>PCSK9*, *Apo-B* (exon 26)	19	1:350 000 Autosomal dominant
Familial chylomicronaemia syndrome (Lipoprotein lipase pathway deficiencies)	*LPL*, *>apoC2*, *apoA5*, *LMF-1, GPIHBP-1*	8, 19,11, 16, 8	1:1 000 000 (homozygous) Autosomal recessive
Tangier disease	*ABC-A1*	9	Very rare Autosomal recessive
Lecithin-cholesterol acyltransferase deficiency	*LCAT*	16	Very rare Autosomal recessive
Apolipoprotein A1 deficiency	*Apo-A1*	1	Very rare Autosomal dominant

in LDLR function, causing accumulation of LDL particles and LDL-C in plasma (**Box 19.2**). There is extensive phenotypic variation, but more severe cases have a 50% CVD mortality rate by age 65 with an approximate four- to nine-fold increase in CVD risk. Even individuals with FH and average levels of LDL-C show a three-fold increase in lifetime CVD risk. Genetic founder effects increase the prevalence in some groups (e.g., Afrikaner South Africans, Lebanese, French-Canadians). Pathognomic signs of FH include cholesterol deposition in extensor tendons such as the Achilles or those of the fingers. Diagnosis is easier if pathognomonic clinical features are present, but these only occur in 10%. Patients often have no physical signs, in which case the diagnosis is made on the presence of very high plasma cholesterol or LDL-C concentrations that are unresponsive to dietary modification and are usually associated with a family history of early CVD before age 60. Lipid deposition in the skin around the eyes (xanthelasma) may be present in some patients but is not diagnostic of FH and is more typical of high triglyceride-low HDL disorders. The diagnosis of FH is confirmed by DNA sequencing. Given its autosomal dominant inheritance, mutation screening is simple to apply, and children of affected individuals have a 50% chance of inheriting the defect.

Homozygous familial hypercholesterolaemia

Homozygous familial hypercholesterolaemia (HoFH) is an ultra-orphan disorder (affecting 1 in 250 000). Affected children have minimal LDLR function in the liver. They have a hugely elevated LDL-C (greater than 13 mmol/L) and massive deposition of lipid in arterial walls, the proximal aorta, and the skin on extensor surfaces (tuberose xanthomata). Most die from ischaemic heart disease by age 35 if untreated. They have a limited response to drugs such as statins, but PCSK9 inhibitors have been used with some success, especially in patients with point mutations. Other options include inhibitors of microsomal transfer protein (e.g., lomitapide) or newer agents such as angiopoietin-like peptide 3 (ANGPTL3) inhibitors (e.g., evinacumab). Repeated fortnightly lipid-specific plasma extraction (apheresis) can be used to remove LDL-C in patients with null mutations. LDLR function can be restored by partial lobe liver transplantation, which normalizes LDL-C levels and causes regression of xanthomata. Some orphan drug therapies are licensed for use in HoFH (**Box 19.3**).

Mutations in the apoB and PCSK9 genes

Mutations in the terminal part of *apoB* affect binding to the LDLR and cause familial defective apolipoprotein B (FDB), another relatively common single-gene disorder resembling FH. The clinical picture resembles classical HeFH but is often milder, with lower LDL-C levels as some LDLR functions are preserved. Patients with FDB form the commonest non-LDLR form of FH, with a prevalence of 1 in 10 000 or up to 10% of FH cases in some areas (e.g., Ireland, Switzerland). Activating mutations in *PCSK9* cause down-regulation of LDLR. This causes a phenocopy of FH and *PCSK9* mutations, which account for 1% of FH cases. In contrast, inactivating mutations in *PCSK9* are associated with enhanced LDLR function and lower cholesterol levels. A specific mutation in apolipoprotein E deleting leucine 167 also gives a phenocopy of FH and may account for 1% of cases. A very small proportion of FH is autosomally recessively inherited and caused by mutations in the low-density lipoprotein receptor adaptor protein 1 (LDLRAP-1). Most cases of autosomal recessive FH have been found in Sardinia. The different subtypes of FH can be distinguished clearly only by genetic tests. The approach to treatment is the same for all.

Box 19.3 Drugs used in the treatment of hyperlipidaemia

Drug/examples[a]	Mechanism of action	Contraindications	Adverse effects	Expected therapeutic effect	Safety
Statins					
Pravastatin Atorvastatin Rosuvastatin (Simvastatin) (Fluvastatin) (Pitavastatin) (Lovastatin)	Inhibit the rate-limiting step in cholesterol synthesis (HMG-CoA reductase)	Active liver disease, pregnancy, breastfeeding	Constant aches/muscle stiffness, myopathy, rhabdomyolysis, derangement of liver biochemistry Gastrointestinal upset Some statins have CYP3A4 interactions (e.g., simvastatin)	Reduce LDL cholesterol (LDL-C) by 20%–60% Have modest lowering effect on triglycerides Little effect on HDL cholesterol (HDL-C) Atorvastatin and rosuvastatin are the most potent	Atorvastatin and rosuvastatin have good long-term safety and extensive outcomes evidence Avoid if possible in women of childbearing age
Cholesterol absorption inhibitors					
Ezetimibe	Inhibits gut absorption of cholesterol from food and bile; inhibits activity of NPC1L1, a lipid transporter found in the duodenum	Breastfeeding	Occasional abdominal discomfort, skin rash	Reduce LDL-C by 20% Reduce triglyceride concentrations by <10% Minimal effect on HDL-C	Mostly act in gut, and recirculate in enterohepatic circulation Some safety and some outcome evidence
Acid-citrate lyase inhibitors					
Bempedoic acid	Inhibits the step in liver cholesterol synthesis before HMG-CoA reductase	Active liver disease, pregnancy, breastfeeding	Gout, diarrhoea, nausea, muscle spasms, anaemia Interaction with simvastatin and pravastatin	Reduce LDL-C by 20% Reduce triglyceride concentrations by 10% Minimal effect on HDL-C	Low rate of muscle adverse effects Increased occurrence of gout
Bile acid sequestrants					
Colesevelam (Colestyramine) (Colestipol)	Bind bile acids in the gut, preventing enterohepatic circulation; the liver makes more bile acids from cholesterol, depleting the cholesterol pool	Biliary obstruction	Gastrointestinal side-effects are very common, e.g., constipation, diarrhoea, bloating, flatulence; may reduce fat-soluble vitamin and drug absorption Other drugs bind to resins and should be taken 1 hour before or 4 hours afterwards	Reduce LDL-C by 15%–25% Increase triglyceride concentration by 5%–15% Have little or no effect on HDL-C	Not systemically absorbed Safety profile is good Appear safe in women of childbearing age Fat-soluble vitamin supplements may be required in children, and pregnant and breastfeeding women
Fibric acid derivatives					
Bezafibrate Fenofibrate (Ciprofibrate) (Gemfibrozil)	Activate peroxisome proliferator-activated nuclear receptors (esp. PPAR-α), producing protean effects on lipid metabolism	Severe hepatic or renal impairment, gall bladder disease, pregnancy	Reversible myositis, nausea, predisposition to gallstones, non-specific malaise, impotence Reversibly increase creatinine by 15% Drug interactions (e.g., statins), an issue with gemfibrozil	Reduce LDL-C by 0%–15% Reduce triglycerides by 25%–70% depending on baseline Increase HDL-C by 10%–30% depending on baseline	No knowledge of effect on developing fetus; however, avoid in women of childbearing age Long-term safety good and moderate beneficial CVD outcomes effects

Box 19.3 Drugs used in the treatment of hyperlipidaemia—cont'd

Drug/examples[a]	Mechanism of action	Contraindications	Adverse effects	Expected therapeutic effect	Safety
Nicotinic acid/niacin (NA) derivatives					
Modified-release niacin (Acipimox)	Unclear: probably inhibit lipid synthesis in the liver by effect on DGAT-2	Pregnancy, breastfeeding	Value limited by frequent side-effects: painful prostaglandin D2- and E2-induced flushing, nausea, malaise, itching, abnormal liver biochemistry Glucose intolerance, hyperuricaemia, dyspepsia; hyperpigmentation may occur	Reduce LDL-C by 15%–20% Reduce triglycerides by 20%–40% Increase HDL-C by 10%–20%	Medium-term safety good but marred by adverse effects Some monotherapy outcomes evidence Modified-release preparations better but still need high-dose aspirin to reduce flushing
Fatty acid compounds					
Omega-3 fatty acids (mixed) Omega-3 fatty acid ethyl esters (e.g., icosapent ethyl)	Reduce hepatic VLDL secretion and inflammation via FFAR-4	Avoid with anticoagulants since bleeding time is prolonged	Occasional nausea and belching	Reduce triglycerides in severe hypertriglyceridaemia No effect on other lipids	Long-term safety is good. High-dose CVD outcomes evidence
MTP inhibitors					
Lomitapide	Inhibits MTP, which is necessary for VLDL assembly and secretion in the liver	Unclear	Abnormal liver biochemistry	Lowers LDL-C by 50% Orphan drug licensed only for homozygous familial hypercholesterolaemia	Long-term safety data very limited
Proprotein convertase subtilisin kexin-9 (PCSK9) inhibitors					
Antibodies: Alirocumab Evolocumab Small interfering RNA: Inclisiran	Inhibit the action of PCSK9 to upregulate liver LDL receptor expression	Needle phobia (injectable agents)	Myalgia (flu-like illness), injection site reactions	Reduce LDL-C by 45%–60% Have modest lowering effect on triglycerides Little effect on HDL-C	Good safety and CVD outcomes evidence

[a]Drugs now rarely used in brackets.

CVD, Cardiovascular disease; *HMG-CoA,* hydroxymethylglutaryl coenzyme A; *LDL,* low-density lipoprotein; *MTP,* microsomal triglyceride transfer protein; *VLDL,* very-low-density lipoprotein.

Polygenic hypercholesterolaemia

This is a term used to lump together the large group of patients without evidence of monogenic cholesterol disorders above but with cholesterol in the upper part of the normal distribution. Panels for diagnosis of polygenic variants have been devised and include genes known to significantly affect cholesterol levels and CVD disease risk such as LDLR, apoA5, apo(a), apo-E, and PCSK9.

Disorders of HDL

Genetic disorders of HDL metabolism are rare and recessively inherited. They include those causing low HDL-C levels including mutations in apo-A1, ABC-A1, and LCAT. Apo-A1 mutations cause low HDL-C, which are associated with amyloid deposition and either atherosclerosis or protection from atherosclerosis, depending on the exact site affected. Very high HDL-C levels are caused by mutations in CETP or endothelial lipase, which affect the metabolism of HDL particles.

Tangier disease

Macrophages are involved in the uptake of excess cholesterol in arterial walls and donate this to HDL particles. If a defect occurs in cholesterol transport, for example, the ABC-A1 transporter, then they accumulate cholesterol (see **Box 19.2**). In Tangier disease, cholesterol accumulates in reticuloendothelial tissue and arteries, causing enlarged, orange-coloured tonsils and hepatosplenomegaly, corneal opacities, and polyneuropathy, as well as CVD.

Combined hyperlipidaemia (hypercholesterolaemia and hypertriglyceridaemia)

The most common patient group with abnormalities of serum lipid concentrations have a polygenic combined hyperlipidaemia. Patients have an increased CVD risk due to high non–HDL-C concentrations, abundant pro-atherogenic TG-rich particles, high LDL-C, and reduction of HDL-C concentrations by hypertriglyceridaemia (low LPL, excess CETP activity).

Familial combined hyperlipidaemia

Familial combined hyperlipidaemia (FCHL) is relatively common, affecting 1 in 200 of the general population. Its lipid phenotype includes hypertriglyceridaemia, hypercholesterolaemia, or both, associated with premature CVD. The genetic basis for the disorder is not clear, though current studies suggest that the primary defect lies in excess adipose tissue fatty acid release. It is diagnosed by finding raised cholesterol and TG concentrations in association with a typical family history. There are no typical physical signs.

Remnant hyperlipidaemia

This is a rare (1 in 5000) cause of combined hyperlipidaemia. It is due to the accumulation of IDL remnant particles and is associated with an extremely high risk of CVD. It may be suspected in a patient with almost equal cholesterol and TG concentrations and by finding lipid deposition in the dermis visible at palmar creases (planar xanthomata, which is diagnostic) and in the skin over the knees and elbows (tuberose xanthomata) (**Fig. 19.5**). Remnant hyperlipidaemia is commonly associated with inheritance of apo-E variant alleles (e.g., homozygosity for apoprotein E2 that has reduced LDLR binding), together with aggravating factor such as another primary hyperlipidaemia. The diagnosis can be confirmed using the TG:apoB-100 ratio, by genotyping apo-E for two specific single nucleotide polymorphisms (cysteine 158 to arginine [apoE2]; cysteine 112 to arginine [apoE4]) or by ultracentrifugation.

Other lipid disorders

Hypolipidaemia

Low lipid levels can be found in severe protein–energy malnutrition. However, some patients with severe anorexia may manifest high cholesterol levels secondary to protein-based gluconeogenesis in the liver. Low lipid levels are also seen occasionally with severe malabsorption and in intestinal lymphangiectasia.

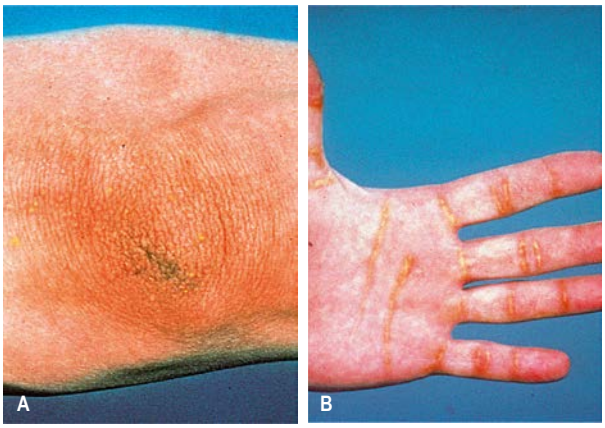


Fig. 19.5 Remnant hyperlipidaemia. **(A)** Tuberous xanthomas behind the elbow. **(B)** Lipid deposits in the hand creases.

Hypobetalipoproteinaemia (see **Box 19.2**) is a benign familial condition with cholesterol levels in the range of 1 to 3.5 mmol/L. Rare causes include mutations in apo-B, PCSK9, and apoC3.

Abetalipoproteinaemia

This is described in Chapter 36.

MANAGEMENT OF HYPERLIPIDAEMIA

Management of patients with abnormal lipid concentrations may require a combination of dietary and non-pharmacological interventions alongside lipid-lowering drugs. Sometimes, lipid abnormalities are so severe that treatment is indicated to prevent life-threatening complications such as pancreatitis in severe hypertriglyceridaemia; sometimes, lipid-lowering may be indicated in patients with only modestly elevated lipid concentrations as part of wider efforts to reduce cardiovascular risk.

The main indications for lipid-lowering are listed below:

Familial hypercholesterolaemia

Individuals with severe elevations of serum cholesterol often require treatment with dietary measures and more than one cholesterol-lowering drug from their teenage years to prevent early development of CVD. Ezetimibe is often added to a statin therapy (**Boxes 19.3** and **19.4**) and PCSK9 inhibitors are used in severe refractory cases.

Hypertriglyceridaemia

A serum TG concentration below 2.0 mmol/L is normal. In the range of 2.0 to 6.0 mmol/L, no specific intervention will be needed unless there are coincident CVD risk factors and, in particular, a strong family history of early CVD or an increased calculated CVD risk. Patients should be told they have a minor lipid problem and offered advice on lifestyle improvement including weight reduction if obese. The commonest causes are associated with MASLD (moderately raised transaminases, raised gamma-glutamyl transferase [GGT], and high TG:HDL-C ratio) or pre-diabetes. Patients with uncontrolled diabetes tend to have raised lipid levels, especially TGs. Some individuals have livers that respond to even moderate degrees of alcohol intake by increased production of VLDL, so a trial of abstinence is advised. Other drugs, including high-dose thiazides, oestrogens, and glucocorticoids, can have similar effects in susceptible patients.

If the TG concentration remains above 6.0 mmol/L despite lifestyle measures, drug therapy is warranted (see **Boxes 19.3** and **19.4**).

Box 19.4 Drug therapy for hyperlipidaemia

Hypertriglyceridaemia (predominantly)	Combined hyperlipidaemia	Hypercholesterolaemia (predominantly)
First-line agents		
Fibrate	Statin	Statin: moderate dose high intensity drug (atorvastatin or rosuvastatin) in primary prevention. Highest tolerated dose in secondary prevention (established CVD)
Additional/alternative agents		
Fish oil capsules including derivatives, e.g., icosapent ethyl (Niacin derivatives)	Fibrate may be required in addition to lower triglycerides adequately (monitor carefully for muscle aches, creatine kinase rise or worsening liver function). Avoid gemfibrozil in combination therapy Ezetimibe (to lower cholesterol further) (Niacin derivative to lower cholesterol and triglycerides further)	Ezetimibe and/or bempedoic acid if statin not tolerated or greater LDL-C lowering needed PCSK9 inhibitor (Cholesterol-binding resin)

LDL-C, Low-density lipoprotein cholesterol.

The severe hypertriglyceridaemia associated with disorders of the LPL pathway may require restriction of dietary fat to 10% to 20% of total energy intake, the use of orphan disease drugs, and the use of special preparations of medium-chain TGs in cooking in place of oil or fat as these are not absorbed via chylomicrons.

Management of combined hyperlipidaemia

Hypertriglyceridaemia increases CVD risk for any cholesterol concentration. Therapy is with diet in the first instance and then with drugs if required. Statins are first-line in patients with TGs less than 8 mmol/L, while fibrates are used first-line for higher levels. Combination therapy is often required (see **Box 19.4**).

Secondary prevention of cardiovascular disease

Statin treatment with high-dose, high-potency agents is warranted for any patient with known macrovascular disease (coronary artery disease, transient ischaemic attack or stroke, peripheral arterial disease), irrespective of the total or LDL-C level. The treatment target is TC below 4.0 mmol/L and LDL-C below 2 mmol/L (ideally <1.5 mmol/L), or alternatively, the maximum tolerated statin dose can be used. If a statin is not tolerated, combinations of other agents are tried (**Box 19.4**).

Primary prevention of cardiovascular disease in people with type 2 diabetes

Lipid-lowering therapy with a moderate dose statin, or alternatives as given previously, is used routinely in asymptomatic individuals with type 2 diabetes alone aged over 40 years, unless the LDL-C is particularly low (<1.8 mmol/L). Calculated CVD risk usually exceeds 10% per decade in these individuals. Good glycaemic control is essential.

Primary prevention of cardiovascular disease in people without diabetes

Lipid-lowering with a moderate dose statin in the first instance is offered to people after an unsuccessful trial of lifestyle intervention whose calculated CVD risk exceeds 10% over the next decade. Cardiovascular risk (below) may be calculated using various scoring systems based on large population studies.

Risk prediction tables

CVD risk is typically multifactorial. An isolated high level of a risk factor may not justify treatment unless levels are extreme (e.g., 95th centile). Epidemiological studies consistently identify age, gender, smoking, presence of type 2 diabetes, hypertension, and the ratio of TC:HDL cholesterol as primary risk factors for CVD. Ancillary risk factors such as family history of premature CVD may add to predictive efficiency in some cases.

CVD risk calculators rely on population studies to predict risk in individuals. This is always an approximate process. The commonest tools calculate either CVD mortality or events based on 10-year, 30-year, or lifetime absolute risk horizons. They are unreliable in populations aged less than 40 years or greater than 75 years, as CVD risk is driven principally by age. Relative risk compared to age and gender-matched controls may be informative outside the core age range.

Generally, drug therapy to lower serum lipids (typically using a statin) is considered mandatory if the cardiovascular risk is 20% or above and may be recommended if the risk is 10% or above. Discussion should be had with patients whose risk is above 5% despite lifestyle intervention. In young individuals, treatment decisions are often based on excess relative risk driven by family history, while in older patients, co-morbidities and life expectancy are significant factors to consider. In some countries imaging techniques such as coronary artery calcium scores are used to refine CVD risk predictions in borderline cases.

Non-pharmacological approaches

Lipid-lowering diet

The main elements of a lipid-lowering diet are similar to those for people with diabetes or hypertension. Additional specific measures are to:

- reduce alcohol consumption, since this may worsen primary lipid disorders at doses that would not affect normal individuals; and
- include foods capable of binding cholesterol such as bran or soya protein.

Undeveloped societies typically have a TC of 3.5 to 4.0 mmol/L with LDL-C of 1.5 to 2.0 mmol/L. A vegan diet can reduce cholesterol and LDL-C by 2 mmol/L and a vegetarian diet by 1 mmol/L compared to the usual Western diet. Several 'nutraceutical' or alternative health products are available that can reduce LDL-cholesterol levels. Plant stanols are analogues of cholesterol that bind to gut cholesterol transporters and reduce the absorption of cholesterol. Products containing sitostanol or sitosterol reduce cholesterol by approximately 0.35 to 0.5 mmol/L, but with a wide inter-individual variation. Red yeast rice is a Chinese medicine that contains a homologue of a statin (monacolin A) capable of reducing LDL-C by 5% to 20%.

Exercise, weight loss, and smoking

Regular exercise, weight loss, and stopping smoking all reduce CVD risk, irrespective of lipid levels. In patients with MASLD, weight loss of 5% will result in at least 10% reduction in serum TG concentration.

Lipid-lowering drugs

See **Box 19.3** for the classes of drugs used to treat hyperlipidaemia and **Box 19.4** for principles of use.

Statins

Statins block HMG-CoA reductase, the rate-limiting step in cholesterol synthesis, leading to compensating effects that increase the expression of LDLRs and hence a reduction in plasma cholesterol concentrations of 20% to 50%. Statins reduce CVD events by 21% per 1 mmol/L reduction in LDL-C. Statins are the first-line treatment for dyslipidaemia in primary and secondary prevention of CVD. In general, their adverse effects are uncommon and reversible; however, the rate of discontinuation of statins ranges between 25% and 50% following the initiation of the treatment. The main cause for the non-adherence is statin intolerance.

Statin intolerance

Statin intolerance can be defined as any adverse effect deemed unacceptable by the patient and/or some laboratory abnormalities (e.g., deranged liver function or elevated creatine kinase [CK]) that are attributed to statins and which improve on discontinuation of the treatment and reappear on re-introduction. Importantly, the nocebo effect (development of adverse effects independent of the active treatment due to negative perceptions or beliefs about the treatment) should be excluded before statin intolerance is confirmed. The nocebo effect is influenced by healthcare beliefs, professional interaction, mainstream media, and social models.

Statin-associated muscle symptoms (SAMSs) are the main reported side-effects, with deranged biomarkers representing the minority of statin-intolerant cases. SAMSs commonly present as generalized muscular aches (myalgia) and occasionally as myositis if CK is elevated; rarely this extends to rhabdomyolysis in severe cases (CK > 5000 IU/L). Pharmacogenetics effects exist for some statins due to variants in the *SLCO1B1* transporter, which increase gut uptake of the statin, resulting in higher levels in plasma of free statin acids and increased toxicity.

Strategies to reduce statin intolerance include having a patient-centred approach and personalized treatment, as well as managing any reversible risk factors that increase the frequency of statin intolerance.

Gut-acting cholesterol-lowering drugs

Ezetimibe is a gut-acting drug binding to the Niemann–Pick C1-like protein 1 (NPC1L1) duodenal cholesterol transporter. This molecule is also the binding site for campesterol and beta-sitosterol. Ezetimibe reduces plasma cholesterol by 20% and has been shown to reduce CVD events when added to statin therapy. It is generally very well tolerated.

Bile acid sequestrants (e.g., colesevelam) are drugs that bind bile acids, preventing micelle formation and thus reducing gut cholesterol uptake. They reduce cholesterol by 15% to 20% and have some CVD outcome evidence from previous trials with monotherapy, but data is lacking on CVD outcomes in the era of statins. They are commonly associated with excess gastrointestinal side-effects and are now rarely used outside the context of liver disorders associated with raised cholesterol with concurrent itch secondary to bile acids.

Acid-citrate lyase inhibitors

Bempedoic acid is a liver-acting drug that inhibits acid-citrate lyase, the step before HMG-CoA reductase. It reduces plasma LDL-C by 20% to 30% and has been shown to reduce CVD events in patients intolerant to statin therapy in line with its effects in reducing LDL-C. It is often co-prescribed with ezetimibe. Little data exists on triple therapy of bempedoic acid with ezetimibe and statins.

PCSK9 inhibitors

PCSK9 is involved in downregulating LDLR expression. Antibodies to PCSK9 (e.g., alirocumab) and small interfering RNA (siRNA) drugs targeting PCSK9 mRNA (e.g., inclisiran) reduce LDL-C by 40% to 60%. Antibodies to PCSK9 have been shown to reduce CVD events, and trials are underway with siRNAs and small molecule PCSK9 inhibitors. These drugs are used as third-line agents in FH, in severe progressing CVD, or for patients intolerant to multiple statins.

Fibrates

Fibrates are gene regulators affecting the peroxisomal-proliferator-activating receptors-alpha subtype (PPAR-α). They reduce TGs by increasing the synthesis of LPL, reducing the synthesis of apoC3 to promote clearance of TG-rich particles (e.g., VLDL), and can increase apo-A1. They have a moderate effect in reducing CVD events, with best evidence in monotherapy. They do not seem to add to the benefits of reducing LDL-C in patients on statins.

Niacin

Niacin (and its related derivative acipimox) are old lipid-lowering drugs. The mechanism of action of niacin is unclear but involves effects in inhibiting diacylglycerol O-acyltransferase 2 (DGAT-2) in the liver to reduce TG and cholesterol transfer to nascent VLDL particles, thus reducing LDL-C. It also inhibits the liver beta-chain adenosine triphosphate synthase receptor to raise plasma apoA1 and HDL-C levels. Its adverse effects include prostaglandin D2 and prostaglandin E mediated flushing and hyperglycaemia. Though effective in reducing CVD events in monotherapy studies, niacin was ineffective when added to statins in recent trials.

Omega-3 fatty acids

These preparations (often termed fish oils) include the essential fatty acids docosahexaenoic acid (DHA) and eicosapentaenoic acid (EPA). They bind to specific free fatty acid receptors (FFAR-4) to affect lipid and inflammation pathways. They exert dose-proportional effects in reducing TGs by 20% to 50%. Their effects in the prevention of CVD are controversial, with conflicting results. Low doses in low-risk patients seem to do little, but high doses of EPA derivatives (e.g., icosapent ethyl) in high-risk patients reduce CVD events when added to statins and are independent of any effect on lipid profiles.

Lipoprotein(a) inhibitors

Epidemiological evidence, cohort, and apheresis studies support a role for reducing Lp(a) levels. Previously, both niacin and recently PCSK9 inhibitors have been shown to reduce Lp(a) levels by 25% to 30%, but subgroup analyses have been unable to clearly differentiate their effects on LDL-C as opposed to reducing Lp(a). Newer antisense therapies are capable of reducing Lp(a) by 70% to 80% and are now in clinical trials in patients with CVD and Lp(a) levels exceeding 150 nmol/L.

Apolipoprotein C3 inhibitors

The clearance of TG-rich particles is promoted by apoE and inhibited by apoC3. Antisense inhibitors of apoC3 have been developed and shown to reduce TGs by 70% to 80% in patients with familial hyperchylomicronaemia syndrome (FCS). This results in a 30% to 50% reduction in pancreatitis events. Volanesorsen, the first in the class, is licensed for the management of FCS but has adverse effects including reducing platelet counts. Newer derivatives (e.g., siRNA) seem to avoid this adverse effect. Studies are being considered to address whether reducing apoC3, and hence residual TGs in statin-treated patients will result in further CVD event reduction as non-HDL-C is a stronger predictor of CVD events than LDL-C levels.

Apheresis

Though drug therapies are the mainstay of lipid management, physical removal of apoB-containing lipoprotein particles (including Lp(a)) is possible by plasmapheresis or, more specifically, by lipid apheresis. This invasive recurrent therapy is used in the management of patients with homozygous FH not responding adequately to drug therapies and in individuals showing rapid progression of atherosclerosis despite reasonable lipid profiles who have high Lp(a) levels or evidence of concurrent inflammation. Plasmapheresis is also used in the management of chylomicronaemia syndromes especially in the later stages of pregnancy to reduce the risk of pancreatitis.

Box 19.5 Classes of drugs in clinical trials for the potential treatment of hyperlipidaemia

Drug class	Mechanism of action
Proprotein convertase subtilisin/kexin type-9 (PCSK9) inhibitors For example: inclisiran; MK-0616	Small interfering RNA drug that promotes PCSK9 mRNA destruction and lowers LDL-C by 50%; 3- to 6-monthly dosing. Phase 3 clinical trial on CVD outcomes (ORION-4) in progress. Small molecule PCSK9 inhibitor: MK-0616 in early-phase trials.
Anti-lipoprotein(a) antisense therapy For example: Apo(a)Lrx	Inhibits formation of lipoprotein(a) in the liver. Reduces Lp(a) by 80%. Phase 3 clinical trials in progress
Apolipoprotein C3 inhibitors For example: volanesorsen; siRNA apoC3	Inhibits the action of apoC3, thus increasing clearance of triglyceride-rich particles. Already used in the management of familial chylomicronaemia syndrome. In development for use in patients with CVD and residual hypertriglyceridaemia.
DNA targeted therapy V-101 PCSK9 base editing AAV8.hLDLR	Introducing a functional gene copy to restore the function of the mutated gene, e.g., LDLR gene or inactivating PCSK9 to compensate for reduced LDLR function in HeFH CRISPR/Cas9 system work as gene editors by cutting part of the gene and altering the function of the transcript to either repair it or inactivate a gene to compensate for the original defect

HDL, High-density lipoprotein; *LDL,* low-density lipoprotein; *Lp(a),* lipoprotein(a); *RNA,* ribonucleic acid.

Novel therapeutics

New classes of lipid-lowering drugs in development are listed in **Box 19.5**. Some of these are developments of existing classes such as oral small-molecule PCSK9 inhibitors to reduce LDL-C or anti–apo C3 drugs with improved hepatic clearance and reduced dosing or formulated through small interfering ribonucleic acid (siRNA) technology to reduce TGs. Novel therapeutic approaches include targeting Lp(a) with siRNA to reduce CVD risk in individuals who have inherited high levels or antagonists to angiopoietin-like peptide 3 (ANGPTL3) to reduce TGs. Gene therapy approaches have included a previous unsuccessful attempt to introduce LDLR into hepatocytes using lentiviruses to treat homozygous FH. An adeno-associated viral (AAV) vector to introduce functional LPL and ameliorate FCS (alipogene tiparvovec) was successful in transiently reducing TGs and in reducing pancreatitis but was withdrawn for commercial reasons. A gene therapy for heterozygous FH caused by LDLR mutations by inactivating PCSK9 using siRNA technology (Verve-101) is in early trials.

Further reading

Cholesterol Treatment Trialists' (CTT) Collaboration. Efficacy and safety of LDL-lowering therapy among men and women: meta-analysis of individual data from 174 000 participants in 27 randomised trials. *Lancet* 2015; 385:1397–1405.
Defesche JC, Gidding SS, Harada-Shiba M et al. Familial hypercholesterolaemia. *Nat Rev Dis Primers* 2017; 3:17093.
Emerging Risk Factors Collaboration. Major lipids, apolipoproteins, and risk of vascular disease. *JAMA* 2009; 302:1993–2000.
Grundy SM, Stone NJ, Bailey AL et al. 2018 AHA/ACC/AACVPR/AAPA/ABC/ACPM/ADA/AGS/APhA/ASPC/NLA/PCNA Guideline on the management of blood cholesterol: Executive summary: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *Circulation* 2018: 139:e1046–e1081.
Silverman MG, Ference BA, Kyungah IM et al. Association Between Lowering LDL-C and Cardiovascular Risk Reduction Among Different Therapeutic Interventions: A Systematic Review and Meta-analysis. *JAMA* 2016; 316:1289–1297.
Van Horn L, Carson JA, Appel LJ et al. Recommended dietary pattern to achieve adherence to the American Heart Association/American College of Cardiology (AHA/ACC) guidelines: a scientific statement from the American Heart Association. *Circulation* 2016; 134:e505–e529.
Visseren FJ, Mach F, Smulders YM et al. 2021 ESC Guidelines on cardiovascular disease prevention in clinical practice. *Eur Heart J* 2021; 42:3227–3337.
Wierzbicki AS, Kim EJ, Esan O, Ramachandran R. Hypertriglyceridaemia: an update. *J Clin Pathol* 2022;75:798–806.

Questions for this chapter can be found on Elsevier eBooks+.

20 Respiratory medicine

Veronica White and Prina Ruparelia

Clinical skills 521
History 521
Examination 522
Function of the respiratory system 523
Anatomy 523
Physiology 525
Clinical approach to the patient with respiratory disease 529
Clinical features of respiratory disease 529
Examination of the respiratory system 530
Investigation of respiratory disease 532
Diseases of the upper respiratory tract 537
Obstructive respiratory disease 540
Asthma 540
Chronic obstructive pulmonary disease 545
Obstructive sleep apnoea 549
Smoking 551
Respiratory infection 552
Pneumonia 552
Tuberculosis 553
Pleural disease 557
Tumours of the respiratory tract 559
Malignant tumours 559
Benign tumours 563
Bronchiectasis 565
Cystic fibrosis 566
Interstitial lung diseases 568
Sarcoidosis 568
Lung and heart–lung transplantation 577
Occupational lung disease 577
Miscellaneous respiratory disorders 578
Disorders of the diaphragm 580
Mediastinal lesions 580

CORE SKILLS AND KNOWLEDGE

Respiratory medicine is an exciting and varied specialty caring for patients with a range of acute and chronic conditions (Box 20.1). One-third of visits to primary care are due to respiratory symptoms, and respiratory conditions are the reason for one-quarter of medical admissions to hospital.

Respiratory physicians provide a large amount of inpatient care for acute problems such as pneumonia, exacerbations of chronic obstructive pulmonary disease, and other chronic respiratory conditions. General respiratory outpatient care undertakes the initial investigation and diagnosis of patients with respiratory symptoms, with specialist clinics providing long-term follow-up of individual conditions. Care is shared with community teams and nurse specialists to promote health and quality of life through education, self-management, and admission-avoidance strategies. Practical procedures performed by respiratory physicians include bronchoscopy, endobronchial ultrasound scanning, and pleural interventions.

Key skills in respiratory medicine include:

- Managing respiratory emergencies: type 1 and type 2 respiratory failure, pleural effusion, pneumothorax, and large-volume haemoptysis
- Interpreting examination findings and respiratory investigations: chest radiographs, arterial blood gas analysis, and spirometry
- Understanding treatments for respiratory conditions: inhaled and nebulized medications, home oxygen therapy, and non-invasive ventilation.

Respiratory medicine can be learned through attending general and specialty respiratory clinics (e.g., lung cancer, tuberculosis, or sleep clinics), observing specialist investigations such as lung function measurement or bronchoscopy, attending clinics and home visits with specialist nurses, and interpreting arterial blood gases to manage oxygen delivery or assisted ventilation systems.

CLINICAL SKILLS

History

The following are relevant in respiratory disease:

- ***Respiratory symptoms:*** cough, sputum production, breathlessness, chest pain, haemoptysis, wheeze. Specify acuity of onset, change over time, change in symptoms with location/work.
- ***Systemic symptoms:*** weight loss, malaise, night sweats.
- ***Occupational exposure:*** previous occupations, with a specific focus on exposure to asbestos, to organic materials such as hay, mushrooms, or cotton, or to animals. Establish any relationship of symptoms to work.

- ***Smoking:*** duration of smoking, number of cigarettes smoked per day, cessation attempts including nicotine replacement and e-cigarettes.
- ***Illicit drug use:*** especially smoked cannabis.
- ***Family history:*** emphysema, bronchiectasis, or cystic fibrosis.
- ***Childhood history:*** preterm delivery, childhood infections such as whooping cough or measles.
- ***Travel history:*** including risk factors for tuberculosis.

Examination

General
- Posture (lying flat or sitting forward)
- Colour (cyanosed)
- Pain on breathing (pleuritic)?
- Fever
- Cachexia
- Anemia

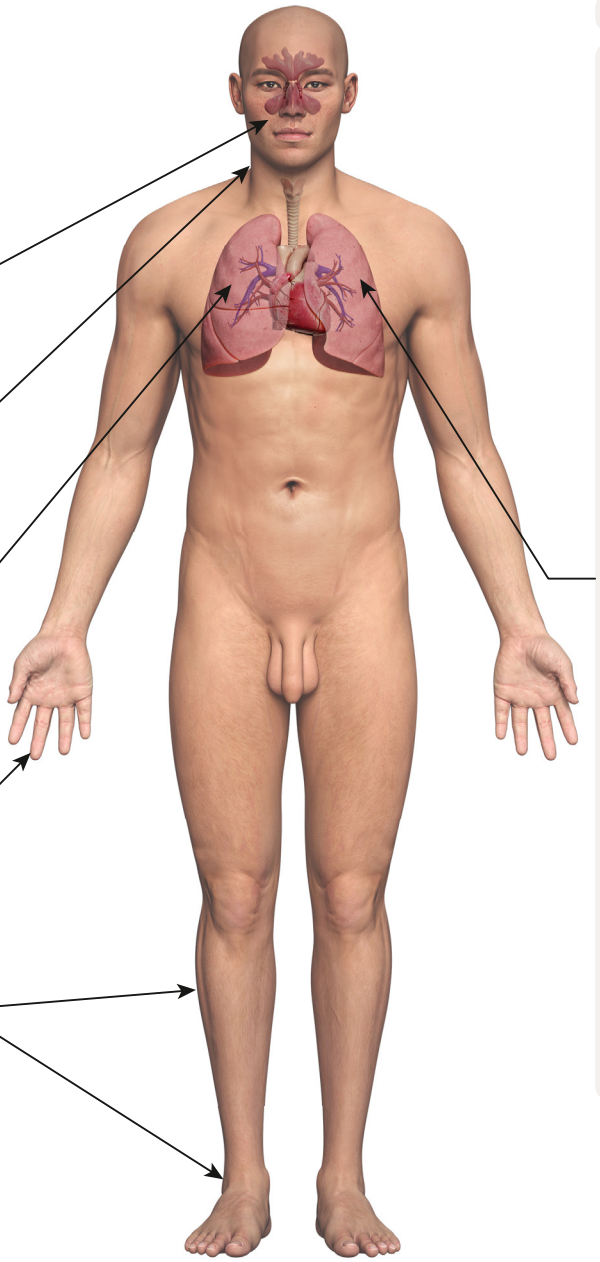
Face
- Horner's syndrome
- Pursed lips
- Beak-like nose
- Central cyanosis

Neck
- Jugular venous pressure (raised, distended)
- Lymphadenopathy

Sputum
- Volume (↑↑ bronchiectasis)
- Mucopurulent (infection)
- Blood-stained (cancer, pulmonary embolism, tuberculosis, bronchiectasis)

Hands
- Clubbing
- Tar staining
- Peripheral cyanosis
- CO_2 retention flap
- Bounding pulse

Peripheries
- Deep vein thrombosis
- Oedema



Chest

Inspection
- Colour
- Breathlessness
- Deformity
- Scars
- Symmetry of movement
- Abdominal paradox
- Hyperinflation
- Prominent veins
- Intercostal indrawing
- Scoliosis
- Respiratory rate
- Use of accessory muscles

Palpation
- Tracheal position
- Cricosternal distance
- Supraclavicular fossa nodes
- Apex beat
- Tenderness
- Liver
  - Position (low lying)
  - Enlargement

Percussion
- Dull
- 'Stony dull' (fluid)
- Hyper-resonant

Auscultation
- Normal = vesicular
- Bronchial breathing
- Wheeze
  - Monophonic (single large airway obstruction)
  - Polyphonic (narrowing of small airways)
- Crackles
  - Coarse
  - Fine, late inspiratory
- Pleural rub
- Vocal resonance

Box 20.1 Core content in respiratory medicine

The 'top 10' respiratory conditions
- Acute and chronic cough
- Airflow obstruction: asthma, chronic obstructive pulmonary disease (COPD), obstructive sleep apnoea
- Pneumonia
- Pulmonary tuberculosis
- Pneumothorax
- Pleural effusion
- Lung cancer
- Bronchiectasis
- Interstitial lung disease

The 'top 10' concepts in respiratory medicine
- Smoking cessation
- Atopy
- The 'treatment ladder' approach to asthma and COPD
- Self-management of chronic conditions
- Admission avoidance
- Home oxygen therapy
- Respiratory failure
- Diagnosis and staging of lung cancer
- Multidisciplinary cancer care
- Management of chest drains

The 'top 10' respiratory medications
- Oxygen
- β_2 agonists
- Antimuscarinics
- Oral and inhaled corticosteroids
- Combination inhalers
- Antihistamines
- Leukotriene receptor antagonists
- Mucolytics
- Monoclonal antibodies
- Antifibrotics

FUNCTION OF THE RESPIRATORY SYSTEM

The principal functions of the respiratory system are to extract oxygen from the external environment and to dispose of carbon dioxide. This requires the lungs to function as efficient bellows, bringing fresh air to the alveoli, and expelling used air at an appropriate rate. Gas exchange is achieved by exposing thin-walled capillaries to alveolar gas and matching ventilation to blood flow through the pulmonary capillary bed. The excretion of carbon dioxide by the lungs is part of acid–base homeostasis.

The lungs expose a large surface area of body tissue to the external environment to achieve gas exchange, hence they can be damaged by dusts, gases, and infective agents. Host defence is therefore a key priority achieved by a combination of structural and immunological mechanisms. The pulmonary circulation acts as a reservoir that can allow response to increased oxygen demands in exercise. The lungs act as a filtration system for small pulmonary emboli. The pulmonary circulation also plays a role in innate immunity, for example, in de-priming neutrophils. Finally, the lungs have a role in speech, as the passage of air through the vocal cords is necessary for phonation.

Anatomy

Nose, pharynx, and larynx

See Chapter 21.

Trachea, bronchi, and bronchioles

The trachea is 10–12 cm in length. It lies slightly to the right of the midline and divides at the carina into right and left main bronchi. The carina lies under the junction of the manubrium sterni and the second right costal cartilage. The right main bronchus is more vertical than the left and inhaled material is therefore more likely to end up in the right lung.

The right main bronchus divides into the upper lobe bronchus and the intermediate bronchus, which further subdivides into the middle and lower lobe bronchi. On the left, the main bronchus divides into upper and lower lobe bronchi only. Each lobar bronchus further divides into segmental and subsegmental bronchi. There are about 25 divisions in all between trachea and alveoli.

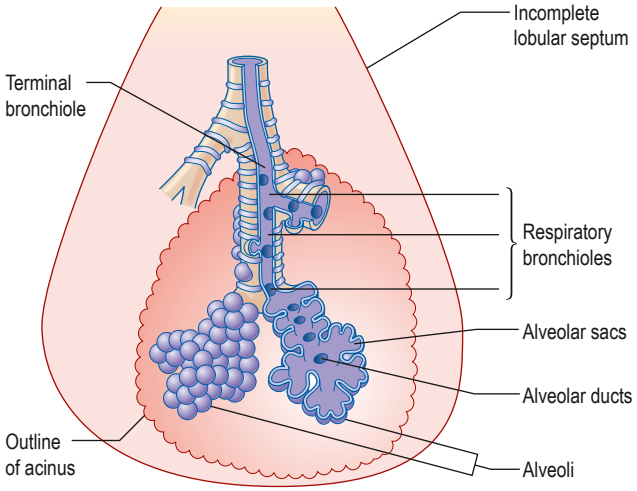


Fig. 20.1 Branches of a terminal bronchiole ending in the alveolar sacs.

The first seven divisions are bronchi that have:
- walls consisting of cartilage and smooth muscle
- an epithelial lining with cilia and goblet cells
- submucosal mucus-secreting glands
- endocrine cells.

The next 16–18 divisions are bronchioles that have:
- no cartilage and a muscular layer that progressively becomes thinner
- a single layer of ciliated cells but very few goblet cells
- granulated Clara cells that produce a surfactant-like substance.

The ciliated epithelium is a key defence mechanism. Each cell bears approximately 200 cilia beating at 1000 beats/min in organized waves of contraction. Mucus, which contains macrophages, cell debris, inhaled particles, and bacteria, is moved by the cilia towards the larynx at about 1.5 cm/min (the 'mucociliary escalator'; see later).

The bronchioles divide within the acinus into smaller respiratory bronchioles that have alveoli arising from the surface (Fig. 20.1). Each respiratory bronchiole supplies approximately 200 alveoli via alveolar ducts. The term 'small airways' refers to bronchioles of less than 2 mm; the average lung contains about 30 000 of these.

Alveoli

There are approximately 300 million alveoli in each lung. Their total surface area is 40–80 m^2. The epithelial lining consists mainly of ***type I pneumocytes*** (Fig. 20.2). These cells have an extremely thin layer of cytoplasm, which only offers a thin barrier to gas exchange. Type I cells are connected to each other by tight junctions that limit the movements of fluid in and out of the alveoli. Alveoli are not completely airtight; holes in the alveolar wall allow communication between alveoli of adjoining lobules (pores of Kohn). ***Type II pneumocytes*** are slightly more numerous than type I but cover less of the epithelial lining. They are found generally in the borders of the alveolus and contain distinctive lamellar vacuoles, which are the source of surfactant. Type I pneumocytes are derived from type II cells. Large alveolar macrophages are present within the alveoli and assist in defence of the lung.

Lungs

The lungs are separated into lobes by invaginations of the pleura, which are often incomplete. The right lung has three lobes, whereas

the left lung has two. The positions of the oblique fissures and the right horizontal fissure are shown in Fig. 20.3. The upper lobe lies mainly in front of the lower lobe and therefore physical signs on the right side in the front of the chest are due to lesions of the upper lobe or the middle lobe.

Pleura

The pleura is a layer of connective tissue covered by a simple squamous epithelium. The visceral pleura covers the surface of the lung, lines the interlobar fissures, and is continuous at the hilum with the parietal pleura, which lines the inside of the hemithorax. At the hilum, the visceral pleura continues alongside the branching bronchial tree for some distance before reflecting back to join the parietal pleura. In health, the pleurae are in apposition, apart from a small quantity of lubricating fluid.

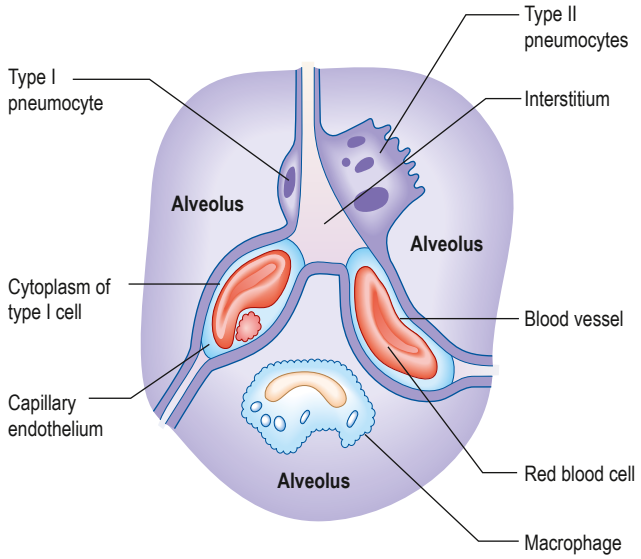


Fig. 20.2 The structure of alveoli, showing the pneumocytes and capillaries.

Diaphragm

The diaphragm is covered by parietal pleura above and peritoneum below. The diaphragmatic muscle fibres arise from the lower ribs and insert into the central tendon. Motor and sensory nerve fibres go separately to each half of the diaphragm via the phrenic nerves. Fifty percent of muscle fibres are of the slow-twitch type with low glycolytic capacity, and are relatively resistant to fatigue.

Pulmonary vasculature and lymphatics

The lung has a dual blood supply made up of deoxygenated blood cells from the right ventricle via the pulmonary artery into which oxygen from the alveoli diffuse, and oxygenated blood via the bronchial artery circulation. The pulmonary artery divides to accompany the bronchi. The arterioles accompanying the respiratory bronchioles are thin-walled and contain little smooth muscle. The pulmonary venules drain laterally to the periphery of the lobules, pass centrally in the interlobular and intersegmental septa, and eventually join to form the four main pulmonary veins. The bronchial circulation arises from the descending aorta. These bronchial arteries supply tissues down to the level of the respiratory bronchiole. The bronchial veins drain into the pulmonary veins, forming a physiological shunt.

Lymphatic channels lie in the interstitial space between the alveolar cells and the capillary endothelium of the pulmonary arterioles.

The tracheobronchial lymph nodes are arranged in five main groups: pulmonary, bronchopulmonary, subcarinal, superior tracheobronchial, and paratracheal. For practical purposes, these form a continuous network of nodes from the lung substance up to the trachea.

Nerve supply to the lung

The innervation of the lung remains incompletely understood. Parasympathetic and sympathetic fibres from the vagus and sympathetic chain, respectively, accompany the pulmonary arteries and the airways. Airway smooth muscle is innervated by vagal afferents, postganglionic muscarinic vagal efferents, and vagally derived non-adrenergic non-cholinergic (NANC) fibres that use a range of neurotransmitters. Three muscarinic receptor subtypes have been identified: M_1 receptors on parasympathetic ganglia, a smaller number of M_2 receptors on muscarinic nerve terminals, and M_3 receptors on airway smooth muscle. The parietal pleura is innervated

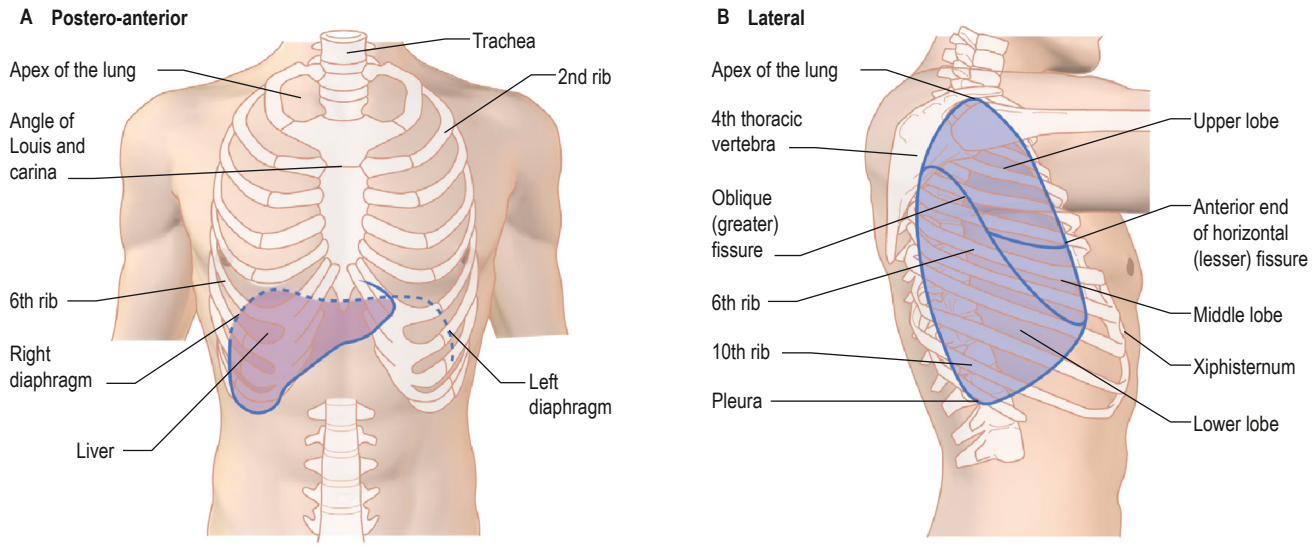


Fig. 20.3 Surface anatomy of the chest. (A) Anterior. (B) Lateral.

from intercostal and phrenic nerves, but the visceral pleura has no innervation.

Further reading

Guyton and Hall. *Textbook of Medical Physiology*. 14th edn. Elsevier: 2020.

Physiology

Nose

The major functions of nasal breathing are:

- to heat and moisten the air
- to remove particulate matter.

Nasal secretions contain immunoglobulin A (IgA) antibodies, lysozyme, and interferons. In addition, the cilia of the nasal epithelium move the mucous gel layer rapidly back to the oropharynx, where it is swallowed. Bacteria have little chance of settling in the nose. Mucociliary protection is less effective against viral infections because viruses bind to receptors on epithelial cells. The majority of rhinoviruses bind to an adhesion molecule, intercellular adhesion molecule 1 (ICAM-1), which is shared by neutrophils and eosinophils. Many noxious gases, such as sulphur dioxide, are almost completely removed by nasal breathing.

Breathing

Lung ventilation consists of:

- the mechanical process of inspiration and expiration
- the control of respiration appropriate for metabolic needs.

Mechanical process

The lungs have an inherent elastic property that causes them to tend to collapse away from the thoracic wall, generating a negative pressure within the pleural space. The strength of this retractive force relates to the volume of the lung: at higher lung volumes, the lung is stretched more, and a greater negative intrapleural pressure is generated. ***Lung compliance*** is a measure of the relationship between this retractive force and lung volume. At the end of a quiet expiration, the retractive force exerted by the lungs is balanced by the tendency of the thoracic wall to spring outwards. At this point, respiratory muscles are resting. The volume of air remaining in the lung after a quiet expiration is called the ***functional residual capacity*** (***FRC***, see Fig. 20.13).

Inspiration from FRC is an active process: a negative intrapleural pressure is created by descent of the diaphragm and movement of the ribs upwards and outwards through contraction of the intercostal muscles. During tidal breathing in healthy individuals, inspiration is almost entirely due to contraction of the diaphragm. More vigorous inspiration requires the use of accessory muscles of ventilation: sternomastoid and scalene. Respiratory muscles are similar to other skeletal muscles but less prone to fatigue. However, inspiratory muscle fatigue contributes to respiratory failure in patients with severe chronic airflow limitation and in primary neurological and muscle disorders. At rest or during low-level exercise, expiration is passive and results from the natural tendency of the lung to collapse. Forced expiration involves activation of accessory muscles, chiefly those of the abdominal wall, which help to push up the diaphragm.

Control of respiration

Coordinated respiratory movements result from rhythmical discharges arising in an anatomically ill-defined group of interconnected neurons in the reticular substance of the brainstem, known as the respiratory centre. Motor discharges from the respiratory centre travel via the phrenic and intercostal nerves to the respiratory musculature. Ventilation is controlled by a combination of neurogenic and chemical factors (Fig. 20.4). In healthy individuals, the main driver for respiration is the arterial pH, which is closely related to the partial pressure of carbon dioxide in arterial blood. Oxygen concentrations in arterial blood are usually above the level that triggers respiratory drive. Typical reference values are shown in Box 20.2.

Breathlessness on physical exertion is normal and is not a symptom unless the level of exertion is very light, such as walking slowly. Surveys of healthy Western populations reveal that over 20% of the general population report breathlessness on relatively minor exertion. Although breathlessness is a very common symptom, the sensory and neural mechanisms underlying it remain obscure. The sensation of breathlessness is derived from at least three sources:

- ***changes in lung volume:*** sensed by receptors in thoracic wall muscles signalling changes in length
- ***tension developed by contracting muscles:*** sensed by Golgi tendon organs
- ***central perception of effort***

Airways of the lungs

From the trachea to the periphery, the airways decrease in size and increase in number. Overall, the cross-sectional area available for airflow increases as the total number of airways increases. The airflow rate is greatest in the trachea and slows progressively towards the periphery since the velocity of airflow depends on the cross-sectional area. In the terminal airways, gas flow occurs solely by diffusion. The resistance to airflow is very low (0.1–0.2 kPa/L in a normal tracheobronchial tree), steadily increasing from the small to the large airways.

Airways expand as lung volume increases. At full inspiration (***total lung capacity***, TLC) they are 30%–40% larger in calibre than at full expiration (***residual volume***, RV). In chronic obstructive pulmonary disease (COPD) the small airways are narrowed but this can be partially compensated by breathing closer to TLC.

Control of airway tone

Bronchomotor tone is maintained by vagal efferent nerves and can be reduced by atropine and β-adrenoceptor agonists. Adrenoceptors on the surface of bronchial muscles respond to circulating catecholamines; there is no direct sympathetic innervation. Airway tone shows a circadian rhythm, which is greatest at 4 am and lowest in mid-afternoon. Tone can be increased transiently by inhaled stimuli acting on epithelial nerve endings, which trigger reflex bronchoconstriction via the vagus. Stimuli include cigarette smoke, solvents, inert dust, and cold air. Airway responsiveness to these stimuli increases following respiratory tract infections, even in healthy subjects. In asthma, the airways are very irritable: as the circadian rhythm remains the same, asthmatic symptoms are usually worse in early morning.

Airflow

Movement of air through the airways results from a difference between atmospheric pressure and pressure in the alveoli; alveolar pressure is negative in inspiration and positive in expiration. During quiet breathing, the pleural pressure is negative throughout the breathing cycle. With vigorous expiratory efforts (e.g., cough) the pleural pressure becomes positive (up to 10 kPa). This compresses the central airways, but the smaller airways do not close off because the driving pressure for expiratory flow (alveolar pressure) is also increased.

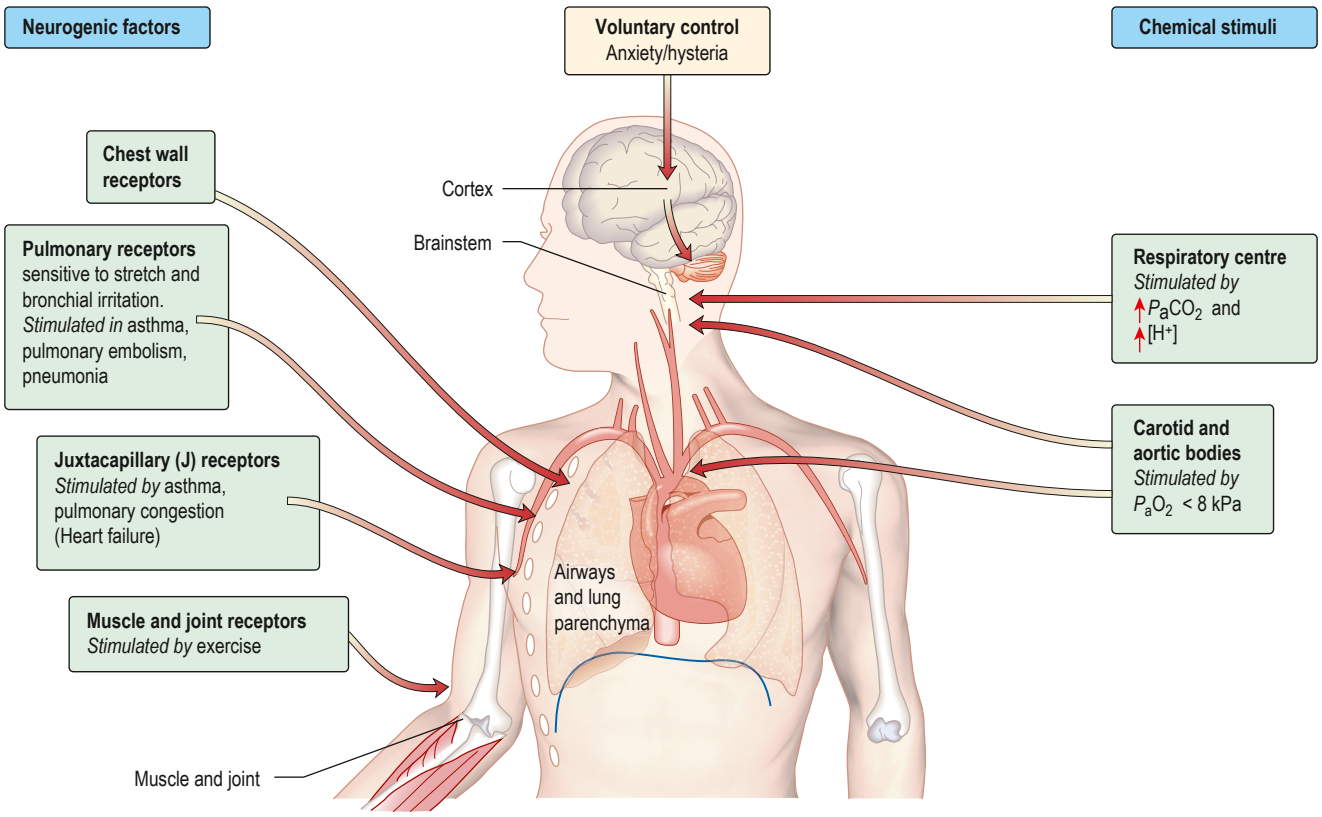


Fig. 20.4 Chemical and neurogenic factors in the control of ventilation. The strongest stimulant to ventilation is a rise in P_aCO_2, which increases $[H^+]$ in cerebrospinal fluid. Sensitivity to this may be lost in chronic obstructive pulmonary disease, when hypoxaemia becomes the chief stimulus to respiratory drive; oxygen treatment may therefore reduce respiratory drive and lead to a further rise in P_aCO_2. An increase in $[H^+]$ due to metabolic acidosis, as in diabetic ketoacidosis, will increase ventilation with a fall in P_aCO_2, causing deep sighing (Kussmaul) respiration. The respiratory centre is depressed by severe hypoxaemia and sedatives (e.g., opiates).

Box 20.2 Normal values for respiratory physiology

In a typical, non-pregnant adult at rest:

- Pulmonary blood flow: 5 L/min carrying 11 mmol/min (250 mL/min) of O_2 to tissues
- Ventilation: 6 L/min removing 9 mmol/min (200 mL/min) of CO_2 from the body
- Pressure of oxygen in arterial blood (P_aO_2): 11–13 kPa
- Pressure of carbon dioxide in arterial blood (P_aCO_2): 4.8–6.0 kPa

Flow–volume loops

The relationship between maximal flow rates and lung volume is demonstrated by the maximal flow–volume (MFV) loop (Fig. 20.5A). In subjects with healthy lungs, maximal flow rates are rarely achieved, even during vigorous exercise. However, in patients with severe COPD, limitation of expiratory flow occurs even during tidal breathing at rest (Fig. 20.5B). To increase ventilation, these patients breathe at higher lung volumes and allow more time for expiration, both of which reduce the tendency for airway collapse. To compensate, they increase flow rates during inspiration, where there is relatively less flow limitation.

The volume that can be forced in from the RV in 1 second (FIV_1) will always be greater than that which can be forced out from TLC in 1 second (FEV_1). Thus, the ratio of FEV_1 to FIV_1 is less than 1. The only exception to this is when there is significant obstruction to the airways outside the thorax, such as tracheal tumour or retrosternal goitre. Expiratory airway narrowing is prevented by tracheal resistance, and expiratory airflow is more effort-dependent. During forced inspiration, this same resistance causes negative intraluminal pressure and tracheal compression by surrounding atmospheric pressure. Inspiratory flow thus becomes less effort-dependent, and the ratio of FEV_1 to FIV_1 is greater than 1. This phenomenon, and the characteristic flow–volume loop, are diagnostic of extrathoracic airways obstruction (Fig. 20.5C).

Ventilation and perfusion relationships

For optimum gas exchange there is a match between (V) and perfusion of the alveoli (Q). In reality there is variation in the $\left(\dot{V}/\dot{Q}\right)$ ratio in both normal and diseased lungs (Fig. 20.6). In the normal lung, ventilation, and perfusion are greater at the bases than at the apices but the gradient for perfusion is steeper, so the net effect is that ventilation exceeds perfusion towards the apices while perfusion exceeds ventilation at the bases. Other causes of mismatch include direct shunting of deoxygenated blood through the lung without passing through alveoli (e.g., the bronchial circulation) and areas of lung deadspace that receive no blood (e.g., bullae, areas of underperfusion during acceleration and deceleration in aircraft and high-performance cars). Increased physiological shunt results in arterial hypoxaemia since it is not possible to compensate for blood being under-oxygenated by increasing ventilation of well-perfused areas. Increased physiological deadspace increases the work of breathing but has less impact on gas exchange in the normal lung as perfused alveoli remain well-ventilated. In lung disease, this compensation

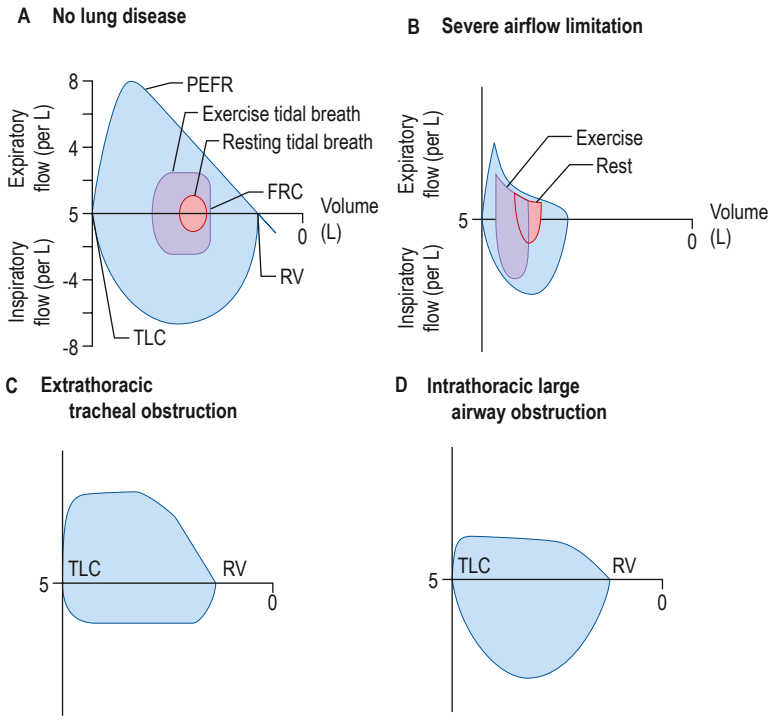


Fig. 20.5 Flow–volume loops. Maximal flow–volume loops, showing the relationship between maximal flow rates on expiration and inspiration. Flow–volume loops during tidal breathing at rest (starting from the functional residual capacity, FRC) and during exercise are also shown **(A)** ***Normal subject***. The highest flow rates are achieved when forced expiration begins at total lung capacity (TLC) and represents the peak expiratory flow rate (PEFR). As air is blown out of the lung, the flow rate decreases until no more air can be forced out, a point known as the residual volume (RV). Because inspiratory airflow is dependent only on effort, the shape of the maximal inspiratory flow–volume loop is different, and inspiratory flow remains at a high rate throughout the manoeuvre. **(B) Severe airflow limitation**. The ventilatory demands of exercise cannot be met, greatly reducing effort tolerance. **(C–D) Flow–volume loops in large airway (tracheal) obstruction**, showing plateauing of maximal expiratory flow. **Extrathoracic tracheal obstruction (C)** has proportionally greater reduction of maximal inspiratory flow rate. In **intrathoracic large airway obstruction (D)** the expiratory plateau is more pronounced.

may not occur, leading to increased alveolar and arterial PCO_2 (P_ACO_2 and P_aCO_2, respectively) and hypoxaemia, which cannot be compensated for by increasing ventilation.

Hypoxaemia occurs more readily than hypercapnia because of the different ways in which oxygen and carbon dioxide are carried in the blood. Carbon dioxide is carried in three forms (in bicarbonate, carbamino compounds, and as simple solution) and the volume carried is proportional to the partial pressure of CO_2. Oxygen is carried in chemical combination with haemoglobin in the red blood cells, with a non-linear relationship between the volume carried and the partial pressure. Alveolar hyperventilation reduces the alveolar PCO_2 and diffusion leads to a proportional fall in the carbon dioxide content of the blood. However, as haemoglobin is already saturated with oxygen, increasing the alveolar PO_2 through hyperventilation will not increase blood oxygen content. This means that hypoxaemia due to physiological shunting cannot be compensated for by hyperventilation.

In individuals who have mild degrees of mismatch, the P_aO_2 and P_aCO_2 will still be normal at rest. Increasing the requirements for gas exchange by exercise will widen the mismatch and the P_aO_2 will fall. Mismatch is by far the most common cause of arterial hypoxaemia.

Alveolar stability

Pulmonary alveoli are polygonal spaces within a sponge-like matrix. Surface tension acting at the curved internal surface tends to cause the alveoli to decrease in size. The surface tension within the alveoli would make the lungs extremely difficult to distend, if not for the presence of surfactant, which reduces surface tension so that alveoli remain stable.

Defence mechanisms of the respiratory tract

Pulmonary disease often results from a failure of normal lung host defence mechanisms (Fig. 20.7). These can be divided into physical, physiological, humoral, and cellular mechanisms.

Physical and physiological mechanisms include

- **Humidification:** prevents dehydration of the epithelium.
- **Particle removal:** >90% of particles of >10 μm in diameter are removed in the nostril or nasopharynx. This includes most pollen grains (typically >20 μm). Particles between 5 and 10 μm become impacted at the carina. Particles capable of reaching the deep lung are those in the 1–5 μm range as they remain airborne.
- **Particle expulsion:** facilitated by coughing, sneezing, or gagging.
- **Respiratory tract secretions:** the mucus of the respiratory tract is gelatinous, consisting of water and highly glycosylated proteins (mucins). The mucus forms a thick gel that is relatively impermeable to water and floats on a liquid or sol layer found around the cilia of the epithelial cells (see Fig. 20.7). The gel layer is secreted from goblet cells and mucus glands as distinct

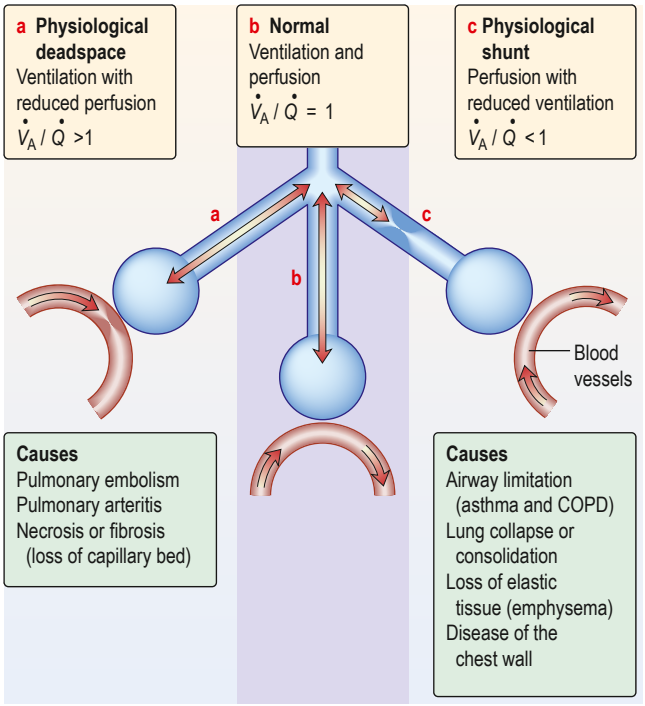


Fig. 20.6 Relationships between ventilation and perfusion: the alveolar–capillary interface. The *centre* **(b)** shows ***normal ventilation*** and perfusion. On the left **(a)** there is a ***block in perfusion (physiological deadspace)***, while on the right **(c)** there is ***reduced ventilation (physiological shunting)***.

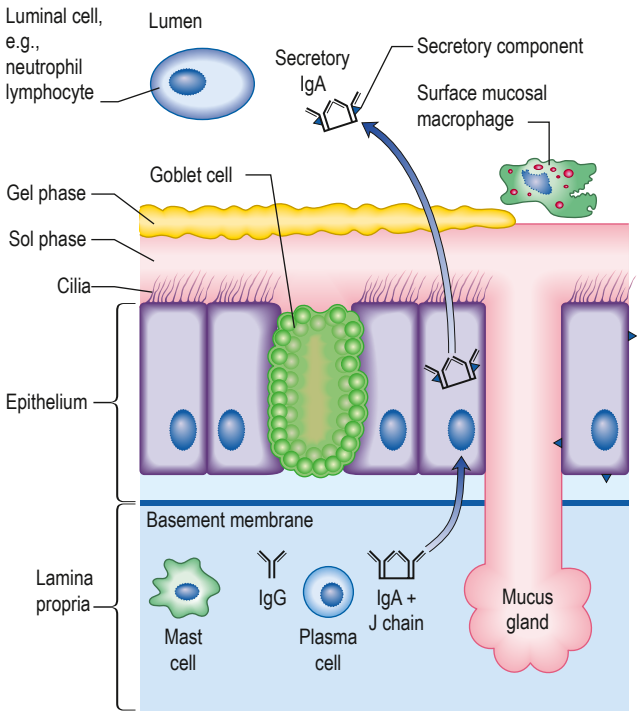


Fig. 20.7 Defence mechanisms present at the epithelial surface.

globules that coalesce increasingly in the central airways to form a continuous mucus blanket. In addition to the mucins, the gel contains various antimicrobial molecules (lysozyme, defensins), specific antibodies (IgA), and cytokines, which are secreted by cells in airways and incorporated into the mucus gel. Bacteria, viruses, and other particles become trapped in the mucus and are either inactivated or simply expelled before they can do any damage. Under normal conditions the tips of the cilia engage with the undersurface of the gel phase and by coordinated movement push the mucus blanket upwards and outwards to the pharynx, where it is either swallowed or coughed up. A major long-term effect of cigarette smoking is reduced mucociliary transport contributing to recurrent infection and prolonged contact with carcinogens. Air pollutants, local and general anaesthetics, and products of bacterial and viral infection also reduce mucociliary clearance. Congenital defects in mucociliary transport (cystic fibrosis [CF] and immotile cilia syndrome) lead to recurrent infections and bronchiectasis.

The respiratory microbiome

There is resident bacterial flora in the lung that is similar to that of the mouth. The composition of the respiratory microbiome is determined by three factors:

- ***microbial immigration:*** by inhalation or microaspiration from the gastrointestinal tract
- ***local growth conditions:*** for bacteria, e.g., temperature, pH, nutrients, concentration and activation of local inflammatory cells, epithelial cell interactions
- ***microbial elimination:*** the mucociliary escalator, coughing, innate, and adaptive humoral mechanisms.

These factors change in acute and chronic lung conditions when there is an increase in pathological bacteria. This respiratory tract dysbiosis causes dysregulation of local immune responses (e.g., prolonged inflammation), which favours growth of bacteria. The composition of the bacterial microbiome in different conditions might favour exacerbation of disease.

Humoral and cellular mechanisms

Non-specific soluble factors

- ***Alpha-antitrypsin*** (α_1-antiprotease) in lung secretions is derived from plasma. It inhibits chymotrypsin and trypsin, and neutralizes proteases including neutrophil elastase.
- ***Antioxidant defences*** include enzymes such as superoxide dismutase and low-molecular-weight antioxidant molecules (ascorbate, urate) in the epithelial lining fluid.
- ***Lysozyme*** is a bactericidal enzyme found in granulocytes.
- ***Lactoferrin*** is synthesized from epithelial cells and neutrophil granulocytes and has bactericidal properties.
- ***Interferons*** are produced by most cells in response to viral infection and are potent modulators of lymphocyte function.
- ***Complement*** in secretions is derived from plasma. In association with antibodies, it plays a major role in cytotoxicity.
- ***Surfactant protein A*** (SPA) is one of four species of surfactant protein that opsonize bacteria or particles, enhancing phagocytosis by macrophages.
- ***Defensins*** are bactericidal peptides present in the azurophil granules of neutrophils.
- ***Dimeric secretory IgA*** targets specific antigens.

Innate and adaptive immunity

These mechanisms act as a defence against microbes, inorganic substances such as asbestos, particulate matter such as dust, and other antigens. They aid opsonization so that macrophages can better ingest foreign material.

With infection, neutrophils migrate out of pulmonary capillaries into the air spaces and phagocytose and kill microbes with

antimicrobial proteins (lactoferrin), degradative enzymes (elastase), and oxidant radicals. In addition, neutrophil extracellular traps ensnare and kill extracellular bacteria. Neutrophils also generate a variety of mediators, such as tumour necrosis factor-alpha (TNF-α), interleukin 1 (IL-1), and chemokines that attract additional inflammatory cells as part of adaptive immunity.

Microbes are detected by host cells by pattern recognition receptors such as toll-like receptors. These act via nuclear factor-kappa B (NF-κB) transcription factors in the epithelial cells to produce adhesion molecules, chemokines, and colony-stimulating factors to initiate inflammation. Inflammation is necessary for innate immunity and host defence but can lead to lung damage; there is a fine line between defence and injury.

Further reading

Hartl D, et al. Innate immunity of the lung: from basic mechanisms to translational medicine. *Innate Immun* 2018; 10:487–501.
Whiteside SA, McGinniss JE, Collman RG. The lung microbiome: progress and promise. *J Clin Invest* 2021; 131(15):e150473.

CLINICAL APPROACH TO THE PATIENT WITH RESPIRATORY DISEASE

Clinical features of respiratory disease

Nasal symptoms

Nasal symptoms are extremely common; 'runny nose' (rhinorrhoea), nasal blockage, and sneezing can be caused by allergic rhinitis (see Chapter 21) and by viral infection. Nasal secretions are usually thin and runny in allergic rhinitis, but thicker and discoloured in viral infections. Nose bleeds and blood-stained nasal discharge are common and rarely indicate serious pathology, unless associated with nasal obstruction and/or pain (see Chapter 21). Nasal polyps typically present with nasal blockage and loss of smell.

Cough

Cough is the most common symptom of lower respiratory tract disease. It is caused by mechanical or chemical stimulation of cough receptors in the epithelium of the pharynx, larynx, trachea, bronchi, and diaphragm. Impulses travel via an afferent pathway to the cough centre in the medulla, with efferent signalling to the expiratory musculature. Smokers often have a morning cough with little sputum. A productive cough is a key feature of chronic bronchitis. Dry coughing, particularly at night, can be a symptom of asthma or acid reflux. Cough also occurs in asthma after mild exertion or forced expiration. Cough can have no definable pathology; psychological causes are postulated with limited evidence.

A worsening cough is the most common presenting symptom of lung cancer. The normal explosive character of the cough is lost when a vocal cord is paralyzed, usually due to cancer infiltrating the left recurrent laryngeal nerve. Cough can be accompanied by stridor in whooping cough or in laryngeal or tracheal obstruction.

Sputum

Approximately 100 mL of mucus is produced daily in a healthy, non-smoking individual. This flows up the airways, through the larynx, and is then swallowed. Excess mucus is expectorated as sputum. Cigarette smoking is the most common cause of excess mucus production.

Mucoid sputum is clear and white but can contain black specks from inhalation of carbon. Yellow or green sputum is due to the presence of cellular material, including bronchial epithelial cells, neutrophils, or eosinophil granulocytes. Yellow sputum is not necessarily due to infection, since granulocytes in the sputum, as seen in asthma, give the same appearance as neutrophils. The production of large quantities of yellow or green sputum is characteristic of bronchiectasis.

Haemoptysis

Haemoptysis (blood-stained sputum) varies from small streaks of blood to massive haemorrhage. The most common cause of mild haemoptysis is infection (**Box 20.3**), particularly in exacerbations of COPD, but this should not be assumed without investigation. Other common causes are pulmonary infarction (e.g., secondary to pulmonary embolism), bronchial carcinoma, and tuberculosis. Pink, frothy sputum is seen in pulmonary oedema, while in bronchiectasis the blood is often mixed in purulent sputum. Massive haemoptysis (>200 mL of blood in 24 h) is usually due to bronchiectasis or tuberculosis, but can be seen in later stages of lung cancer.

Although a diagnosis can often be made from a chest radiograph (e.g., bronchiectasis, tuberculosis), a normal film does not exclude serious disease. However, if the chest radiograph is normal, CT scanning and bronchoscopy are diagnostic in only about 5% of patients with haemoptysis.

Breathlessness

Dyspnoea (breathlessness) is a sense of awareness of increased respiratory effort that is unpleasant and recognized by the patient as inappropriate. Patients often experience tightness in the chest, which must be differentiated from angina. (**Box 20.4**). The degree of breathlessness should be assessed in relation to lifestyle. For example, a moderate degree of breathlessness will be disabling if the patient has to climb many flights of stairs to reach home. Breathlessness can be graded using the Medical Research Council grading of dyspnoea (**Box 20.5**).

Tachypnoea and ***hyperpnoea*** are, respectively, an increased rate of breathing and an increased level of ventilation. These may be appropriate responses (e.g., during exercise).

Hyperventilation is inappropriate over-breathing. This may occur at rest or on exertion, and results in a lowering of alveolar and arterial PCO_2.

Orthopnoea is breathlessness on lying down. While it is classically linked to heart failure, the weight of abdominal contents

? Box 20.3 Causes of haemoptysis

- Malignant and benign lung tumours, including lung metastases
- Pulmonary infection, including bacterial pneumonia, tuberculosis, lung abscess, and fungal infection
- Bronchiectasis, including cystic fibrosis
- Pulmonary emboli
- Congestive heart failure
- Pulmonary fibrosis
- Pulmonary vasculitis, e.g., Goodpasture's syndrome, microscopic polyangiitis
- Severe pulmonary hypertension
- Arteriovenous malformation
- Chest trauma and foreign body
- Endometriosis
- Anticoagulation, coagulopathy (look for underlying pathology)
- Drugs, e.g., cocaine, thrombolytics

Box 20.4 Causes of breathlessness

Acute
- Airways obstruction
- Anaphylaxis
- Asthma
- Pneumothorax
- Pulmonary embolus
- Myocardial infarction
- Pulmonary oedema
- Arrhythmia
- Anxiety

Subacute
- Pneumonia
- Exacerbation of COPD
- Angina
- Cardiac tamponade
- Metabolic acidosis
- Pain
- Pontine haemorrhage

Chronic
- COPD
- Pleural effusion
- Malignancy
- Chronic pulmonary emboli
- Restrictive lung disease
- Congestive cardiac failure
- Valvular dysfunction
- Cardiomyopathy
- Anaemia
- Neuromuscular disorders
- Deconditioning
- Breathing pattern disorder with no underlying pathology

Box 20.5 Medical Research Council grading of dyspnoea (breathlessness scale)

1. Not troubled by breathlessness except on strenuous exercise
2. Short of breath when hurrying or walking up a slight hill
3. Walks slower than contemporaries on the level because of breathlessness, or has to stop for breath when walking at own pace
4. Stops for breath after about 100 m or after a few minutes on the level
5. Too breathless to leave the house, or breathless when dressing or undressing

pushing the diaphragm up into the thorax may contribute; thus patients with diaphragm paralysis may present with symptoms of orthopnoea. Such patients may also become breathless on bending over.

Paroxysmal nocturnal dyspnoea describes acute episodes of breathlessness at night, typically due to heart failure.

Wheezing

Wheezing is a common symptom resulting from airflow limitation due to any cause. It is not diagnostic of asthma; other causes include vocal cord dysfunction, bronchiolitis, and COPD. Conversely, wheeze may be absent in early-stage asthma.

Chest pain

Common chest pain reported in respiratory disease is a localized, sharp pain termed 'pleuritic'. It is made worse by deep breathing or coughing. Localized anterior chest wall pain with tenderness of a costochondral junction is caused by costochondritis. Shoulder tip pain suggests irritation of the diaphragmatic pleura, and subdiaphragmatic pathology should be considered for example, liver abscess. Central chest pain radiating to the neck and arms is likely to be cardiac. Retrosternal soreness occurs in tracheitis. Malignant invasion of the chest wall causes a constant, severe, dull pain.

Examination of the respiratory system

Examination of the nose and throat are detailed in Chapter 21.

Inspection

Observe the patient as they enter the room or move around the bed. Are they breathless at rest/exertion? Do they have a cough

Box 20.6 Signs to look for in the hands

- Clubbing
- Pallor
- Warm, well-perfused palms (CO_2 retention)
- Cyanosis
- Flap
- Tremor
- Tobacco staining
- Bruising and/or thin skin
- Pulse: rate and character

or wheeze? Assess mental alertness, cyanosis, use of accessory muscles, and chest wall shape, deformity, scars, and movement on both sides. Kyphosis and scoliosis of the spine can cause asymmetry of the chest. CO_2 intoxication causes coarse tremor or flap of the outstretched hands. Prominent veins on the chest may imply obstruction of the superior vena cava. The patient's face may reveal signs of anaemia, or there may be a Horner's syndrome due to a Pancoast tumour.

Respiratory rate and rhythm may be altered; the normal respiratory rate is 14–16 breaths per minute. Tachypnoea is an increased respiratory rate. Apnoea is the absence of breathing; some patients have episodes of apnoea during sleep.

Hands should be inspected for evidence of tobacco staining on the fingers, clubbing, or a fine tremor (Box 20.6).

Cyanosis is a dusky discoloration of the skin and mucous membranes, due to the presence of more than 50 g/L of desaturated haemoglobin. When it has a central cause, cyanosis is visible on the tongue (especially the underside) and lips. Patients with central cyanosis will also be cyanosed peripherally. Peripheral cyanosis without central cyanosis is caused by a reduced peripheral circulation and is noted on the fingernails and skin of the extremities, with associated coolness of the skin.

Finger clubbing is present when the normal angle between the base of the nail and the nail fold is lost (Fig. 20.8). The base of the nail is fluctuant owing to increased vascularity, and there is an increased curvature of the nail in all directions, with expansion of the end of the digit (Box 20.7). Clubbing is not a feature of uncomplicated COPD.

Palpation and percussion

The position of the trachea and apex beat should be checked. The supraclavicular fossa, cervical chains, and axilla are examined for enlarged lymph nodes. The distance between the sternal notch and the cricoid cartilage (3–4 finger-breadths in full expiration) is reduced in patients with severe airflow limitation reflecting hyperinflated lungs. Chest expansion should be checked; a tape measure may be used if precise or serial measurements are needed, such as in ***ankylosing spondylitis***. Local discomfort over the sternochondral joints suggests costochondritis. In rib fractures, compression of the chest laterally and anteroposteriorly produces localized pain. On percussion, liver dullness is usually detected anteriorly at the level of the sixth right-sided rib. Liver and cardiac dullness disappear when the lungs are over-inflated.

Auscultation

The patient is asked to take deep breaths through the mouth. Inspiration should be longer than expiration. Normal breath sounds are caused by turbulent flow in the larynx and sound harsher anteriorly over the upper lobes (particularly on the right). Healthy lungs filter out most of the high-frequency component,

and the resulting sounds are termed vesicular. If the lung is consolidated or collapsed, the high-frequency hissing components of breath are not attenuated and can be heard as 'bronchial breathing'. Similar sounds may be heard over areas of localized fibrosis or bronchiectasis. Bronchial breathing is accompanied by whispering pectoriloquy (whispered, high-pitched sounds heard distinctly through a stethoscope).

Fig. 20.8 Clubbing deformity. The finger on the right is clubbed compared with the normal-shaped finger on the left. *(From Hochberg MC, Gravallese EM, Silman AJ et al.* Rheumatology, *7th edn. Elsevier Inc.; 2019; Fig. 213.2.)*

? Box 20.7 Some causes of finger clubbing

Respiratory
- Bronchial carcinoma, including hypertrophic pulmonary osteoarthropathy
- Chronic suppurative lung disease:
  - Bronchiectasis
  - Lung abscess
  - Empyema
- Idiopathic lung fibrosis
- Pleural and mediastinal tumours (e.g., mesothelioma)
- Cryptogenic organizing pneumonia

Cardiovascular
- Cyanotic heart disease
- Subacute infective endocarditis
- Atrial myxoma

Miscellaneous
- Congenital
- Cirrhosis
- Inflammatory bowel disease
- Thyroid acropachy

Added sounds

Wheeze. Wheeze results from vibrations in the collapsible part of the airways when the large and medium-sized bronchi become constricted. It is usually heard during expiration and is commonly, but not invariably, present in asthma and COPD. In acute severe asthma, wheeze may not be heard, as airflow may be insufficient to generate the sound. Wheezes may be monophonic (single large airway obstruction) or polyphonic (narrowing of many small airways). An end-inspiratory wheeze or 'squeak' may be heard in obliterative bronchiolitis.

Crackles. Brief crackling sounds are probably produced by opening of previously closed bronchioles; early inspiratory crackles are associated with diffuse airflow limitation, while late inspiratory crackles are characteristically heard in pulmonary oedema, lung fibrosis, and bronchiectasis.

Pleural rub. A creaking or groaning sound (said to sound like a foot crunching through fresh snow), which is usually well localized. It indicates inflammation and roughening of pleural surfaces, which no longer glide silently over one another, and is heard in lung infection and consolidation.

Vocal resonance. Healthy lung attenuates high-frequency notes, as compared to the lower-pitched components of speech. Low-frequency sounds such as 'ninety-nine' are well transmitted across healthy lung (vocal resonance/fremitus). Tactile vocal fremitus is the palpation of sound vibration, usually by placing the edge of the hand on the chest wall. Consolidated lung transmits sound more easily than air, thereby resulting in increased vocal resonance/fremitus, whereas pleural fluid severely dampens or obliterates the vibrations altogether. Consolidated lung also transmits high frequencies, so that spoken word has a bleating quality and whispered (high-pitched) speech can be clearly heard over consolidated areas compared to healthy lung (whispered pectoriloquy).

Stridor: see Chapter 21.

Respiratory examination findings in different diseases are detailed in **Box 20.8**.

i Box 20.8 Physical signs of respiratory disease

Pathological process	Mediastinal displacement	Percussion note	Breath sounds	Vocal resonance	Added sounds
Consolidation (i.e., lobar pneumonia)	None	Dull	Bronchial	Increased	Fine crackles
Collapse					
Major bronchus	Towards lesion	Dull	Diminished or absent	Reduced or absent	None
Peripheral bronchus	Towards lesion	Dull	Bronchial	Decreased	Fine crackles
Fibrosis					
Localized	Towards lesion	Dull	Bronchial	Increased	Coarse crackles
Diffuse (e.g., idiopathic lung fibrosis)	None	Normal	Vesicular	Increased	Fine crackles
Pleural effusion (>500 mL)	Away from lesion (in massive effusion)	Stony dull	Vesicular reduced or absent	Reduced or absent	None
Large pneumothorax	Away from lesion	Normal or hyperresonant	Reduced or absent	Reduced or absent	None
Asthma	None	Normal	Vesicular Prolonged expiration	Normal	Expiratory polyphonic wheeze
Chronic obstructive pulmonary disease	None	Normal	Vesicular Prolonged expiration	Normal	Expiratory polyphonic wheeze and coarse crackles
Bronchiectasis	None	Normal	Vesicular	Normal	Coarse crackles

Additional bedside tests

Review measured oxygen saturations and the concentration of additional oxygen received. Inspect any sputum pots. Many patients with respiratory disease have airflow limitation, so airflow should be routinely measured using a peak flow meter or spirometer, which is a more useful assessment of airflow limitation than any physical sign.

Investigation of respiratory disease

Imaging

Imaging is essential in the investigation of most chest symptoms. Some diseases, such as tuberculosis or lung cancer, may be undetectable on clinical examination but may be obvious on a chest radiograph. Conversely, asthma or chronic bronchitis may be associated with a normal chest radiograph. Try to obtain previous images for comparison.

Plain radiograph imaging of the chest is addressed in Chapter 8.

Computed tomography

Computed tomography (CT) provides excellent images of the lungs and mediastinal structures (Fig. 20.9). It is essential in staging bronchial carcinoma by demonstrating tumour size, nodal involvement, metastases, and invasion of mediastinum, pleura, or chest wall. CT-guided needle biopsy allows samples to be obtained from peripheral masses. Staging scans should assess liver and adrenals, which are common sites for metastatic disease. Mediastinal structures are shown more clearly after injecting intravenous contrast medium.

High-resolution CT (HRCT) samples lung parenchyma with 1–2 mm thickness scans at 10–20 mm intervals and is used to assess diffuse inflammatory and infective parenchymal processes. HRCT scanning does not require intravenous contrast. It is valuable in:

- evaluation of diffuse disease of the lung parenchyma, including sarcoidosis, hypersensitivity pneumonitis (HP), occupational lung disease, and any other form of interstitial pulmonary fibrosis
- diagnosis of bronchiectasis, having a sensitivity and specificity of >90%
- distinction of emphysema from both diffuse parenchymal lung disease and pulmonary vascular disease as a cause of a low gas transfer factor with otherwise normal lung function
- suspected opportunistic lung infection in immunocompromised patients
- diagnosis of lymphangitis carcinomatosa.

Multidetector CT scanners can produce detailed images in two or three dimensions in any plane. This allows spaced, contiguous, and/or overlapping HRCT images to be reconstructed. Pulmonary nodules and airway disease can be easily defined.

CT pulmonary angiography (CTPA) is used to investigate for pulmonary embolism and enables visualization of the pulmonary arteries. Contrast is injected and images are taken in timed fashion while the contrast agent is in the pulmonary circulation.

Magnetic resonance imaging

Magnetic resonance imaging (MRI) with electrocardiography (ECG) gating allows accurate imaging of the heart and aortic aneurysms. MRI has been used in staging lung cancer and assessing tumour invasion in the mediastinum, chest wall, and at the lung apex because it produces good images in the sagittal and coronal planes. Vascular structures can be clearly differentiated, as flowing blood produces a signal void on MRI.

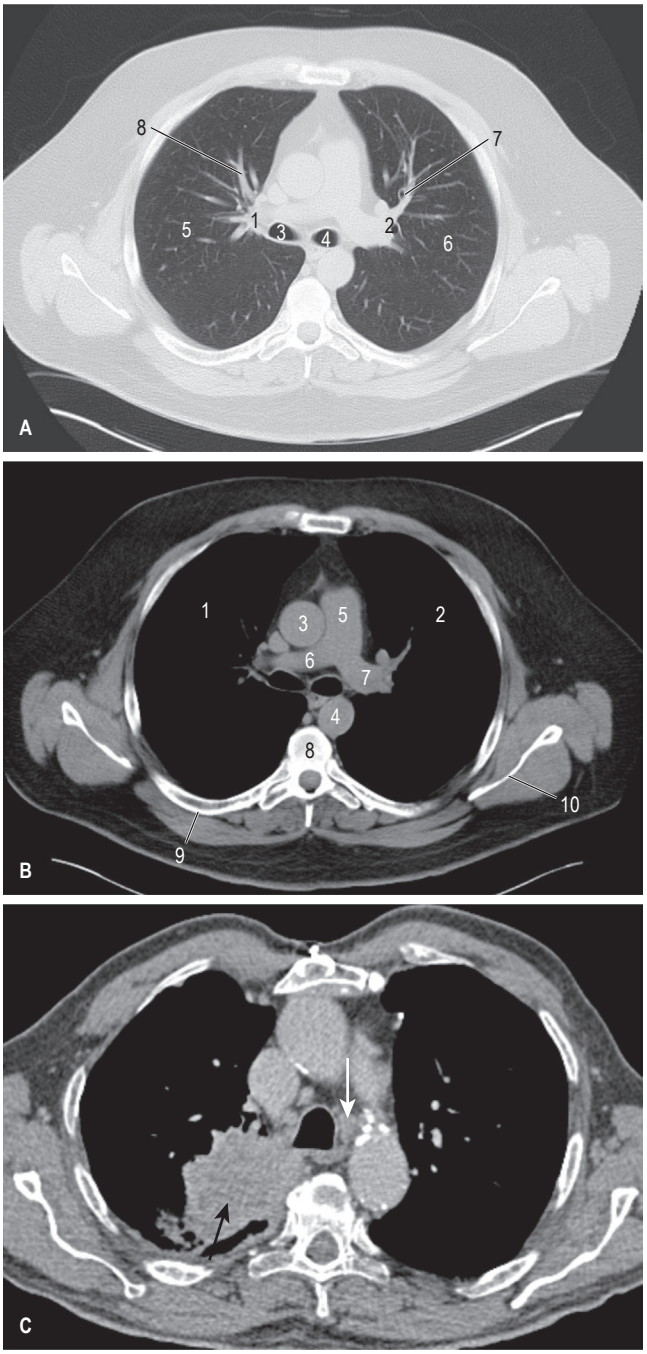


Fig. 20.9 Computed tomography scans of the lung. **(A)** Axial CT image of the thorax on lung settings. (1) Right hilum; (2) left hilum; (3) right main bronchus; (4) left main bronchus; (5) right lung; (6) left lung; (7) bronchus; (8) blood vessel. **(B)** Axial CT image of the thorax on soft tissue settings. (1) Right lung; (2) left lung; (3) ascending aorta; (4) descending aorta; (5) pulmonary trunk; (6) right pulmonary artery; (7) left pulmonary artery; (8) vertebra; (9) rib; (10) scapula. **(C)** CT chest demonstrates a right upper lobe carcinoma that is invading the mediastinum *(black arrow)* and prominent mediastinal lymph nodes *(white arrow)*.

Positron emission tomography–computed tomography

PET-CT combines a CT scan and positron emission tomography. Positron-emitting isotopes such as ^{18}fluorodeoxyglucose (^{18}FDG) are used as contrast and are taken up rapidly by metabolically

active tissue such as lung cancers. PET imaging is used for lung cancer staging prior to curative treatment such as surgery or radiotherapy. FDG-PET can also be useful in determining an appropriate site for a biopsy and can assist with differentiating benign from malignant tumours. Inflammatory lesions may also be FDG-avid.

Scintigraphic imaging

Isotopic ventilation/perfusion (V/Q) lung scans were widely used for the detection of pulmonary emboli but are now performed less often, owing to the widespread use of CTPA. They may be used in pregnancy, where plain chest radiography is normal and other diagnoses are considered unlikely, to reduce radiation exposure to the maternal breast.

Single photon emission computed tomography (SPECT) is increasingly used. It is a three-dimensional nuclear medicine imaging technique that combines information from scintigraphy with that of CT. This offers better detail, contrast, and spatial information than isolated planar nuclear imaging.

Ultrasound

Two-dimensional transthoracic ultrasound is a technique used for assessing a pleural effusion. Ultrasound confirms the presence of pleural fluid and provides details about the nature of the effusion, such as whether it is a simple pleural effusion (single collection), heavily loculated with adhesions or organized (more gelatinous). Ultrasound assists in determining the best site for aspiration and it is recommended that any invasive pleural procedure, such as pleural aspiration and intercostal chest drain placement, is performed with ultrasound screening. Ultrasound-guided biopsy is used for lung masses that abut the pleura or pleural masses, if appropriate. It is also used in bronchoscopy (endobronchial ultrasound, EBUS) to stage and sample mediastinal lymph nodes (see 'Fibreoptic bronchoscopy' section).

Respiratory function tests

Airflow limitation can be assessed by a variety of tests that have good intra-subject repeatability (**Box 20.9**). Results must be compared with predicted values for healthy subjects, as reference intervals vary with sex, age, height, and ethnicity. Even then, there remains considerable variation between healthy individuals of the same size and age; the standard deviation for PEFR is approximately 50 L/min, and for FEV_1 approximately 0.4 L. Repeated measurements of lung function are useful for assessing disease progression.

Spirometry

The patient takes maximum inspiration followed by a forced expiration (for as long as possible) into the spirometer. The spirometer measures the 1-second forced expiratory volume (FEV_1) and the total volume of exhaled gas (forced vital capacity, FVC). Both FEV_1 and FVC are related to height, age, sex, and ethnicity. These measures differentiate between an obstructive and a restrictive pattern of respiratory compromise (**Fig. 20.10**; **Box 20.10**).

In chronic airflow limitation (particularly COPD and asthma), the TLC is usually increased, yet there is nearly always some reduction in the FVC. This is because collapse of small airways causes obstruction to airflow before the normal RV is reached. Trapping of air within the lung is a characteristic feature of these diseases.

Peak expiratory flow rate

Peak expiratory flow rate (PEFR) is a simple, cheap, and reproducible test. Subjects take a full inspiration to TLC and then blow out forcefully into the peak flow meter (**Fig. 20.11**). The best of three attempts is recorded. PEFR is mainly dependent on flow in larger airways and may be falsely reassuring in moderate airflow limitation. PEFR is mainly used to diagnose and monitor asthma, including response to treatment. Regular measurements of peak flow rates on

Box 20.9 Respiratory function and exercise tests

Test	Use	Advantages	Disadvantages
PEFR	Monitoring changes in airflow limitation in asthma	Portable Can be used at the bedside	Effort-dependent Poor measure of chronic airflow limitation
FEV, FVC, FEV_1/FVC	Assessment of airflow limitation (the best single test)	Reproducible Relatively effort-independent	Bulky equipment but smaller portable machines available
Flow–volume curves	Assessment of flow at lower lung volumes Detection of intra- and extrathoracic large-airway obstruction (e.g., tracheal stenosis, tumour)	Flow–volume curve patterns for different diseases	Sophisticated equipment needed for full test but expiratory loop possible with compact spirometry
Airways resistance	Assessment of airflow limitation	Sensitive	Technique difficult to perform
Lung volumes	Differentiation between restrictive and obstructive lung disease	Effort-independent, complements FEV_1	Sophisticated equipment needed
Gas transfer	Assessment and monitoring of extent of interstitial lung disease and emphysema	Non-invasive	Sophisticated equipment needed
Blood gases	Assessment of respiratory failure	Can detect early lung disease when measured during exercise	Invasive
Pulse oximetry	Postoperative, sleep studies and respiratory failure	Continuous monitoring Non-invasive	Measures saturation only
Exercise test (6-min walk)	Practical assessment for disability and effects of therapy	No equipment required	Time-consuming Learning effect At least two walks required
Cardiorespiratory assessment	Early detection of lung/heart disease Fitness assessment	Differentiates breathlessness due to lung or heart disease	Sophisticated equipment needed

CT, Computed tomography; *FEV*, forced expiratory volume; *FEV_1*, forced expiratory volume in 1 sec; *FVC*, forced vital capacity; *PEFR*, peak expiratory flow rate.

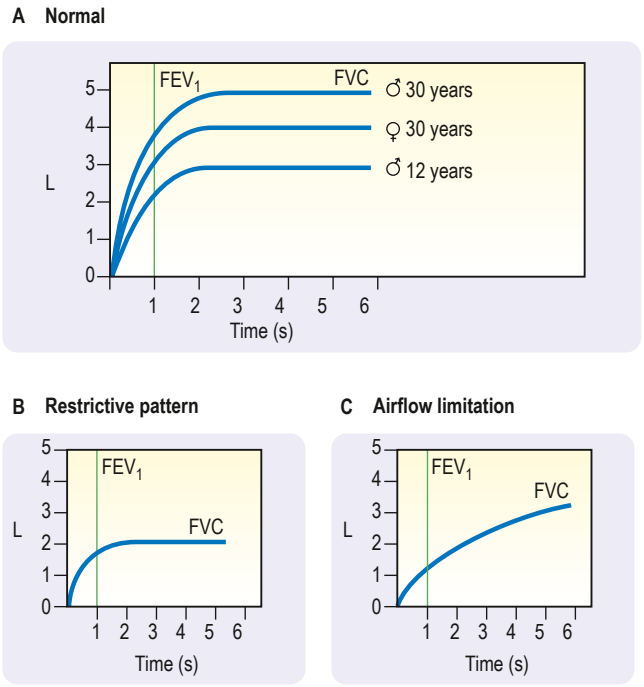


Fig. 20.10 Spirometry: volume–time curves. **(A)** Normal patterns for age and sex. **(B)** Restrictive pattern (FEV_1 and FVC reduced). **(C)** Airflow limitation (only FEV_1 reduced). *FEV_1,* Forced expiratory volume in 1 sec; *FVC,* forced vital capacity.

Box 20.10 FEV_1/FVC ratio

- ***Normal:*** approximately 75%
- ***Airflow obstruction:*** reduced
- ***Restrictive lung disease:*** normal or increased

FEV_1, Forced expiratory volume in 1 sec; *FVC,* forced vital capacity.

waking, during the afternoon, and before going to bed demonstrate the wide diurnal variations in airflow limitation that characterize asthma and allow objective assessment of response to treatment (Fig. 20.12).

Plethysmography

Measurement of airways resistance in a body box is more sensitive than PEFR but the equipment is expensive and necessary manoeuvres are exhausting for many patients with chronic airflow limitation.

Flow–volume loops

Plotting flow rates against expired volume (flow–volume loops, see Fig. 20.5) shows the site of airflow limitation within the lung. At the start of expiration from TLC, maximum resistance is from large airways, which affects flow rate for the first 25% of the curve. Obstruction of large airways includes tracheal narrowing for example, due to tumour or retrosternal goitre. As air is exhaled, lung volume reduces and flow rate becomes dependent on resistance of smaller airways. In COPD, which mainly affects the smaller airways, expiratory flow rates at 50% or 25% of the vital capacity are disproportionately reduced when compared with flow rates at larger lung volumes.

Lung volumes

The subdivisions of lung volume are shown in Fig. 20.13. Tidal volume and vital capacity can be measured using a simple spirometer but alternative techniques are needed to measure TLC and RV. TLC is measured by inhaling air containing a known concentration of helium and measuring its dilution in the exhaled air. TLC measurements using this technique are inaccurate if there are large cystic spaces in the lung because helium cannot diffuse into them. Under these circumstances the thoracic gas volume can be measured more accurately using body plethysmograph (see earlier). RV can be calculated by subtracting the vital capacity from the TLC. The

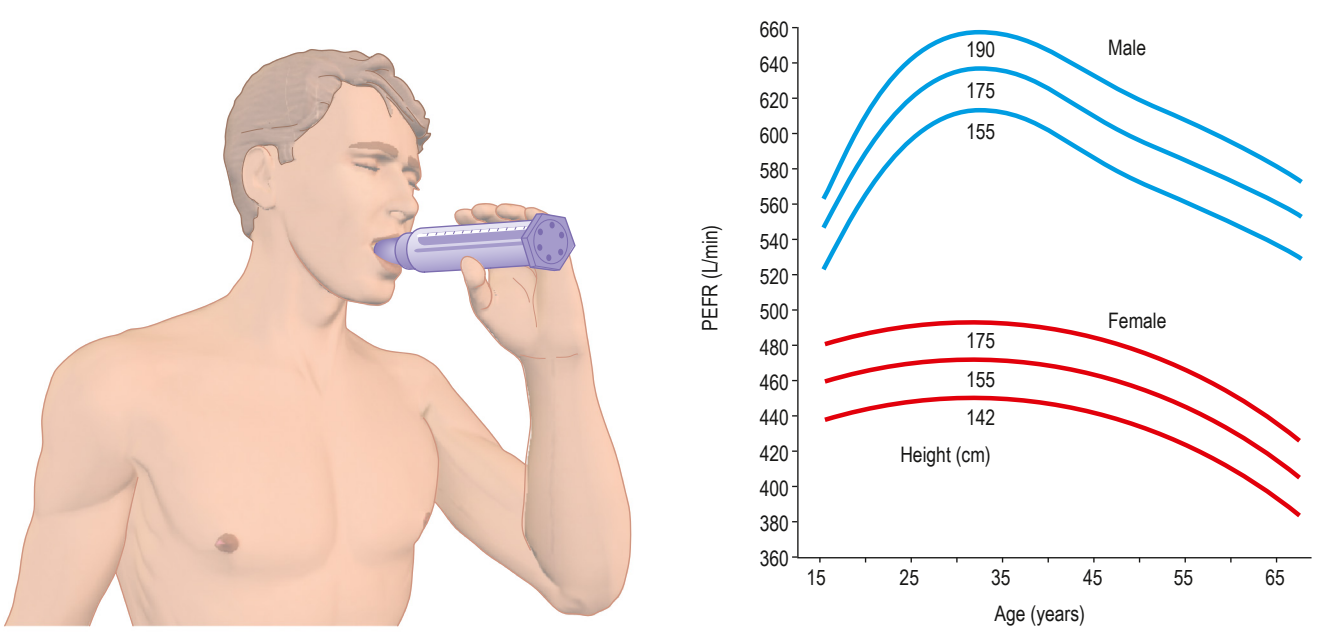


Fig. 20.11 Peak flow measurement. **(A)** Peak flow measurement: lips should be tight around the mouthpiece. **(B)** References values: vary with age, height, and sex. *PEFR,* Peak expiratory flow rate.

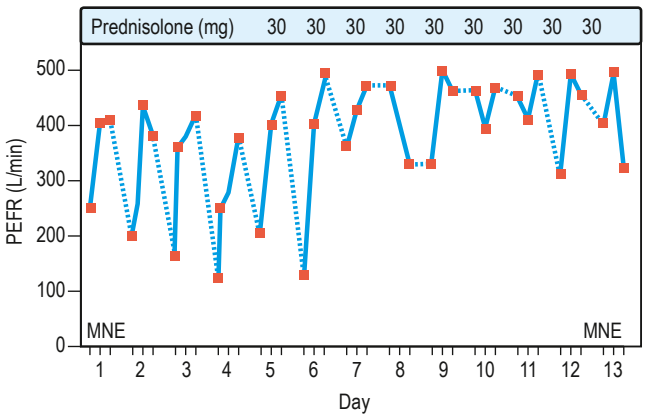
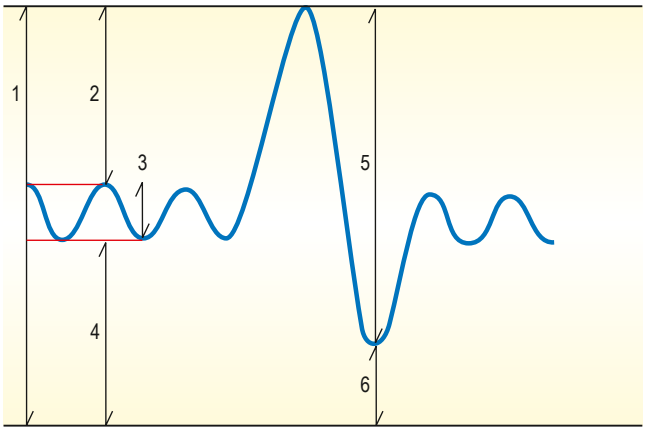


Fig. 20.12 Diurnal variability in peak expiratory flow rate (PEFR) in asthma, showing the effect of 30 mg prednisolone.
MNE, Morning, noon, evening.



1 Total lung capacity
2 Inspiratory reserve volume
3 Tidal volume
4 Functional residual capacity
5 Vital capacity
6 Residual volume

Fig. 20.13 The subdivisions of the lung volume.

difference between measurements made by these two methods shows the extent of non-communicating air space within the lungs.

Gas transfer

In normal lungs, transfer factor reflects how efficiently oxygen diffuses from alveolar air into blood, and depends on the thickness of the alveolar–capillary membrane. In lung disease, this diffusing capacity (D_{CO}) is also affected by ventilation and perfusion. Carbon monoxide is used as a surrogate marker to measure this, as it has similar diffusion to oxygen. A low concentration of carbon monoxide is inhaled, and rate of absorption is calculated. To control for differences in lung volume, the uptake of carbon monoxide is expressed relative to lung volume as a transfer coefficient (K_{CO}). Gas transfer is reduced in patients with severe emphysema and fibrosis but also in heart failure and anaemia. Although relatively non-specific, gas transfer is particularly useful in the detection and monitoring of diseases affecting the lung parenchyma (e.g., idiopathic pulmonary fibrosis [IPF], sarcoidosis, and asbestosis). The K_{CO} is raised in pulmonary haemorrhage.

Peripheral oxygen saturation (S_pO_2) can be continuously measured using an oximeter with either ear or finger probe. Pulse oximetry has become an essential part of routine monitoring of patients in hospital and clinics. It is also useful in exercise testing and reduces the need to measure arterial blood gases.

Measurement of blood gases

Measurement of the partial pressures of oxygen and carbon dioxide in arterial blood is essential in managing respiratory failure, when repeated measurements guide therapy (see **Boxes 20.2**, **11.11**, and **20.18**).

Fractional exhaled nitric oxide (FeNO)

Nitric oxide (NO) is produced by the bronchial epithelium and increases in eosinophilic asthma and other forms of airway inflammation. A portable device can be used to measure the concentration of exhaled nitric oxide and guide therapy in asthma.

Six-minute walk test

In this validated test, the distance walked in metres in 6 minutes is recorded. Oxygen saturations and heart rate are also measured.

Cardiopulmonary exercise testing

This provides a functional assessment of cardiopulmonary reserve and provides information about cardiorespiratory and metabolic muscle function. It is usually performed on a treadmill or a cycle ergometer. Oxygen consumption, carbon dioxide production, and ventilation can be calculated. Indications include the investigation of unexplained breathlessness, establishing prognosis in respiratory illnesses, and predicting risk in perioperative assessment.

Nocturnal polygraphy

This multichannel sleep study records heart rate, oxygen saturations, nasal flow, body position, and thoracic and abdominal wall movement. It is useful in the investigation of sleep-disordered breathing.

Haematological and biochemical tests

It is useful to measure:

- ***haemoglobin:*** detects anaemia or polycythaemia
- ***packed cell volume:*** secondary polycythaemia occurs with chronic hypoxia
- ***routine biochemistry:*** abnormalities may be detected in lung cancer and infection
- ***D-dimer:*** a negative test may be helpful when there is low clinical suspicion of pulmonary embolism.

Other blood investigations sometimes required include α_1-antitrypsin concentration, *Aspergillus* antibodies, viral and mycoplasma serology, autoantibody profiles, immunoglobulin concentrations, and specific IgE measurements.

Sputum

Sputum should be inspected for colour:

- Yellowish green indicates inflammation (infection or allergy).
- Blood suggests bronchiectasis, lung cancer, or pulmonary infarction.

Gram stain and culture are rarely helpful in upper respiratory tract infections or in acute or chronic bronchitis. Conversely, they are of value in:

- pneumonia
- tuberculosis (a specific request for acid-fast bacilli [AFB] is required)
- bronchiectasis
- cystic fibrosis.

Sputum cytology

This may be an adjunct in the management of asthma, and is used in research studies. Advantages are speed, cost, and its non-invasive nature. A reliable cytologist is needed. Sputum can be induced by inhalation of nebulized hypertonic saline (5%), although better samples can be obtained by bronchoscopy and bronchial washings.

Pleural aspiration

Diagnostic aspiration may be necessary to determine the aetiology of a pleural effusion, especially if unilateral. This is usually done under ultrasound guidance. A needle is inserted under local anaesthesia through an intercostal space towards the top of the fluid area identified on ultrasound. Fluid is withdrawn and the presence of any blood is noted. Samples are sent for cytology and biochemical analysis (protein estimation, lactate dehydrogenase [LDH], and bacteriological examination, including culture and Ziehl–Neelsen/auramine staining for tuberculosis). Therapeutic aspiration of a large pleural effusion to help relieve breathlessness.

Pleural biopsy

A pleural biopsy may be necessary for a unilateral exudative pleural effusion or suspicious pleural thickening. Pleural biopsies may be obtained using CT or ultrasound guidance. An alternative technique is video-assisted thoracoscopy (VATS). A scope is introduced into the pleural space, allowing direct visualization and biopsy of the parietal pleura. A pleural effusion can be drained during this procedure.

Intercostal drain placement

An intercostal chest drain is an indwelling drain placed in the pleural space using ultrasound guidance (see Box 11.7). Occasionally, a ***long-term pleural drain*** may be needed for recurrent malignant effusions.

Fibreoptic bronchoscopy

This endoscopic procedure performed using local anaesthesia and sedation, allows direct visualization of the endobronchial tree down to the subsegmental level (Fig. 20.14).

Indications for bronchoscopy (Box 20.11) are:

- ***Visualization and biopsy of an endobronchial lesion*** i.e., suspected of malignancy.
- ***Collapsed lung or lobe*** to determine the nature of the obstruction. Causes include aspirated foreign body or mucus plug (removed during the bronchoscopy by suction, forceps, or a dormier basket) and endobronchial tumours (which can be biopsied). Airway patency may be restored by cryotherapy, laser therapy, or stenting.
- ***Microbiological sampling*** in the context of unresolved infection or suspected tuberculosis. Distal lesions may be sampled by washing or blind brushing.
- ***Cytological assessment*** of bronchoalveolar lavage samples may be helpful to diagnose malignancy. Differential cell counts may be useful in differentiating parenchymal lung disease.
- ***Diagnosis of diffuse inflammatory processes*** by bronchoalveolar lavage and transbronchial biopsy. Diagnostic yield is best in sarcoidosis, lymphangitis carcinomatosa, and HP. (Other fibrotic lung diseases may require open or thoracoscopic lung biopsy.)

Endobronchial ultrasound (EBUS) A dedicated scope is used with an ultrasound probe at the distal end, which allows detailed visualization of the mediastinal nodes, under local anaesthesia and sedation. The lymph nodes can be aspirated for cytological and microbiological examination. EBUS is performed when CT imaging shows enlarged mediastinal lymph nodes, for which the differential diagnosis includes sarcoid, tuberculosis, lung cancer, lymphoma, and metastatic disease.

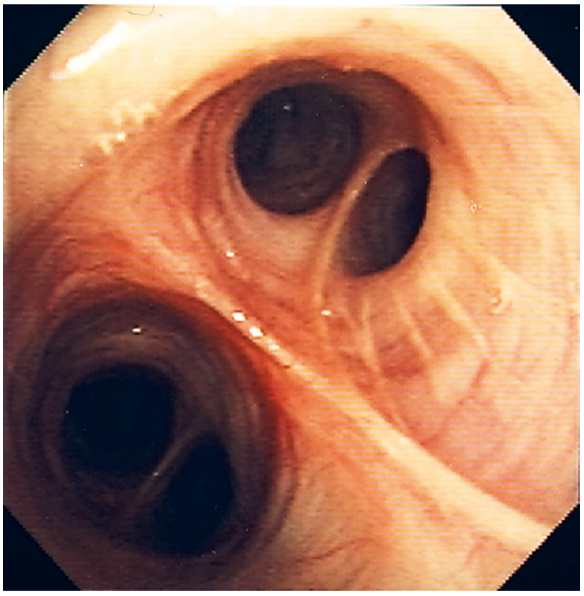


Fig. 20.14 Normal endobronchial appearances seen at fibreoptic bronchoscopy.

Box 20.11 Fibreoptic bronchoscopy

Indications

- Lesions requiring biopsy seen on chest radiograph
- Haemoptysis
- Stridor
- Positive sputum cytology for malignant cells with no chest radiograph abnormality
- Collection of bronchial secretions for bacteriology, especially tuberculosis
- Recurrent laryngeal nerve paralysis of unknown aetiology
- Infiltrative lung disease (to obtain a transbronchial biopsy)
- Investigation of collapsed lobes or segments and aspiration of mucus plugs

Disadvantages

- Sedation required
- Minor and transient cardiac dysrhythmias in up to 40% of patients on passage of the bronchoscope through the larynx. Monitoring is required
- Oxygen supplementation required in patients with P_aO_2 <8 kPa
- Risk of pneumothorax
- Massive bleeding may occur after biopsy, e.g., vascular lesions or carcinoid tumours. Rigid bronchoscopy may be required to allow adequate access for haemostasis

Navigational bronchoscopy is an advanced bronchoscopy technique which enables the operator to reach peripheral pulmonary lesions not accessible by traditional bronchoscopy. It utilizes a traditional bronchoscope, electromagnetic navigation, and software with preloaded CT images which enable guided targeting of small peripheral lesions. It is an alternative to a CT-guided lung biopsy and may be preferable in selected cases.

Mediastinoscopy

This surgical procedure is performed under general anaesthesia. A rigid scope is passed from just above the sternum into the anterior

mediastinum. Mediastinoscopy is used in the diagnosis and staging of nodal disease. Use has declined due to availability of EBUS.

Video-assisted thoracoscopic lung biopsy

VATS lung biopsy has largely replaced open thoracotomy when lung biopsy is required.

Skin-prick tests

Allergen solutions are placed on the skin (usually the volar surface of the forearm) and the epidermis is broken using a 1 mm tipped lancet. A separate lancet should be used for each allergen. Positive and negative controls are also tested. Weal development reveals sensitivity with diameter measured after 10 minutes. A weal of at least 3 mm diameter is regarded as positive, provided the control test is negative. History informs interpretation. Skin tests are not affected by bronchodilators or corticosteroids, but antihistamines should be discontinued at least 48 hours before testing.

Bronchial provocation testing

Airway hyper-responsiveness (AHR) is a characteristic feature of asthma and can be demonstrated by asking patients to inhale gradually increasing concentrations of histamine or methacholine. This induces transient airflow limitation in susceptible individuals (approximately 20% of the population). Severity is graded according to the provocation dose (PD) or concentration (PC) of the agonist that produces a 20% fall in FEV_1. Patients with asthma respond to very low doses of methacholine. AHR can also be assessed by exercise testing or inhalation of cold/dry air, mannitol, or hypertonic saline.

Further reading

Bohadana A, Izbicki G, Kraman SS. Fundamentals of lung auscultation. *N Engl J Med* 2014; 370:744–751.
Padilla-Ortiz AL, Ibarra D. Lung and heart sounds analysis: state-of-the-art and future trends. *Crit Rev Biomed Eng* 2018; 46(1):33–52.

DISEASES OF THE UPPER RESPIRATORY TRACT

Rhinitis

Rhinitis is defined clinically as sneezing attacks, nasal discharge, or blockage occurring for more than an hour on most days.

Viral rhinitis

The common cold (acute coryza) is a highly infectious illness (see Chapter 21) caused by several respiratory viruses including rhinoviruses (most common), coronaviruses, and adenoviruses. Infectivity from personal contact (nasal mucus on hands) or droplets is high in early infection. There are at least 100 different antigenic strains of rhinovirus, making it difficult for the immune system to confer ongoing protection. The incubation period varies from 12 hours to 5 days. Clinical features are tiredness, pyrexia, malaise, and sore nose and pharynx. Sneezing and profuse, watery nasal discharge are followed by thick mucopurulent secretions that may persist for up to a week. Secondary bacterial infection occurs in a minority of cases.

Other forms of rhinitis

Rhinitis can be subdivided by the frequency with which symptoms occur:

1. for a limited period of the year (seasonal or intermittent rhinitis)
2. throughout the whole year (perennial or persistent rhinitis).

Seasonal rhinitis (intermittent)

This is the most common allergic disorder. It is often called 'hay-fever', but as this implies that only grass pollen is responsible, it is better described as seasonal (or intermittent) allergic rhinitis. Common seasonal allergens include tree and grass pollens and mould spores. Since the pollination patterns vary from country to country, seasonal rhinoconjunctivitis and accompanying wheeze may occur at different times of year in different regions. Worldwide prevalence varies from 2%–20%. Prevalence is maximal in the second decade, and up to 30% of UK teenagers and young adults are affected.

Nasal irritation, sneezing, and watery rhinorrhoea occur, but many also suffer from itching of the eyes and soft palate. Itching of the ears can occur due to the common innervation of the pharyngeal mucosa and the ear. Approximately 20% have associated seasonal wheeze.

Perennial rhinitis (persistent)

In about 50% symptoms of sneezing and watery rhinorrhoea predominate, while the other 50% complain mostly of nasal blockage. The patient may lose sense of smell and taste, but rarely has eye or throat symptoms. Sinusitis occurs in about 50% of cases, due to mucosal swelling that obstructs drainage from the sinuses. Perennial rhinitis is most frequent in the second and third decades, decreasing with age; there are a number of common subtypes:

Perennial allergic rhinitis

The most common cause is allergy to the faecal particles of the house-dust mite, *Dermatophagoides pteronyssinus* or *D. farinae* (Fig. 20.15). Mites live off desquamated human skin scales and the highest concentrations (4000 mites/g of surface dust) are found in human bedding. The next most common allergens are proteins derived from the skin or urine or saliva spread over the surface of domestic pets (especially cats). Allergy to urinary proteins from small mammals is a major cause of morbidity among laboratory workers. Industrial dust, vapours, and fumes cause occupationally related perennial rhinitis more often than asthma. The presence of perennial rhinitis makes the nose more reactive to non-specific stimuli, such as cigarette smoke, washing powders, household detergents, strong perfumes, and traffic fumes. Although patients often think they are allergic to these stimuli, these are irritant responses and do not involve antibodies.

Perennial non-allergic rhinitis with eosinophilia

No extrinsic allergic cause can be identified, either by history or skin testing, but eosinophilic granulocytes are present in nasal secretions. Most of these patients are intolerant of aspirin/non-steroidal anti-inflammatory drugs (NSAIDs).

Vasomotor rhinitis

Rhinitis occurs without demonstrable allergy or nasal eosinophilia. Watery secretions and nasal congestion are triggered by, for example, cold air, smoke, perfume, or newsprint, possibly because of an imbalance of the autonomic nerves controlling the erectile tissue (sinusoids) in the nasal mucosa.

Investigations and diagnosis

Allergic factors causing rhinitis can usually be identified from the history. Skin-prick testing supports the history. A positive test does not mean that an allergen is the cause of respiratory disease, but it is likely to be relevant if there is a compatible clinical history.

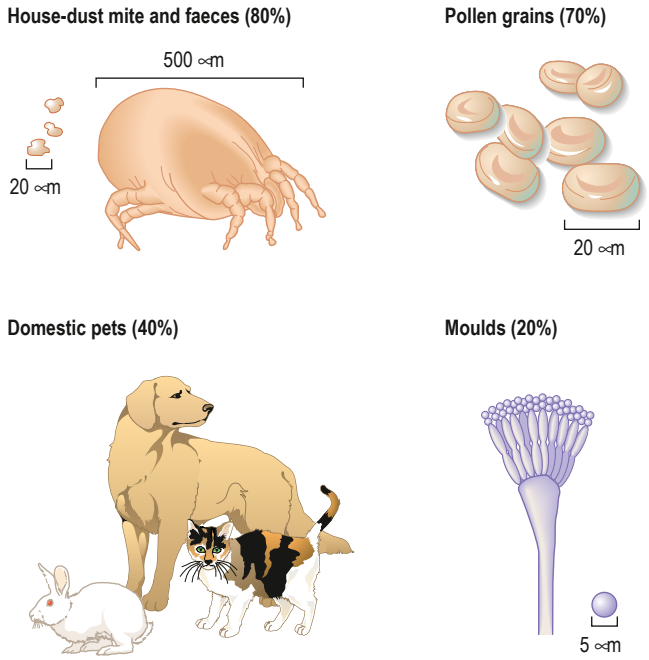


Fig. 20.15 Common allergens causing allergic rhinitis and asthma. These include the house-dust mite, faeces of house-dust mites, pollen grains, domestic pets, and moulds. Percentages indicate the proportion of positive skin-prick tests to these allergens in patients with allergic rhinitis.

Allergen-specific IgE antibodies can be measured in blood but such tests are much more expensive than skin tests and should be used only in patients who cannot be skin-tested for some reason (e.g., dermatographism, eczema, inability to stop antihistamines).

Management

Allergen avoidance

Removal of a household pet or total enclosure of industrial processes releasing sensitizing agents can lead to cure of rhinitis and, indeed, asthma. However, pollen avoidance is impossible. Contact may be diminished by wearing sunglasses, driving with car windows shut, avoiding countryside walks (particularly in late afternoon, when pollen grain number at ground level is highest), and keeping bedroom windows shut at night, those these measures are rarely sufficient in themselves to control symptoms. Exposure to pollen is generally lower in coastal regions, where sea breezes carry pollen grains inland. Exposure to mite allergen can be reduced by enclosing bedding in fabric specifically designed to prevent its passage, while allowing water vapour through. This is comfortable and reduces symptoms. Acaricides (substances toxic to mites) are less effective and not recommended. Increased room ventilation and reduced soft furnishings help to reduce the mite load.

H_1 antihistamines

Antihistamines remain the most common therapy for rhinitis. In the UK, most can be purchased without prescription. They are particularly effective against sneezing and itching of the eyes and palate, but are less effective against rhinorrhoea, and have little influence on nasal blockage. First-generation antihistamines (chlorphenamine, hydroxyzine) cause sedation and loss of concentration in all patients and are no longer used. Second-generation drugs include loratadine (10 mg once daily), desloratadine (5 mg daily), cetirizine (10 mg daily), and fexofenadine (120 mg daily) are at least as potent and do not cause sedation.

Decongestants

Drugs with sympathomimetic activity (α-adrenergic agents) are widely used to treat nasal obstruction. They may be taken orally or, more commonly, as nasal drops or sprays (e.g., ephedrine nasal drops). Xylometazoline and oxymetazoline are widely used because they have a prolonged action and tachyphylaxis does not develop. Secondary nasal hyperaemia can occur some hours later as a rebound effect and *rhinitis medicamentosa* can develop if patients take increasing quantities of decongestant to try and treat this phenomenon. Although local decongestants are an effective treatment for vasomotor rhinitis, patients must be warned about rebound nasal obstruction and should use the drugs carefully. Ideally, preparations should be prescribed for only a limited period to open the nasal airways and allow better access for other local therapy, such as topical corticosteroids.

Anti-inflammatory drugs

Sodium cromoglicate and nedocromil sodium act by blocking an intracellular chloride channel and influence mast cell and eosinophil activation and nerve function. Topical sodium preparations can be very effective in allergic conjunctivitis but are of limited value in allergic rhinitis.

Corticosteroids

The most effective treatment for rhinitis is a topical corticosteroid preparation (e.g., beclometasone, fluticasone propionate, fluticasone furoate). Topical steroids should be started before the beginning of seasonal symptoms. The combination of a topical corticosteroid with a non-sedating antihistamine taken regularly is particularly effective. Patients should be carefully instructed in nasal steroid device use to achieve optimal drug deposition. In selected cases an α-adrenergic agonist may help to decongest the nose prior to taking the topical corticosteroid. Nasal steroids can cause epistaxis but the amount used is insufficient to cause systemic steroid side-effects.

If other therapy has failed, a short course of corticosteroids could be considered with oral prednisolone 0.5 mg/kg for 5 to 10 days in adults. Nasal polyps may respond to oral corticosteroids and their recurrence may then be prevented by continuous use of topical corticosteroids.

Leukotriene antagonists

If there is no response to antihistamines or topical steroids, a leukotriene antagonist (e.g., montelukast 10 mg daily in the evening) may be helpful, especially in patients with a history of NSAID sensitivity or concomitant asthma.

Immunotherapy

This is used for patients with seasonal allergic rhinitis who have not responded to standard drugs (see Chapter 21).

Nasal polyps

These are round, smooth, soft, semitranslucent, pale or yellow, glistening structures attached to the sinus mucosa by a relatively narrow stalk or pedicle, occurring in patients with allergic or vasomotor rhinitis. The mechanism(s) of their formation is not known. They contain mast cells, eosinophils, and mononuclear cells in large numbers and cause nasal obstruction, loss of smell and taste, and mouth breathing. Sneezing is rare since the mucosa of the polyp is largely denervated.

Pharyngitis

The most common viruses causing pharyngitis are adenoviruses. Endemic adenovirus infection causes sore throat, in which the oropharynx and soft palate are reddened, and tonsils are inflamed and

swollen. Within 1–2 days tonsillar lymph nodes enlarge. Over several decades the proportion of sore throats due to bacterial infections, such as haemolytic streptococcus, has fallen. Pharyngitis is usually self-limiting and requires symptomatic treatment, with no role for antibiotics. Persistent and severe tonsillitis may be managed with phenoxymethypenicillin 500 mg four times daily, cefalexin 500 mg three times daily, or clarithromycin 500 mg bd. Amoxicillin and ampicillin should be avoided if there is a possibility of infectious mononucleosis (see Chapter 42), due to drug rash development.

Acute laryngotracheobronchitis

Acute laryngitis is an occasional but striking complication of upper respiratory tract infections, particularly those caused by parainfluenza viruses and measles. The condition is most severe in children under the age of 3 years. Inflammatory oedema extends to the vocal cords and the epiglottis, causing narrowing of the airway, with associated ***tracheitis*** or ***tracheobronchitis***. The voice becomes hoarse, and there is a barking cough (croup) and audible stridor. Progressive airways obstruction may occur, with recession of the soft tissue of the neck and abdomen during inspiration and, in severe cases, central cyanosis. Steam inhalations are not helpful. Nebulized adrenaline (epinephrine) gives short-term relief. Oral or intramuscular corticosteroids (e.g., dexamethasone) should be given, with oxygen and fluids if clinically indicated. Endotracheal intubation and tracheostomy are rarely required.

Acute epiglottitis

Haemophilus influenzae type b (Hib) can cause life-threatening infection of the epiglottis, usually in children under 5 years of age. High fever occurs and severe airflow obstruction may develop rapidly. This is a life-threatening emergency and requires urgent endotracheal intubation and intravenous antibiotics (e.g., ceftriaxone, chloramphenicol). The epiglottis, which is red and swollen, should not be inspected until facilities to maintain the airways are available.

Other manifestations of Hib infection are meningitis, septic arthritis, and osteomyelitis. A highly effective vaccine is routinely offered in the UK to infants, given at 2, 3, and 4 months with a booster at about 12 months of age. Vaccination has reduced death rates from Hib infections virtually to zero.

Influenza

The influenza virus belongs to the orthomyxovirus group and exists in two main forms, A and B. Influenza B is associated with localized outbreaks of mild disease, whereas influenza A causes worldwide pandemics (see Chapter 42).

Clinical features

The incubation period of influenza is usually 1–3 days. The illness starts abruptly with a fever, shivering, and myalgia. This is associated headache, sore throat, and a dry cough that can persist for several weeks. Diarrhoea occurs in 70% of cases of H5N1 ('bird flu').

Complications

Secondary bacterial infection is common, particularly with *Streptococcus pneumoniae* and *H. influenzae*. Secondary pneumonia caused by *Staphylococcus aureus* is rarer but more severe, with mortality of up to 20%. Post-infectious encephalomyelitis is rare. Influenza infection can be followed by a prolonged period of debility and depression.

Diagnosis and management

Laboratory diagnosis is not usually necessary in the community, but definitive diagnosis can be established via nasopharyngeal swabs. Testing is undertaken for hospital inpatients to facilitate appropriate infection control.

Management is supportive and includes rest and paracetamol; antibiotics should be considered if secondary infection is suspected, particularly in vulnerable patients such as those with chronic bronchitis or cardiac or kidney disease. Neuraminidase inhibitors, such as oseltamivir and zanamivir, are recommended in high-risk individuals, and, if possible, should be given within 48 hours of first symptoms. Inpatients should have respiratory isolation measures in place to avoid cross-infection.

Prophylaxis

Protection by influenza vaccines is effective in about 70% of people and lasts for about a year. New vaccines are prepared annually according to predicted viral antigenicity. Routine vaccination is recommended in the UK for those aged 65 years or more, and for younger people at risk of severe disease including those with heart, lung, and kidney disease, diabetes, and any who are immunosuppressed. Front-line healthcare workers are also offered vaccination annually. Egg-free vaccines are available for those with allergy to egg proteins.

Inhalation of foreign bodies

Children inhale foreign bodies more commonly than adults. Risk factors for inhalation include excess alcohol and general anaesthesia (loose teeth or dentures). When the foreign body is large, it may impact the trachea. The person chokes and then becomes silent; death ensues unless the material is quickly removed. Guidelines for the management of choking should rapidly be followed (**Box 20.12**). More often, impaction occurs in the right main bronchus and produces choking and persistent monophonic wheeze. It may lead to a suppurative pneumonia and/or lung abscess.

Cough

Cough is one of the most common respiratory symptoms. Acute cough lasts less than 3 weeks, chronic cough persists for more than 8 weeks, with the term subacute cough used between 3 and 8 weeks.

Acute cough

Acute cough is often associated with upper respiratory tract infection, particularly in the winter months (**Box 20.13**). Any patient with worrying symptoms or signs including haemoptysis, breathlessness, fever, weight loss, night sweats, chest pain, or foreign body inhalation, should have an urgent chest radiograph to exclude

Box 20.12 Sequence of steps for managing adult foreign body airway obstruction

Suspect: suddenly unable to talk, particularly if eating
Cough: encourage to cough
Back blows: if cough becomes ineffective, give up to five back blows:
- Lean the person forwards
- Apply blows between the shoulder blades using the heel of one hand

Abdominal thrusts: if back blows are ineffective, give up to 5 abdominal thrusts
- Stand behind the person and put both your arms around their upper abdomen
- Lean the person forwards
- Clench your fist and place it between umbilicus and ribcage
- Grab your fist with the other hand and pull upwards

Alternate: if choking not relieved after 5 abdominal thrusts, continue alternating 5 back blows with 5 abdominal thrusts until it is relieved, or the person becomes unresponsive.
CPR: if the person becomes unresponsive, start CPR

CPR, Cardiopulmonary resuscitation.
Reproduced with the kind permission of Resuscitation Council UK.

Box 20.13 Causes of cough

Cause	Example
Respiratory disease	Viral or bacterial infection, bronchospasm, chronic obstructive pulmonary disease, non-asthmatic eosinophilic asthma, bronchiolitis, malignancy, parenchymal disease (e.g., interstitial lung disease, bronchiectasis, cystic fibrosis, sarcoidosis), pleural disease, aspiration
Upper airways disease	Postnasal drip, sinusitis, inhaled foreign body, tonsillar enlargement
Cardiovascular disease	Heart failure, mitral stenosis
Gastro-oesophageal disease	Gastro-oesophageal reflux disease, also associated with laryngopharyngeal reflux
Neurological disease	Aspiration
Drugs and irritants	Angiotensin-converting enzyme (ACE) inhibitors, cigarette smoke

malignancy, pneumonia, and tuberculosis. Most acute cough is self-limiting, and antibiotics are not indicated.

Chronic cough

All patients with a chronic cough should have a chest radiograph and those with abnormalities should be investigated. If the chest radiograph is normal, then the most likely diagnoses are undiagnosed asthma, postnasal drip, gastro-oesophageal reflux disease, recent upper respiratory tract infection, or smoking. Spirometry should be performed, and reversible airways obstruction treated as asthma. It is important to identify treatable traits that may cause cough. An individual may have more than one treatable trait. If spirometry is normal, patients should have a trial of treatment for the most likely cause based on their history:

Potential treatable traits:

- **angiotensin-converting enzyme (ACE) inhibitor:** convert to angiotensin receptor antagonist
- ***nocturnal cough or wheeze:*** trial of oral or inhaled steroids
- ***acid reflux:*** trial of a proton pump inhibitor
- ***postnasal drip:*** trial of topical nasal steroids
- ***smoking:*** referral to the local smoking cessation service
- ***obstructive sleep apnoea:*** assessment and weight loss, consider CPAP therapy
- ***cough hypersensitivity:*** cough control, speech, and language therapy, low dose opiates or gabapentin
- ***anxiety and depression:*** reassurance, psychological intervention as necessary.

If a chronic cough does not settle after empirical treatment and the aetiology remains unclear, further investigations might include CT chest, bronchoscopy, bronchial provocation testing, oesophageal manometry, and sinus imaging. Referral to a specialist cough clinic may be available.

Further reading

Bousquet J, Anto JM, Bachert C et al. Allergic rhinitis. *Nat Rev Dis Primers* 2020; 6:95.

National Institute for Health and Care Excellence. *NICE Guideline 120: Cough (Acute): Antimicrobial Prescribing*. NICE 2019; https://www.nice.org.uk/guidance/ng120.

Scadding GK, Kariyawasam HH, Scadding G et al. BSACI guidelines for the diagnosis and management of allergic and non-allergic rhinitis. *Clin Exp Allergy* 2017; 47:856–889.

https://www.worldallergy.org. World Allergy Organization: education programmes and allergic disease resource centre.

https://www.who.int/news-room/fact-sheets/detail/influenza-(seasonal). *Influenza (Seasonal)*

OBSTRUCTIVE RESPIRATORY DISEASE

Asthma

Asthma is a common chronic condition, the cause of which is incompletely understood. Symptoms include wheeze, chest tightness, cough, and shortness of breath, often worse at night. Asthma commonly starts in childhood between the ages of 3 and 5 years and may either worsen or improve during adolescence. Classically, asthma has three characteristics:

- ***airflow limitation:*** usually reversible spontaneously or with treatment
- ***airway hyper-responsiveness:*** to a wide range of stimuli
- ***bronchial inflammation:*** T-lymphocytes, mast cells, eosinophils with plasma exudation, oedema, smooth muscle hypertrophy, matrix deposition, mucus plugging, epithelial damage.

In chronic asthma, inflammation leads to irreversible airflow limitation as a result of large and small airway wall remodelling and mucus impaction.

Prevalence

In many countries, the prevalence of asthma has increased since the mid-1980s, particularly in children and young adults, with up to 15% of the population affected. Asthma is more common in high-income countries, with some of the highest prevalence in the UK, New Zealand, and Australia, and lower rates in East Asia, Africa, and Eastern Europe. Long-term follow-up in low- and middle-income countries suggests that asthma prevalence may increase with adoption of a 'Westernized' lifestyle though the environmental factors accounting for this remain unknown. Studies of occupational asthma suggest that a large proportion of the workforce (15%–20%) may become asthmatic if exposed to potent sensitizers. Worldwide, over 300 million people have asthma and the disease kills approximately 1000 people per day.

Classification

Asthma is a complex disorder with symptoms caused by several different mechanisms. Asthma can be classified according to trigger factors, age of onset, inflammatory subtypes, or response to therapy, with considerable overlap.

Childhood-onset asthma is commonly a wheezing illness with inhaled allergic triggers. Up to 90% of children and 70% of adults with persistent asthma have positive skin-prick tests to common inhalant allergens such as dust mite, animal danders, pollens, and fungi. Childhood-onset asthma is often accompanied by eczema.

Adult-onset asthma is often triggered by respiratory infections. Nevertheless, many patients show positive allergen skin tests and, on close questioning, some of these will give a history of childhood respiratory symptoms suggesting allergic asthma. Non-atopic individuals may develop asthma in middle age from extrinsic causes, such as sensitization to occupational agents like toluene diisocyanate, intolerance to NSAIDs, or prescription of β-adrenoceptor-blocking agents that block the protective effect of endogenous catecholamines. Extrinsic causes must be considered in all asthma and avoided if possible.

Asthma can also be classified as eosinophilic or non-eosinophilic. Eosinophilic airway inflammation is often present in asthma particularly in patients with more severe symptoms.

Aetiology

The major factors involved in the development of asthma and stimuli that can precipitate attacks are shown in **Fig. 20.16**.

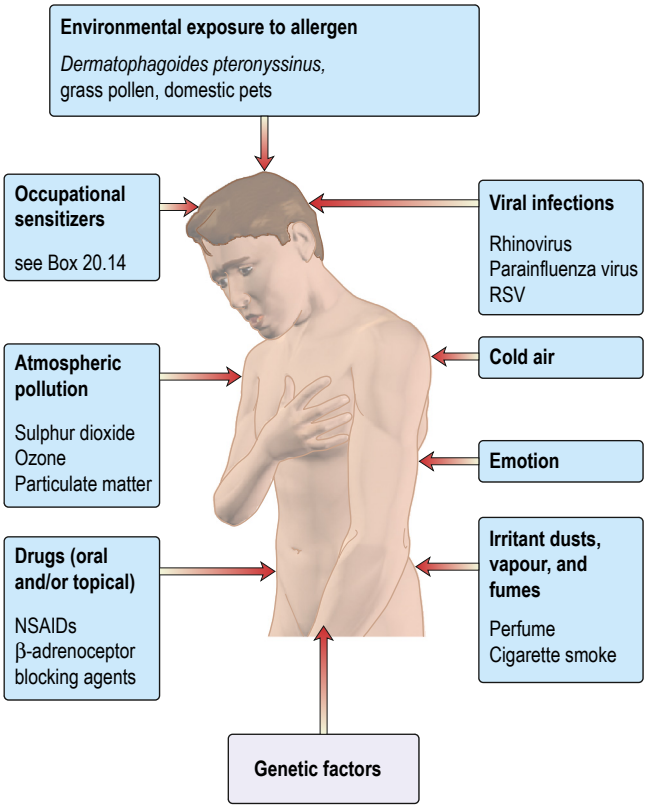


Fig. 20.16 Causes and triggers of asthma. *NSAIDs,* Non-steroidal anti-inflammatory drugs; *RSV,* respiratory syncytial virus.

Atopy and allergy

The term 'atopy' describes a group of disorders, including asthma and hayfever, that run in families, have positive skin-prick tests to common inhalant allergens, and have circulating allergen-specific antibodies. Allergen-specific IgE is present in 30%–40% of the UK population, and elevated serum IgE concentrations are linked to airway hyper-responsiveness and the prevalence of asthma. Serum total IgE concentrations are affected by genetic and environmental factors.

Genetic factors: There is no single gene for asthma. Several genes appear to influence asthma development, in combination with environmental factors, including genes involved cytokine and IgE synthesis.

Environmental factors: Early childhood exposure to allergens and maternal smoking influence IgE production. Intestinal bacteria and childhood infections are hypothesized to shape the immune system in early life. Growing up in a relatively 'clean' environment may predispose to an IgE response to allergens ('hygiene hypothesis'). Conversely, growing up in a 'dirtier' environment may prevent the immune system from developing allergic responses, with early-life exposure to inhaled and ingested products of microorganisms reducing subsequent risk of allergy and asthma.

The allergens involved in allergic asthma are similar to those in rhinitis, although pollens are less implicated in asthma. Most patients with allergic asthma are sensitized to house-dust mite allergens. Allergens from furry pets (especially cats) are increasingly common. Cockroach allergy has been implicated in asthma in inner-city children in the USA. Fungal spores from *Aspergillus fumigatus* cause a range of lung disorders, including asthma.

? Box 20.14 Occupational asthma

Cause	Source/occupation
Low molecular weight (non–IgE-related)	
Isocyanates	Polyurethane varnishes Industrial coatings Spray painting
Colophony fumes	Soldering/welders Electronics industry
Wood dust	
Drugs	
Bleaches and dyes	
Complex metal salts, e.g., nickel, platinum, chromium	
High molecular weight (IgE-related)	
Allergens from animals and insects	Farmers, workers in poultry and seafood processing industry; laboratory workers
Antibiotics	Nurses, health industry
Latex	Health workers
Proteolytic enzymes	Manufacture (but not use) of 'biological' washing powders
Complex salts of platinum	Metal refining
Acid anhydrides and polyamine hardening agents	Industrial coatings

Occupational sensitizers

Over 250 materials encountered in the workplace can cause occupational asthma, which accounts for up to 15% of all asthma (**Box 20.14**). Patients in insurable employment are eligible for statutory compensation, provided they apply within 10 years of leaving the occupation in which the asthma developed.

Asthma can be due to:

- low-molecular-weight compounds, e.g., reactive chemicals such as isocyanates and acid anhydrides that bond chemically to epithelial cells to activate them, as well as providing haptens recognized by T cells
- high-molecular-weight compounds, e.g., flour, organic dusts, and other large protein molecules involving specific IgE antibodies.

The proportion of employees developing occupational asthma depends primarily on the level of exposure. Enclosure of industrial processes and appropriate ventilation reduce risk. Atopic individuals develop occupational asthma more rapidly when exposed to agents triggering development of specific IgE antibody. Non-atopic individuals can also develop asthma when exposed to such agents, but after a longer period of exposure. Smoking increases the risk of developing some forms of occupational asthma.

Non-specific factors

Due to their AHR (airway hyper-responsiveness), patients with asthma will respond to a wide variety of non-specific direct and indirect stimuli, as well as reacting to specific allergens.

Cold air and exercise

Most asthmatics wheeze after prolonged exercise or inhalation of cold, dry air. Typically, the attack does not occur while exercising but afterwards. Exercise-induced wheeze is driven by release of histamine, prostaglandins (PGs), and leukotrienes (LTs) from mast cells, as well as stimulation of neural reflexes.

Atmospheric pollution and irritant dusts, vapours, and fumes

Many patients with asthma experience worsening of symptoms on exposure to tobacco smoke, car exhaust fumes, solvents, strong perfumes, or high concentrations of airborne dust. Major epidemics have been recorded when large amounts of allergens are released into the air, and asthma exacerbations increase during summer and winter air pollution episodes associated with climatic temperature inversions: in the presence of high concentrations of ozone, particulates, and NO_2 in the summer and particulates, NO_2 and SO_2 in the winter.

Diet

Increased intake of fresh fruit and vegetables has been shown to be protective, possibly owing to antioxidants or other protective molecules such as flavonoids. Genetic variation in antioxidant enzymes is associated with more severe asthma.

Emotion

Emotional factors influence both acute and chronic asthma. Patients at high risk of life-threatening attacks are understandably anxious, but there is no evidence patients with asthma have more psychological pathology than peers without asthma.

Drugs

NSAIDs, particularly aspirin, indomethacin, and ibuprofen, are implicated in triggering asthma in approximately 5% of patients. NSAID intolerance is especially prevalent in those with nasal polyps and asthma, and is often associated with rhinitis and flushing on drug exposure.

Beta-blockers. Parasympathetic innervation of the airways leads to bronchoconstriction. There is no direct sympathetic innervation of bronchial smooth muscle, so antagonism of parasympathetically induced bronchoconstriction is critically dependent on circulating adrenaline (epinephrine) acting through β_2-receptors on the surface of smooth muscle cells. Inhibition by non-selective β-adrenoceptor-blocking drugs (e.g., propranolol, carvedilol) leads to bronchoconstriction and airflow limitation in asthmatic subjects. Selective β_1-adrenergic-blocking drugs (e.g., atenolol, bisoprolol) may also induce attacks of asthma.

Clinical features

The principal symptoms of asthma are wheezing attacks and episodic shortness of breath. Symptoms are usually worst during the night. Cough is a frequent symptom that sometimes predominates, especially in children, in whom nocturnal cough can be a presenting feature. Attacks vary greatly in frequency and duration and may be precipitated by a wide range of triggers (see **Fig. 20.16**). Asthma is a major cause of impaired quality of life and has an impact on work and recreation, affecting both physical activities and emotions.

Investigations

There is no single satisfactory diagnostic test for all patients with asthma.

Lung function tests

PEFR measurements on waking, before bed, and prior to and after taking a bronchodilator are useful in demonstrating the variable airflow limitation that characterizes asthma (see **Figs 20.11 and 20.12**). Diurnal variation in PEFR is a good measure of asthma activity and aids longer-term assessment of the disease and response to treatment.

Spirometry is useful, especially in assessing reversibility. Asthma can be diagnosed by demonstrating a greater than 15% improvement in FEV_1 or PEFR following inhalation of a bronchodilator. Carbon monoxide (CO) transfer is normal in asthma.

Interpretation of results

In asthma there is evidence of airflow obstruction (FEV_1:FVC ratio <75%) AND evidence of variation. Variation means that lung function tests can range from normal to severely obstructed in a single individual.

Evidence of variation may include

- Post-bronchodilator improvement of FEV_1 >12% or 200 mL compared to pre-bronchodilator values (measure spirometry pre and post 200–400 mcg inhaled salbutamol). An increase of >15% or 400 mL provides greater assurance.
- Average diurnal variation in peak flow of >10% (measured twice daily over a 2-week period)
- Significant increase in lung function after 4 weeks of anti-inflammatory treatment—improvement of FEV_1 >12% or 200 mL compared to pre-bronchodilator values
- Positive bronchial challenge test: drop in FEV_1 >20% with standard doses of inhaled methacholine or >15% with mannitol
- Variation of FEV_1 of >12% or >200 mL between visits (in absence of infection).

Fractional exhaled nitric oxide (FeNO)

This test is a measure of airway inflammation and corticosteroid response. The trend in the values can help to determine a response to, or failure of, treatment. FeNO is measured in parts per billion (ppb). FeNO <25 ppb is considered normal in a treatment naïve patient, or controlled inflammation in a patient on steroid treatment. FeNO 25–50 ppb is intermediate. FeNO >50 ppb is high and indicative of airways inflammation.

Blood and sputum tests

Patients with asthma sometimes have increased numbers of eosinophils in peripheral blood (>0.4 × 10^9/L) but sputum eosinophilia is a more specific diagnostic finding.

Chest radiograph

There are no diagnostic features of asthma on chest radiography, although overinflation is characteristic during an acute episode or in chronic severe disease. A chest radiograph excludes a pneumothorax, which can occur as a complication of asthma, and in detection the pulmonary infiltrates associated with allergic bronchopulmonary aspergillosis.

Skin tests

Skin-prick tests should be performed in all asthma to help identify allergic trigger factors. Allergen-specific IgE can be measured in serum if skin-prick test facilities are not available, the patient is taking antihistamines, or no suitable allergen extracts are available.

Allergen provocation tests

Allergen inhalation challenge is a useful research tool. It is required when investigating patients with suspected occupational asthma but not in ordinary asthma.

Management

The ***aims of treatment*** are to:

- abolish symptoms
- restore normal or best possible lung function
- reduce the risk of severe attacks
- enable normal growth in children
- minimize absence from school or employment.

This involves:

- patient and family education about asthma
- patient participation in treatment
- avoidance of identified triggers where possible
- the lowest effective doses of convenient medications to minimize short-term and long-term side-effects.

Many people with asthma join self-help groups to improve understanding of disease and to foster self-confidence and fitness.

Control of extrinsic factors

Where specific allergen triggers are identified, these should be avoided if possible. Sublingual allergen immunotherapy (SLIT) with house-dust mites reduces the number of asthma attacks in children and is now being developed for management of asthma in adults. Active and passive smoking should be avoided, as should beta-blockers in both tablet and eye-drop form. Individuals intolerant to aspirin should avoid NSAIDs, although they may tolerate cyclo-oxygenase-2 (COX-2) inhibitors. About one-third of individuals sensitized to occupational agents may be cured if they permanently avoid exposure.

Drug treatment

The mainstay of asthma therapy is use of inhaled therapeutic agents, delivered as aerosols or powders directly into the lungs (**Box 20.15**). The advantages of this method are drug delivery direct to the airways avoiding first-pass metabolism in the liver; thus lower doses are necessary and systemic side-effects are minimized. To help those who cannot coordinate aerosol activation and inhalation, breath-activated and dry powder devices have been developed. Patients vary in ability to use such devices. Care should be taken in device selection and education.

Box 20.15 Inhaled therapy for asthma

Patients should be taught how to use inhalers and their technique should be checked regularly.

Use of a metered-dose inhaler

1. The canister is shaken.
2. The patient exhales to functional residual capacity (not residual volume), i.e., normal expiration.
3. The aerosol nozzle is placed to the open mouth.
4. The patient simultaneously inhales rapidly and activates the aerosol.
5. Inhalation is completed.
6. The breath is held for 10 sec if possible.

Even with good technique, only 15% of the contents is inhaled; 85% is deposited on the wall of the pharynx and ultimately swallowed.

Spacers

- Plastic cones/spheres inserted between the patient's mouth and the inhaler, or in-built extension to inhaler.
- Designed to reduce particle velocity so less drug is deposited in the mouth.
- Diminish the need for coordination between aerosol activation and inhalation.
- They are useful in all who find inhaler technique difficult, e.g., children, elderly.
- Reduced risk of oral candidiasis.

Several national and international guidelines have been published on the treatment of asthma (**Fig. 20.17**) based on three principles:

- Asthma should be self-managed with an agreed, individualized, documented treatment plan including monitoring using PEFR.
- Asthma is an inflammatory disease, so anti-inflammatory (controller) therapy should be started, even in mild cases.
- Short-acting inhaled bronchodilators (e.g., salbutamol and terbutaline) should be used only to relieve breakthrough symptoms. Increased use of bronchodilator treatment to relieve increasing symptoms is an indication of deteriorating disease.

A list of drugs used in asthma is shown in **Box 20.16**. These are used in a stepwise fashion, as indicated in **Fig. 20.17**.

Beta$_2$-adrenoceptor agonists

Beta$_2$-adrenoceptor agonists are selective for the respiratory tract and do not stimulate the β_1 adrenoceptors of the myocardium. These drugs relax bronchial smooth muscle and are effective in relieving symptoms, but do not affect underlying airways inflammation.

- ***Short-acting β agonists (SABAs):*** salbutamol 100 μg (albuterol in the USA) or terbutaline 250 μg. SABAs without inhaled corticosteroids are no longer recommended in the management of asthma.
- ***Long-acting β_2-adrenoceptor agonists (LABAs):*** salmeterol or formoterol are effective by inhalation for up to 12 hours and are given once or twice daily. They are used in combination with an inhaled corticosteroid as fixed-dose combinations (e.g., salmeterol/fluticasone and formoterol/budesonide) in the same inhaler.

SABA treatment should no longer be used alone, even for management of mild symptoms. The Global Initiative for Asthma 2023 recommends that the first step in management of chronic asthma is a combination inhaled corticosteroid-formoterol inhaler as needed, or inhaled corticosteroids whenever a SABA is taken as needed.

Inhaled corticosteroids

Inhaled corticosteroids (ICS) are now recommended for all patients with asthma. Beclometasone dipropionate (BDP) is the most widely used inhaled steroid and is available in doses of 50, 100, 200, and 250 μg per puff. Other inhaled steroids include budesonide, fluticasone propionate, fluticasone furoate, mometasone furoate, and triamcinolone. Side-effects of inhaled steroids include oral thrush and hoarseness; patients should be instructed to rinse their mouths after use. Subcapsular cataract formation is rare but can occur in the elderly. Osteoporosis is less likely than with oral steroids but can occur with high-dose inhaled corticosteroids (beclometasone or budesonide >800 μg daily). In children, inhaled corticosteroid doses above 400 μg daily retard short-term growth, although final heights are not affected. Inhaled corticosteroid use should be stepped down to the minimum effective dose once asthma is under control.

Oral corticosteroids and steroid-sparing agents

Oral corticosteroids are needed for acute exacerbations. Acute exacerbations of asthma in adults should be treated with 40 to 50 mg prednisolone for 5 to 7 days. The benefit of short-term treatment with prednisolone 30 mg daily as measured by PEFR is shown in **Fig. 20.12**. Steroids are rarely used longer term, and only when other drug regimens do not control symptoms. The dose should be kept as low as possible to minimize side-effects. Steroid-sparing agents should be considered to allow discontinuation of oral steroid use. These include monoclonal antibodies, low-dose methotrexate, and ciclosporin.

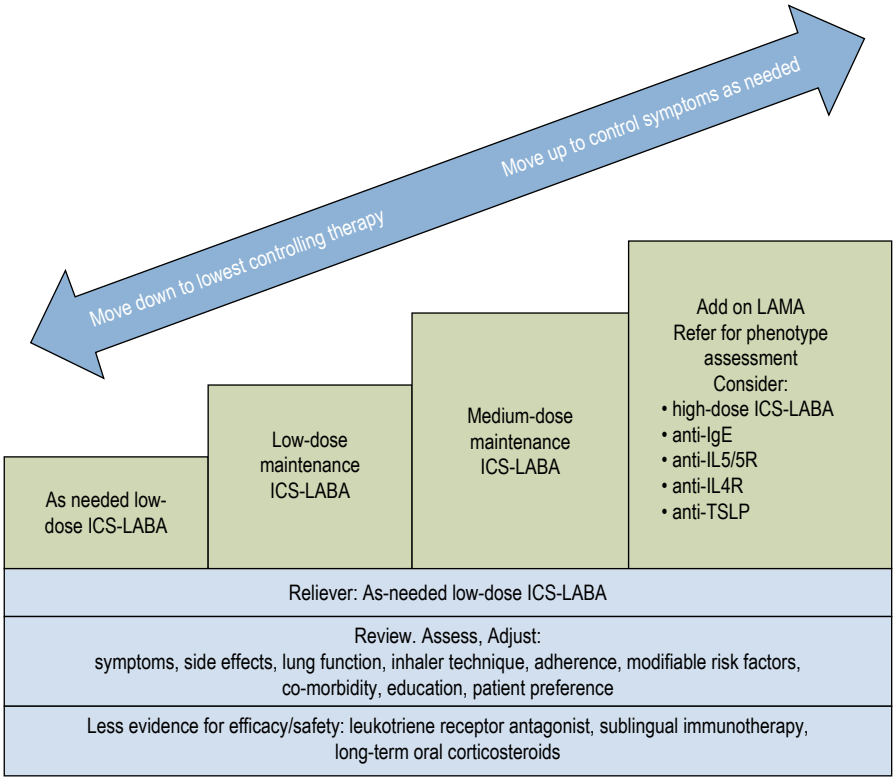


Fig. 20.17 Summary of asthma management in adults. *ICS,* Inhaled corticosteroid; *LABA,* long-acting β agonist; *LTRA,* leukotriene receptor antagonist; *MART,* maintenance and reliever therapy. *(Global Initiative for Asthma. Global Strategy for Asthma Management and Prevention, 2023. Updated July 2023. Available from www.ginasthma.org.)*

Box 20.16 Drugs used in asthma

Short-acting relievers
- Inhaled β_2 agonists (e.g., salbutamol [albuterol in the USA], terbutaline)

Long-acting relievers/disease controllers
- Inhaled long-acting β_2 agonists (e.g., salmeterol, formoterol)
- Inhaled corticosteroids (e.g., beclometasone, budesonide, fluticasone)
- Compound inhaled long-acting β_2 agonists and corticosteroids (e.g., salmeterol and fluticasone)
- Sodium cromoglicate
- Leukotriene modifiers (e.g., montelukast, zafirlukast, zileuton)

Other agents with bronchodilator activity
- Inhaled antimuscarinic agents (e.g., ipratropium, oxitropium, aclidinium)
- Theophylline preparations

Other
- Oral corticosteroids (e.g., prednisolone 40 mg daily)

Steroid-sparing agents
- Anti-IgE monoclonal antibody: omalizumab
- Anti-interleukin monoclonal antibody: mepolizumab, benralizumab, dupilumab, reslizumab
- Ciclosporin
- Methotrexate
- Intravenous immunoglobulin

Leukotriene receptor antagonists

Therapy targets the cysteinyl LT1 receptor. Montelukast, pranlukast, and zafirlukast are given orally. They are effective in some patients with asthma, though prediction of effectiveness is difficult: a 4-week trial of leukotriene receptor antagonist (LTRA) therapy is recommended before a decision is made to continue or stop. LTRAs should be tried if disease is not controlled on low to medium doses of inhaled steroids; action is additive to that of LABAs. LTRAs are particularly useful in patients with aspirin-intolerant asthma. These drugs are also helpful in asthma combined with rhinitis, and in young children with asthma and/or virus-associated wheezing

Antimuscarinic bronchodilators

Muscarinic receptors are found in the respiratory tract: large airways contain mainly M_3 receptors, and the peripheral lung tissue contains M_3 and M_1 receptors. Nebulized short-acting ipratropium is used in acute severe exacerbations of asthma (**Box 11.21** in Acute medicine chapter). There is no benefit to long-term use of inhaled, short-acting antimuscarinic agents. Longer-acting, inhaled antimuscarinics (tiotropium, aclidinium) are considered in the latest guidelines at Step 4 if symptoms are not controlled despite medium to high dose ICS-LABA.

Anti-inflammatory drugs

Sodium cromoglicate and nedocromil sodium prevent activation of inflammatory cells, but are no longer recommended in the treatment of asthma due to their low efficacy.

Monoclonal antibodies

Omalizumab, a recombinant humanized monoclonal antibody directed against IgE, chelates free IgE and downregulates number and activity of mast cells and basophils. It is given subcutaneously every 2–4 weeks, depending on total serum IgE concentration and body weight. It is cost-effective in patients with frequent exacerbations requiring oral corticosteroids. Mepolizumab, reslizumab and

benralizumab target interleukin-5 (IL-5) or its receptor. Dupilumab (targets interleukin-4 and interleukin-13 signalling) and tezepelumab (targets thymic stromal lymphopoietin [TSLP], reducing a broad spectrum of biomarkers and cytokines) are other alternatives. They are effective in eosinophilic asthma and are cost-effective in patients with recurrent exacerbations despite high-dose inhaled corticosteroids.

Antibiotics

There is little evidence that antibiotics are helpful in managing patients with acute asthma. During acute exacerbations, yellow or green sputum containing eosinophils and bronchial epithelial cells may be coughed up. This is usually due to viral rather than bacterial infection and antibiotics are not required. There is mixed evidence in severe asthma for long-term treatment with the macrolide antibiotic azithromycin, which has both anti-inflammatory and antibacterial actions.

Bronchial thermoplasty

Bronchial thermoplasty is an approach for patients who still have symptoms despite optimization of treatment and referral to a specialist asthma centre. This bronchoscopic procedure uses radiofrequency radiation to heat the bronchial wall and reduce the mass of airway smooth muscle, decreasing bronchoconstriction. Long-term risks and benefits are under evaluation. With the advent of biologic therapies this procedure is not performed often.

Asthma attacks

Exacerbations of asthma are most associated with reduced adherence to treatment, respiratory virus infection, or exposure to an allergen or triggering drug. Whenever possible, patients should have a personalized documented plan to implement in anticipation, or at the start, of an exacerbation, which includes early use of a short course of oral corticosteroids. If the PEFR is >150 L/min, patients may improve dramatically with nebulized therapy and may not require hospital admission. Regular treatment should be increased, to include treatment with prednisolone 40 to 50 mg for 5 to 7 days, followed by inhaled corticosteroids. Short courses of oral prednisolone can be stopped without tailing down the dose.

The diagnosis and management of acute severe asthma is detailed in the Acute medicine chapter (see Chapter 11).

Further reading

Global Initiative for Asthma. *Global Strategy for Asthma Management and Prevention*, 2023. Updated July 2023. Available from www.ginasthma.org.

Acute bronchitis

Acute bronchitis is an acute inflammation of the bronchi, usually due to viral infection. It is most commonly a mild self-limiting condition with symptoms lasting a few days, which does not cause long-term complications. Asthma in contrast is a relapsing and remitting condition due to genetic and environmental factors.

Acute bronchitis illness begins with an irritating, non-productive cough, together with discomfort behind the sternum. There may be associated chest tightness, wheezing, and shortness of breath. The cough becomes productive, with yellow or green sputum. Wheeze with occasional crackles can be heard on auscultation. There is a mild fever and a neutrophil leucocytosis. In otherwise healthy adults, the disease improves spontaneously in 4–8 days without serious illness.

Acute bronchitis in those without underlying lung disease is most commonly viral. In most cases, antibiotics are not indicated. Secondary bacterial infection with *S. pneumoniae* or *H. influenzae* is more likely in cigarette smokers and those with COPD.

Chronic obstructive pulmonary disease

Chronic obstructive pulmonary disease (COPD) is characterized by airflow limitation that is not fully reversible. Airflow limitation is usually progressive. It is associated with and abnormal inflammatory response of the lungs to noxious particles or gases, with destruction of lung parenchyma. COPD is an overarching diagnosis that includes a number of clinical syndromes: emphysema, small airways disease, and chronic bronchitis. There is hyperinflation of the lungs, ventilation/perfusion mismatch, increased work of breathing, and breathlessness. Association with co-morbidities (ischaemic heart disease, hypertension, diabetes, heart failure, cancer), means that it may be part of a generalized, systemic, inflammatory process.

Epidemiology and aetiology

COPD is caused by long-term exposure to toxic particles and gases. In high-income countries, cigarette smoking accounts for over 90% of cases. Only 10%–20% of smokers develop COPD, suggesting underlying individual susceptibility. In low-income countries, inhalation of smoke from biomass heating fuels and cooking in poorly ventilated areas may contribute. The effects of urbanization, air pollution, deprivation, and occupation are difficult to separate from smoking. Animal studies suggest that diet could be a risk factor for COPD but there are limited human data.

The economic burden of COPD is considerable. In the UK, COPD causes approximately 20 million lost working days annually, accounting for about 7% of sick-leave days. Nevertheless, the number of admissions to UK hospitals due to COPD has been falling since the mid-1980s.

Pathophysiology

Airways inflammation and structural changes within the airways and the lung parenchyma are evident. Increased mucus-producing goblet cells in the bronchial mucosa lead to chronic bronchitis (**Figs. 20.18 and 20.19**) and development of airflow limitation. Acute and chronic inflammation is seen on endobronchial biopsies with a predominance of neutrophils, CD8-predominant lymphocytes, and macrophages. Chronic inflammation results in scarring and fibrosis of small airways. Inflammatory destruction of the alveolar walls results in emphysema. The clinical phenotype of the disease differs depending on the predominance of small airways disease, emphysema, or chronic bronchitis.

Emphysema

Emphysema is defined as abnormal and permanent enlargement of air spaces distal to the terminal bronchiole, accompanied by destruction of their walls. It is classified according to distribution:

- ***Centri-acinar emphysema*.** Distension and damage of lung tissue are concentrated around the respiratory bronchioles, while more distal alveolar ducts and alveoli tend to be well preserved. This form of emphysema is extremely common.
- ***Pan-acinar emphysema*.** Distension and destruction affect the whole acinus, and in severe cases the lung is just a collection of bullae. Severe airflow limitation and mismatch occur. This is

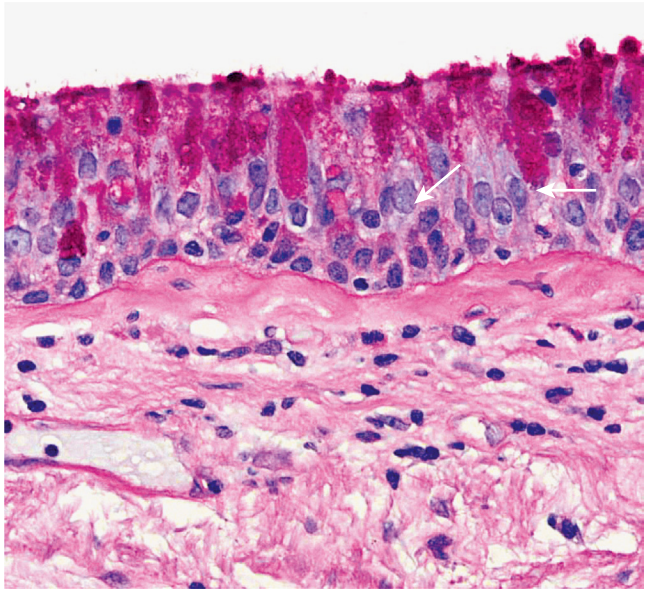


Fig. 20.18 Chronic obstructive pulmonary disease (COPD). Section of bronchial mucosa stained for mucus glands by periodic acid–Schiff (PAS) showing an increase in mucus-secreting goblet cells *(arrowed)*. (*Courtesy of Dr J Wilson and Dr S Wilson, University of Southampton.*)

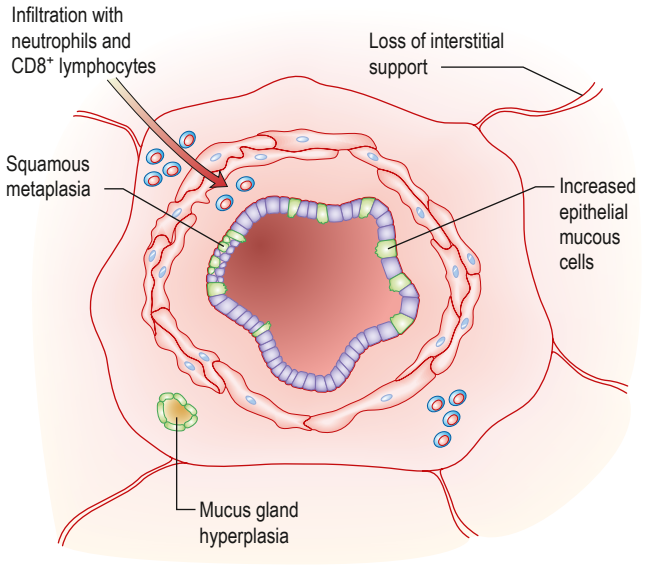


Fig. 20.19 Pathological changes in the airways in chronic obstructive pulmonary disease.

less common than the centri-acinar disease. α_1-antitrypsin deficiency causes pan-acinar disease (see later).

- ***Irregular emphysema*.** There is scarring and damage that affect the lung parenchyma patchily, independent of acinar structure.

Emphysema leads to expiratory airflow limitation and air trapping. The loss of lung elastic recoil results in an increase in TLC. Premature closure of airways limits expiratory flow while the loss of alveoli decreases gas transfer.

The historical ***Fletcher and Peto studies*** (Fig. 20.20) showed a FEV_1 loss of 50 mL per year in patients with COPD compared with 20 mL per year in healthy people. More recent data shows a 40 mL loss per year in only 38% of patients studied. Biomarkers to predict rate of lung function decline have not yet been identified.

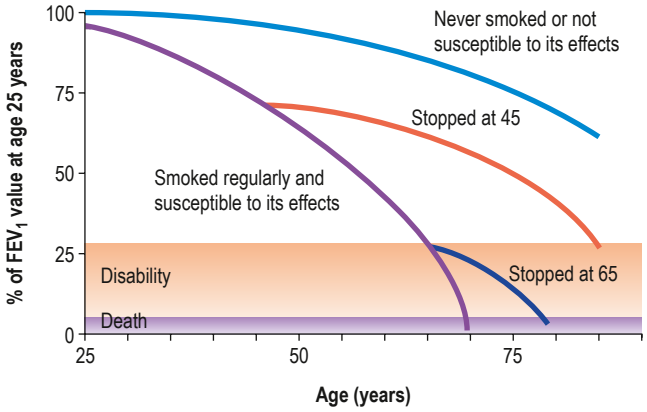


Fig. 20.20 Influence of smoking on airflow limitation. FEV_1, Forced expiratory volume in 1 sec. *(From Fletcher CM, Peto R. The natural history of chronic airflow obstruction*. Bri Med J *1977; 1:1645.)*

Chronic bronchitis

This clinical syndrome of COPD is defined as a daily productive cough for 3 months per year for 2 consecutive years.

Pathogenesis

Cigarette smoking

Bronchoalveolar lavage and airway biopsies in smokers show an increased number of neutrophil granulocytes. These granulocytes release elastases and proteases; an imbalance between protease and antiprotease activity is a factor in development of emphysema. Mucous gland hypertrophy in larger airways is thought to be a direct response to persistent irritation from inhalation of cigarette smoke. An adverse effect on surfactant favours over-distension of the lungs.

Infections

Respiratory infections are a common precipitant of acute exacerbations of COPD. It is less clear whether infection contributes to disease progression. Prompt use of antibiotics and routine vaccination against influenza, SARS-CoV-2, and pneumococci are appropriate.

Alpha$_1$-antitrypsin deficiency

Alpha$_1$-antitrypsin is a proteinase inhibitor produced in the liver and secreted into blood with diffusion into the lung. It inhibits proteolytic enzymes such as neutrophil elastase, which can destroy alveolar wall connective tissue. In α_1-antitrypsin deficiency, accumulation in the liver leads to low concentrations in the lung. More than 75 alleles of the α_1-antitrypsin gene have been described. The three main phenotypes are MM (normal), MZ (heterozygous deficiency), and ZZ (homozygous deficiency). Hereditary deficiency of α_1-antitrypsin accounts for about 2% of emphysema in the UK.

Clinical features

Symptoms

Characteristic symptoms are productive cough with white or clear sputum, wheeze, and breathlessness. Individuals are prone to lower respiratory tract infections. Systemic effects include hypertension, osteoporosis, depression, weight loss, reduced muscle mass and weakness, and right heart failure.

Signs

In mild COPD there may be no signs or quiet wheeze throughout the chest. In severe disease the patient is tachypnoeic, with

prolonged expiration and accessory muscle use. Intercostal indrawing on inspiration and pursing of the lips on expiration is seen. The cricosternal distance is reduced. Chest expansion is reduced. Lungs are hyperinflated with loss of the normal cardiac and liver dullness.

Patients who maintain a respiratory response to CO_2 are usually breathless and rarely cyanosed. Heart failure and oedema are also rare, except as terminal events. In contrast, patients whose respiratory drive is insensitive to CO_2 are often oedematous and cyanosed, but not particularly breathless. Hypercapnia leads to peripheral vasodilation, a bounding pulse, and a coarse flapping tremor of the outstretched hands. Severe hypercapnia causes confusion and progressive drowsiness. Papilloedema may be present but it is specific or sensitive for diagnosis. Patients in later-stage disease develop respiratory failure, pulmonary hypertension, and cor pulmonale.

Degree of breathlessness may be recorded using the Medical Research Council (MRC) dyspnoea score, while the COPD Assessment Test (CAT) is a patient-scored symptom tool that measures the disease's impact on individual health and wellbeing (**Box 20.17**).

Diagnosis

Diagnosis is usually clinical, based on history of breathlessness and sputum production in a chronic smoker. In the absence of cigarette smoking, asthma, and α_1-antitrypsin deficiency should be considered. No specific clinical feature is diagnostic. Lung hyperinflation also occurs in other diseases such as asthma and centri-acinar emphysema may be present without signs of over inflation. A 'barrel-shaped' but this can also occur in osteoporosis.

Investigations

- ***Lung function tests*** show evidence of airflow limitation (see **Fig. 20.10**). Post-bronchodilator FEV_1:FVC ratio is reduced to less than 0.7 (in contrast to asthma) and the PEFR is low. Airflow limitation is not fully reversible (change in FEV_1 <15%). Lung volumes may be normal or increased. Carbon monoxide gas transfer factor is low when significant emphysema is present. The severity of COPD is classified by the degree of airflow limitation (see **Box 11.39** in Acute medicine chapter).
- ***Chest radiograph*** is often normal, even in advanced disease. Possible features include over-inflation of the lungs with low, flattened diaphragms, and sometimes the presence of large bullae. Blood vessels may be 'pruned', with large proximal vessels and relatively little blood visible in peripheral lung fields.
- ***CT*** may be useful to confirm emphysema if there is diagnostic uncertainty. It may also be helpful to determine the distribution of emphysema and to look for coexistent pathology such as bronchiectasis.
- **A nuclear perfusion scan** may be useful in determining suitability for lung volume reduction surgery or endobronchial valve placement.
- ***Haemoglobin concentration and packed cell volume*** can be increased because of persistent hypoxaemia (see secondary polycythaemia, Chapter 38).
- ***Blood eosinophil count*** >300 cells per microlitre is associated with a greater risk of exacerbations and identifies benefit from inhaled corticosteroid therapy.
- ***Blood gases*** for evidence of respiratory failure.
- ***Sputum examination.*** Many acute episodes have a viral trigger. *S. pneumoniae*, *H. influenzae*, and *Moraxella catarrhalis* can also cause infective exacerbation.
- ***ECG*** is often normal. Pulmonary hypertension secondary to COPD leads to a tall P-wave (P pulmonale), right bundle branch block, and right ventricular hypertrophy.
- ***Echocardiography*** is indicated to assess cardiac function where there is disproportionate dyspnoea.
- ***α_1-Antitrypsin*** concentrations and genotype are measured in premature disease and life-long non-smokers.
- ***Bone densitometry*** is considered in those at risk of osteoporosis, e.g., frequent courses of oral corticosteroids.

Box 20.17 COPD Assessment Test (CAT)

The score ranges from 0–40. The higher the score, the greater the impact on the patient:

<10 low impact, 10–20 medium impact, 20–30 high impact, >30 very high impact. The CAT score helps to identify what symptoms may need to be treated and may guide the treatment plan.

(From Jones PW, Harding G, Berry P et al. Development and first validation of the COPD Assessment Test. Eur Respir J *2009; 34:648–654.)*

Management

See **Fig. 20.21** for management strategies.

Smoking cessation

The single most useful treatment is smoking cessation. Even in advanced disease, this may slow down the rate of deterioration and prolong time before disability and death (see **Fig. 20.20**). Smoke from burning biomass fuels in poorly ventilated homes should be reduced whenever possible.

Drug therapy

This is used both for short-term management of exacerbations and long-term relief of symptoms.

Bronchodilators

Inhaled bronchodilator therapy helps manage symptoms. When choosing the device, it is important to check inhaler technique to ensure appropriate delivery. Pressurized metered dose inhalers contribute to greenhouse gas emissions and should be avoided if an appropriate alternative can be used. If using a dry powder inhaler, it is important to assess peak inspiratory flow to ensure the necessary inspiratory flow to achieve effective drug delivery.

- ***β-Adrenoceptor agonists:*** Many patients with mild COPD feel less breathless after inhaling a β-adrenergic agonist such as salbutamol (200 μg every 4–6 h). In more severe airway limitation (moderate and severe COPD) a long-acting β_2 agonist (LABA) should be used.
- ***Antimuscarinic drug:*** Regular use of a long-acting antimuscarinic agent (LAMA) such as inhaled tiotropium improves lung function, symptoms of dyspnoea, and quality of life. Use of a LAMA does not prevent decline in FEV_1.
- ***Combination inhalers:*** LABA/LAMA inhalers are available and may aid adherence.
- ***Theophyllines:*** Long-acting preparations of theophylline are of little benefit in COPD.

Phosphodiesterase type 4 inhibitors

Roflumilast is a phosphodiesterase inhibitor with anti-inflammatory properties. It is used as an adjunct to bronchodilators for

maintenance treatment in those patients with an FEV_1 of less than 50% and chronic bronchitis.

Corticosteroids

Inhaled corticosteroids (ICS) are recommended in patients with frequent exacerbations or a FEV_1 of less than 50% predicted. Demonstration of a blood eosinophilia may identify patients more likely to have a beneficial response. High-dose inhaled steroids are not advised, as use is linked to increased rates of pneumonia. Triple therapy combination inhalers which include LABA/LAMA/ICS are available. Oral corticosteroids are only prescribed in the context of an acute exacerbation.

Antibiotics

Prompt antibiotic treatment shortens bacterial exacerbations and use may prevent hospital admission and further lung damage. Patients can be given antibiotics to keep at home, starting them as soon as sputum turns yellow or green. In patients who experience frequent exacerbations, long-term treatment with macrolide antibiotics such as azithromycin has been shown to reduce exacerbations and improve quality of life.

Mucolytic agents

These reduce sputum viscosity. Meta-analysis data shows that mucolytics such as carbocysteine are useful in preventing exacerbations in those who experience them frequently.

Biologic therapy

Currently, biologic therapies are not licensed for the treatment of COPD. Promising results in Phase III trials suggest they may emerge as a treatment option in the future.

Oxygen therapy

Two controlled trials (predominantly in males) showed improved survival with continuous administration of oxygen at 2 L/min via nasal prongs to achieve an oxygen saturation of more than 90% for large proportions of the day and night. Survival curves from these two studies are shown in Fig. 20.22. Only 30% of those not receiving long-term oxygen therapy survived for more than 5 years. A fall in pulmonary artery pressure was achieved if oxygen was given for 15

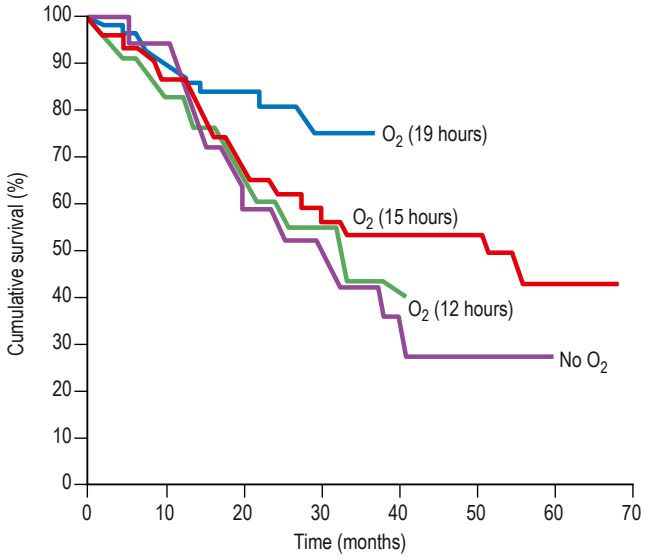


Fig. 20.22 Cumulative survival curves for patients with COPD receiving domiciliary oxygen. Oxygen doses are in hours per day.

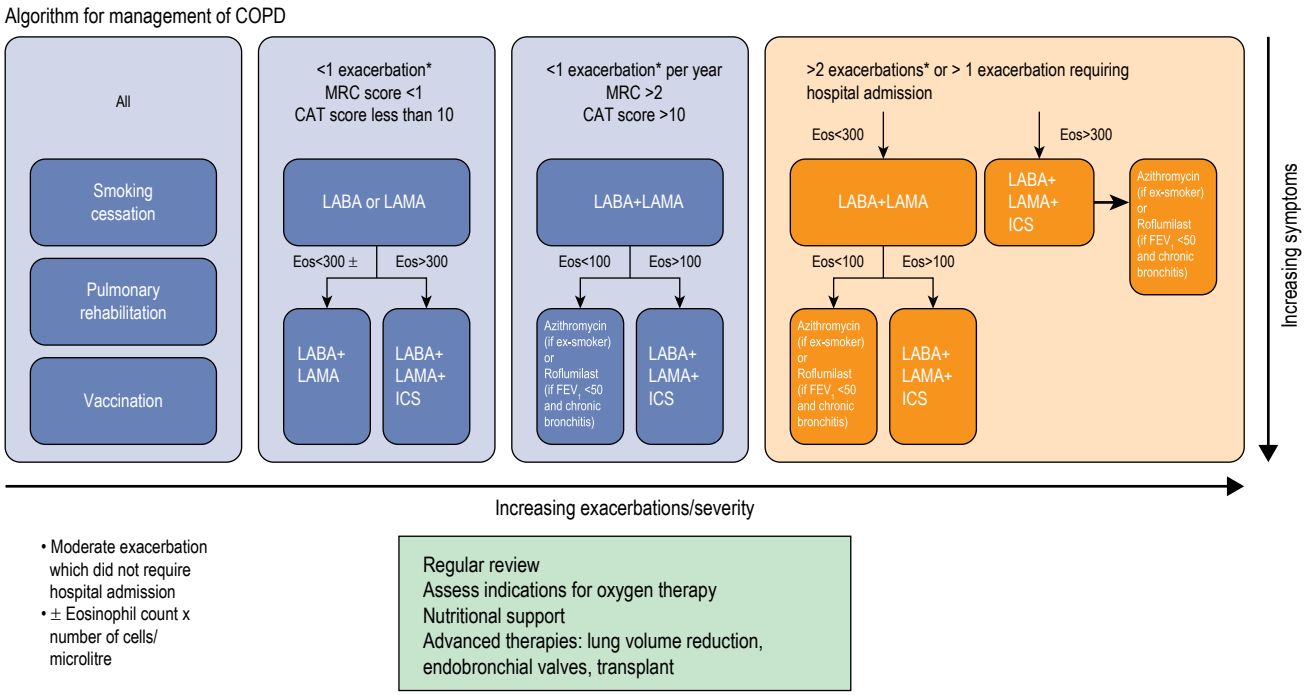


Fig. 20.21 Algorithm for the treatment of chronic obstructive pulmonary disease. The various components of management based on dyspnoea scores, exacerbation frequency, and eosinophil counts. (*Adapted from Pharmacological treatment of stable COPD, in Global Initiative for Chronic Obstructive Lung Disease (GOLD). Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease report 2023.*)

hours daily, but substantial improvement in mortality was achieved only by administration of oxygen for 19 hours daily. Domiciliary oxygen is best provided via an oxygen concentrator. Long-term continuous domiciliary oxygen therapy is available for patients with:

- P_aO_2 of <7.3 kPa (55 mmHg) when breathing air. Measurements should be taken on two occasions at least 3 weeks apart after appropriate bronchodilator therapy (Box 20.18)
- P_aO_2 of <8 kPa with secondary polycythaemia, nocturnal hypoxaemia, peripheral oedema, or evidence of pulmonary hypertension
- carboxyhaemoglobin of <3% (i.e., patients who have stopped smoking).

Pulmonary rehabilitation

Pulmonary rehabilitation courses are an essential part of management in COPD. Randomized control trials have shown benefits in improving symptoms of fatigue and dyspnoea, as well as exercise tolerance.

Additional measures

- ***Vaccines*.** Patients with COPD should be offered a single dose of the polyvalent pneumococcal polysaccharide vaccine and annual influenza vaccinations. SARS-CoV-2 vaccination is offered in line with national programmes.
- ***Nutrition*.** Additional metabolic requirements due to systemic inflammation and increased work of breathing should be assessed.
- ***α_1-antitrypsin replacement*.** Weekly or monthly infusions of α_1-antitrypsin are recommended for patients with serum concentrations <310 mg/L and abnormal lung function. Whether this modifies long-term progression has not been determined.
- ***Heart failure*.** This should be treated according to best practice (see Chapter 17). Systematic reviews and population studies that have assessed the use of beta-blockers in COPD suggest a reduction in mortality. Cardioselective beta-blockers have not been shown to affect lung function tests. Beta-blockers should be prescribed in line with cardiovascular risk factors with observation for breathlessness.
- ***Secondary polycythaemia*.** Venesection if packed cell volume >55%.
- ***Sensation of breathlessness*.** Breathing techniques and fan therapy may be helpful. Short-acting sedation such as sublingual lorazepam or opiates may be a helpful palliative measure for intractable dyspnoea.
- ***Air travel*.** Commercial aircraft are pressurized to the equivalent of 2000–2400 m altitude. In those without lung disease, this causes P_aO_2 to fall from 13.5–10 kPa and a 3% drop in oxygen saturation. Patients with moderate COPD may desaturate significantly. Air travel can be simulated by breathing 15% oxygen at sea level and if saturation drops below 85% within 15 min, supplemental oxygen is required for air travel.
- ***Surgery*.** Some patients have large emphysematous bullae that reduce lung capacity. Bullectomy theoretically enables adjacent areas of collapsed lung to re-expand, improving lung function. Lung volume reduction surgery in carefully selected patients with severe COPD (FEV_1 <1 L) has shown improved ventilation and a reduced sensation of breathlessness but no mortality benefit. A controlled trial in severe emphysema found no disease improvement and increased mortality. ***Single lung transplantation*** (see 'Lung and heart–lung transplantation' section) is considered in end-stage emphysema.
- ***Endobronchial valves*.** These occlude the airways of hyperinflated emphysematous lungs, effectively achieving lung volume reduction. Studies have shown an improvement in quality of life and exercise tolerance in selected patients.
- ***Palliative care*.** Specialist palliative care input can aid in the management of breathlessness and advanced care planning.

Acute exacerbation of COPD

The diagnosis and management of an acute exacerbation of COPD is addressed in Chapter 11 ('Exacerbation of chronic obstructive pulmonary disease [COPD]' section).

Prognosis of COPD

Poor prognostic factors include increasing age and worsening airflow limitation (decreasing FEV_1). The BODE (***B***ody mass index, degree of airflow ***O***bstruction, ***D***yspnoea, and ***E***xercise capacity) predictive index is shown in Box 20.19. A patient with a BODE index of 0–2 has a 4-year mortality rate of 10%, compared with 80% in someone with a BODE index of 7–10. This scoring tool may be useful in timing of referral for consideration of lung transplantation.

Box 20.18 Guidelines for home oxygen use in adults

- Chronic obstructive pulmonary disease with P_aO_2 <7.3 kPa when breathing air during a period of clinical stability
- Chronic obstructive pulmonary disease with P_aO_2 7.3–8 kPa in the presence of secondary polycythaemia, nocturnal hypoxaemia, peripheral oedema, or evidence of pulmonary hypertension
- Severe chronic asthma with P_aCO_2 <7.3 kPa or persistent disabling breathlessness
- Diffuse lung disease with P_aO_2 <8 kPa or P_aO_2 >8 kPa with disabling dyspnoea
- Cystic fibrosis with P_aO_2 <7.3 kPa or P_aO_2 7.3–8 kPa in the presence of secondary polycythaemia, nocturnal hypoxaemia, pulmonary hypertension, or peripheral oedema
- Pulmonary hypertension without parenchymal lung involvement with P_aO_2 <8 kPa
- Obstructive sleep apnoea despite continuous positive airways pressure therapy, after specialist assessment
- Pulmonary malignancy or other terminal disease with disabling dyspnoea
- Heart failure with a daytime P_aO_2 <7.3 kPa on air or with nocturnal hypoxaemia

(*British Thoracic Society guidelines, June 2015*)

Obstructive sleep apnoea

Obstructive sleep apnoea (OSA) is a form of sleep-disordered breathing characterized by upper airway collapse, resulting in obstructive apnoeas and hypopnoeas with desaturation. Prevalence is 3%–5% of the population, most common in overweight,

Box 20.19 BODE index: B = *B*ody mass index, O = degree of airflow *O*bstruction, D = *D*yspnoea, E = *E*xercise capacity

Variable	Points on BODE index			
	0	1	2	3
FEV_1 (% of predicted)	≥65	50–64	36–49	≤35
Distance walked in 6 min (m)	≥350	250–349	150–249	≤149
MMRC dyspnoea scale[a]	0–1	2	3	4
Body mass index	>21	≤21		

[a]Scores on the modified Medical Research Council (MMRC) dyspnoea scale range from 0 to 4, a score of 4 indicating that the patient is breathless when dressing.

Box 20.20 Symptoms in obstructive sleep apnoea (OSA)

Symptom	% of patients with OSA
Loud snoring	95
Daytime sleepiness	90
Unrefreshed sleep	40
Restless sleep	40
Morning headache	30
Nocturnal choking	30
Reduced libido	20
Morning 'drunkenness'	5
Ankle swelling	5

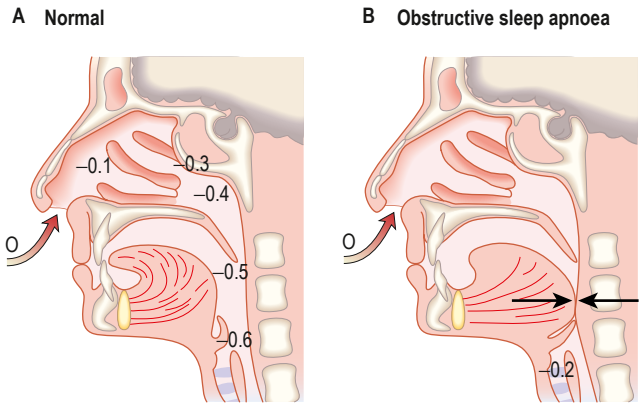


Fig. 20.23 Section through head, showing pressure changes in normal subjects and patients with obstructive sleep apnoea (OSA). **(A)** There is a pressure drop during inspiration as air is sucked through the turbinates. Changes are shown in kPa. **(B)** In OSA this is sufficient to collapse the pharynx *(arrowed on right)*, obstructing inspiration.

middle-aged males. Prevalence increases with age, menopause, obesity, and endocrine conditions such as acromegaly and hypothyroidism. OSA prevalence is predicted to rise with increasing obesity in the population. OSA can occur in children, particularly those with enlarged tonsils or trisomy 21. Major symptoms of OSA and their frequency are listed in Box 20.20.

Pathophysiology

During sleep, activity of respiratory muscles is reduced, especially during rapid eye movement (REM) sleep when the diaphragm is virtually the only active muscle. Apnoeas occur when the oropharyngeal airway closes in inspiration during sleep. This is prevented during waking by upper airway muscles (genioglossus and palatal muscles), but these become hypotonic during sleep (Fig. 20.23). Partial narrowing results in snoring, critical narrowing causes hypopnoea, and complete occlusion causes apnoea. Apnoea leads to hypoxia and increasingly strenuous respiratory efforts to overcome resistance. The combination of central hypoxic stimulation and the effort to overcome obstruction wakes the patient from sleep. These awakenings are so brief that patients may be unaware of them but may be woken hundreds of times per night, leading to sleep deprivation, daytime sleepiness, and impaired intellectual performance.

Correctable factors occur in about one-third of cases and include:

- ***encroachment on pharynx:*** obesity, acromegaly, enlarged tonsils
- ***nasal obstruction:*** nasal deformities, rhinitis, polyps, adenoids
- ***respiratory depressants:*** alcohol, sedatives, strong analgesia.

Box 20.21 Epworth sleepiness scale

How likely are you to doze off or fall asleep in the following situations, in contrast to just feeling tired? This refers to your usual way of life in recent times. Even if you have not done some of these things recently, try to work out how they would have affected you. Use the following scale to choose the most appropriate number for each situation.

0 = would never doze
1 = slight chance of dozing
2 = moderate chance of dozing
3 = high chance of dozing

Situation	Chance of dozing
Sitting and reading	_____
Watching TV	_____
Sitting and inactive in a public place (theatre or meeting)	_____
As a passenger in a car for an hour without a break	_____
Lying down to rest in the afternoon when circumstances permit	_____
Sitting and talking to someone	_____
Sitting quietly after lunch (without alcohol)	_____
In a car, while stopped for a few minutes in the traffic	_____
TOTAL	_____

Normal <9

Excessive daytime somnolence >9. Causes include obstructive sleep apnoea, environmental disruption, anxiety/depression, epilepsy, Parkinson's disease, narcolepsy, restless leg syndrome, periodic limb movement disorder, and sedating medication.

Box 20.22 STOP BANG tool

Question	Answer
S*noring*—do you snore?	Yes/No
T*iredness*—do you often feel tired, fatigued, or sleepy?	Yes/No
O*bserved apnoeas*	Yes/No
P*ressure*—do you have high blood pressure or are you on treatment for it?	Yes/No
B*ody mass index* >35	Yes/No
A*ge* >50 years	Yes/No
N*eck size* (≥male 43 cm, female ≥41 cm)	Yes/No
G*ender/sex*—male	Yes/No

Score: 0–2 = low risk, 3–4 = possible risk, 5–8 = high risk.

Diagnosis

A collateral history of a snore–silence–snore cycle may be given. Patients experience poor concentration and wake feeling unrefreshed. The Epworth Sleepiness Scale (Box 20.21) measures daytime sleepiness. The STOP BANG tool (Box 20.22) is a useful screening tool for OSA, facilitating appropriate referral for further investigation and may help discriminate OSA from simple snoring.

Investigations

If the diagnosis is suspected, investigation for sleep-disordered breathing is necessary. Many of these investigations can be

performed at home including overnight pulse oximetry and heart rate monitoring. Oximetry will demonstrate desaturations in a cyclical manner, which give a typical sawtooth appearance. The oximetry desaturation index (ODI) measures the number of desaturations per hour. Overnight oximetry will not differentiate central and obstructive apnoeas. Multichannel sleep studies (nocturnal polygraphy) measure body posture and movements, breathing rate, and electroencephalography alongside pulse oximetry. Polysomnography yields the ***apnoea–hypopnea index (AHI),*** the number of episodes of apnoea or hypopnea per hour, which determines severity:

- AHI <5: normal
- AHI 5–15: mild OSA
- AHI 15–30: moderate OSA
- AHI >30: severe OSA.

Management

Management consists of correction of treatable factors, including weight loss and alcohol reduction. Patients who have associated daytime somnolence should be offered continuous positive airway pressure (CPAP) during sleep. CPAP splints the upper airway such that it cannot occlude. It improves symptoms, quality of life, daytime alertness, and survival. In the event that an individual is unable to tolerate CPAP therapy, second-line treatment options considered on a case-by-case basis include custom-made mandibular advancement splints, ENT surgery, and hypoglossal nerve stimulation.

Chronic ventilatory failure

Chronic ventilatory failure is type II respiratory failure due to failure of alveolar ventilation leading to hypercapnia. This can be the sequela of several conditions (**Box 20.23**). Symptoms of hypercapnia include morning headaches, daytime somnolence, confusion, memory impairment, and unsteadiness.

Domiciliary non-invasive ventilation can be considered if a patient has persistent hypercapnia. Patients typically use a portable ventilator overnight, which provides positive pressure support at different levels during inspiration and expiration. Home mechanical ventilation improves survival in patients with COPD and persistent type II respiratory failure following an acute admission. There are no randomized trials of non-invasive ventilation in kyphoscoliosis; however, survival is improved with use. In patients with neuromuscular disorders, home non-invasive ventilation improves quality of life and prolongs survival.

? Box 20.23 Causes of chronic ventilatory failure

Airflow obstruction
- Chronic obstructive pulmonary disease

Restrictive lung disease
- Obstructive sleep apnoea/obesity hypoventilation syndrome
- Idiopathic pulmonary fibrosis
- Kyphoscoliosis
- Thoracoplasty

Neuromuscular disease
- Post-polio syndrome
- Diaphragm palsy
- Motor neuron disease
- Myotonic dystrophy

Central causes
- Brain injury
- Multiple system atrophy

Further reading

Boucher RC. Muco-obstructive lung diseases. *New Engl J Med* 2019; 380:1941–1953.
Murphy PB, Rehal S, Arbane G et al. Effect of home non-invasive ventilation with oxygen therapy vs oxygen therapy alone on hospital readmission or death after an acute COPD exacerbation: a randomised controlled trial. *JAMA* 2017; 317:2177–2186.
Suissa S and Drazen JM. Making sense of Triple Inhaled Therapy for COPD. *New Engl J Med* 2018; 1723–1724.
Zeng Z, Yang D, Huang X et al. Effect of carbocisteine on patients with COPD: a systematic review and meta-analysis. *Int J Chron Obstruct Pulmon Dis* 2017; 12:2277–2283.
https://goldcopd.org/. Global Initiative for Chronic Obstructive Lung Disease (2023 Report).
https://www.brit-thoracic.org.uk. *British Thoracic Society guidelines for home oxygen use in adults (2015)*.
https://www.catestonline.org. *COPD Assessment Test (CAT)*.

SMOKING

Prevalence

There are an estimated 1.3 billion tobacco smokers worldwide. Cigarette smoking is declining in the Western world but remains a leading cause of preventable death and remains prevalent in low- and middle-income countries. The World Health Organization (WHO) predicts that tobacco is directly responsible for the death of 8 million people each year. Information from the World Health Organization in 2020 reported that 36.7% of males and 7.8% of females worldwide used tobacco. The UK annual population survey of 2021 showed that 13.3% of adults smoke.

Toxic effects

Cigarette smoke contains polycyclic aromatic hydrocarbons and nitrosamines, which are potent carcinogens. Smokers have an increase in neutrophils and macrophages in the airways. These inflammatory cells release proteases that can destroy elastin and lead to lung damage. Pulmonary epithelial permeability increases, even in symptomless cigarette smokers, and correlates with the concentration of carboxyhaemoglobin in blood. This altered permeability may allow carcinogens easier access into the blood stream.

Dangers

Cigarette smoking is addictive and harmful to health (**Box 20.24**). Smoking 20 cigarettes daily for 20 years increases the lifetime

i Box 20.24 Dangers of cigarette smoking

General
- Cancer: lung, oesophagus, larynx, tongue, bladder
- Chronic obstructive pulmonary disease (COPD)
- Ischaemic heart disease
- Peripheral vascular disease
- Abnormal spermatozoa
- Cognitive impairment

Maternal smoking
- Fetal growth restriction
- Preterm delivery
- Childhood asthma

Passive smoking
- Asthma, pneumonia, and bronchitis in children of smoking parents
- Increased symptoms in asthma and COPD
- Cancer

risk of lung cancer by about ten times compared to a life-long non-smoker. Smoking and asbestos exposure are synergistic risk factors; the amount of asbestos exposure and smoking increase the risk of lung cancer risk considerably. Environmental tobacco smoke ('passive smoking') increases frequency and severity of asthma in children and may also increase incidence. It is also associated with a small but definite increase in lung cancer. Worldwide, second-hand smoke is estimated to cause premature death in 1.3 million people per annum. Surveys completed by the Global Youth Tobacco Survey in 142 countries over the period 2010–2018 showed a worldwide prevalence of second-hand smoke exposure (in adolescents) in any place was 62.9%. By 2022, 74 countries benefited from comprehensive smoke-free policies, which represents 38% of countries worldwide. Partial bans are in place in 35% countries, with no smoke-free legislation in 27%.

Smoking cessation

Smoking cessation has a huge effect on healthcare. National campaigns, bans on cigarette advertising, a substantial increase in the cost of cigarettes, and smoking bans in workplaces and public spaces contribute on a population level. Meanwhile, active encouragement to stop smoking is a useful approach for individuals: ask about smoking status, advise to quit, and refer to a smoking cessation programme for behavioural support. WHO reports state that 74 countries have comprehensive national smoke-free law which protects a quarter of the world's population.

Recommended approaches to smoking cessation include the '5A' approach:

1. Ask about smoking status at every healthcare visit
2. Advise: provide personalized clear advice
3. Assess willingness to quit
4. Assist in quitting
5. Arrange follow-up/support.

Pharmacological therapies have been shown to facilitate smoking cessation. Nicotine replacement therapies including gum, lozenges, inhalator, nasal spray, and patches have been shown to improve cessation over placebo. Varenicline is an oral partial agonist of the $\alpha_4\beta_2$ subtype of the nicotinic acetylcholine receptor. It stimulates the nicotine receptor, reducing withdrawal symptoms and craving for cigarettes. A 12-week course doubles the chances of smoking cessation. Therapy should be changed after 3 months if abstinence is not achieved.

The role of e-cigarettes remains unclear, and their long-term effects are unknown. Bupropion and nortriptyline have been shown to improve sustained quit rates and can be used as part of a smoking cessation intervention programme.

Further reading

Christiani DC. Vaping induce acute lung injury. *N Engl J Med* 2020; 382:960–962.
Colditz GA. Smoke alarm—tobacco control remains paramount. *N Engl J Med* 2015; 372:665–666.
Chuanwei Ma, Emerald G Heiland, Zilin Li et al. Global trends in the prevalence of secondhand smoke exposure among adolescents aged 12–16 years from 1999 to 2018: an analysis of repeated cross-sectional surveys. *Lancet Glob Health* 2021; 9:e1667–e1678.
Olivieri M, Murgia N, Carsin AE et al. Effects of smoking bans on passive smoking exposure at work and at home. The European Community respiratory health survey. *Indoor Air* 2019; 29(4):670–679.
Štěpánek L, Ševcíková J, Horáková D, Patel MS, Durd'áková R. Public health burden of secondhand smoking: case reports of lung cancer and a literature review. *Int J Environ Res Public Health* 2022.
Tobacco: preventing uptake, promoting quitting and treating dependence. NICE guideline Published: 30 November 2021 Last updated: 16 January 2023.
World Health Organization report on the Global Tobacco Epidemic 2023.

RESPIRATORY INFECTION

Pneumonia

The presentation, management, and complications of pneumonia are detailed in Chapter 11 ('Pneumonia' section).

COVID-19 pneumonitis

SARS-CoV-2 is a novel coronavirus which emerged in 2019 leading to COVID-19 disease. The virus mutated rapidly and produced several variants. It primarily causes respiratory symptoms in adults. In some, it causes coryzal-like symptoms with high temperature, myalgia, cough, and runny nose. In the early waves of the pandemic, the virus classically caused loss of smell (anosmia) and loss of taste (ageusia) which could last for months after general recovery. The initial variants showed a predisposition to develop a rapidly progressive pneumonitis. Certain groups of patients were much more susceptible to progressing to pneumonitis and respiratory failure: older age, people from minority ethnic backgrounds, patients with diabetes, hypertension, and kidney disease, those living with obesity, and those with immunodeficiency of any cause.

Symptoms and signs of severe COVID-19 pneumonitis are cough, progressive shortness of breath, and severe hypoxia in a conscious and talking patient; the clinical state can rapidly change to unconsciousness and death from acute and rapid respiratory failure. Chest radiograph appearances showed widespread confluent shadows throughout the lung fields which clear completely on recovery. COVID-19 pneumonitis is more likely to be associated with thromboembolic disease and, when slow to recover, a superadded bacterial infection should be considered. Patients can be left with residual pulmonary fibrosis, which appears to be more common in those needing mechanical ventilation. Treatment for respiratory failure is predominantly supportive. The Recovery trial demonstrated that in those that required oxygen therapy, treatment with dexamethasone for 10 days reduced 28-day mortality. The Recovery RS trial found that CPAP therapy reduced the need for invasive ventilation in those with respiratory failure. Additional management strategies included appropriate thromboprophylaxis, and judicious fluid balance. In selected patients, additional treatment options such as antiviral or IL-6 inhibitors may be appropriate.

Respiratory symptoms associated with long COVID are unresolved breathlessness and chronic cough in the presence of otherwise normal imaging and respiratory function tests. Cough can be helped by low-dose steroid inhalers and gradually settles. Patients with breathlessness should be encouraged to build up from a gentle exercise regimen and respiratory physiotherapy is the main stay of treatment if no other cause is found.

Further reading

Perkins GD et al. Effect of noninvasive respiratory strategies on intubation or mortality among patients with acute hypoxemic respiratory failure and COVID-19: The RECOVERY-RS randomized clinical trial. *JAMA* 2022; 327(6):546–558.
The RECOVERY collaborative group. Dexamethasone in Hospitalized patients with COVID-19. *N Engl J Med* 2021; 384:693–704.

Pneumocystis jirovecii pneumonia

Pneumocystis pneumonia is one of the most common opportunistic infections encountered in clinical practice. It affects patients on immunosuppressant therapy, including prolonged corticosteroids, monoclonal antibodies, methotrexate, anti-rejection medication after a solid organ or haemopoietic stem cell transplantation; and those infected with HIV. Individuals with CD4 counts of less than

200/mm^3 are at particular risk. *Pneumocystis jirovecii* is found in the air, and pneumonia arises from re-infection rather than reactivation of persisting organisms acquired in childhood.

Pneumocystis pneumonia is associated with a high fever, breathlessness, and dry cough. Rapid desaturation on exercise or exertion is characteristic. The typical radiographic appearance is diffuse bilateral alveolar and interstitial shadowing beginning in the perihilar regions and spreading out in a butterfly pattern. Other chest radiograph appearances include localized infiltration, nodules, cavitation, or pneumothorax. Empirical treatment is justified in sick, high-risk patients. Whenever possible, diagnosis should be confirmed by indirect immunofluorescence on induced sputum or PCR of bronchoalveolar lavage fluid. Increasingly measurement of serum beta-D-glucan (BDG) may assist in the diagnosis of PCP. BDG is raised in mycological infections as it is present in the cell wall of pathological fungi. First-line treatment of *Pneumocystis* pneumonia is with high-dose co-trimoxazole, with adjunctive corticosteroids in patients with HIV infection.

Further reading

Mansell A, Niederman LA. Aspiration pneumonia. *N Engl J Med* 2019; 380:651–663.
National Institute for Health and Care Excellence. *Clinical Guideline 191: Pneumonia in Adults: Diagnosis and Management*. NICE 2014 (update 2022); https://www.nice.org.uk/guidance/cg191.
Wunderick RG, Waterer G. Advances in the causes and management of community acquired pneumonia in adults. *BMJ* 2017; 358:j2471.

Tuberculosis

Tuberculosis (TB) is one of the world's most common infectious diseases. It is caused by the bacterium *Mycobacterium tuberculosis*. It is estimated that one-third of the world's population is infected with tuberculosis, with most cases (around 65%) in Africa and Asia. There is a growing incidence of multidrug-resistant and extremely drug-resistant strains, which, together with HIV co-infection, places a huge health burden on resource-poor nations, with a high mortality from the two coexistent diseases. TB was responsible for 1.5 million deaths worldwide in 2020 and 20% of these were in HIV co-infected individuals. There are several factors affecting prevalence and risk of TB (**Box 20.25**).

Pathogenesis

M. tuberculosis is an aerobic, intracellular pathogen. Due to their relative impermeability to acid-based dyes in the laboratory, these organisms are often termed 'acid-fast bacilli'. TB is an airborne infection spread by coughing via respiratory droplets. Only a small number of bacteria need to be inhaled for infection to develop, but not all those who are infected develop active disease. On initial contact with infection, less than 5% of patients develop active disease. This increases to 10% within the first year of exposure.

Primary tuberculosis

'Primary TB' describes the first infection with TB. When the bacteria reach the alveolar macrophages, they are ingested and the subsequent inflammatory reaction results in tissue necrosis and formation of a granuloma. These granulomatous lesions consist of a central area of necrotic material called caseation, surrounded by epithelioid cells and Langhans giant cells. Subsequently, the caseated areas heal and many calcify. Some of these calcified nodules contain bacteria, which are contained by the immune system (and the hypoxic acidic environment created within the granuloma) and are capable of lying dormant for many years. This is the primary focus or the 'Ghon focus' of the disease. On a chest radiograph, the Ghon focus can be evident as a small, calcified nodule.

Box 20.25 Risk factors for tuberculosis in high-income countries

Contact with high-risk groups
- Origination from a high-incidence country (defined as >40/100 000)
- Frequent travel to high-incidence areas

Immune deficiency
- HIV infection
- Corticosteroids or immunosuppressant therapy
- Chemotherapeutic drugs
- Nutritional deficiency
- Diabetes mellitus
- Chronic kidney disease
- Malnutrition

Lifestyle factors
- Drug/alcohol misuse
- Homelessness/hostels/overcrowding
- Prison inmates

Genetic susceptibility
- Twin studies of gene polymorphisms

Box 20.26 Factors implicated in the reactivation of latent tuberculosis

- HIV co-infection
- Immunosuppressant therapy: corticosteroids, chemotherapy, monoclonal antibody treatment
- Diabetes mellitus
- Kidney failure requiring renal replacement
- Malnutrition
- Old age

Reactivation tuberculosis

In most people infected by *Mycobacterium* spp., the immune system contains the infection, and the patient develops cell-mediated immune memory of the bacteria. This is termed 'latent TB'. Most active TB cases are due to reactivation of latent infection, where the initial contact usually occurred many years earlier. Most patients are young and previously healthy but may have one or more risk factors for development of active disease (**Box 20.26**). In patients with HIV infection, newly acquired TB is common. The majority of active TB occurs in the lung, but extrapulmonary infection occurs with spread to the lymph nodes, particularly cervical and intrathoracic chains where it causes active disease in 20%–25% in the UK, and also via the bloodstream to more distant sites including brain, bones, and skin.

Clinical features and diagnosis

Cardinal symptoms of TB are cough, haemoptysis, and systemic features including fevers, night sweats, and weight loss. With infection in extrapulmonary sites, respiratory symptoms are often absent, and systemic symptoms may be neglected by both patients and healthcare practitioners.

In all cases of suspected TB, it is essential, depending on the site of disease, to obtain sputum, biopsy samples, or fluid for microscopy, smear, and culture, for information on sensitivity. Tissue samples should also be sent for histopathological examination.

Pulmonary TB

Patients are frequently symptomatic with a productive cough and, occasionally haemoptysis, along with systemic symptoms. Where

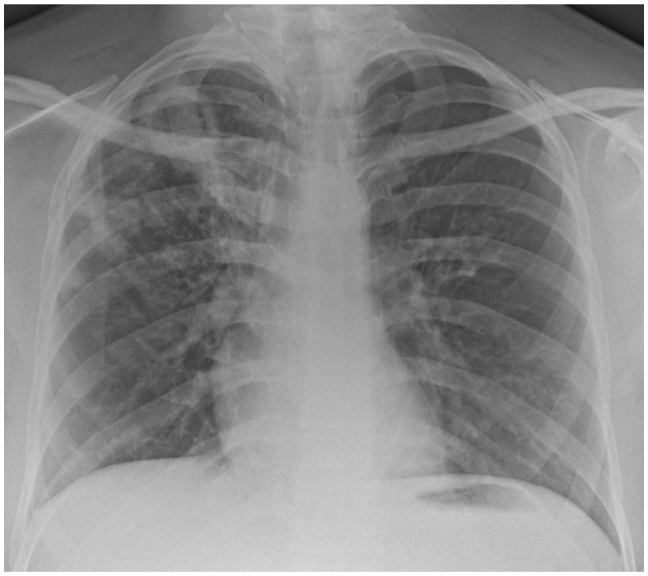


Fig. 20.24 Chest radiograph demonstrating patchy right mid and upper lobe consolidation and a cavity at the right apex in a patient with tuberculosis.

there is laryngeal involvement, a hoarse voice and a severe cough are found. If disease involves the pleura, pleuritic pain is frequent.

The chest radiograph (Fig. 20.24) can show consolidation with or without cavitation, pleural effusion, or thickening or widening of the mediastinum caused by hilar or paratracheal adenopathy.

Serial sputum samples should be collected on at least three occasions (ideally, first thing in the morning). If the patient is unable to produce sputum, induced sputum or bronchoscopy may be needed to obtain samples.

Patients whose sputum is smear-positive for TB are considered infectious and should be isolated when in hospital. Those who are smear-negative but subsequently culture-positive are less infectious and generally do not need to be isolated, although care should be taken when contacts include immunocompromised individuals.

Lymph node TB

The lymph nodes are the second most commonly affected organs. Extrathoracic nodes are more commonly involved than intrathoracic or mediastinal nodes. Usually, presentation is with firm, non-tender enlargement of a cervical or supraclavicular node. The node becomes necrotic centrally and can liquefy and be fluctuant if peripheral. The overlying skin is frequently indurated and there can be sinus tract formation with purulent discharge. Characteristically there is no erythema ('cold abscess'). Nodes are typically enlarged for several months prior to diagnosis. On CT imaging, the central area appears necrotic (Box 20.27). Samples should be obtained via either ultrasound-guided fine needle aspiration (FNA), core biopsy, or node excision. All samples must be sent for AFB, smear and culture, with cytology to help exclude malignancy. EBUS can be used to biopsy intrathoracic nodes.

Miliary TB

Miliary disease occurs through haematogenous spread of the bacilli to multiple sites, including the central nervous system (CNS) in 20% of cases. It often presents with systemic symptoms. The chest radiograph demonstrates multiple nodules, which appear like millet seeds, hence the term 'miliary'. Other findings are liver and splenic microabscesses with deranged liver enzymes, cholestasis, and gastrointestinal symptoms. All patients should have brain imaging to look for evidence of cerebral disease, which can present as an asymptomatic brain tuberculoma.

Microbiological diagnosis

Once samples have been taken, the rapid identification of the presence of mycobacterium by stains is essential and should be performed within 24 hours. Culture allows determination of antibiotic sensitivity.

Stains

Auramine–rhodamine staining is more sensitive and therefore more widely used than Ziehl–Neelsen staining. It requires fluorescence microscopy and highlights bacilli as yellow–orange on a green background.

Culture

The majority of high-income countries uses liquid/broth culture of mycobacteria in addition to solid media (Lowenstein–Jensen slopes or Middlebrook agar), as time to culture is shorter (1–3 weeks) for liquid culture compared to solid culture (3–8 weeks). Using liquid culture in the presence of antimycobacterial drugs (usually first-line therapy) establishes the sensitivity.

Nucleic acid amplification

Nucleic acid amplification testing (NAAT) is increasingly used for rapid identification of MTb complex and is useful in differentiating between *M. tuberculosis complex* and non-tuberculous mycobacteria (NTM), as well as identifying TB in smear-negative sputum specimens. It works by using the polymerase chain reaction (PCR) to replicate and then identify mycobacterium DNA. PCR is useful only at the initial stage of diagnosis, as it frequently remains positive despite treatment, due to detection of dead organisms. Culture and staining are still necessary.

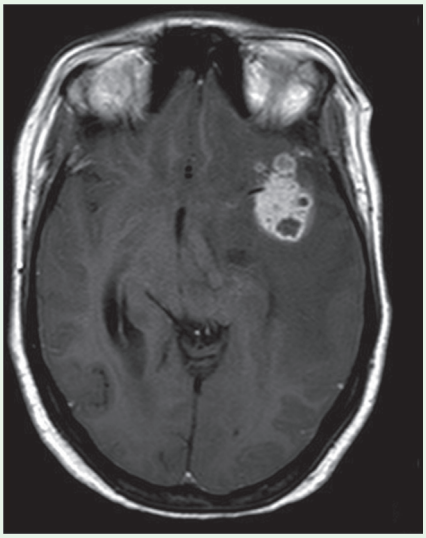
Identification of mycobacterial DNA is useful in facilitating rapid treatment including identification of drug resistance. Genetic mutations conferring rifampicin resistance are highly predictive of drug resistance. The development of a highly specific probe designed to detect this mutation thereby allows efficient identification of resistant disease and commencement of appropriate therapy sooner than waiting for culture results. Commercial kits, such as GeneXpert, reliably perform molecular testing in the field and take less than 2 hours. They are now widely used in low- and middle-income countries to detect mutations associated with rifampicin resistance.

Whole-genome sequencing (WGS) can identify different strains of mycobacterium and also detect drug resistance. This is starting to replace routine culture and phenotypic testing.

Management

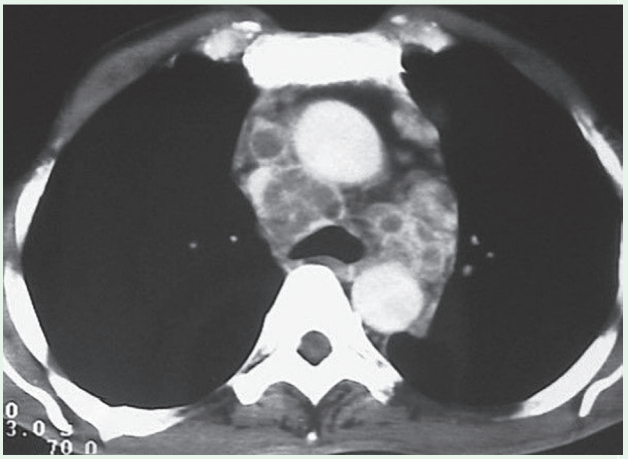
All patients should have routine blood tests, a viral hepatitis screen, and be offered testing for HIV before treatment. Patients with active viral hepatitis have an increased risk of fatal drug-induced hepatitis and need careful monitoring and counselling. Those with fully sensitive TB require 6 months of treatment; extended to 12 months for CNS infection. Isoniazid, rifampicin, pyrazinamide, and ethambutol are the first-line TB drugs, known as quadruple therapy. In CNS and pericardial disease, corticosteroids are used as an adjunct at treatment initiation to reduce long-term complications. Box 20.28 summarizes standard recommended regimens. Enhanced case management, together with ***directly observed therapy (DOT)***, is recommended where there are concerns about adherence (Box 20.29).

Box 20.27 TB infection: radiological findings and diagnostic investigations

Radiology	Investigations
Pulmonary, pleural and laryngeal TB	
Chest radiograph of suspected pulmonary TB	Smear and culture of: • Sputum: ≥2 samples increases diagnostic yield • Induced sputum: e.g., inhaled hypertonic saline, diagnostic yield comparable to bronchoscopy • Bronchoalveolar lavage: if cough unproductive and induced sputum not possible • Aspiration of pleural fluid and pleural biopsy • Gastric aspirates: useful in paediatric disease • Laryngeal disease: naso/endoscopic examination and biopsy
Miliary TB	
Chest radiograph with miliary pulmonary changes	Blood cultures for AFB Bronchoalveolar lavage fluid: usually smear-negative but culture-positive MRI head and lumbar puncture in all cases, unless contraindicated, to assess for central nervous system involvement (affects treatment duration) Sampling of other involved organs as indicated
Central nervous system TB	
 MRI head with lesion in right cerebellum	Neurology/neurosurgery MDT discussion Lumbar puncture if no contraindication: • CSF protein very high (usually >2–3 g/L) • CSF glucose <½ blood glucose • CSF lymphocytosis Biopsy if concerns about diagnosis

(Continued)

Box 20.27 TB infection: radiological findings and diagnostic investigations—cont'd

Radiology	Investigations
Lymph node TB	
 CT thorax with left anterior mediastinal nodes with necrotic centre	Fine needle aspiration/biopsy of lymph node: culture, smear, and histopathological examination. Mediastinal nodal sampling: EBUS, transbronchial needle aspiration, mediastinoscopy/mediastinotomy

CSF, Cerebrospinal fluid; *EBUS,* endobronchial ultrasound; *MDT,* multidisciplinary team.

Box 20.28 Usual treatment and duration in fully sensitive tuberculosis

Site of disease	Duration of therapy	Drug choice
Pulmonary[a] **Extrapulmonary** (excluding CNS) **Miliary** (excluding CNS)	6 months	2 months HRZE then 4 months HR[a]
CNS TB	12 months	2 months HRZE then 10 months HR Plus corticosteroids, e.g., dexamethasone 0.3–0.4 mg/kg/day weaned over 8–12 weeks
Latent TB	3 or 6 months	3 months RH OR 6 months H

[a]For pulmonary TB: 3 times weekly dosing for 4 months HR, is acceptable.
CNS, Central nervous system; *E,* ethambutol; *H,* isoniazid; *R,* rifampicin; *Z,* pyrazinamide.

Box 20.29 Criteria for implementation of directly observed therapy (DOT) for tuberculosis

- Adherence factors
  - History of non-adherence to TB or other therapies
  - Previous TB treatment
  - Mental illness impacting adherence
- Multidrug-resistant TB
- Street-, shelter-, or hostel-dwelling homelessness
- Prison history
- Substance misuse

The most common side-effects of quadruple TB therapy are nausea, vomiting, rash, and itching. An antiemetic or antihistamine can be prescribed to alleviate these symptoms, although in some cases treatment may need to be interrupted. If liver function tests become deranged and there is concern about a drug-induced hepatitis, it is often necessary to stop all four drugs and re-introduce one at a time. Drugs should be stopped only if the serum bilirubin becomes elevated or if transaminases are more than three times greater than the upper reference limit.

Isoniazid can cause a polyneuropathy due to B_6 deficiency, as isoniazid interacts with pyridoxal phosphate. Vitamin B_6 (pyridoxine 10–25 mg daily) should be prescribed concomitantly to prevent this. Occasionally, isoniazid gives rise to allergic reactions, such as a skin rash and fever. Hepatitis occurs in less than 1% of cases but may lead to liver failure, transplantation, or death if the drug is continued.

Rifampicin induces liver enzymes, which may be transiently elevated in the serum of many patients. Concomitant drug treatment metabolized by the same enzymes may therefore be less effective and a careful review of other medications is required, particularly antidepressants, anticoagulants, and antiepileptics. Oral and subdermal contraception may be less effective, so alternative birth-control methods should be used (e.g., depot-medroxyprogesterone acetate, intrauterine device). Rifampicin stains body secretions red/pink and patients should be warned of change in colour of urine, tears (affecting contact lenses), and sweat. Thrombocytopenia has been reported.

Pyrazinamide has common side-effects including itch, rash, and arthralgia. Renal excretion of urate is reduced and hyperuricaemic gout may develop. Hepatic toxicity is rare.

Ethambutol can cause a dose-related optic retrobulbar neuritis that presents with colour blindness for green, reduction in visual acuity, and a central scotoma. Patients should have visual acuity and colour vision checked prior to treatment using Snellen and Ishihara charts. This condition usually reverses provided the drug is stopped when symptoms develop; patients should therefore be warned of its effects. An urgent ophthalmology opinion should be sought if optic symptoms develop.

Drug resistance

Mono- or multidrug resistance arises due to incomplete or incorrect drug treatment and can be spread from person to person. Isoniazid monoresistance exists in approximately 10% of TB cases in the UK. A risk assessment for drug resistance should be routine (**Box 20.30**). The incidence of multidrug resistance in TB (resistance to both rifampicin and isoniazid, termed MDR-TB) is relatively low in high-income countries (around 1%). Extensively drug-resistant disease (XDR-TB) is

Box 20.30 Factors associated with an increased risk of drug-resistant tuberculosis

- History of prior drug treatment of TB (particularly if unsupervised/self-administered)
- Co-infection with advanced HIV
- Infection acquired in a region with high rates of drug resistance
- Contact with a known case of resistant TB
- Failure to respond to empirical TB therapy despite documented adherence
- Exposure to multiple courses of fluoroquinolone antibiotics for presumed community-acquired pneumonia
- Healthcare workers exposed to cases of resistant TB

defined as high-level resistance to rifampicin, isoniazid, any fluoroquinolone, and one additional WHO Group A drug (drugs effective in MDR-TB including levofloxacin or moxifloxacin, bedaquiline, and linezolid).

TB in special situations

Mycobacterium bovis infection

Mycobacterium bovis infection occurs in humans who have consumed unpasteurized milk, farmers working with infected cows, and abattoir workers. TB due to *M. bovis* does not differ from ordinary TB in the chest but extrapulmonary sites of infection, such as lymph nodes, are more frequently involved. Immunosuppression is also a risk factor. The tuberculin test is positive. Treatment is with isoniazid, rifampicin, and ethambutol; pyrazinamide resistance is common.

HIV co-infection

Increased TB prevalence over recent decades has occurred in association with HIV co-infection, especially in sub-Saharan Africa, the Indian subcontinent, and parts of Eastern Europe and Russia. The incidence of HIV infection in TB worldwide is around 10% and TB is responsible for about one-quarter of acquired immunodeficiency syndrome (AIDS)-related deaths. Alongside the increased morbidity and mortality of co-infection, there are specific issues relating to the treatment of TB with HIV co-infection including drug interactions, intolerability, and increased treatment toxicity, drug resistance, and immune reconstitution reactions. TB/HIV infection should be managed by experts in TB (respiratory or infectious disease physicians) alongside HIV specialists.

Latent TB infection

Latent TB infection (LTBI) is diagnosed by demonstrating immune memory to mycobacterial proteins. Two tests are available:

1. ***Tuberculin skin test ('Mantoux test').*** A positive result is indicated by a delayed hypersensitivity reaction (raised, indurated lesion) 48–72 h after intradermal injection of purified protein derivative (PPD). False-negative (anergic) results are common in patients with immunosuppression due to HIV infection ($CD4^+$ <200/mm^3), those taking immunosuppressant medications (chemotherapy, anti-TNF therapy, steroids), those at the extremes of age, and those with active disease. False-positives occur due to cross-reactivity with NTM and bacille Calmette–Guérin (BCG) vaccination.
2. ***Interferon-gamma release assays (IGRAs)*** detect T-cell secretion of IFN-γ following exposure to *M. tuberculosis*–specific antigens (ESAT-6, CFP-10). Where a person has been previously infected or is currently infected with TB, activated T cells within their extracted whole blood secrete quantifiable levels of IFN-γ in response to re-exposure to TB-specific antigens. The test does not differentiate between active and latent infection. However, it is highly specific compared to tuberculin skin testing, with similar or better sensitivity, and requires only a single patient visit.

In certain groups with LTBI, ***chemoprophylaxis*** is offered to reduce the risk of active infection, depending on whether a high-burden or low-burden country:

- household and close workplace contacts of patients with pulmonary and laryngeal TB
- health workers
- recent new migrant entrants from high-risk countries
- patients who are immunocompromised
- those about to commence treatment with biologic agents (such as infliximab)
- those due to have stem cell or solid organ transplants.

LTBI ***treatment*** is either with isoniazid and rifampicin for 3 months or with isoniazid monotherapy for 6 months.

BCG vaccination

BCG is a live attenuated vaccine derived from *M. bovis* that has lost its virulence. Though it is no longer offered routinely to all due to the lack of cost efficacy, it is recommended to some high-risk populations. In the UK, it is offered to neonates and children living in high-prevalence areas in the UK or immigrating from high-prevalence areas, and to close contacts of those with infectious TB.

Non-tuberculous mycobacterial infection

Non-tuberculous mycobacteria (NTM) occur in soil and water, and are not usually pathogenic due to lack of virulence. However, where there is a breach of normal host defence mechanisms, certain strains have the potential to become pathogenic (**Box 20.31**). Factors associated with an increased risk of pulmonary NTM infection include structural lung disease such as COPD, bronchiectasis and CF, and immunosuppression. Treatment is indicated if there is a compatible clinical context, or isolation of the organism from an invasive sample or more than one sputum sample obtained at different times.

Further reading

Conradie F, Diacon AH, Ngubane N et al. Treatment of highly drug-resistant pulmonary tuberculosis. *N Engl J Med* 2020; 382:893–902.
National Institute for Health and Care Excellence. *NICE Clinical Guideline NG33: Tuberculosis*. NICE 2016 (updated 2019); https://www.nice.org.uk/guidance/ng33.
Tiberi S, du Plessis N, Walzl G et al. Tuberculosis: progress and advances in development of new drugs, treatment regimens, and host-directed therapies. *Lancet Infect Dis* 2018; 18:e199–e210.
Walzl G, McNerney R, Du Plessis N et al. Tuberculosis: advances and challenges in development of new diagnostics and biomarkers. *Lancet Infect Dis* 2018; e183–e198.
World Health Organization. *Global Tuberculosis Report 2022*; http://www.who.int/.

PLEURAL DISEASE

Pleural effusion

A pleural effusion is an excessive accumulation of fluid in the pleural space. It can be detected on chest radiography when 300 mL or more of fluid is present, and clinically when there is 500 mL or more. Chest radiograph appearances (**Fig. 20.25**) range from obliteration of the costophrenic angle to dense homogeneous shadows occupying part or all of the hemithorax. Fluid below the lung (a subpulmonary effusion) can simulate a raised hemidiaphragm. Fluid in a lung fissure may resemble an intrapulmonary mass. The physical signs are described in **Box 20.8**.

Diagnosis

Pleural aspiration (see Box 11.7) is usually done under ultrasound guidance. Fluid may be a transudate or an exudate.

Transudates

Transudates can be bilateral but are often larger on the right. A pleural transudate is defined by pleural protein content of less than 30 g/L (provided serum protein concentration is normal), LDH less than 200 IU/L, and fluid to serum LDH ratio less than 0.6. If the serum protein is less than 30 g/L, then Lights criteria (see **Box 20.32**) should be used to determine if the sample is a transudate or an exudate.

Causes include:

- heart failure
- hypoproteinaemia (e.g., nephrotic syndrome)
- constrictive pericarditis
- hypothyroidism
- ovarian tumours producing right-sided pleural effusion (Meigs' syndrome).

Box 20.31 Some non-tuberculous mycobacteria strains implicated in disease

Strain	Site of disease
M. avium intracellulare complex (MAC)	Pulmonary
	Disseminated (usually in immunosuppression/HIV)
	Hypersensitivity pulmonary disease ('hot-tub lung')
	Lymphadenitis in children
M. kansasii	Pulmonary
	Disseminated disease (in immunosuppression/HIV)
M. abscessus	Skin, soft tissue, and bone disease
	Pulmonary
M. chelonae	Skin, bone, and soft tissue
	Pulmonary: similar to *M. abscessus*
M. fortuitum	Pulmonary: similar to *M. abscessus*
M. gordonae	Rarely pathogenic: can be significant in immunocompromised host
M. xenopi	Pulmonary: fibrocavitary disease in chronic obstructive pulmonary disease
	Contaminated surgical instruments causing bone/soft tissue infection
M. malmoense	Pulmonary
	Lymph node
M. marinum	Soft tissue, skin, and bone
M. szulgai	Pulmonary: similar to *M. tuberculosis*
M. chimaera	Pulmonary, prosthetic valve, bone, disseminated—linked to contaminated heater coolers used in heart bypass surgery

Exudates

The protein content of exudates is over 30 g/L and the LDH is more than 200 IU/L. Causes include:

- bacterial pneumonia (common)
- carcinoma of the bronchus—fluid may be blood-stained (common)
- pulmonary infarction
- TB
- autoimmune connective tissue diseases
- mesothelioma
- post-myocardial infarction syndrome (rare)
- acute pancreatitis (high amylase content) (rare)
- sarcoidosis (very rare)
- yellow-nail syndrome (effusion due to lymphoedema, very rare)
- familial Mediterranean fever (very rare).

Pleural biopsy may be necessary if diagnosis is not established by simple aspiration and clinical features.

Box 20.32 Light's criteria to diagnose an exudative effusion

- Pleural fluid protein: serum protein >0.5
- Pleural fluid LDH: serum LDH >0.6
- Pleural fluid LDH >2/3 upper limit of normal for serum (105–333 IU/L)

(Adapted from Light RW, Macgregor MI, Luchsinger PC et al. Pleural effusions: the diagnostic separation of transudates and exudates. Ann Intern Med *1972; 77:507–513).*

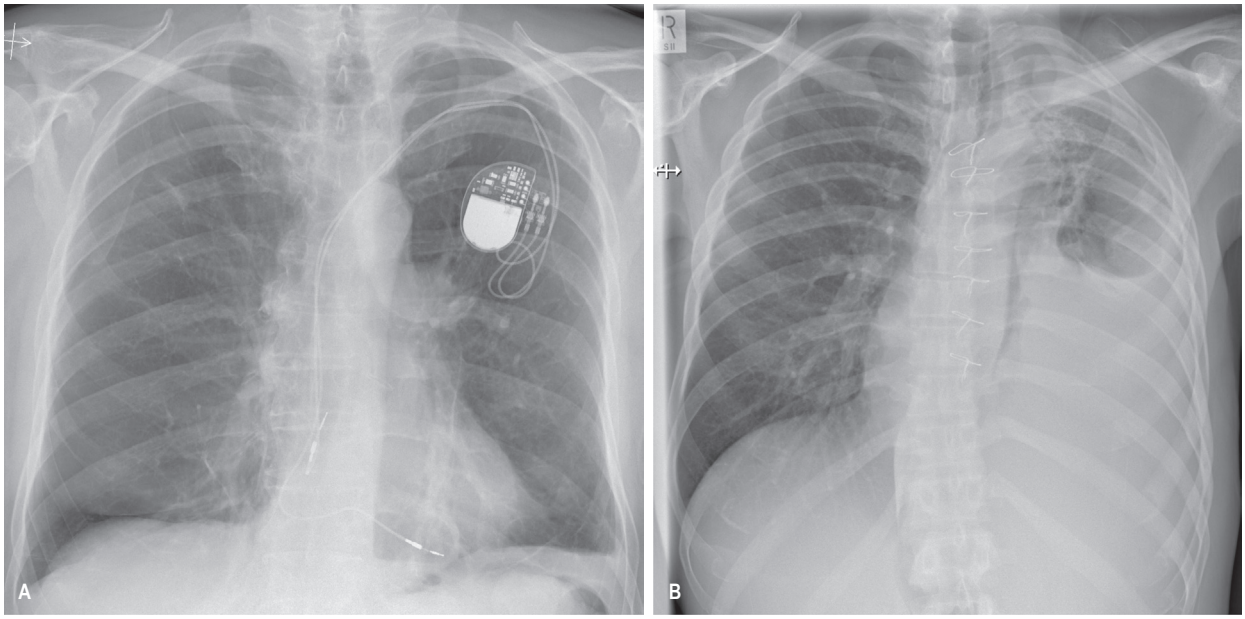


Fig. 20.25 Chest radiograph demonstrating pleural effusion. **(A)** Blunting of the left costophrenic angle due to a small left pleural effusion. There is a dual-lead pacemaker *in situ*. **(B)** Large left-sided pleural effusion.

Management

Treatment is of underlying condition.

Empyema

The presence of pus in the pleural space can be a complication of pneumonia. It requires urgent drainage.

Malignant pleural effusion

Malignant pleural effusions that reaccumulate and are symptomatic can be aspirated to dryness followed by the instillation of a sclerosing agent such as talc. Effusions should be drained slowly since rapid shift of the mediastinum causes pain and occasionally shock. For longer-term relief, indwelling pleural catheters (IPC) are being increasingly utilized for recurrent malignant pleural effusions in the community and can be managed by patients, their relatives, and/or community nursing teams.

Chylothorax

This is caused by accumulation of lymph in the pleural space, usually resulting from leakage from the thoracic duct following trauma or infiltration by carcinoma.

TUMOURS OF THE RESPIRATORY TRACT

Malignant tumours

Bronchial carcinoma

Bronchial carcinoma is the most common malignant tumour worldwide, causing around 1.76 million deaths annually. It is the third most common cancer in the UK, and the most common cause of cancer-related death. Cigarette smoking (including passive smoke exposure) accounts for 72% of lung cancer cases. There remains a higher incidence of bronchial carcinoma in urban compared with rural areas, even when allowance is made for cigarette smoking. Other aetiological factors include:

- ***Environmental:*** radon exposure, asbestos, polycyclic aromatic hydrocarbons, ionizing radiation, occupational exposure to arsenic, chromium, nickel, petroleum products, and oils.
- ***Host factors:*** pre-existing lung disease, e.g., pulmonary fibrosis, HIV infection, genetic factors.

Pathophysiology

Historically, lung cancers have been divided into small-cell and non–small-cell carcinoma, based on histological appearance of the cells within the tumour. This distinction is helpful with respect to prognosis and determining best treatment. Non–small-cell carcinoma is further divided into a number of cell types (adenocarcinoma, squamous cell carcinoma, and large-cell carcinoma; **Box 20.33**). A number of molecular characteristics have been described in the subtypes of cancer, which confer potential prognostic benefit and may result in the ability to deliver a more personalized therapy with targeted agents. The most common of these are activating mutations within epidermal growth factor receptor *(EGFR)*, most commonly encountered in non-smokers, females, and those of Asian origin; and the presence of anaplastic lymphoma kinase *(ALK)* fusion oncogene, more commonly found in non- or ex-smokers and younger patients.

Box 20.33 Lung cancer cell types and clinical features

Cell type	% of lung cancers in the UK	Features
Non–small-cell carcinoma		
Squamous cell carcinoma	25–30	Arises from epithelial cells, associated with production of keratin
		Occasionally cavitates with central necrosis
		Causes obstructing lesions of bronchus with post-obstructive infection
		Local spread common, metastasizes relatively late
Adenocarcinoma	40%	Increasing incidence since 2005 possibly linked to low-tar cigarettes
		Originates from mucus-secreting glandular cells
		Most common cell type in non-smokers
		Often causes peripheral lesions on chest radiograph/CT
		Subtypes include bronchoalveolar cell carcinoma (associated with copious mucus secretion, multifocal disease)
		Metastases common: pleura, lymph nodes, brain, bones, adrenal glands
Large-cell carcinoma	10–15	Often poorly differentiated
		Metastasizes relatively early
Small-cell carcinoma	15–20	Arises from neuroendocrine cells (APUD cells)
		Often secretes polypeptide hormones
		Often arises centrally and metastasizes early

Clinical features

The presentation and clinical course vary between the different cell types (see **Box 20.33**). Symptoms and signs may be different, depending on the extent and site of disease. Common presenting features can be divided into those caused by direct/local tumour effects, metastatic spread, and non-metastatic extrapulmonary features.

Local effects

- ***Cough.*** The most common symptom in lung cancer. Due to evidence that cough is neglected by both patients and healthcare professionals, campaigns in the UK have highlighted the '3-week cough' as a symptom that merits a chest radiograph.
- ***Breathlessness.*** Central tumours occlude large airways, resulting in collapse and breathlessness on exertion. Many patients with lung cancer have coexistent COPD, which is also a cause of breathlessness. Patients may develop a pleural effusion due to metastatic involvement of the pleura.

- ***Haemoptysis.*** Fresh or old blood is coughed up because of tumour bleeding into an airway.
- ***Chest pain.*** Peripheral tumours invade the well-innervated chest wall or pleura, resulting in sharp pleuritic pain. Large-volume mediastinal nodal disease can present with a characteristic dull central chest ache.
- ***Wheeze.*** A monophonic wheeze is due to partial airway obstruction by tumour.
- ***Hoarse voice.*** Mediastinal nodal disease or direct tumour invasion of the mediastinum results in compression of the left recurrent laryngeal nerve.
- ***Nerve compression.*** Pancoast tumours in the apex of the lung invade the brachial plexus, causing C8/T1 palsy with small muscle weakness and wasting in the hand, and weakness and pain radiating down the arm. An associated Horner's syndrome also occurs due to compression of the sympathetic chain causing miosis, ptosis, and anhidrosis.
- ***Recurrent infections.*** Tumour causing partial obstruction of an airway result in post-obstructive pneumonia.
- ***Direct invasion of the phrenic nerve.*** Bronchial carcinoma invading the phrenic nerve causes paralysis of the ipsilateral hemidiaphragm. Involvement of the oesophagus produces progressive dysphagia. Involvement of the pericardium leads to pericardial effusion and malignant dysrhythmias.
- Superior vena caval obstruction. See Chapter 11.
- ***Tracheal tumours.*** These cause progressive dyspnoea and stridor. Flow–volume curves show reduced inspiratory flow (see Fig. 20.5C).

Metastatic effects

Bronchial carcinoma commonly spreads to mediastinal, cervical, axillary, and intra-abdominal nodes. In addition, the liver, adrenal glands, bones, brain, and skin are frequent sites for metastases:

- ***Liver.*** Common symptoms are anorexia, nausea, and weight loss. Right upper quadrant pain radiating across the abdomen due to liver capsular pain.
- ***Bone.*** Bone pain and pathological fractures occur. If the spine is involved, there is a risk of spinal cord compression (see Chapter 11), which requires urgent treatment.
- ***Adrenal glands.*** Metastases to the adrenals do not usually result in adrenal insufficiency and are generally asymptomatic.
- ***Brain.*** Metastases present as space-occupying lesions with mass effect and signs of raised intracranial pressure. Carcinomatous meningitis with cranial nerve defects, headache, and confusion is less common.
- ***Malignant pleural effusion.*** This presents with breathlessness and is commonly associated with pleuritic pain.

Non-metastatic extrapulmonary manifestations

Haematological manifestations of lung cancer including normocytic anaemia and thrombocytosis are common. Finger clubbing and hypertrophic pulmonary osteoarthropathy (HPOA) occur. Approximately 10% of small-cell tumours produce ectopic hormones, giving rise to paraneoplastic syndromes (Box 41.10).

Investigations

Investigations are indicated to:

- stage the extent of disease
- make a tissue diagnosis (to differentiate small-cell from non–small-cell lung cancer and detail molecular characteristics for targeted biologic agents and immunotherapy)
- assess fitness to undergo treatment.

Staging and diagnosis

Chest radiograph

Plain chest radiographs may show obvious evidence of lung cancer or non-specific appearances (Box 20.34). In some cases, the initial film appears normal, either because the lesion is small or because disease is confined to central structures.

Computed tomography

CT indicates the extent of disease. Imaging should include liver and adrenal glands, which are common sites for metastases. The International Association for the Study of Lung Cancer (IASLC) has devised the most widely used staging definitions, based on CT imaging of tumour size (T), nodal involvement (N), and metastases (M), along with prognostic data (Box 20.35). Using CT criteria, lymph nodes <1 cm in diameter are not classed enlarged, yet they can contain malignant cells. With increasing node size, the positive predictive value of CT in detecting malignancy grows, though not all enlarged nodes are malignant. PET-CT and biopsy may be indicated. If cerebral metastases are suspected, CT imaging of the brain should be performed. Bone metastases that lie within the staging field of view may be seen.

Other imaging modalities

MRI is not useful for the diagnosis of primary lung tumours other than Pancoast tumours with nerve invasion or the assessment of chest wall involvement prior to surgery. MRI spine is required if there is clinical suspicion of spinal cord compression. MRI brain may also be required to assess cerebral metastases.

Bone is a common site for metastatic deposits. Bone scan of sites of bone pain not included in other imaging may be helpful to demonstrate bony deposits.

Obtaining histology and cytology

Investigations for this purpose are listed in Box 20.36.

Other investigations

These include a full blood count for the detection of anaemia, and biochemistry to assess for liver involvement, hypercalcaemia, and hyponatraemia. Investigations for non-metastatic extrapulmonary manifestations of cancer may be indicated (Box 20.37)

Assessing fitness for treatment

The Eastern Cooperative Oncology Group (ECOG) performance status should be recorded for all patients with malignancy (Box 20.38). Depending on the treatment proposed and co-morbidity, assessment of fitness may include lung function testing with transfer capacity, cardiopulmonary exercise testing, stress echo, or other assessment of coronary perfusion.

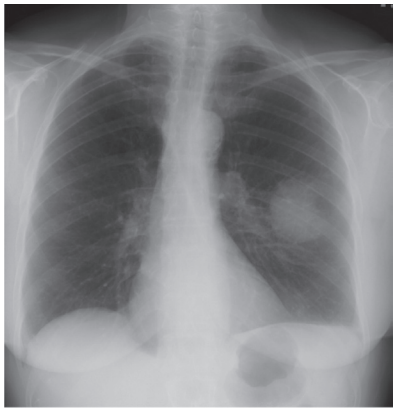
Management

The management of lung malignancy is detailed in Chapter 41.

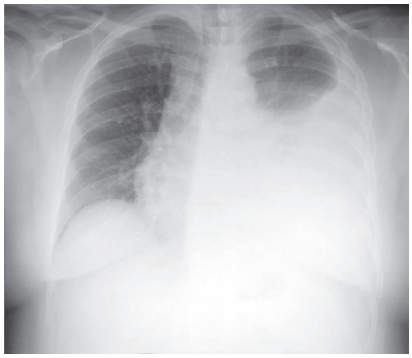
Mesothelioma

Mesothelioma is a malignant tumour arising from the parietal or visceral mesothelial lining of the lung. These tumours are almost always related to asbestos exposure and typically develops from pre-existing pleural plaque disease. The number of cases of mesothelioma has increased progressively since the mid-1980s, now causing 2500 deaths per year in the UK. The most common presentation of mesothelioma is a pleural effusion. Persistent chest wall pain should raise the index of suspicion even if the initial pleural

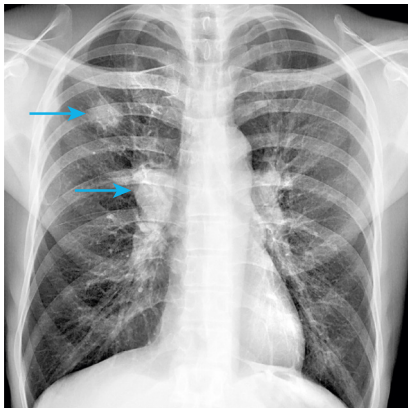
Box 20.34 Lung cancer presentations on chest X-ray



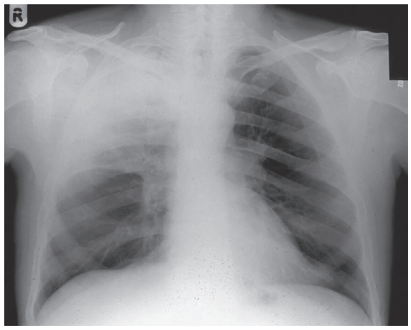
Mass lesion
Visible if >1 cm in diameter. Spiculated, cavitating, or smooth-edged. Often an incidental finding, usually asymptomatic if small. Chest radiograph is almost always abnormal if symptomatic mass lesion.



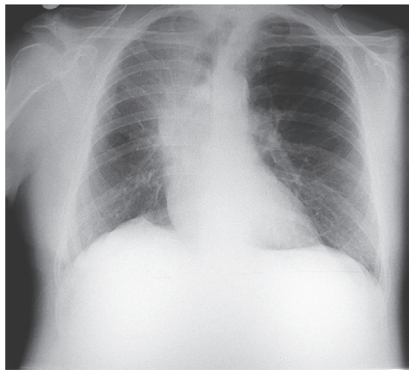
Pleural effusion
Usually unilateral. Can obscure an underlying mass or pleural tumour. Differential includes mesothelioma and metastatic disease.



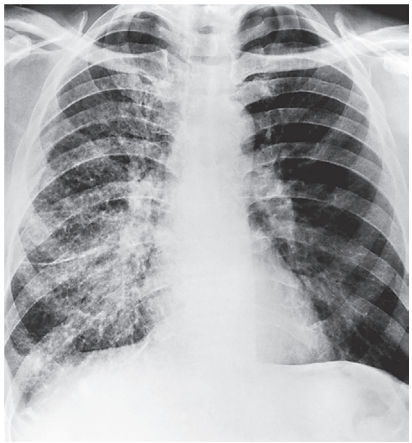
Mediastinal widening or hilar adenopathy
Splayed carina, hilar enlargement, paratracheal shadowing



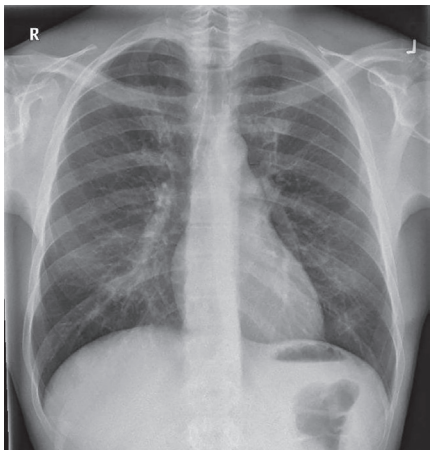
Slow-resolving consolidation
Tumour causes partial obstruction of a bronchus, resulting in retention of secretions, bacterial overgrowth, and infection. Radiograph changes persist due to tumour in right upper lobe.



Collapse
Endoluminal tumour causes complete collapse of a lung and associated mediastinal shift, or collapse of a lobe/segment with volume loss on the affected side: raised hemidiaphragm/deviated trachea.



Reticular shadowing
Carcinoma spreads through the lymphatic channels of the lung to give rise to lymphangitis carcinomatosa. In bronchial carcinoma this is usually unilateral and associated with striking dyspnoea. Bilateral lymphangitis should prompt investigation for a primary site other than lung, such as breast, stomach, or bowel.



Normal
A normal film does not rule out an underlying tumour. A minority of tumours are confined to the central airways and mediastinum without obvious change on plain chest radiograph. Normal chest radiograph should not discourage further investigation in symptomatic smokers over 40 years.

Box 20.35 TNM staging system for lung cancer

Notation	Description		
T—primary tumour			
TX	Primary tumour cannot be assessed, or tumour proven by the presence of malignant cells in sputum or bronchial washings, but not visualized by imaging or bronchoscopy		
T0	No evidence of primary tumour		
Tis	Carcinoma *in situ*[a]		
T1	Tumour 3 cm or less in greatest dimension, surrounded by lung or visceral pleura, without bronchoscopic evidence of invasion more proximal than the lobar bronchus (i.e., not in the main bronchus); subtypes mi, a, b, and c exist according to the tumour size.		
T2	Tumour more than 3 cm but not more than 5 cm; or tumour with any of the following features[b]: • Involves main bronchus, regardless of distance to the carina but without involvement of the carina • Invades visceral pleura • Associated with atelectasis or obstructive pneumonitis that extends to the hilar region involving either part of or the entire lung		
T2a	Tumour more than 3 cm but not more than 4 cm in greatest dimension		
T2b	Tumour more than 4 cm but not more than 5 cm in greatest dimension		
T3	Tumour more than 5 cm but not more than 7 cm in greatest dimension or one that directly invades any of the following: parietal pleura, chest wall (including superior sulcus tumours), phrenic nerve, parietal pericardium; or tumour nodule(s) in the same lobe as the primary		
T4	Tumour more than 7 cm or of any size that invades any of the following: diaphragm, mediastinum, heart, great vessels, trachea, recurrent laryngeal nerve, oesophagus, vertebral body, carina; separate tumour nodule(s) in a different ipsilateral lobe to the primary		
N—regional lymph nodes			
NX	Regional lymph nodes cannot be assessed		
N0	No regional lymph node metastasis		
N1	Metastasis in ipsilateral peribronchial and/or ipsilateral hilar lymph nodes and intrapulmonary nodes, including involvement by direct extension		
N2	Metastasis in ipsilateral mediastinal and/or subcarinal lymph node(s)		
N3	Metastasis in contralateral mediastinal, contralateral hilar, ipsilateral or contralateral scalene, or supraclavicular lymph node(s)		
M—distant metastasis			
M0	No distant metastasis		
M1	Distant metastasis		
M1a	Separate tumour nodule(s) in a contralateral lobe; tumour with pleural nodules or malignant pleural or pericardial effusion[c]		
M1b	Single extrathoracic metastasis in a single organ[d]		
M1c	Multiple extrathoracic metastasis in a single or multiple organs		
Resultant stage groupings			
Occult carcinoma	TX	N0	M0
Stage 0	Tis	N0	M0
Stage IA[e]	T1	N0	M0
Stage IB	T2a	N0	M0
Stage IIA	T2b	N0	M0
Stage IIB	T1, T2a, b	N1	M0
	T3	N0	M0
Stage IIIA	T1, T2a, b	N2	M0
	T3	N1	M0
	T4	N0, N1	M0
Stage IIIB	T1, T2a, b	N3	M0
	T3, T4	N2	M0
Stage IIIC	T3, T4	N3	M0
Stage IV	Any T	Any N	M1
Stage IVA	Any T	Any N	M1a, M1b
Stage IVB	Any T	Any N	M1c

[a]Tis includes adenocarcinoma *in situ* and squamous carcinoma *in situ*.
[b]T2 tumours with these features are classified T2a if 4 cm or less or if size cannot be determined, and T2b if greater than 4 cm but not larger than 5 cm.
[c]Most pleural (pericardial) effusions with lung cancer are due to tumour. In a few patients, however, multiple microscopic examinations of pleural (pericardial) fluid are negative for tumour, and the fluid is non-bloody and is not an exudate. Where these elements and clinical judgement dictate that the effusion is not related to the tumour, the effusion should be excluded as a staging descriptor.
[d]This excludes involvement of a single non-regional node.
[e]Stage IA is split into three subclasses 1, 2, and 3 according to tumour size.
(Based on data from Goldstraw P, Crowley J, Chansky K et al.; International Association for the Study of Lung Cancer International Staging Committee; Participating Institutions. The IASLC Lung Cancer Staging Project: proposals for the revision of the TNM stage groupings in the forthcoming (8th) edition of the TNM classification of malignant tumours. J Thorac Oncol *2016; 2:706–714.)*

Box 20.36 Investigations to obtain histology and cytology

- Fibreoptic bronchoscopy (**Fig. 20.26**)
- Endobronchial ultrasound (**Fig. 20.27**)
- CT/ultrasound-guided biopsy of lung lesions
- Ultrasound-guided biopsy of lymph nodes, liver metastases, or skin lesions
- Ultrasound-guided pleural aspiration
- Medical or video-assisted thoracoscopy
- Mediastinoscopy

Box 20.37 Non-metastatic extrapulmonary manifestations of bronchial carcinoma

Metabolic
- Loss of weight
- Lassitude
- Anorexia

Endocrine (10%) (usually small-cell carcinoma)
- Ectopic adrenocorticotrophin syndrome
- Syndrome of inappropriate secretion of antidiuretic hormone (SIADH)
- Hypercalcaemia (usually squamous cell carcinoma)
- Rarer: hypoglycaemia, thyrotoxicosis, gynaecomastia

Neurological (2%–16%)
- Encephalopathy—including subacute cerebellar degeneration
- Myelopathies—motor neuron disease
- Neuropathies—peripheral sensorimotor neuropathy
- Muscular disorders—polymyopathy, myasthenic syndrome (Eaton–Lambert syndrome)

Vascular and haematological
- Venous thromboembolic disease
- Thrombophlebitis migrans
- Non-bacterial thrombotic endocarditis
- Microcytic and normocytic anaemia
- Disseminated intravascular coagulopathy
- Thrombotic thrombocytopenic purpura
- Haemolytic anaemia

Skeletal
- Clubbing (30%)
- Hypertrophic osteoarthropathy (± gynaecomastia) (3%)

Cutaneous (rare)
- Dermatomyositis
- Acanthosis nigricans
- Herpes zoster

Box 20.38 Eastern Cooperative Oncology Group (ECOG) Performance Status

Grade	Description
0	Fully active, able to carry out all pre-disease performance without restriction
1	Restricted in physically strenuous activity but ambulatory and able to carry out work of a light or sedentary nature, e.g., light house work, office work
2	Ambulatory and capable of all self-care but unable to carry out work activities. Up and about more than 50% of waking hours
3	Capable of limited self-care only. Confined to bed or chair for more than 50% of waking hours
4	Disabled, cannot carry out any self-care, confined to bed or chair

fluid or biopsy samples are non-diagnostic. CT/ultrasound-guided biopsy or VATS pleural biopsy is often needed to obtain sufficient tissue for diagnosis. Clinical trials of chemotherapy, sometimes combined with surgery, are under way but the outlook for most patients remains very limited.

Screening for lung cancer

A large trial carried out in the USA demonstrated a 20% mortality benefit from low-dose helical CT screening for lung cancer in high-risk populations of smokers and ex-smokers aged 55–74. A similar trial has been undertaken in the Netherlands and the UK. Targeted lung screening is therefore being implemented in the UK. The programme invites 'at risk' individuals for low-dose CT scanning.

Secondary tumours

Metastases in the lung are very common. They are usually detected on chest radiography or CT in patients already diagnosed as having cancer, but can be the first presentation. Typical sites for the primary tumour include the kidney, prostate, breast, bone, gastrointestinal tract, cervix, or ovary. Single pulmonary metastases can be removed surgically but exclusion of additional metastases usually on CT or CT-PET scanning is essential before surgery.

Cancer (particularly stomach, pancreas, and breast) can involve mediastinal glands and spread along the lymphatics of both lungs, leading to progressive and severe breathlessness (lymphangitis carcinomatosis). This may be seen on a chest X-ray as bilateral lymphadenopathy with streaky basal shadowing fanning out over both lung fields.

Solitary pulmonary nodules

A solitary pulmonary nodule is defined as a discrete lesion of less than 3 cm in diameter. The differential diagnoses for a solitary pulmonary nodule are:

- primary bronchial carcinoma
- pulmonary metastases
- inflammatory lesions, e.g., rounded pneumonia or abscess
- granuloma
- benign tumour of the lung, e.g., hamartoma
- rheumatoid nodules
- hydatid cyst.

With the increased use of CT scanning for other conditions, there has been greater incidental detection of asymptomatic, small, sub-centimetre nodules. The majority of these are benign; however, radiological follow-up should be arranged at intervals in line with recommended guidelines, determined by the size of the nodule in millimetres and the risk of malignancy. Scoring tools such as the Brock score are useful in assessing the risk of malignancy based on nodule characteristics and size.

Bronchial carcinoid tumours

These rare tumours arising from neuroendocrine cells in the bronchial epithelium and account for approximately 1% of all bronchial tumours. Many of these are asymptomatic. Some patients will present with symptoms related to obstruction, recurrent infection, or haemoptysis. They are typically slow-growing and low-grade although aggressive, atypical tumours can occur. Surgery is usually the treatment of choice, with long-term surveillance. As foregut derivatives, bronchial carcinoids can produce adrenocorticotrophic hormone (ACTH) leading to clinical Cushing's syndrome in a minority. They do not usually produce 5-hydroxytryptamine seen in midgut or hindgut carcinoid tumours.

Benign tumours

Pulmonary hamartoma

This is the most common benign tumour of the lung and is usually seen on chest radiography as a well-defined round lesion 1–2 cm

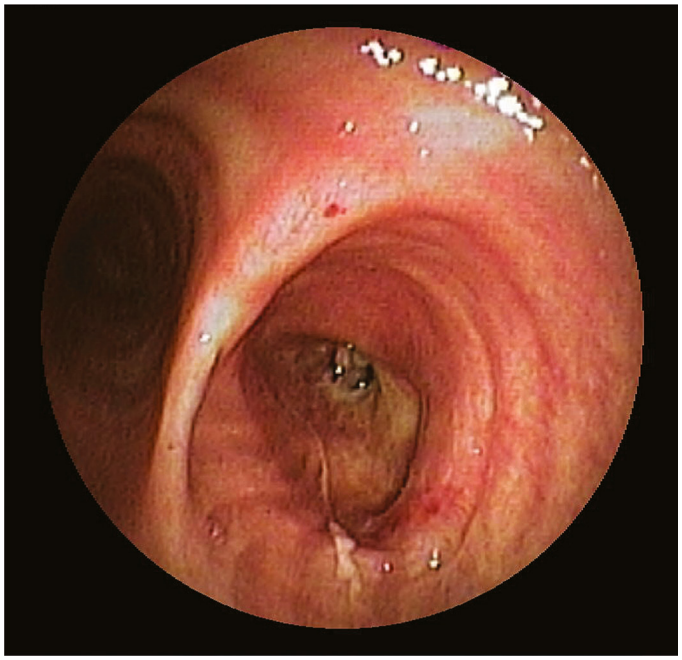


Fig. 20.26 Rigid bronchoscopy. A large obstructing tumour mass is well visualized. Similar endobronchial appearance can be seen at fibreoptic bronchoscopy. (*Optimal processing of diagnostic lung specimens. Practical Pulmonary Pathology: A Diagnostic Approach, Fourth Edition. pii:B9780323795470000126/f003-018-9780323795470.*)

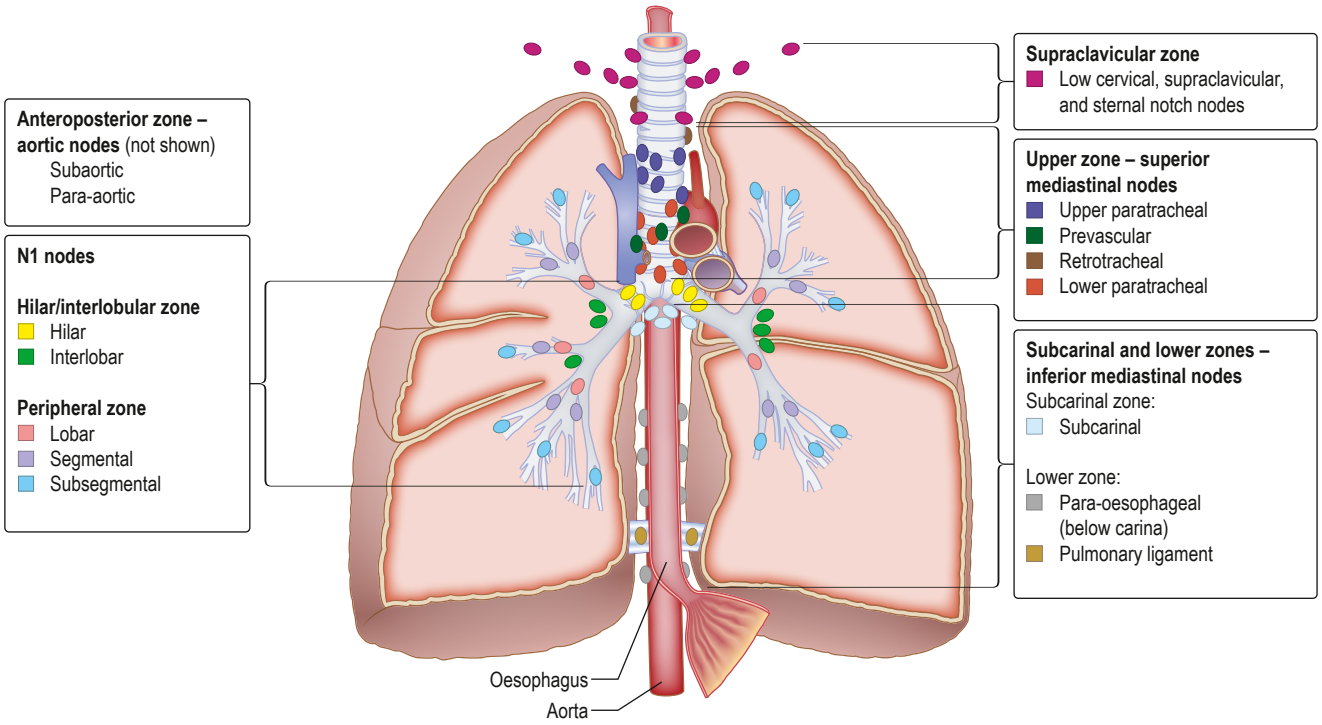


Fig. 20.27 Lymph nodes commonly involved in lung cancer. Histopathology and molecular analysis help determine staging and optimum treatment.

in diameter in the periphery of the lung. Growth is extremely slow, but the tumour can reach several centimetres in diameter. Rarely, it arises from a major bronchus and causes obstruction.

Bronchial adenoma

This diverse group of benign tumours arises from mucus glands and ducts of the windpipe.

Cylindroma, chondroma, and lipoma

These extremely rare tumours grow in the bronchus or trachea, causing obstruction.

Tracheal tumours

Benign tumours include squamous papilloma, leiomyoma, and haemangiomas.

Further reading

Callister MEJ, Baldwin DR, Akram AR; British Thoracic Society. BTS guidelines for the investigation and management of pulmonary nodules. *Thorax* 2015; 70(suppl 2).
International Association for the Study of Lung Cancer. *8th Edition of the TNM Classification for Lung Cancer*. IASLC 2017; www.iaslc.org.
National Institute for Health and Care Excellence. *NICE Guideline 122: Lung Cancer: Diagnosis and Management*. NICE 2019; http://www.nice.org.uk/guidance/ng122.
Patz EF Jr, Greco E, Gatsonis C et al. Lung cancer incidence and mortality in National Lung Screening Trial participants who underwent low-dose CT prevalence screening: a retrospective cohort analysis of a randomised, multicentre, diagnostic screening trial. *Lancet Oncol* 2016; 17:590–599.
Ru Zhao Y, Xie X, de Koning HJ et al. NELSON lung cancer screening study. *Cancer Imaging* 2011; 11(A):S79–S84.

BRONCHIECTASIS

Bronchiectasis describes abnormal and permanently dilated airways. The disease is characterized by a cycle of neutrophilic inflammation, recurrent infection, and airway damage which further impairs mucociliary clearance.

Bronchiectasis is associated with a number of diseases but is a cause found in only around 50% of cases. Little is known about the epidemiology and there is a wide variation in reported incidence. Bronchiectasis related to CF is generally considered a separate entity.

Aetiology

The causes of bronchiectasis are listed in **Box 20.39**. Globally, TB is the leading cause. Bronchiectasis associated with other lung diseases, particularly COPD, is increasingly recognized.

Clinical features

These are shown in **Box 20.40**.

Investigations

The aims of investigation are to confirm the diagnosis, rule out an underlying cause, look for reversible factors, and exclude/prevent complications.

- ***CT scan*** is the investigation of choice. Characteristically, non-tapering 'tram track' airways and an increased bronchoarterial ratio termed the 'signet ring' sign are seen (**Fig. 20.28**).
- ***Chest radiograph*** is often normal but tram track airways, ring shadows, and cysts may be seen.
- ***Sputum examination*** is useful for targeted antibiotic treatment, as well as exclusion of non-tuberculous mycobacterial disease. Extended microbiological culture is often required and needs to be specifically requested.
- ***Immune assessment*** includes immunoglobulin concentrations and response to Hib, tetanus, and pneumococcal vaccines. More detailed immunological assessment by an immunologist may be necessary.
- ***Sweat test and CF genetic assessment*** should be carried out in all patients under 40 years and any patient in where there is a clinical suspicion.
- ***Nasal nitric oxide*** is a useful test for screening for primary ciliary dyskinesia (PCD), in which it is very low. Further ciliary investigation in a specialist centre may be required.
- ***Total IgE and* Aspergillus*-specific IgE or* Aspergillus *skin-prick testing*** should be done to exclude allergic bronchopulmonary aspergillosis.

Management

Therapy can be divided into airway clearance, anti-inflammatories, and treatment of infection and complications.

? Box 20.39 Causes of bronchiectasis

Congenital
- Deficiency of bronchial wall elements
- Pulmonary sequestration

Mechanical bronchial obstruction

Intrinsic
- Foreign body
- Inspissated mucus
- Post-tuberculous stenosis
- Tumour

Extrinsic
- Lymph node
- Tumour

Postinfective bronchial damage
- Bacterial and viral pneumonia, including pertussis, measles, and aspiration pneumonia

Granuloma
- Tuberculosis
- Sarcoidosis

Diffuse diseases of the lung parenchyma
- Idiopathic pulmonary fibrosis

Pathological immunological response
- Allergic bronchopulmonary aspergillosis
- Post-lung transplant
- Graft-versus-host disease

Immune deficiency

Primary
- Panhypogammaglobulinaemia
- Selective immunoglobulin deficiencies (IgA and IgG_2)

Secondary
- Human immunodeficiency virus
- Immunosuppressive medication

Mucociliary clearance defects

Genetic
- Primary ciliary dyskinesia (Kartagener's syndrome with dextrocardia and situs inversus)
- Cystic fibrosis

Acquired
- Young's syndrome—azoospermia, sinusitis

i Box 20.40 Symptoms and signs in bronchiectasis

- ***Cough:*** usually persistent
- ***Sputum production:*** large amounts and purulent
- ***Breathlessness:*** as disease progresses
- ***Haemoptysis:*** usually a sign of infection; streaking is common, massive haemoptysis is a rare medical emergency
- ***Infection:*** usually characterized by increased sputum volume and increased purulence
- ***Pleuritic chest pain:*** can be a feature of infection
- ***Coarse crackles:*** heard on auscultation but examination can be normal
- ***Clubbing:*** especially in cystic fibrosis

Airway clearance

Daily airway clearance therapies are advised. Active cycle breathing techniques, autogenic/self-drainage and postural drainage are used. Several devices are available to assist, including Flutter or Acapella, providing positive expiratory pressure with or without airway oscillation. Nebulized hypertonic saline is approved as a mucoactive agent.

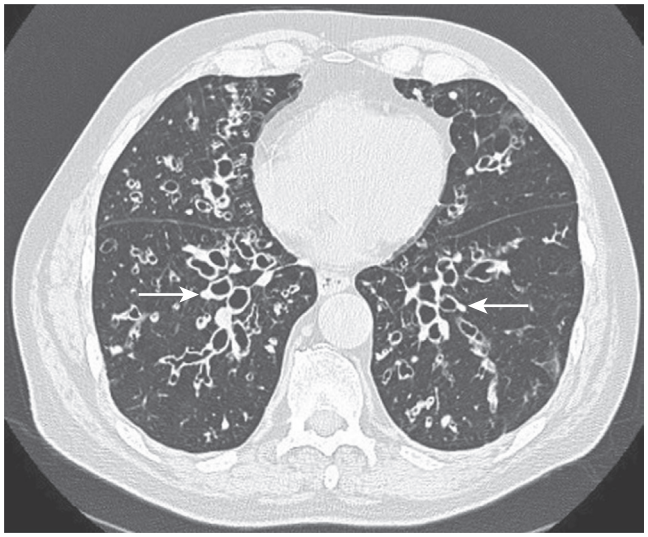


Fig. 20.28 CT scan showing bronchiectasis. Note the dilated bronchi with thickened wall, which are larger than adjacent arteries, giving a signet ring appearance *(white arrows)*.

Anti-inflammatories

Long-term azithromycin has an immunomodulatory effect and has been demonstrated to reduce exacerbation frequency. Inhaled corticosteroids are beneficial to some patients.

Treatment of infection

Treatment of exacerbations usually lasts 2 weeks and is initially based on previously obtained microbiological information. High-dose ciprofloxacin (750 mg twice daily) is a useful oral drug for treatment of *Pseudomonas*, though dual therapy is often required due to multiresistant pathogens. *H. influenzae* infection is common and usually responds to amoxicillin, co-amoxiclav, or doxycycline. Multiresistant species may need intravenous cephalosporin treatment. Experience in CF has promoted the use of aggressive antibiotic strategies in bronchiectasis, with eradication therapy and chronic suppressive nebulized therapy with colistimethate or an aminoglycoside for *P. aeruginosa*. Rotating oral antibiotic regimens and long-term quinolones should be avoided.

Treatment of complications

- Pulmonary rehabilitation should be offered to patients with a reduced exercise capacity and breathlessness.
- Massive haemoptysis is a life-threatening medical emergency. Treatment is resuscitation with airway protection until bronchial artery embolization can be performed. If this is not successful or available, surgery may be required.
- Treatment of *Aspergillus* lung disease and NTM is covered in 'Diseases caused by *Aspergillus fumigatus*' section.
- Respiratory failure should be treated with oxygen and non-invasive ventilation. Suitable patients should be referred for transplantation.
- Surgery may be considered for localized disease.

Prognosis

Prognosis is variable, depending on disease severity. A low FEV_1 and infection with *P. aeruginosa* are associated with poorer outcomes.

Box 20.41 CFTR abnormalities in CF

Class	CFTR effect	Mutation example
I	Protein synthesis defect. No functional CFTR	G542X, R553X
II	Folding/trafficking defect. Does not reach apical membrane	F508del
III	Gating defect	G551D
IV	Conductance impairment due to narrow channel	R117H
V	Splicing defect	3849+10kbC>T
VI	Reduced stability	4326delTC

CYSTIC FIBROSIS

Cystic fibrosis (CF) is an autosomal recessive condition. In the UK, the birth prevalence is 1 in 2500 and the carrier rate is 1 in 25. Rates are lower in non-Caucasian populations. Survival has improved dramatically. The current expected median survival is now over 50 years of age and is likely to continue to improve with new developments in treatments. CF is a multisystem disease, although respiratory problems are usually the most prominent. A cycle of mucus stasis, inflammation, and infection leads to progressive respiratory failure. Most individuals with CF have pancreatic insufficiency.

Pathogenesis

The CF gene is located on the long arm of chromosome 7 at position 31.2. Mutations lead to abnormalities in the production of the cystic fibrosis transmembrane conductance regulator (CFTR) protein. This protein is expressed in the apical membrane of epithelial cells and acts as a chloride channel. Over 1000 mutations have been identified, most of them rare. The F508del mutation is the most common, accounting for around 70% of cases. Mutations have been divided into different classes, depending on their effect on CFTR (Box 20.41). This classification is used therapeutically, with new CF treatments aimed at improving CFTR function. Ivacaftor was the first drug available for CF that improves CFTR function.

In the lungs, CFTR dysfunction leads to dehydrated airway surface liquid, mucus stasis, airway inflammation, and recurrent infection. This process originates in the small airways, leading to progressive airway obstruction and bronchiectasis (Fig. 20.29).

Clinical features and complications

CF is a multisystem disease but in 90% of cases morbidity and mortality are related to respiratory disease. Pancreatic insufficiency occurs in the majority of patients and CF-related diabetes is becoming increasingly common as survival improves, occurring in up to 50% of older adults with CF. Liver disease occurs in around 20%, leading to cirrhosis in around 2%. Clinical manifestations and complications are summarized in Box 20.42.

Diagnosis

Most new CF diagnoses are currently made at newborn screening. The test involves measuring immunoreactive trypsinogen at the time of the neonatal heel-prick test. If the concentration is raised, formal testing is performed. Aside from neonatal screening, diagnosis for children and adults is based on a combination of:

- Common clinical features (see Box 20.42).
- CFTR functional testing. The ***sweat test (pilocarpine iontophoresis)*** measures chloride concentration. The normal range is <30

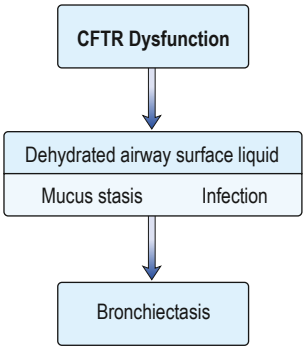


Fig. 20.29 **Cystic fibrosis: pathogenesis of lung disease.**

Box 20.42 Clinical features and complications of cystic fibrosis

Respiratory
- Recurrent respiratory infection
- Chronic daily cough and sputum production
- Breathlessness
- Nasal polyps
- Haemoptysis (sometimes massive)
- Pneumothorax
- Respiratory failure
- Cor pulmonale
- Recurrent sinusitis

Gastrointestinal
- Failure to thrive in infancy, low body mass index in adults
- Meconium ileus in infancy
- Distal intestinal obstruction syndrome
- Steatorrhoea secondary to pancreatic insufficiency
- Cystic fibrosis-related liver disease and cirrhosis
- Increased risk of gastrointestinal malignancy

Other
- Cystic fibrosis-related diabetes
- Male and female infertility
- Osteoporosis
- Arthropathy

mmol/L, with concentrations of 30–60 mmol/L found in milder 'atypical' phenotypes. Where there is diagnostic difficulty, nasal potential difference can be measured.
- Confirmatory genetic testing.

Management

Patients with CF should be managed in a specialist centre by a multidisciplinary team of healthcare professionals. They should have a clinical review at least every 3 months and a detailed annual review. Lung function (FEV_1) and body mass index (BMI) should be recorded at every appointment, as they have prognostic importance.

Treatment of pulmonary disease

The aim of chronic pulmonary therapies is to slow the decline in FEV_1 and reduce daily symptoms and exacerbation frequency.
- ***Airway clearance techniques*** are a vital part of CF treatment, taught to patients and their caregivers by specialist respiratory physiotherapists. Techniques include percussion, vibration, deep breathing, and forced coughing. There is no consensus on the best technique and patient choice is an important factor.
- ***Nebulized therapy:***
  - Recombinant human DNase (rhDNase) works by lysing bacterial DNA and reducing sputum viscosity. Routine treatment is advised from early childhood (regardless of disease status) due to evidence of improvement in lung disease and possible survival benefit.
  - Hypertonic saline works as an osmotic agent to draw water to the cell surface and reduce sputum viscosity.
  - Inhaled mannitol increases mucociliary clearance.
- ***Anti-inflammatory treatment*** with long-term azithromycin is widely used in CF and has an immunomodulatory effect.

Treatment of respiratory infection

Respiratory infection is dangerous for patients with CF. Clinics are microbiologically cohorted, patients are managed in single rooms, and no patient social events are organized. *P. aeruginosa* infection is common and associated with accelerated decline in lung function. Eradication is aim of treatment. A combination of oral, nebulized, and intravenous antibiotics can be given, together with long-term nebulized suppression therapy. Other organisms, such as *Burkholderia cepacia*, methicillin-resistant *S. aureus* (MRSA), and *Stenotrophomonas maltophilia*, have been associated with worsening respiratory outcomes and so eradication regimens for these bacteria are also used. Non-tuberculous mycobacterial disease, in particular *M. abscessus*, can be associated with rapid decline and active infection may preclude transplantation. In exacerbations, intravenous antibiotic therapy is based on previous infection history; *in vitro* sensitivities are less useful. Many patients have a permanently implanted venous access device for delivery of intravenous therapy.

CFTR modulators

CFTR modulators (Kaftrio, Kalydeco, and Symdeko) are licensed in the UK. Ninety percent of patients are eligible for this highly effective combination therapy which is already changing outcomes and life expectancy.

Advanced disease

Respiratory failure is treated with oxygen and NIV. Patients are referred for consideration of lung transplantation when FEV_1 falls to around 30% predicted. In end-stage disease, palliative care is an essential. This can be challenging when patients are on an active waiting list for a lung transplant.

Non-respiratory complications

Pancreatic enzymes and vitamin supplements are used to treat patients with pancreatic insufficiency. Close attention is paid to diet and calorie supplementation. Overnight gastric feeding may be required to maintain BMI in some patients.

CF-related diabetes is screened for at annual review. It is distinct from type 1 and type 2 diabetes. It commonly requires treatment with insulin. Osteoporosis is screened for and treated. Fertility treatment is available for males with CF who are infertile. Females with CF should be offered pre-pregnancy counselling, and have close maternal and fetal surveillance in pregnancy, usually delivering in a maternal medicine centre.

Further reading

Chambers JD, Chotirmall SH. Bronchiectasis: new therapies and new perspectives. *Lancet Respir Med* 2018; 6:715–726.
Despotes, KA and Donaldson SH. Current state of CFTR modulators for treatment of cystic fibrosis. *Curr Opin Pharmacol* 2022; 65:102239.
Flume PA, Chalmers JD, Olivier KN. Advances in bronchiectasis: endotyping, genetics, microbiome, and disease heterogeneity. *Lancet* 2018; 392:880–890.
King JA, Nichols AL, Bentley S, Carr SB, Davies JC. An update on CFTR modulators as new therapies for cystic fibrosis. *Paediatr Drugs* 2022; 24:321–333.
Ong T, Ramsey MD. Cystic fibrosis, a review. *JAMA* 2023; 329(21):1859–1871.
Shteinberg M et al. Cystic fibrosis. *Lancet* 2021; 397(10290):2195–2211.

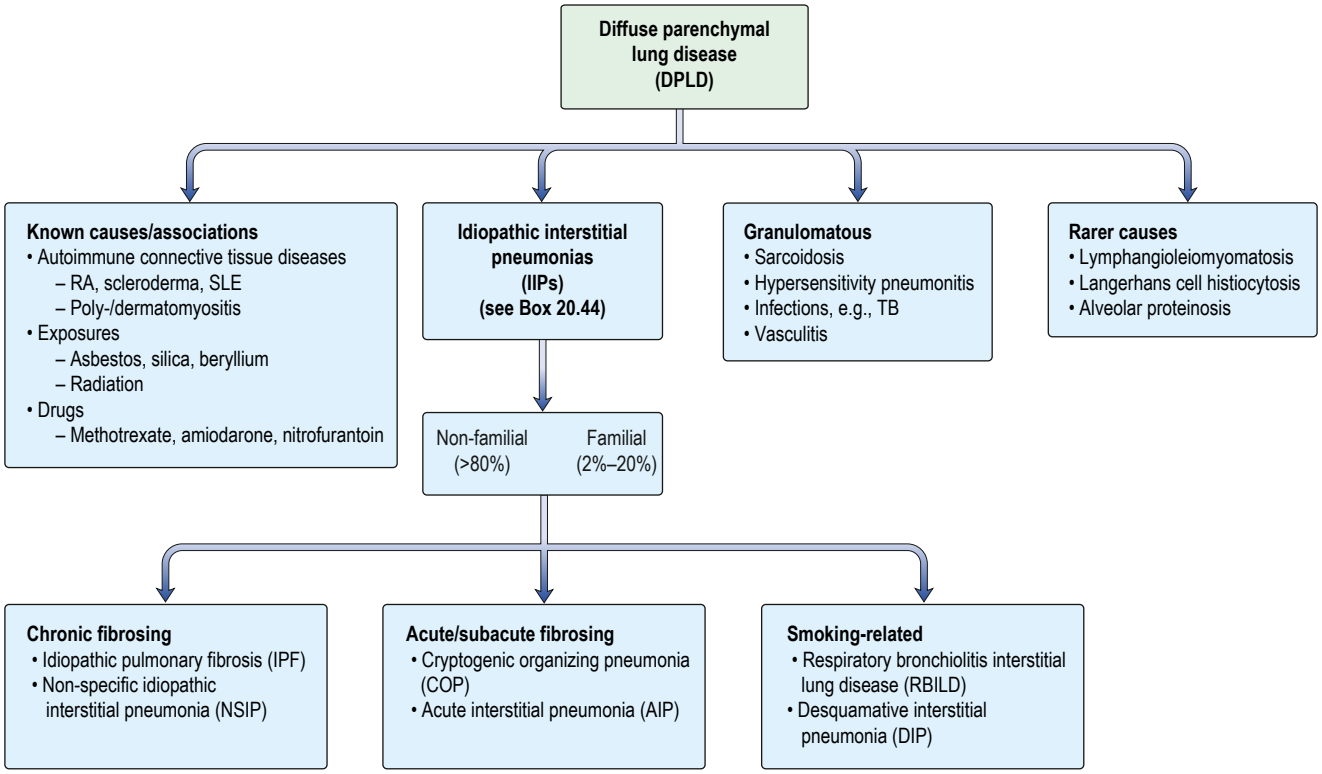


Fig. 20.30 Classification of the interstitial lung diseases. *RA,* Rheumatoid arthritis; *SLE,* systemic lupus erythematosus; *TB,* tuberculosis.

INTERSTITIAL LUNG DISEASES

This heterogeneous group of conditions is also referred to as ***diffuse parenchymal lung diseases*** and accounts for about 15% of respiratory clinical practice. They are characterized by varying degrees of inflammation and fibrosis, initially affecting the interstitium of the lung and typically presenting with exertional dyspnoea, with or without cough. A classification is shown in Fig. 20.30.

Sarcoidosis

Sarcoidosis is a multisystem granulomatous disorder; it commonly affects young adults and typically presents with bilateral hilar lymphadenopathy, pulmonary infiltration, and skin or eye lesions.

Epidemiology and aetiology

Sarcoidosis is a disease of unknown aetiology, often detected on routine chest radiography. It is most common in Northern Europe (annual incidence 5–40/100 000) and uncommon in Japan (incidence 1–2/100 000). It occurs more frequently in Afro-Caribbean patients, who are also more likely to develop extrapulmonary and chronic disease. There is a female preponderance with a peak incidence in the third and fourth decades. There is no relation to any histocompatibility antigen, but first-degree relatives (particularly Caucasians) have an increased risk of developing sarcoidosis.

Immunopathology

- Typical sarcoid granulomas consist of focal accumulations of epithelioid cells, macrophages, and lymphocytes, mainly T cells.
- Depressed cell-mediated reactivity to tuberculin; the Mantoux test is usually negative. Circulating T lymphocytes are low but B cells are slightly increased, with overall lymphopenia.
- Bronchoalveolar lavage shows an increase in the overall cell number; a lymphocytosis (particularly $CD4^+$ T-helper cells) is common.
- Transbronchial biopsies show infiltration of the alveolar walls and interstitial spaces with leucocytes, mainly T cells, prior to granuloma formation.

Clinical features

Sarcoidosis can affect any organ (Box 20.43) but has a predilection for the lungs (up to 90%). Presentation may be with respiratory symptoms, but it is not unusual for the diagnosis to be made incidentally on chest radiography. Common extrathoracic manifestations include eye, skin, or lymph node involvement. Constitutional upset with fatigue is frequent and often refractory.

Löfgren's syndrome is the typical combination of bilateral hilar lymphadenopathy, erythema nodosum, arthralgia, and fever. It usually resolves spontaneously without the need for treatment.

In less classic presentations, HRCT may suggest sarcoidosis, but tissue biopsy of an affected organ is required for definitive diagnosis (non-caseating granuloma). Skin biopsy is relatively non-invasive, and careful examination for infiltration of scars and tattoos amenable to sampling should be undertaken.

Pulmonary manifestations

Cough, exertional breathlessness, and vague chest discomfort are common presentations. The chest is often clear to auscultation, although wheeze may be evident if there is significant endobronchial

disease. There are four radiological stages of lung involvement, which help inform prognosis:

- ***Stage 1.*** Bilateral hilar lymphadenopathy (BHL) alone: 55%–90% spontaneous remission.
- ***Stage 2.*** Pulmonary infiltrates with BHL: 40%–70% spontaneous remission.
- ***Stage 3.*** Pulmonary infiltrates without BHL: 10%–20% spontaneous remission.
- ***Stage 4.*** Pulmonary fibrosis: once established, tends to be irreversible

Patients may present with any stage of disease and do not necessarily progress through the stages sequentially. Moreover, the extent of disease on chest radiography does not correlate with impairment on pulmonary function testing.

Box 20.43 Clinical presentations of sarcoidosis

Constitutional
- Fever
- Weight loss
- Fatigue

Reticuloendothelial
- Splenomegaly
- Lymphadenopathy

Respiratory
- Stage 1–4 pulmonary involvement
- Cough
- Dyspnoea
- Wheeze

Hepatic
- Deranged liver function
- Hepatomegaly

Renal
- Interstitial nephritis and chronic kidney disease

Ocular
- Anterior uveitis
- Keratoconjunctivitis sicca

Hypercalcaemia/ hypercalciuria
- Nephrocalcinosis
- Calculi
- Tubulointerstitial nephritis

Cutaneous
- Erythema nodosum
- Lupus pernio

Neurological
- Cognitive dysfunction
- Headache
- Cranial nerve palsies
- Mononeuritis multiplex
- Peripheral neuropathy
- Seizures

Cardiac
- Arrhythmias
- Heart block
- Cardiomyopathy
- Sudden death

Bilateral hilar lymphadenopathy

Symmetrical BHL is a characteristic feature of sarcoidosis and is usually asymptomatic. Occasionally, it is associated with a dull ache in the chest, malaise, and a mild fever. The differential diagnosis of BHL includes:

- ***Lymphoma:*** rarely affects the hilar lymph nodes in isolation.
- ***Pulmonary TB:*** asymmetrical enlargement is usual.
- ***Carcinoma of the bronchus with lymph node spread:*** rarely symmetrical.

Pulmonary infiltration

Although the lung fields may appear normal on plain chest radiography, the lung parenchyma is frequently involved as shown by CT scanning (Fig. 20.31), transbronchial biopsy, and bronchoalveolar lavage. Symptoms may be minimal (or even absent) despite marked radiographic abnormalities. Progressive disease leads to irreversible fibrosis in up to 20% of cases. The principal differential diagnoses are TB, pneumoconiosis, IPF, and HP.

Extrapulmonary manifestations

Sarcoidosis can affect any organ. Cutaneous sarcoidosis and ocular sarcoidosis are the most common extrapulmonary presentations, but cardiac, CNS, and renal involvement are the most important clinically.

Skin

Occur in 10%–30% of cases. Sarcoidosis is the most common cause of erythema nodosum (see Chapter 45, 'Erythema nodosum' section). Lupus pernio (indurated erythematous or violaceous papules or plaques, Fig. 20.32) may be seen over the face and ears.

Eye

Anterior, or less commonly posterior, uveitis presents with misting of vision, a painful, red eye, or progressive loss of vision in 5% of patients. Asymptomatic uveitis may be found in up to 25% of patients, and ophthalmological assessment should be considered in all patients with a new diagnosis. Keratoconjunctivitis sicca and lacrimal gland enlargement also occur.

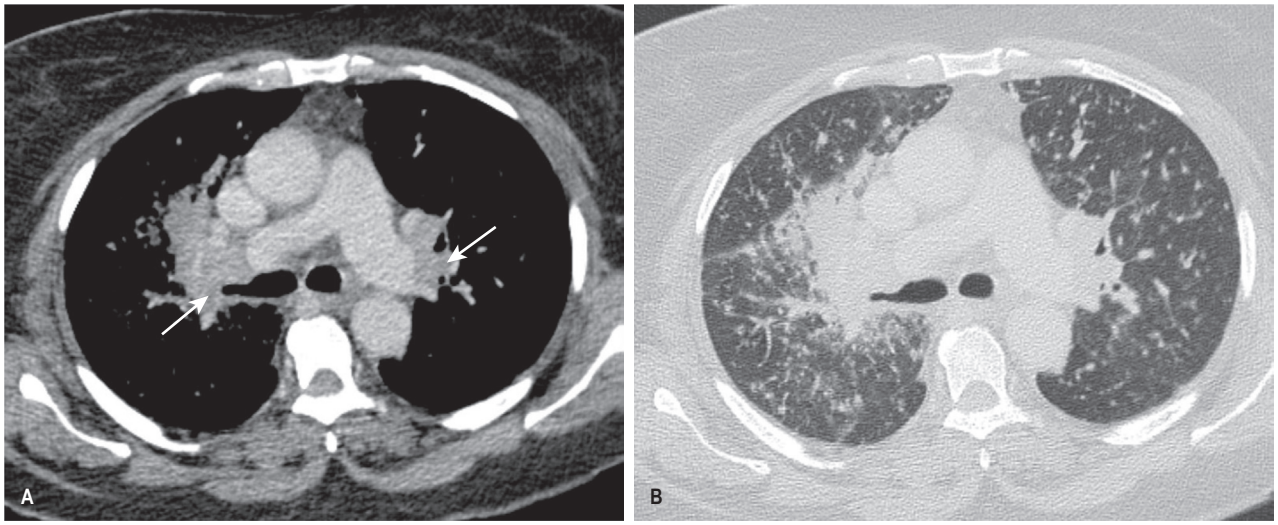


Fig. 20.31 CT chest showing sarcoidosis. **(A)** Soft tissue settings demonstrate bilateral hilar lymphadenopathy *(arrowed)*. **(B)** Lung settings demonstrate multiple small nodules and reticulation in both lungs.

Fig. 20.32 **The lupus pernio form of sarcoidosis.**

Metabolic manifestations

Hypercalcaemia and hypercalciuria can lead to the development of renal calculi, nephrocalcinosis, chronic kidney disease, and kidney failure.

Central nervous system

CNS involvement is rare (2%) but can lead to severe neurological disease (see Chapter 28).

Bone and joint involvement

Arthralgia without erythema nodosum is seen in 5% of cases. Bone cysts with associated swelling, particularly affecting the digits, may be seen on plain radiographs.

Hepatosplenomegaly

Mild derangement of liver function tests is common, and granulomas are seen in most biopsy specimens, although these findings are rarely of clinical consequence. Progression to portal hypertension or liver failure is uncommon.

Renal involvement

Sarcoidosis classically causes a granulomatous tubulointerstitial nephritis, although kidney damage secondary to hypercalcaemia and stone formation may also be seen.

Cardiac involvement

Ventricular dysrhythmias, conduction defects, and cardiomyopathy with congestive cardiac failure are rare (3%). Severity ranges from benign rhythm disturbance to sudden cardiac death. All patients with sarcoidosis should have an ECG and echo at presentation.

Investigations

- ***Imaging*.** Chest radiography is the initial modality for staging, followed by HRCT for assessment of parenchymal involvement. This may show nodules of up to 10 mm diameter that form a characteristic 'beading' appearance along airways, vessels, and fissures. Nodules may aggregate into larger nodules or masses of up to 3 cm. There may be increased reticulation due to septal thickening, with honeycombing in severe cases fibrotic.
- ***Full blood count*.** Mild, normochromic, normocytic anaemia with raised ESR.
- ***Biochemistry*.** Kidney involvement is occasionally found and biopsy may be indicated. Hypercalcaemia occurs in 10%–20% and hypercalciuria in 30%–50%. Activated macrophages in lung and lymph nodes are able to hydroxylate vitamin D directly (independent of parathyroid hormone levels), leading to increased intestinal absorption of dietary calcium. Urinary calcium excretion should be performed at presentation.
- ***Serum ACE concentration*.** This is elevated in over 75% of patients with untreated sarcoidosis. Raised (but lower) concentrations are also seen in patients with lymphoma, pulmonary TB, asbestosis, and silicosis limiting specificity. The utility of serum ACE in monitoring disease activity and response to treatment is contentious.
- ***Cardiac tests*.** ECG and echocardiogram should be performed at presentation. If these raise concern about underlying cardiac sarcoidosis, then further investigation is with cardiac MRI.
- ***EBUS and bronchoscopy*.** Both EBUS and bronchoscopy can be used for diagnosis. EBUS has been shown to give a higher diagnostic yield. Transbronchial biopsy of the lung parenchyma is positive in up to 80% of cases of stage 2 and 3 disease. Bronchoalveolar lavage typically shows a lymphocytosis with raised CD4:CD8 ratio.
- ***Lung function tests*.** These show a restrictive lung defect with reduced gas transfer in patients with parenchymal infiltration or fibrosis. However, an obstructive defect may be seen in endobronchial disease, and a mixed pattern is also possible. Lung function is usually normal in patients with isolated hilar adenopathy or extrapulmonary disease.

Prognosis and management

The disease course is unpredictable and varies from spontaneous remission to progression and death in a small number (1%–5%). Even if remission occurs, relapses are common. Worse outcomes are seen in patients of Afro-Caribbean and Asian ethnicities, and those presenting with extrathoracic disease. Systemic treatment is indicated for hypercalcaemia and extrathoracic major organ involvement, particularly neurological, cardiac, and renal, and in ocular disease resistant to topical therapy. Treatment of pulmonary sarcoidosis is less clear-cut as spontaneous resolution is frequently seen, typically within the first 6 months. Moreover, although corticosteroids improve radiographic appearances, this is not consistently reflected in improved lung function. Treatment is therefore reserved for patients with troublesome symptoms, deteriorating lung function, or radiological evidence of disease progression. First-line treatment is with prednisolone (or equivalent) 0.5 mg/kg for 4–6 weeks, gradually tapering to a maintenance dose for at least 12 months. Alternative immunosuppressants, including methotrexate, azathioprine, and hydroxychloroquine, have been used in place of, or in addition to, prednisolone. Relapses are common on withdrawal of therapy. Lung transplantation should be considered for suitable patients with stage IV disease and respiratory failure.

Idiopathic interstitial pneumonias

The terminology used to describe the idiopathic interstitial pneumonias can be confusing, but distinction must be made between subgroups due to significant differences in prognosis and treatment. Clinical patterns usually link to particular histological subtypes; classification is shown in **Box 20.44**.

Idiopathic pulmonary fibrosis

Idiopathic pulmonary fibrosis (IPF) is the most common of the idiopathic interstitial pneumonias. It is a progressive and ultimately fatal disease of unknown cause. There is significant worldwide variation in reported prevalence, but incidence appears to be increasing. The onset is usually in the seventh decade, and is rare below the age of 50. Males are twice as likely to be affected.

Box 20.44 Classification of idiopathic interstitial pneumonias

Clinical diagnosis	Pathological pattern
Idiopathic pulmonary fibrosis (IPF)	Usual interstitial pneumonia (UIP)
Desquamative interstitial pneumonia (DIP)	Desquamative interstitial pneumonia (DIP)
Respiratory bronchiolitis interstitial lung disease (RBILD)	Respiratory bronchiolitis interstitial lung disease (RBILD)
Acute interstitial pneumonia (AIP)	Diffuse alveolar damage (DAD)
Non-specific interstitial pneumonia (NSIP)	Non-specific interstitial pneumonia (NSIP)
Cryptogenic organizing pneumonia (COP)	Organizing pneumonia (OP)

(From ATS/ERS. An official American Thoracic Society/European Respiratory Society statement: update of the international multidisciplinary classification of the idiopathic interstitial pneumonias. Am J Respir Crit Care Med *2013; 188:733–748.)*

Pathology

Usual interstitial pneumonia (UIP) is the histological finding in IPF. The key feature is a heterogenous appearance with areas of normal lung punctuated by areas of marked fibrosis, honeycombing mainly in subpleural areas, and fibroblastic foci (dense proliferations of fibroblasts and myofibroblasts). The terms IPF and UIP are often used interchangeably but they are *not* synonymous, as UIP is also the primary histological finding in several other diffuse parenchymal lung diseases (e.g., pulmonary autoimmune rheumatic disease).

Pathogenesis

It is thought that repetitive injury to the alveolar epithelium, caused by currently unidentified environmental stimuli, leads to the activation of several pathways responsible for repair of the damaged tissue. However, in IPF, the wound healing mechanisms become uncontrolled, leading to over-production of fibroblasts and deposition of increased extracellular matrix in the interstitium with little inflammation. The structural integrity of the lung parenchyma is therefore disrupted: there is loss of elasticity and the ability to perform gas exchange is impaired, leading to progressive respiratory failure.

Clinical features

Patients typically present with insidious onset of progressive dyspnoea that may be accompanied by cough, with or without sputum production. Bi-basal end-inspiratory crackles are found on examination and therefore it is not uncommon for patients to be mistakenly treated for heart failure or recurrent chest infections before diagnosis of IPF is made. Finger clubbing is seen in 25%–50% of cases.

Progressive respiratory failure may be complicated by pulmonary hypertension. Stepwise deterioration can occur due to pneumothorax, pulmonary embolism, or intercurrent infection, but acute exacerbations with no identifiable cause are well recognized and are associated with increased mortality.

Investigations

- ***Respiratory function tests*** usually show a restrictive pattern (FEV_1/FVC ratio >70%), with reduced lung volumes and gas transfer. However, spirometry may be normal in early disease and lung volumes can be preserved in the presence of coexisting emphysema.

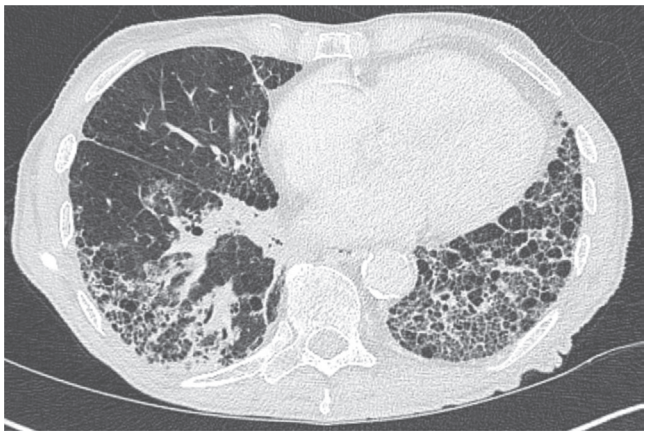


Fig. 20.33 CT demonstrating idiopathic pulmonary fibrosis.
There is subpleural reticulation and honeycombing of both lungs.

- ***Blood tests.*** There is no specific serological test for IPF. Antinuclear antibodies (ANA) and rheumatoid factor (RF) are performed to exclude autoimmune connective tissue disease.
- ***Chest radiograph*** shows small-volume lungs with increased reticular shadowing at the bases. It may be normal in early disease.
- ***CT*** is the imaging modality of choice. A confident diagnosis of IPF may be made in patients with:
  - ***Basal distribution:*** abnormalities are more pronounced at lung bases
  - ***Subpleural reticulation:*** reticulation most evident in lung peripheries
  - ***Traction bronchiectasis:*** the fibrotic process distorts normal lung architecture, pulling the airways open and causing bronchiectasis
  - ***Honeycombing:*** basal layers of small, cystic airspaces with irregularly thickened walls composed of fibrous tissue (Fig. 20.33).
- ***Bronchoalveolar lavage*** is performed to exclude malignancy and infection. A differential cell count may support other diagnoses: lymphocytosis is suggestive of HP or sarcoidosis, and eosinophilia occurs in eosinophilic pneumonia.
- ***Histological confirmation*** is necessary in some patients. Surgical lung biopsy, usually via VATS, is the most reliable method for obtaining diagnostic histological samples; transbronchial biopsy can be undertaken bronchoscopically but only obtains small samples. Cryobiopsies are undertaken in some centres.

Differential diagnosis

The main differential diagnoses for the UIP pattern of fibrosis are vasculitis, HP, connective tissue disease, and asbestos exposure. Differentials for the chest radiograph appearance include interstitial pulmonary oedema, infection, and lymphangitis carcinomatosa.

Prognosis and management

The median survival time for patients with IPF is 2–5 years. Serial lung function testing is used to monitor disease progression and a 10% decline in FVC or 15% decline in gas transfer (TLCO) in the first 6–12 months confers a worse prognosis. Periods of stability may be interspersed with spells of more accelerated decline but failure to recover back to baseline following these episodes is common. Mortality is increased following acute exacerbations.

Pirfenidone and nintedanib, two licensed antifibrotic agents, have been shown to slow the rate of FVC decline. The most common

Box 20.45 Other idiopathic interstitial pneumonias (IIPs)

IIP	Presentation	HRCT	Pathology	Treatment[a]
Desquamative interstitial pneumonia (DIP)	Middle-aged smokers Males 2:1 females Dyspnoea and cough over weeks to months	Widespread ground-glass opacification	Alveolar spaces filled with pigmented macrophages (due to tobacco smoke)	Smoking cessation—may remit spontaneously Corticosteroids in severe or progressive disease ± additional immunosuppressants Response generally good
Respiratory bronchiolitis interstitial lung disease (RBILD)	Current or ex-smokers Similar to DIP	Centrilobular nodules Ground-glass opacification	Pigmented macrophages in lumen of respiratory bronchioles	Smoking cessation No clear benefit to corticosteroids Outcome more favourable than in DIP
Acute interstitial pneumonia (AIP)	Dyspnoea and progressive respiratory failure over days to weeks Often preceded by viral prodrome	Ground-glass opacification Traction bronchiectasis Consolidation Septal thickening	Diffuse alveolar damage (DAD)	Pulsed i.v. methylprednisolone for 3 days followed by maintenance oral corticosteroids Additional immunosuppressants may be required Mortality 50%–80%
Non-specific interstitial pneumonia (NSIP)	Similar to IPF but more indolent course May be associated with connective tissue disease	Similar to IPF but increased ground-glass opacification, minimal honeycombing	Uniform inflammatory infiltrate with or without fibrosis (fibrotic vs cellular NSIP)	Smoking cessation Corticosteroids ± additional immunosuppressants e.g., azathioprine, cyclophosphamide Prognosis better with cellular form Outcome more favourable than in IPF
Cryptogenic organizing pneumonia (COP)	Influenza-like symptoms, dyspnoea, cough over weeks to months Secondary OP may be related to connective tissue, autoimmune disease, or drugs	Bilateral flitting/migratory peripheral consolidation Variable ground-glass opacification	Buds of connective tissue (Masson bodies) in alveoli and alveolar ducts	Usually rapidly responsive to corticosteroids but relapses common

[a]Nintedanib is recommended as an additional therapy in all fibrotic interstitial lung disease with a progressive phenotype (symptoms, lung function, or radiology), despite adequate treatment of underlying cause.
HIV, Human immunodeficiency virus; *HRCT,* high-resolution computed tomography; *IPF,* idiopathic pulmonary fibrosis.

side-effects are a reversible photosensitive rash and gastrointestinal disturbance. Gastro-oesophageal reflux disease should be treated if symptomatic. Immunosuppression is avoided in IPF and steroids are not recommended.

Even with treatment, IPF is a life-limiting disease and transplant assessment should be undertaken in accordance with local guidelines. All patients should have their need for supportive care evaluated with respect to oxygen therapy, pulmonary rehabilitation, and palliative care input.

Other idiopathic interstitial pneumonias

These are described in **Box 20.45**.

Hypersensitivity pneumonitis

Hypersensitivity pneumonitis (HP) is an allergic reaction affecting the small airways and alveoli in response to an inhaled antigen or occasionally following ingestion of a causative drug. See **Box 20.46**. One of the most common causes worldwide is farmer's lung, which affects up to 9% of farmers in humid climates. Cigarette smokers have a lower risk of developing HP due to decreased antibody reaction to antigen, but once established, smoking may lead to a more chronic or severe disease course.

Pathogenesis

Histological features include chronic inflammatory infiltrates and poorly defined interstitial granulomas, together with interstitial fibrosis and honeycomb change in chronic disease. The allergic response to the inhaled antigen involves both cellular immunity and deposition of immune complexes, causing foci of inflammation through activation of complement via the classical pathway.

Clinical features

HP can be categorized according to the time course of symptoms, determined by duration and intensity of exposure. Symptoms include weight loss, malaise, dyspnoea, and cough. Auscultation reveals inspiratory squeaks due to bronchiolitis, and bilateral fine crackles. Wheeze is uncommon.

- ***Acute:*** Symptom onset 4–6 h following exposure. Fever is common and patients may be mistakenly diagnosed with infection. Resolution occurs 24–48 h after inciting antigen removal.
- ***Subacute:*** Usually occurs with intermittent or lower-level exposure. Improvement is seen in weeks to months following antigen removal.
- ***Chronic:*** Insidious onset of respiratory and constitutional symptoms is typical, with no history of preceding acute symptoms.

Box 20.46 Some causes of hypersensitivity pneumonitis

Disease	Situation	Antigens
Farmer's lung	Forking mouldy hay/vegetable material	Thermophilic actinomycetes, e.g., *Micropolyspora faeni* Fungi, e.g., *Aspergillus umbrosus*
Bird fancier's lung	Handling pigeons, cleaning lofts or budgerigar cages	Proteins present in the 'bloom' on the feathers and in excreta
Maltworker's lung	Turning germinating barley	*Aspergillus clavatus*
Humidifier fever	Contaminated humidifying systems in air conditioners or factories (especially printing works)	Possible variety of bacterium or amoeba (e.g., *Naegleria gruberi*) Thermophilic actinomycetes
Mushroom worker's lung	Turning mushroom compost	Thermophilic actinomycetes
Cheese washer's lung	Mouldy cheese	*Penicillium casei* *Aspergillus clavatus*
Winemaker's lung	Mould on grapes	Botrytis

Finger clubbing may be present. Progression to irreversible fibrosis is associated with increased mortality.

Investigations

A high index of suspicion and detailed exposure history are required. Identification of a culprit antigen in the context of typical clinical and radiological findings often makes lung biopsy unnecessary.

- ***Chest radiograph*** may be normal in acute and subacute disease. When present, abnormalities include diffuse small nodules and increased reticular shadowing.
- ***CT*** shows nodules with ground-glass opacity and evidence of air trapping. Increased reticulation and honeycomb change are seen in advanced disease. Abnormalities are most marked in mid or upper zones.
- ***Lung function tests*** are not diagnostic. A restrictive ventilatory defect with decreased carbon monoxide gas transfer is seen in chronic disease.
- ***Precipitating antibodies*** are present in the serum. One-quarter of pigeon fanciers have precipitating IgG antibodies against pigeon protein and droppings in their serum, but only a small proportion have lung disease. Precipitating antibodies are therefore evidence of exposure, not disease.
- ***Bronchoalveolar lavage*** shows a lymphocytosis. A low CD4:CD8 ratio can help differentiate HP from sarcoidosis.
- ***Lung biopsy*** demonstrates a lymphocyte-rich infiltrate with varying degrees of fibrosis, depending on chronicity.

Differential diagnosis

Although HP due to inhalation of the spores of *Micropolyspora faeni* is common among farmers (farmer's lung), it is probably more usual for these individuals to suffer from asthma related to inhalation of antigens from the variety of mites that infest stored grain and other vegetable material, such as *Lepidoglyphus domesticus, L. destructor,* and *Acarus siro*.

Management

Avoidance of exposure to the inciting antigen (if known) is key and may require change in work practice. Pigeon fancier's lung is more difficult to control, as affected individuals remain strongly attached to their hobby. Prednisolone should be initiated in patients whose symptoms persist despite withdrawal from the causative antigen, and in severe disease. Immunosuppressants such as azathioprine and mycophenolate (MMF) may be used as second-line steroid-sparing agents. Established fibrosis will not resolve and, in some patients, the disease may progress inexorably to respiratory failure despite intensive therapy. Farmer's lung is a recognized occupational disease in the UK and sufferers are entitled to compensation, depending on the degree of disability.

Rare interstitial lung diseases

Langerhans cell histiocytosis

This rare disease is characterized histologically by proliferation of Langerhans cells. There is a wide variation in clinical presentation, from isolated lytic bone lesions to multisystem disease involving skin, lymph nodes, and major organs (more common in young children). Pulmonary involvement occurs in 10% of cases and is strongly associated with cigarette smoking. Recurrent spontaneous pneumothorax is seen in up to 25% and is a common mode of presentation. CT shows characteristic interstitial thickening, nodules, cysts, and honeycombing with mid- and upper-zone predominance. This may be sufficient for diagnosis in a young smoker (typical age 20–40 years). Smoking cessation is essential. Various treatment strategies including corticosteroids, chemotherapy agents, and the purine analogue cladribine, have been used with variable success. Lung transplantation may be considered in advanced disease. Outcome varies from spontaneous remission to progressive end-stage fibrosis. Overall, 5-year survival is 75%.

Pulmonary lymphangioleiomyomatosis

Pulmonary lymphangioleiomyomatosis (LAM) is a rare disorder of premenopausal females, causing hamartomatous smooth muscle infiltration of the lungs. Extrapulmonary involvement, especially with renal angiomyolipomas (hamartomas), is common. Some 15% of patients with pulmonary LAM have tuberous sclerosis. Presentation is with dyspnoea, chylous pleural effusions, and pneumothorax. CT shows diffuse thin-walled cysts scattered throughout the lungs. Treatment with hormonal manipulation or oophorectomy has shown a variable response. Sirolimus (rapamycin) can be effective but lung transplantation may be necessary.

Pulmonary alveolar proteinosis

In this rare disease, lipoproteinaceous material accumulates within the alveoli. It can be congenital but most cases are acquired and appear to have an autoimmune basis, with antibodies directed against cytokine granulocyte-macrophage colony-stimulating factor (GM-CSF). The disease mostly affects males and presents with progressive exertional dyspnoea and cough. Inspiratory crackles are present in 50%. Diagnosis is made by bronchoalveolar lavage,

which reveals a milky appearance and many large, foamy macrophages but few other inflammatory cells. Initial therapy is with whole-lung lavage.

Small-vessel vasculitides

The vasculitides are a group of autoimmune diseases that cause inflammation of blood vessels. The small-vessel vasculitides associated with ***anti-neutrophil cytoplasmic antibody (ANCA)*** include granulomatosis with polyangiitis (GPA), microscopic polyangiitis (MPA), and eosinophilic granulomatosis with polyangiitis (EGPA, formerly called Churg–Strauss syndrome). Staining for ANCA shows either a diffuse pattern (c-ANCA) with antibodies directed against proteinase 3 (PR3), or a perinuclear pattern (p-ANCA) with antibodies to myeloperoxidase (MPO). The respiratory tract and kidneys are frequently involved and the ESR is often markedly elevated (>100 mm/h).

Granulomatosis with polyangiitis

GPA typically affects older adults and is more common in Caucasians. c-ANCA is usually positive with elevated PR3 antibodies. The ears and upper respiratory tract are frequently affected with bloody nasal discharge, crusting and destruction, sinusitis, and otitis media. Evidence of glomerulonephritis should be sought on urinalysis (see Chapter 26). Respiratory symptoms include cough, dyspnoea, and pleuritic chest pain. Diffuse alveolar haemorrhage occurs in up to 45% and haemoptysis can be life-threatening. Thoracic imaging characteristically shows multiple nodules that often cavitate, areas of consolidation, and ground-glass opacification (which may be due to pulmonary haemorrhage). Diagnosis should be confirmed with biopsy of the active site but it is sometimes necessary to initiate empirical treatment in acutely unwell patients. Biopsy has highest yield from kidneys, if renal involvement, or nasal sinuses if sinus involvement. Lung biopsy has lower yield. Initial immunosuppressant therapy is with a combination of glucocorticoids, cyclophosphamide, rituximab, or mycophenolate.

Microscopic polyangiitis

MPA is primarily associated with p-ANCA positivity. Presentation and treatment are similar to those of granulomatosis with polyangiitis. It is diagnosed on tissue biopsy, where the absence of granuloma formation differentiates it from GPA.

Eosinophilic granulomatosis with polyangiitis

EGPA classically presents in early adulthood with allergic rhinitis, asthma that is often difficult to control, and peripheral blood eosinophilia (>10%). ANCA is positive in up to 60% (usually p-ANCA). Systemic vasculitis subsequently develops, sometimes many years later. Involvement of skin (tender subcutaneous nodules, petechiae, or purpuric lesions), peripheral nerves (mononeuritis multiplex), heart, kidneys, and gastrointestinal tract may occur. Chest radiography shows migratory patchy opacities that may be accompanied by nodules and pleural effusions. EGPA generally responds well to corticosteroids, although additional immunosuppressants are required for severe or refractory disease. Occasionally, EGPA is 'unmasked' when oral steroids are withdrawn in patients being treated for asthma.

Anti-glomerular basement membrane disease (Goodpasture's syndrome)

Anti-glomerular basement membrane (anti-GBM) disease is characterized by the triad of pulmonary haemorrhage, glomerulonephritis, and the presence of circulating antibodies directed against an antigen intrinsic to the basement membrane of both kidney and lung. Respiratory symptoms may precede the onset of glomerulonephritis by weeks or months. Chest radiography shows transient patchy shadows due to intrapulmonary haemorrhage. Haemoptysis can vary from negligible to life-threatening. Carbon monoxide gas transfer is increased due to the presence of haemoglobin in the alveoli. The extent of renal recovery depends on early detection and treatment. Treatment is with plasmapheresis to remove circulating antibodies and immunosuppression (prednisolone and cyclophosphamide) to prevent further antibody production.

Diffuse alveolar haemorrhage

Bleeding into the alveolar spaces is associated with certain drugs, infections, and autoimmune rheumatic diseases including vasculitis, but can also occur without an identifiable precipitating cause. Haemoptysis may be minimal (or even absent), despite significant blood loss into the lungs. It is one of the relatively few causes of raised carbon monoxide gas transfer (K_{CO}). Gas exchange is impaired due to the presence of blood in the alveoli and patients may present with severe respiratory failure requiring intensive care support. Treatment is directed at the underlying cause, if known.

Pulmonary manifestations of autoimmune diseases

Rheumatoid lung disease

The lungs can be affected by rheumatoid arthritis (RA) (Fig. 20.34), and also by some of the drugs used in its treatment.

- ***Pleural effusions*** are often unilateral and tend to be chronic. Low glucose content is typical but not specific.
- ***Pulmonary fibrosis*** in RA has similar clinical features to the idiopathic form of the disease but often follows a more chronic course. In patients taking methotrexate, it is often impossible to determine whether fibrosis is due to the drug or the underlying disease so methotrexate is substituted for an alternative agent.
- ***Rheumatoid nodules*** appearing on chest radiography may be single or multiple, ranging in size from a few millimetres to a few

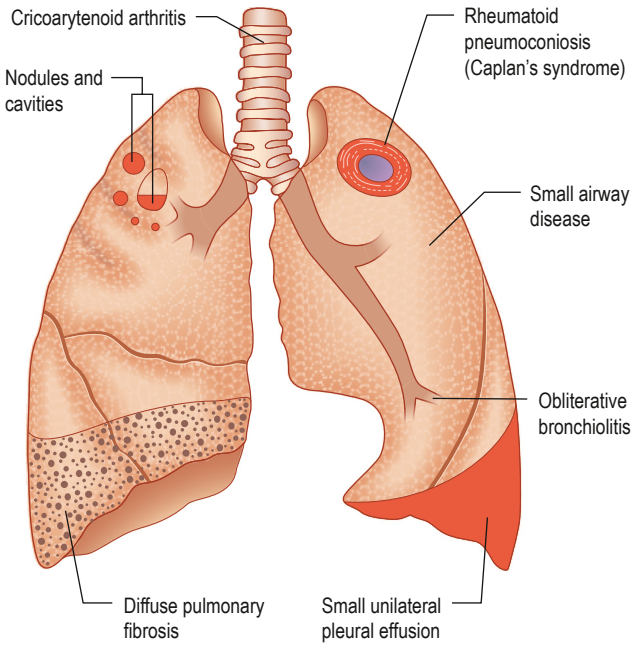


Fig. 20.34 **Respiratory manifestations of rheumatoid disease.**

centimetres, with frequent cavitation. They usually produce no symptoms but can give rise to a pneumothorax or pleural effusion.

- ***Obliterative bronchiolitis*** causing concentric narrowing of the bronchioles is a rare disorder characterized by progressive breathlessness and irreversible airflow limitation. Response to immunosuppressive therapy is generally poor. Macrolide antibiotics may have a role.
- ***Cricoarytenoid joint involvement*** in RA gives rise to dyspnoea, stridor, and hoarseness. Occasionally, severe obstruction necessitates tracheostomy.
- ***Caplan's syndrome*** is due to occupational dust inhalation in patients with RA. It occurs particularly in coal worker's pneumoconiosis but can be caused by exposure to other dusts, such as silica and asbestos. Typically, the chest radiograph shows rounded nodules 0.5–5.0 cm in diameter but progressive fibrosis can sometimes occur. These lesions may precede the development of arthritis. Rheumatoid factor is positive in the majority of patients.
- ***Drugs*** used in the treatment of RA can cause pulmonary disease:
  - pneumonitis with methotrexate, gold, and NSAIDs
  - fibrosis with methotrexate
  - bronchospasm with NSAIDs
  - infections with corticosteroids and other immunosuppressants
  - reactivation of TB with anti-TNF therapy.

Systemic lupus erythematosus

The most common respiratory manifestation is pleurisy, occurring in up to two-thirds of cases, with or without an effusion, which is usually small and bilateral. Pneumonia also occurs, either because of infection or because of the disease process itself. In contrast to RA, diffuse pulmonary fibrosis is uncommon.

Systemic sclerosis

Some degree of lung involvement is present in most cases of systemic sclerosis, and pulmonary complications are the leading cause of death. Interstitial fibrosis and pulmonary arterial hypertension are the most common pathologies. Serial lung function testing can aid early detection. Other complications include bronchiectasis and aspiration pneumonitis secondary to oesophageal dysfunction.

Pulmonary infiltration with eosinophilia

This is a group of allergic respiratory conditions. Common types and characteristics are shown in **Box 20.47**. They range from simple pulmonary eosinophilia to the often-fatal hypereosinophilic syndrome. ***Simple pulmonary eosinophilia*** is a relatively mild illness, with a slight fever and cough, and usually lasts for less than 2 weeks. If symptoms become more prolonged and there is an eosinophilia in the blood, it is then called ***chronic pulmonary eosinophilia***. In both conditions, the chest radiograph shows either localized or diffuse opacities. The simple form is probably due to a transient allergic reaction in the alveoli. Many allergens have been implicated including *Ascaris lumbricoides*, *Ancylostoma*, *Trichuris*, *Trichinella*, *Taenia*, and *Strongyloides*. Drugs such as aspirin, penicillin, nitrofurantoin, and sulphonamides have also been implicated. Often, however, no allergen is identified. The disease is self-limiting and no treatment is required, apart from withdrawing the identified cause. In the more chronic form, all unnecessary treatment should be withdrawn and corticosteroid therapy may be indicated, with resolution over weeks.

Asthmatic bronchopulmonary eosinophilia is characterized by the presence of asthma, transient fleeting shadows on the chest X-ray, and blood or sputum eosinophilia. By far the most common cause worldwide is allergy to *A. fumigatus* (see later), although *Candida albicans* and other mycoses may be the inciting allergen in a small number of patients. In many, no allergen can be identified. Whether these cases are intrinsic or driven by an unidentified extrinsic factor is uncertain. ***Tropical pulmonary eosinophilia*** is the term reserved for an allergic reaction to microfilaria from *Wuchereria bancrofti*.

Hypereosinophilic syndrome is characterized by eosinophilic infiltration in various organs, sometimes associated with an eosinophilic arteritis. The heart muscle is particularly involved, but pulmonary involvement in the form of a pleural effusion or interstitial lung disease occurs in about 40% of cases.

Diseases caused by *Aspergillus fumigatus*

The various types of lung disease caused by *A. fumigatus* are illustrated in **Fig. 20.35**.

Spores (diameter 5 mm) are readily inhaled and are present in the atmosphere throughout the year. *Aspergillus* can be grown from sputum in up to 15% of patients with chronic lung disease, in whom it does not produce active fungal disease. They are a cause of extrinsic asthma in atopic individuals.

Allergic bronchopulmonary aspergillosis (asthmatic pulmonary eosinophilia)

This rare disease is caused by a hypersensitivity reaction when the bronchi are colonized by *Aspergillus*. It can complicate asthma and CF. Proximal bronchiectasis occurs.

Episodes of eosinophilic pneumonia present with wheeze, cough, fever, and malaise, associated with expectoration of firm sputum plugs containing fungal mycelium. Occasionally, large mucus plugs obliterate the bronchial lumen, causing collapse of the lung. Left untreated, repeated episodes of eosinophilic pneumonia can result in progressive pulmonary fibrosis that usually affects the upper zones.

The peripheral blood eosinophil count is usually raised and total levels of IgE are usually extremely high, at more than 1000 ng/mL (both specific to *Aspergillus* and non-specific). Sputum may show eosinophils and mycelia. Treatment is with 0.5 mg/kg of prednisolone daily for 2 weeks followed by a tapering regimen, which causes rapid clearing of pulmonary infiltrates. Antifungal agents should be

Box 20.47 Common types and characteristics of pulmonary infiltration with eosinophilia

Disease	Symptoms	Blood eosinophils (%)	Multisystem involvement	Duration	Outcome
Simple pulmonary eosinophilia	Mild	10	None	<1 month	Good
Prolonged pulmonary eosinophilia	Mild/moderate	>20	None	>1 month	Good
Asthmatic bronchopulmonary eosinophilia	Moderate/severe	5–20	None	Years	Fair
Tropical pulmonary eosinophilia	Moderate/severe	>20	None	Years	Fair
Hypereosinophilic syndrome	Severe	>20	Always	Months/years	Poor

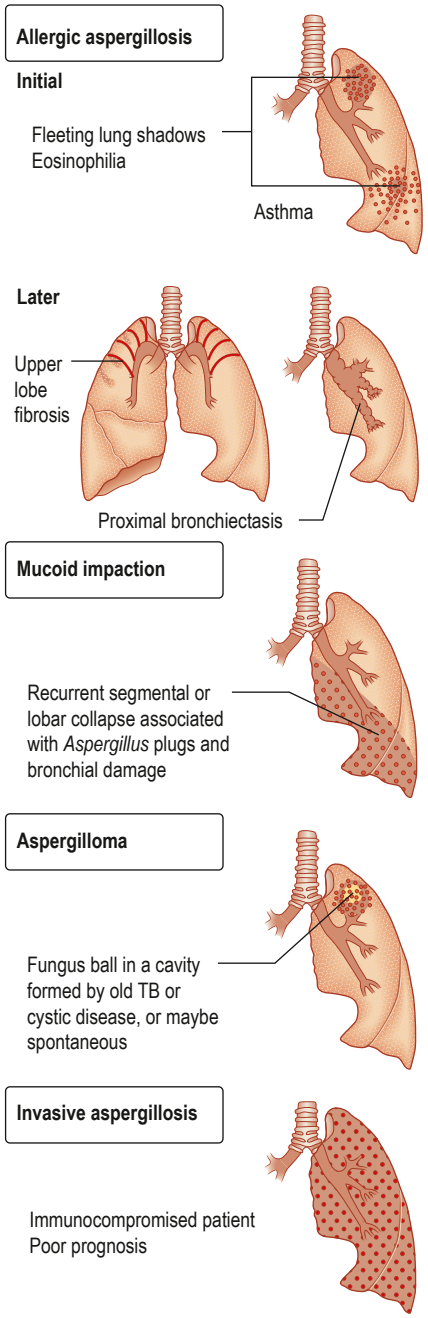


Fig. 20.35 **Diseases caused by *Aspergillus fumigatus*.**

used in patients on high doses of steroids. The asthma component responds to inhaled corticosteroids, although these do not influence the occurrence of pulmonary infiltrates.

Aspergilloma and invasive aspergillosis

Aspergilloma is the growth of *A. fumigatus* within previously damaged lung tissue, where it forms a ball of mycelium within lung cavities. Typically, the chest radiograph shows a round lesion with an air 'halo' above it. The aspergilloma itself causes little trouble, though occasionally haemoptysis may occur, requiring resection of the area of damaged lung containing the aspergilloma. Treatment is with oral antifungal agents, although invasive aspergillosis is a recognized complication of immunosuppression and often requires intravenous antifungal therapy.

Box 20.48 Some drug-induced respiratory reactions

Bronchospasm
- Penicillins, cephalosporins
- Sulphonamides
- Aspirin/NSAIDs
- Monoclonal antibodies, e.g., infliximab
- Iodine-containing contrast media
- β-Adrenoceptor-blocking drugs (e.g., propranolol)
- Non-depolarizing muscle relaxants
- Intravenous thiamine
- Adenosine

Interstitial lung disease and/or fibrosis
- Amiodarone
- Anakinra (IL-1 receptor antagonist)
- Nitrofurantoin
- Paraquat (weedkiller)
- Continuous oxygen
- Cytotoxic agents: particularly busulfan, CCNU, bleomycin, methotrexate

Pulmonary eosinophilia
- Antibiotics:
  - Penicillin
  - Tetracycline
  - Sulphonamides, e.g., sulfasalazine
- NSAIDs
- Cytotoxic agents

Acute lung injury
- Paraquat

Pulmonary hypertension
- Fenfluramine, dexfenfluramine, phentermine

SLE-like syndrome including pulmonary infiltrates, effusions, and fibrosis
- Hydralazine
- Procainamide
- Isoniazid
- Phenytoin
- Anti-TNF agents, particularly etanercept and infliximab

Drug- and radiation-induced respiratory reactions

Drugs affecting the respiratory system are shown in **Box 20.48**, together with the types of reaction they produce. Pulmonary infiltrates with fibrosis may result from a number of cytotoxic drugs used in cancer treatment. The most common causative drug is bleomycin, which causes dose-related pulmonary damage. The most sensitive test is a decrease in carbon monoxide gas transfer, and therefore gas transfer should be measured repeatedly during treatment with the drug. The use of corticosteroids may help resolution.

Irradiation of the lung during radiotherapy can cause radiation pneumonitis. Patients experience breathlessness and a dry cough. Radiation pneumonitis results in a restrictive lung defect. Corticosteroids should be given in the acute stage.

Further reading

Flaherty KR et al. Nintedanib in progressive fibrosing interstitial lung diseases. *N Engl J Med* 2019; 381:1718–1727.
Lamoth F, Calandra T. Pulmonary aspergillosis: diagnosis and treatment. *Eur Resp Rev* 2022; 31:220114.
Muthiah MP, El Gamal A. Sarcoidosis. *BMJ Best Practice*; https://bestpractice.bmj.com/topics/en-gb/109.
Raghu G, Remy-Jardin M, Richeldi L, et al. Idiopathic pulmonary fibrosis (an update) and progressive fibrosis in adults: an official ATS/ERS/JRS/ALAT clinical practice guideline. *Am J Respir Crit Care Med* 2022; 205:e18–e47.
Thillai, M et al. BTS clinical statement on pulmonary sarcoidosis. *Thorax* 2019.

LUNG AND HEART–LUNG TRANSPLANTATION

Indications and donor selection

The main diseases treated by transplantation are:

- pulmonary fibrosis
- primary pulmonary hypertension
- cystic fibrosis
- bronchiectasis
- emphysema (particularly α_1-antitrypsin inhibitor deficiency)
- Eisenmenger's syndrome.

Patients selected for transplantation are usually under 60 years, have an untreated life expectancy of less than 18 months, with no underlying cancer and no serious systemic disease. Organs are taken from donors under 40 years, with good cardiac and lung function, and chest measurements slightly smaller than those of the recipient. Matching for ABO blood group is essential but Rhesus blood group compatibility is not necessary. Since donor material is limited, single-lung transplantation is preferred to double-lung or heart–lung transplantation. It can be successfully undertaken in pulmonary fibrosis, pulmonary hypertension, and emphysema. Bilateral lung transplantation is needed in infective conditions to prevent spillover of bacteria from the diseased lung to a single transplanted lung. Eisenmenger's syndrome requires heart–lung transplantation.

Immunosuppression is with ciclosporin (the inhaled formulation has shown benefit), tacrolimus, azathioprine, mycophenolate mofetil, and prednisolone.

Complications and their treatment

- ***Early post-transplant pulmonary oedema*** requires diuresis and ventilatory support.
- ***Infections*** are common, particularly within the first 3 months, and need prompt treatment:
  - bacterial pneumonia: antibiotics
  - cytomegalovirus infection: ganciclovir or valganciclovir (prophylaxis given in the early post-transplantation period)
  - herpes simplex virus: aciclovir
  - *Pneumocystis jirovecii:* co-trimoxazole (prophylaxis given in the early post-transplantation period).
- ***Rejection:***
  - Early (first few weeks): high-dose intravenous corticosteroids.
  - Late (after 3 months): often produces the histological pattern of obliterative bronchiolitis. High-dose intravenous corticosteroids are sometimes effective.
- ***Post-transplant lymphoproliferative disease*** refers to a range of lymphomas seen in recipients of solid organ transplants; they may respond to rituximab, an anti–B-cell monoclonal antibody, or other forms of chemotherapy.

Prognosis

Several studies show a major improvement in quality of life after transplantation. One-year survival rates are around 80%, with mortality rate thereafter of about 10%/year. Death is mainly due to obliterative bronchiolitis. Overall survival varies with the original diagnosis but median survival is approximately 4 years.

OCCUPATIONAL LUNG DISEASE

Exposure to dusts, gases, vapours, and fumes at work can cause several different types of lung disease:

- ***acute bronchitis*** and ***pulmonary oedema*** from irritants such as sulphur dioxide, chlorine, ammonia, or oxides of nitrogen
- ***pulmonary fibrosis*** caused by mineral dust
- ***occupational asthma*** (see **Box 20.14**) is now the most common industrial lung disease in high-income countries
- ***hypersensitivity pneumonitis*** (see **Box 20.46**)
- ***bronchial carcinoma*** due to industrial agents, e.g., asbestos, polycyclic hydrocarbons, and radon in mines.

Box 20.49 Taking an occupational history

- Proceed chronologically. Most people cannot randomly remember, for example, what they might have been doing 20 years ago, or indeed, if asked in isolation, when they worked in a particular job. But if you start at the beginning of their life and work forward they find it much easier to remember (try it on yourself, starting with school).
- Start by asking the patient how old they were when they left school, and what job or further education they went to; then ask them to continue through their life to the present day.
- Particularly for those who went on to further education, ask about holiday jobs (you might be surprised at their responses) and if they travelled overseas with their employment, especially if they were in the armed forces.
- Do not assume that all 80 year olds are retired or that all young patients are in employment.

The degree of fibrosis that follows inhalation of mineral dust varies. While iron (siderosis), barium (baritosis), and tin (stannosis) lead to dramatic, dense, nodular shadowing on chest radiography, their effect on lung function and symptoms is minimal. In contrast, exposure to silica or asbestos leads to extensive fibrosis and disability. Coal dust has an intermediate fibrogenic effect and historically accounted for 90% of all compensated industrial lung disease in the UK.

Box 20.49 outlines how to take an occupational history in lung disease.

Coal-worker's pneumoconiosis

Caused by coal dust particles approximately 2–5 μm in diameter and retained in the small airways and alveoli. Incidence is related to total dust exposure, which is highest at the coal face, particularly if ventilation and dust suppression are poor. Improved ventilation and working conditions reduce disease risk.

Two very different syndromes result from the inhalation of coal.

Simple pneumoconiosis

Due to deposition of coal dust in the lung, which produces fine micronodular shadowing on the chest radiographs. It is graded on the chest radiograph appearance according to standard categories set by the International Labour Office:

- ***Category 1:*** small round opacities definitely present but few in number
- ***Category 2:*** numerous small round opacities but normal lung markings still visible
- ***Category 3:*** very numerous small round opacities and normal lung markings partly or totally obscured.

There is considerable dispute about the effects of simple pneumoconiosis on respiratory function and symptoms. In many cases, symptoms may be due to COPD related to coexisting cigarette smoking. Changes to UK workers' compensation legislation mean that coal miners who develop COPD are compensated for disability regardless of chest radiograph appearance.

Simple pneumoconiosis can progress to the development of progressive massive fibrosis (see next section). The latter virtually never occurs on a background of category 1 simple pneumoconiosis but does arise in about 7% of category 2 disease and 30% of category 3 disease. Miners with category 1 pneumoconiosis are

unlikely to receive compensation unless they also have evidence of COPD. Those with more extensive radiographic changes are compensated solely on the basis of chest radiograph appearances.

Progressive massive fibrosis

In progressive massive fibrosis (PMF), patients develop round, fibrotic masses several centimetres in diameter, almost invariably situated in the upper lobes, sometimes with necrotic central cavities. Rheumatoid factor and antinuclear antibodies are both often present in the serum. Pathologically, there is apical destruction and disruption of the lung, resulting in emphysema and airway damage. Lung function tests show a mixed restrictive and obstructive defect with loss of lung volume, irreversible airflow limitation, and reduced gas transfer.

PMF leads to considerable effort dyspnoea, usually with a cough. Sputum may be black. The disease can progress (or even develop) after exposure to coal dust has ceased and may lead to respiratory failure.

Silicosis

Silicosis is caused by inhalation of silica (silicon dioxide). While uncommon, it may still be encountered in stonemasons, sandblasters, pottery and ceramic workers, and foundry workers involved in fettling (removing sand from metal castings made in sand-filled moulds). The dust is highly fibrogenic; a coal miner can remain healthy with 30 g of coal dust in the lungs but 3 g of silica is sufficient to kill. Silica seems particularly toxic to alveolar macrophages and readily initiates fibrogenesis. The chest radiograph appearances and clinical features of silicosis are similar to those of PMF but distinctive thin streaks of calcification may be seen around the hilar lymph nodes ('eggshell' calcification).

Diseases caused by asbestos

Asbestos is a mixture of silicates of iron, magnesium, nickel, cadmium, and aluminium, and has the unique property of occurring naturally as a fibre. It is remarkably resistant to heat, acid, and alkali, and has been widely used for roofing, insulation, and fireproofing. Several different types of asbestos are recognized: about 90% is chrysotile, 6% crocidolite, and 4% amosite.

Chrysotile (white asbestos) is the softest asbestos fibre. Each fibre is often as long as 2 cm but only a few microns thick. It is less fibrogenic than crocidolite.

Crocidolite (blue asbestos) is particularly resistant to chemical destruction and exists in straight fibres up to 50 cm in length and 1–2 μm in width. Crocidolite is the type of asbestos most likely to produce asbestosis and mesothelioma.

Amosite (brown asbestos) was used in cement and pipe insulation and has sharp, needle-like fibres. Exposure carries a higher risk of cancer compared to chrysotile asbestos.

Exposure to asbestos occurred particularly in shipbuilding yards and in power stations, but it was used so widely that low levels of exposure were very common. There is a considerable time lag between exposure and development of disease, particularly mesothelioma (20–40 years). Regulations in the UK prohibit the use of crocidolite and severely restrict the use of chrysotile. Careful dust control measures are enforced, which should eventually abolish the problem.

Risk of primary lung cancer (usually adenocarcinoma) is increased in people exposed to asbestos, even non-smokers. This risk is about five- to sevenfold greater in those who have parenchymal asbestosis and about 1.5-fold in those with pleural plaques without parenchymal fibrosis. A synergistic relationship exists between asbestosis and cigarette smoking for the risk of bronchial carcinoma, depending on the amount smoked and degree of asbestos exposure.

Diseases caused by asbestos are summarized in **Box 20.50**. Bilateral diffuse pleural thickening, asbestosis, mesothelioma, and asbestos-related carcinoma of the bronchus are all eligible for industrial injury benefits in the UK.

Asbestosis

Asbestosis is defined as fibrosis of the lungs caused by asbestos dust, which may or may not be associated with fibrosis of the parietal or visceral layers of the pleura. It is a progressive disease characterized by breathlessness, accompanied by finger clubbing and bilateral basal end-inspiratory crackles. Minor degrees of fibrosis that are not seen on chest radiography are often revealed on CT scan. No treatment is known to alter progress, though corticosteroids are often prescribed.

Byssinosis

This disease is caused by cotton dust. It occurs worldwide but is declining where the number of people employed in cotton mills is falling. Typically, symptoms start on the first day back at work after a break (Monday sickness), with improvement as the week progresses. Tightness in the chest, cough, and breathlessness occur within the first hour in dusty areas of the mill. The exact nature of the disease and its aetiology remain disputed. Pure cotton does not cause the disease, and cotton dust has some effect on airflow limitation in all those exposed. Individuals with asthma are particularly affected by exposure to cotton dust. The most likely aetiology is constriction of the airways of the lung caused by endotoxins from bacteria present in raw cotton. There are no changes on chest radiographs.

Berylliosis

Beryllium–copper alloy has a high tensile strength and is resistant to metal fatigue, high temperature, and corrosion. It is used in the aerospace industry, atomic reactors, and many electrical devices. When beryllium is inhaled, it can cause a systemic illness with a clinical picture similar to that of sarcoidosis. Clinically, there is progressive dyspnoea with pulmonary fibrosis. Strict controls have made this disease a rarity.

Further reading

Cullinan P, Muñoz X, Suojalehto H et al. Occupational lung disease: from old and novel exposures to effective preventive strategies. *Lancet Respir Med* 2017; 5:445–455.
European Respiratory Society. *Occupational Lung Disease*; https://www.erswhitebook.org.

MISCELLANEOUS RESPIRATORY DISORDERS

Lung cysts

These can be congenital, bronchogenic, or the result of a sequestrated pulmonary segment (dysplastic lung tissue with no communication with the rest of the tracheobronchial tree). Lung cysts therefore have a wide differential diagnosis, informed by clinical presentation. Causes include:

- hydatid disease: fluid-filled cysts
- lung abscesses: thin-walled cysts, found particularly in staphylococcal pneumonia
- tuberculosis: cavitation

Box 20.50 Effects of asbestos on the lung

Effect	Exposure	Chest radiograph	Lung function	Symptoms	Outcome
Asbestos bodies	Light	Normal	Normal	None	Evidence of asbestos exposure only
Pleural plaques	Light	Pleural thickening (parietal pleura) and calcification (also in diaphragmatic pleura)	Mild restrictive ventilatory defect	Rare, occasional mild effort dyspnoea	No other sequelae
Effusion	First two decades following exposure	Effusion	Restrictive	Pleuritic pain, dyspnoea	Often recurrent
Bilateral diffuse pleural thickening	Light/moderate	Thickening of both parietal and visceral pleura >5 mm and extending over more than ¼ of chest wall	Restrictive ventilatory defect	Effort dyspnoea	May progress in absence of further exposure
Mesothelioma	Light: interval of 20–40 years from exposure to disease	Pleural effusion, usually unilateral	Restrictive ventilatory defect	Pleuritic pain, increasing dyspnoea	Median survival 2 years
Asbestosis	Heavy: interval of 5–10 years from exposure to disease	Diffuse bilateral streaky shadows, honeycomb lung	Severe restrictive ventilatory defect and reduced gas transfer	Progressive dyspnoea	Poor, progression in some cases after exposure
Asbestos-related carcinoma of the bronchus	The features of asbestosis, bilateral diffuse pleural thickening, or bilateral pleural plaques plus those of bronchial carcinoma				Fatal

- septic pulmonary infarction
- primary bronchogenic carcinoma or cavitating metastatic neoplasms
- paragonimiasis caused by the lung fluke *Paragonimus westermani*
- systemic conditions associated with pulmonary cysts, e.g., Birt–Hogg–Dubé syndrome.

Trauma

Trauma to the thoracic wall can cause penetrating wounds and lead to pneumothorax or haemothorax.

Rib fractures are caused by trauma or coughing (particularly in the elderly), and can occur in osteoporosis. Pathological rib fractures are due to metastatic spread (most often bronchus, breast, kidney, prostate, or thyroid cancer). Ribs can also become involved in mesothelioma. Fractures may not be readily visible on a postero-anterior chest radiograph and so lateral radiograph and oblique views may be necessary. Pain prevents adequate chest expansion and coughing, and this can lead to pneumonia. Treatment is with adequate oral analgesia, or by local infiltration or an intercostal nerve block. Two fractures in one rib can lead to a flail segment with paradoxical movement: that is, part of the chest wall moves inwards during inspiration. This can produce inefficient ventilation and may require intermittent positive-pressure ventilation, especially if several ribs are similarly affected.

Rupture of the trachea or a major bronchus can occur during deceleration injuries, leading to pneumothorax, surgical emphysema, pneumomediastinum, and haemoptysis. Surgical emphysema is caused by air leaking into the subcutaneous connective tissue; this can also arise after insertion of an intercostal drainage tube. A pneumomediastinum occurs when air leaks from the lung inside the parietal pleura and extends along the bronchial walls.

Rupture of the oesophagus (see Chapter 22, 'Oesophageal rupture' section) leads to mediastinitis, usually with mixed bacterial infection. This is a serious complication of external injury, endoscopic procedures, bougienage, carcinoma, or necrotic tuberculous lymph nodes and requires broad-spectrum antibiotics.

Lung contusion causes widespread fluffy shadows on the chest radiography due to intrapulmonary haemorrhage. It may give rise to acute respiratory distress syndrome.

Kyphoscoliosis

Kyphoscoliosis may be congenital, due to disease of the vertebrae such as TB or osteomalacia, or due to neuromuscular disease such as Friedreich's ataxia or poliomyelitis. The respiratory effects of severe kyphoscoliosis are often more pronounced than might be expected and respiratory failure and death often occur in the fourth or fifth decade. The abnormality should be corrected at an early stage if possible. Positive airway pressure ventilation delivered through a tightly fitting nasal mask is the treatment of choice for respiratory failure.

Ankylosing spondylitis

Limitation of chest wall movement is often well compensated by diaphragmatic movement and so the respiratory effects of this disease are relatively mild. It is occasionally associated with upper lobe fibrosis.

Pectus excavatum and pectus carinatum

Pectus excavatum causes few problems other than the aesthetic of a deep vertical furrow in the chest, which can be corrected surgically. The heart is seen to lie well to the left on chest radiography.

Pectus carinatum (pigeon chest) is often the result of rickets. No treatment is required.

Pleurisy

Pleurisy is pain arising from any disease of the pleura. Localized inflammation produces sharp localized pain, which is worse on deep inspiration, coughing, and occasionally on twisting and bending movements. Common causes are pneumonia, pulmonary infarct, and carcinoma. Rarer causes include rheumatoid arthritis and systemic lupus erythematosus.

Epidemic myalgia (Bornholm disease) is caused by infection with Coxsackie B virus. It is common in young adults in late summer and autumn. It is characterized by upper respiratory tract illness followed by pleuritic pain in the chest and upper abdomen with tender muscles. The chest radiograph remains normal and spontaneous resolution occurs within a week.

DISORDERS OF THE DIAPHRAGM

Diaphragmatic fatigue

The diaphragm can become fatigued if the force of contraction during inspiration exceeds 40% of the force it can develop in a maximal static effort. When this occurs acutely, in patients with exacerbations of COPD or CF, or in those with quadriplegia, positive-pressure ventilation is required. Breathing against resistance for 30 minutes a day increases strength and endurance.

Unilateral diaphragmatic paralysis

This is common and symptomless. The affected diaphragm is usually elevated and moves paradoxically on inspiration. It can be diagnosed by ultrasound when a sniff causes the paralyzed diaphragm to rise and the unaffected diaphragm to descend. Causes include:

- surgery
- carcinoma of the bronchus with involvement of the phrenic nerve
- neurological disease, including poliomyelitis and herpes zoster
- trauma to the cervical spine, birth injury, or subclavian vein puncture
- infection, such as TB, syphilis, or pneumonia.

Bilateral diaphragmatic weakness or paralysis

This causes breathlessness in the supine position and may lead to sleep apnoea, with daytime headaches and somnolence. Tidal volume is decreased and respiratory rate is increased. Vital capacity is substantially reduced when lying down, and sniffing causes a paradoxical inward movement of the abdominal wall, best seen in the supine position. Causes include viral infections, multiple sclerosis, motor neuron disease, poliomyelitis, Guillain–Barré syndrome, quadriplegia after trauma, and rare myopathies. Treatment is either diaphragmatic pacing or night-time assisted ventilation.

Complete eventration of the diaphragm

This is a congenital condition (invariably left-sided) in which muscle is replaced by fibrous tissue. It presents as marked elevation of the left hemidiaphragm, sometimes associated with gastrointestinal symptoms. Partial eventration, usually on the right, causes a hump (often anteriorly) on the diaphragmatic shadow on chest radiography.

Diaphragmatic hernia

These are most commonly through the oesophageal hiatus but occasionally occur anteriorly through the foramen of Morgagni, posterolaterally through the foramen of Bochdalek, or at any site following traumatic tear.

Hiccups

Hiccups are due to involuntary diaphragmatic contractions with closure of the glottis and are extremely common. Occasionally, patients present with persistent hiccups. This can be as a result of diaphragmatic irritation (e.g., subphrenic abscess) or may have a metabolic cause (e.g., uraemia). The underlying cause should be treated, if known. Gabapentin or pregabalin may be tried.

Further reading

McCool ED, Tzelepis GE. Dysfunction of the diaphragm. *N Engl J Med* 2012; 366:932–942.

MEDIASTINAL LESIONS

The mediastinum is defined as the region between the pleural sacs. It is additionally divided as shown in Fig. 20.36. Tumours affecting the mediastinum are rare. Masses are detected accurately on CT or MRI (Fig. 20.37), and can be localized to the anterior, middle, or posterior mediastinum. The position and characteristics of the lesion will help to determine the underlying aetiology.

Retrosternal or intrathoracic thyroid

The most common mediastinal mass is a retrosternal or intrathoracic thyroid, which is nearly always an extension of the thyroid present in the neck. Enlargement of the thyroid by a colloid goitre or malignant disease, or, rarely, in thyrotoxicosis, can cause displacement of the trachea and oesophagus to the opposite side. Symptoms of compression develop insidiously before producing

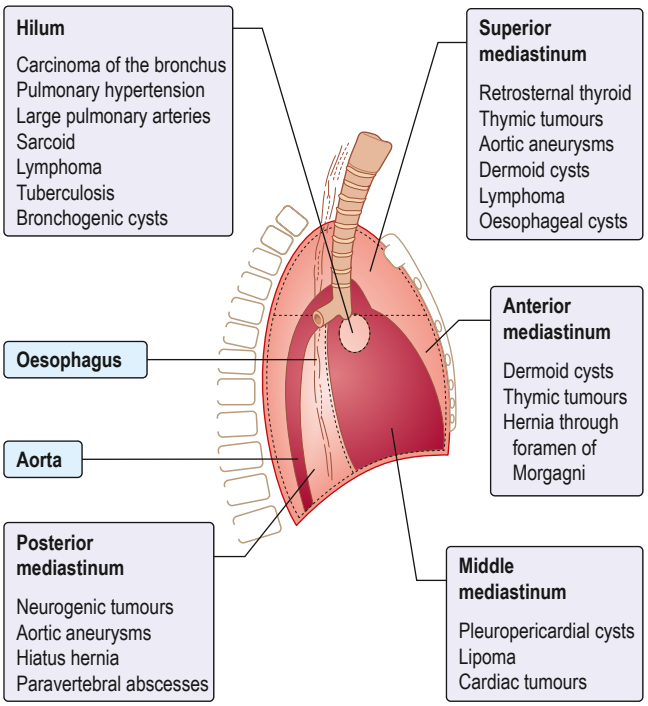


Fig. 20.36 **Subdivisions of the mediastinum and mass lesions.**

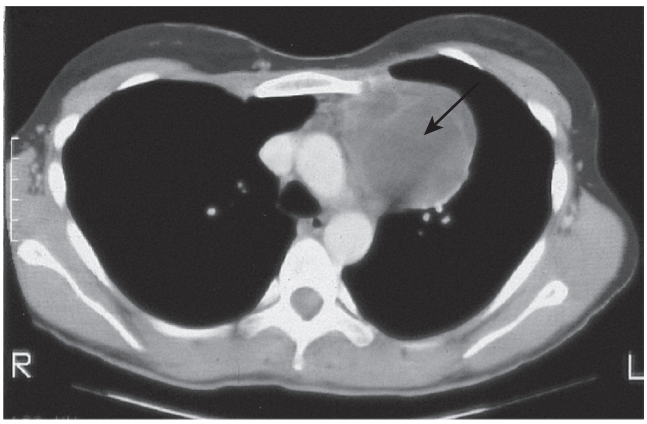


Fig. 20.37 **CT scan of a dermoid cyst *(arrowed)* in the mediastinum.**

dyspnoea. Flow–volume loops are useful to assess physiological impact. Very occasionally, an intrathoracic thyroid may cause dysphagia or hoarseness and vocal cord paralysis due to stretching of the recurrent laryngeal nerve. The treatment is surgical removal.

Thymic tumours (thymomas)

The thymus is large in childhood and occupies the superior and anterior mediastinum. It involutes with age but may be enlarged by cysts, which are rarely symptomatic, or by tumours that compress the trachea, or rarely, the oesophagus. Surgery is the treatment of choice. Approximately half of patients presenting with a thymic tumour have myasthenia gravis. Good's syndrome, a combined defect of humoral and cellular immunity, is seen in 10% of thymomas.

Middle mediastinum pathology

- ***Bronchogenic cyst***, a benign embryological remnant.
- ***Mediastinal lymphadenopathy*** due to a number of conditions including metastatic lesions, primary lung cancer, sarcoid, and infection such as tuberculosis. Both Hodgkin and non-Hodgkin lymphoma commonly present with enlargement of middle mediastinal nodes (see Chapter 39).
- ***Pericardial cysts***, which may be up to 10 cm in diameter, are filled with clear fluid. Up to 70% are situated anteriorly in the cardiophrenic angle on the right side. Infection is rare and malignant change does not occur. The diagnosis may be made on MRI, with needle aspiration in diagnostic uncertainty. No treatment is required but patients should be followed up, as an increase in cyst size suggests an alternative pathology and surgical excision is then advised.
- ***Vascular abnormalities*** include aortic aneurysm and aortic dissection.
- ***Tracheal tumours*.**

Posterior mediastinum pathology

- ***Embryological remnants*** include neurogenic and neuroenteric cysts.
- ***Oesophageal abnormalities***, e.g., oesophageal tumour, hiatus hernia

Significant websites

http://www.asthma.org.uk *Asthma UK.*
http://www.brit-thoracic.org.uk *British Thoracic Society.*
http://www.nhs.uk/live-well/quit-smoking *Good site for those wanting to stop smoking or to help patients to stop.*
http://www.thoracic.org *American Thoracic Society.*

Questions for this chapter can be found on Elsevier eBooks+.

21 Ear, nose, and throat

Elinor Warner and Kay Seymour

Clinical skills 583
The ear 583
Anatomy 583
Clinical approach to the patient with a disorder of the ear 585
Examination 585
Investigations 585
Common disorders of the ear 585
The painful ear (otalgia) 585
The discharging ear (otorrhoea) 587
Hearing loss 588
Vertigo 589
Tinnitus 591
The nose 591
Anatomy and physiology 591
Clinical approach to the patient with a disorder of the nose 591
Examination 591
Common disorders of the nose 591
Nasal obstruction 591
Rhinorrhoea (runny nose) 593
Anosmia 593
Epistaxis 593
The throat 595
Anatomy and physiology 595
Clinical approach to the patient with a disorder of the throat 596
Examination 596
Common disorders of the throat 596
Odonophagia (pain on swallowing) 596
Dysphagia 597
Hoarseness (dysphonia) 598
Stridor 599
Stertor (snoring) 600

CORE SKILLS AND KNOWLEDGE

Ear, nose, and throat (ENT) is a broad specialty that, in addition to covering diseases of the ears, nose, and throat, is also concerned with diseases of the trachea and skull base. It is responsible for head and neck cancer, facial plastic surgery, sleep-disordered breathing, and congenital head and neck anomalies.

One in every five primary care consultations relates to ENT conditions, and while only 3% of emergency department visits are related to ENT diagnoses, they are often life-threatening (e.g., acute airway obstruction, major upper airway haemorrhage). Common emergency presentations include epistaxis, upper airway compromise (e.g., due to infection, malignancy), ingested or inhaled foreign bodies, and penetrating neck injuries.

Surgical treatments for ENT conditions often involve technological aids such as endoscopes, microscopes, powered instruments, lasers, and more recently robotics. Technology is also used in treatment (e.g., cochlear and brainstem auditory implants). The head and neck are complicated anatomical regions with multiple vital structures near each other. While being mindful of this complicated anatomy, surgery varies from the fine (e.g., stapedotomy) to expansive (e.g., laryngopharyngectomy). Nearly three-quarters of the operative load is now performed as day case surgery, reducing inpatient burden.

Multidisciplinary working is important with many conditions managed in collaboration with other specialists including oncologists and plastic surgeons for head and neck cancers, endocrinologists for thyroid disease, neurosurgeons for skull base disease, and respiratory physicians for tracheal disease, sleep-disordered breathing, and allergy. Collaboration with allied health professionals includes speech and language therapists for speech and swallowing, electrophysiologists for voice assessment, audiologists for hearing and balance, teachers of the deaf for those with cochlear implants, and psychologists for body dysmorphic disorders.

Key skills in ENT include:

- Developing a diagnostic approach for core symptoms
- Identifying and managing ENT emergencies including life-threatening airway obstruction and large-volume epistaxis
- Understanding symptoms of ENT malignancy, which should prompt urgent referral
- Interpretation of imaging including CT radiographs of the head and neck and endoscopic images of the ear, nose, or larynx.

CLINICAL SKILLS

When a patient presents with a location-specific symptom, it is wise to ask about other symptoms specific to that site. Common symptoms depend on the anatomical location of the lesion (Fig. 21.1). The nose communicates with the ear through the Eustachian tube and with the throat through the nasopharynx, so symptoms can cross from one location to another (e.g., patients with chronic rhinosinusitis often experience pressure in their ears). In addition, dual innervation by multiple branches of the same nerve can lead to referred symptoms, for example, ear pain with throat or dental infection.

The nature, character, and chronological features of symptoms should be systematically explored. Concomitant medical problems, medication use, and occupational and environmental factors can be the cause of ENT disorders and impact treatment, so a detailed history is essential. It is important to be aware of the red flag symptoms related to the ENT (Box 21.1).

Box 21.1 Red flags for head and neck cancer by anatomical site

Red flag symptoms for malignancy by anatomical site

Ear
Otalgia—particularly severe and night pain
Bloody ear discharge
Facial nerve palsy
Visible mass or growth in or around ear
Neck lumps
Risk factors—smoking, sun exposure, immunosuppression

Nose
Unilateral symptoms
Pain—especially severe and persistent
Blood-stained rhinorrhea
Unilateral glue ear
Neck lumps
Numbness/paraesthesia of the cheek
Proptosis or cheek swelling
Visible mass in nose
Epiphora (watery eye)
Risk factors—smoking, alcohol, occupational exposure to metal including nickel, leather and woodworkers

Throat
Otalgia—especially if unilateral and persistent
Neck lump
Stridor and dyspnoea
Hoarseness
Odynophagia
Dysphagia
Non-healing painless ulcer or tonsillar asymmetry
Risk factors—smoking, alcohol, or human papilloma virus exposure

THE EAR

Anatomy

The ear is divided into three parts: outer, middle, and inner (Fig. 21.2).

The ***outer ear*** has a skin-lined tube 2.5 cm long leading to the tympanic membrane (ear drum). The outer third is cartilaginous and contains hair and sebaceous and ceruminous glands; the inner two-thirds are bony. The outer ear is self-cleaning as the skin is migratory, so there are no indications to clean with cotton buds. Wax should only be seen in the outer third.

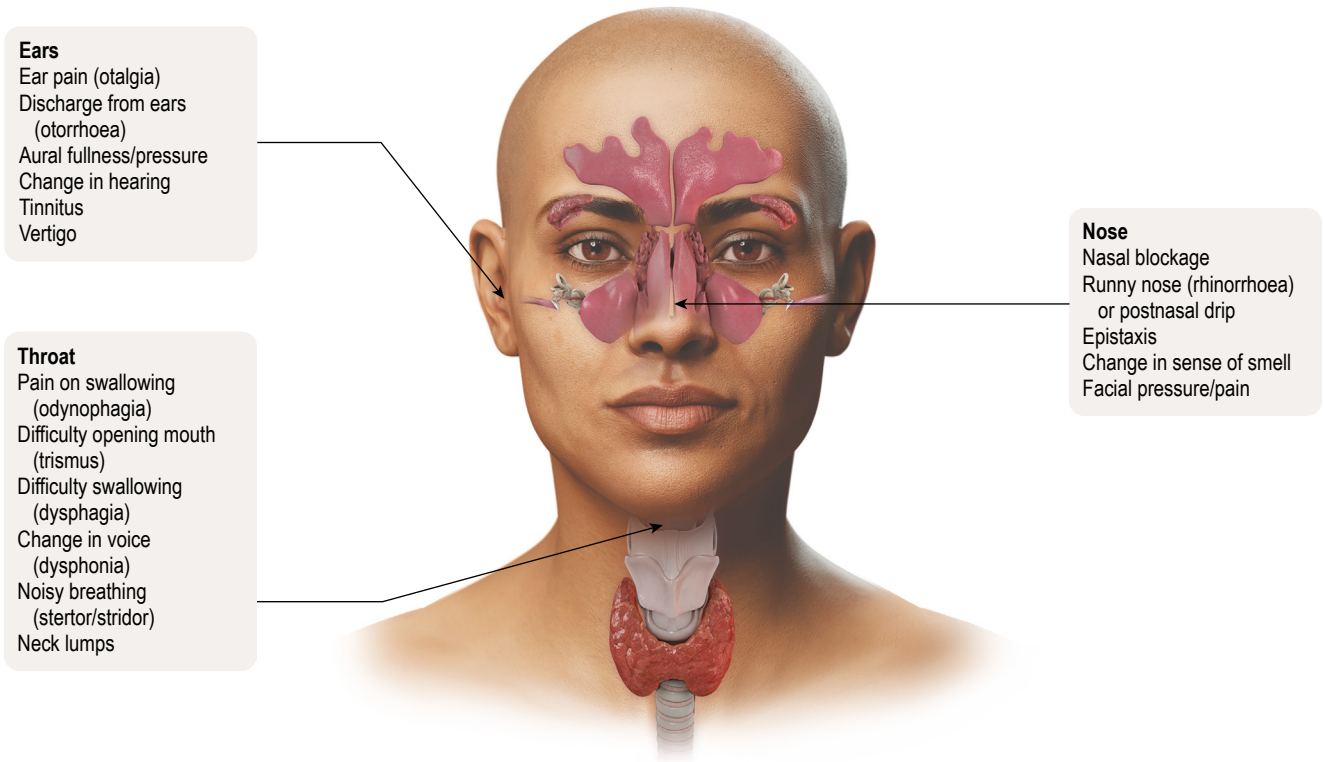


Fig. 21.1 Symptoms of ear, nose, and throat disorders.

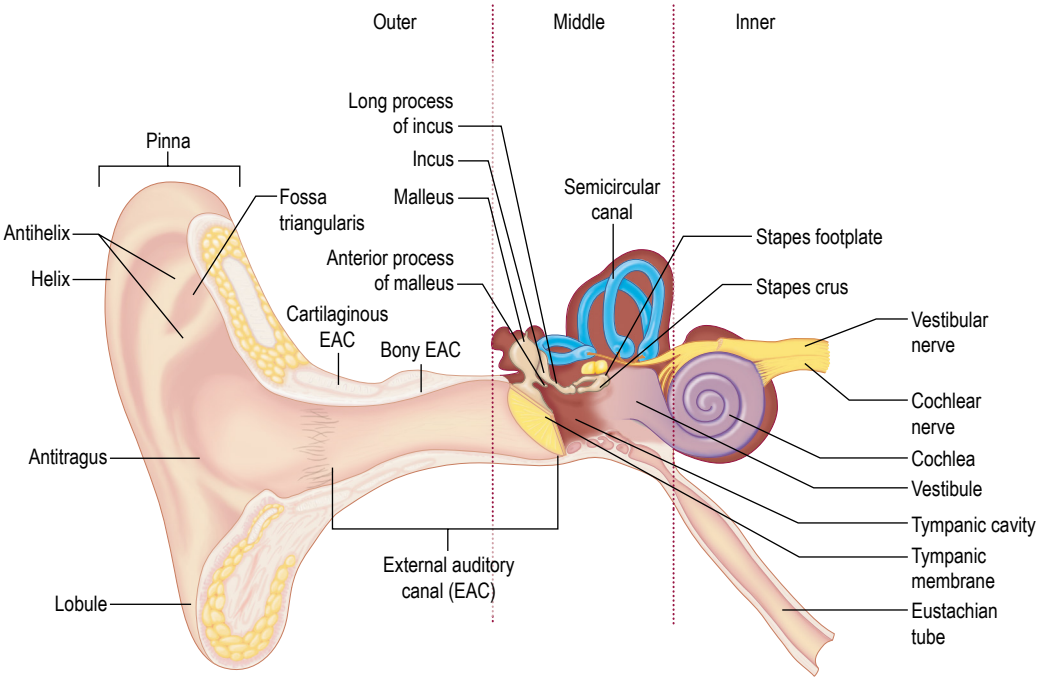


Fig. 21.2 The anatomy of the ear. The ear is anatomically divided into outer, middle, and inner ear, as indicated at the top. The outer ear consists of the pinna and external auditory meatus and funnels sound energy to the middle ear, selectively amplifying sounds in the 2 to 4 kHz range by 15 decibels (dB, speech sounds). When sound energy passes from air in the outer ear to fluid in the inner ear, it loses 30 dB due to energy that is reflected at the interface of the two mediums. The middle ear acts as a transformer, increasing the pressure of the acoustic wave to account for the energy lost at this air–fluid interface. As the acoustic wave moves across the tunnel of the inner ear, passing from the oval window to the round window, it causes maximal vibration of the basilar membrane in a location that is specific to its resonant frequency. This vibration causes the hair cell to open its ion gates and an action potential to be propagated along a fibre of the cochlear nerve. The location of the nerve fibre and the strength of its action potential allow the central nervous system to construct meaning from the sound we hear.

The ***middle ear*** is an air-containing cavity derived from the branchial clefts. It communicates with the mastoid air cells superiorly and connects to the nasopharynx medially via the Eustachian tube. The Eustachian tube ventilates the middle ear and maintains equal air pressure across the tympanic membrane. It is normally closed but opens via the action of the palatal muscles to allow air entry when swallowing or yawning. A defect in this mechanism, for example, cleft palate, will prevent air from entering the middle ear cleft, which may then fill with fluid (otitis media with effusion). Lying within the middle ear cavity are three ossicles, namely malleus, incus, and stapes, which transmit sound from the tympanic membrane (Fig. 21.3) to the inner ear. On the medial wall of the cavity is the horizontal segment of the facial nerve, which can be damaged during surgery or by direct extension of infection in the middle ear.

The ***inner ear*** contains the cochlea for hearing and the vestibule and semicircular canals for balance. There is a semicircular canal arranged in each body plane, and these canals are stimulated by rotatory movement. The facial, cochlear, and vestibular nerves emerge from the inner ear and run through the internal acoustic meatus to the brainstem.

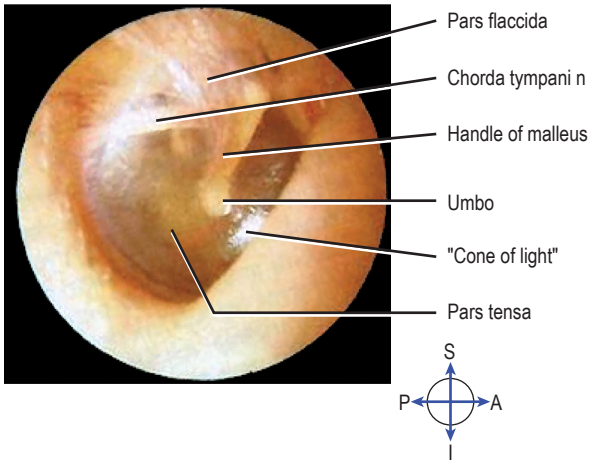


Fig. 21.3 An otoendoscopic view of a normal right tympanic membrane. *(From The Anatomical Basis of Dentistry, Fifth Edition. pii:B9780323824057000060/f06-118-9780323824057.)*

Physiology of hearing

The ossicles, in the middle ear, transmit sound waves from the tympanic membrane to the cochlea. The ossicles amplify sound by approximately 1.5 times using a lever system, while the large surface area of the tympanic membrane compared to the oval window is responsible for amplification by a factor of 20. Together, these amplification mechanisms compensate for the loss of sound waves moving from the air-filled middle ear to the fluid-filled cochlea. Hair cells in the basilar membrane of the cochlea detect vibrations and transduce these into nerve impulses, which pass to the cochlear nucleus and then to the superior olivary nuclei of the pons; thus

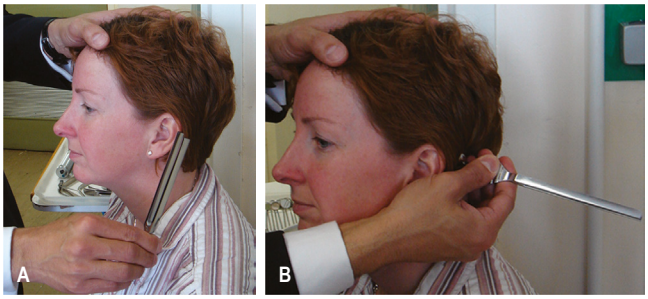


Fig. 21.4 Rinne test. Comparing conduction of sound. **(A)** Air conduction. **(B)** Bone conduction.

lesions medial to the cochlear nucleus do not cause unilateral hearing loss. If the ossicles are diseased, sound can reach the cochlea by vibration of the temporal bone (bone conduction).

CLINICAL APPROACH TO THE PATIENT WITH A DISORDER OF THE EAR

Examination

The pinna and post-auricular region should be examined for erythema, scars, and swelling. An otoscope is used to examine the external ear canal while the pinna is retracted backwards and upwards. Check for wax, discharge, and foreign bodies. The tympanic membrane should always be visualized, with a light reflex seen anteroinferiorly. Previous repeated infections may cause a thickened, whitish drum due to chronic inflammation leading to calcium deposition (tympanosclerosis). Fluid in the middle ear may show as dullness of the drum, a yellow or orange appearance of fluid behind the drum, or air bubbles. Perforations can be described as marginal if they involve the annulus (thickened peripheral rim of the tympanic membrane). Retraction pockets may occur in the pars tensa or pars flaccida of the tympanic membrane (Fig. 21.3), indicating longstanding middle ear disease (chronic otitis media), which may lead to perforations or ossicular erosion with resulting hearing loss, recurrent infections, and cholesteatoma formation.

Investigations

Tuning fork tests and interpretation

Rinne test

1. Rinne test positive (normal): A tuning fork (512 Hz) is heard louder if held next to the ear (air conduction) than if placed on the mastoid bone (bone conduction).
2. Rinne test negative (abnormal): The tuning fork is perceived as louder when placed on the mastoid (bone conduction) due to a defect in the conducting mechanism of the external or middle ear (Fig. 21.4).

Weber test

A tuning fork placed on the forehead or vertex of a patient with normal hearing (or symmetrical hearing loss) is perceived centrally by the patient. A patient with unilateral conductive hearing loss will hear the sound loudest in the affected ear, whereas a patient with unilateral sensorineural hearing loss will report the sound to be loudest in the unaffected ear (Fig. 21.5).

Fig. 21.5 Weber test.

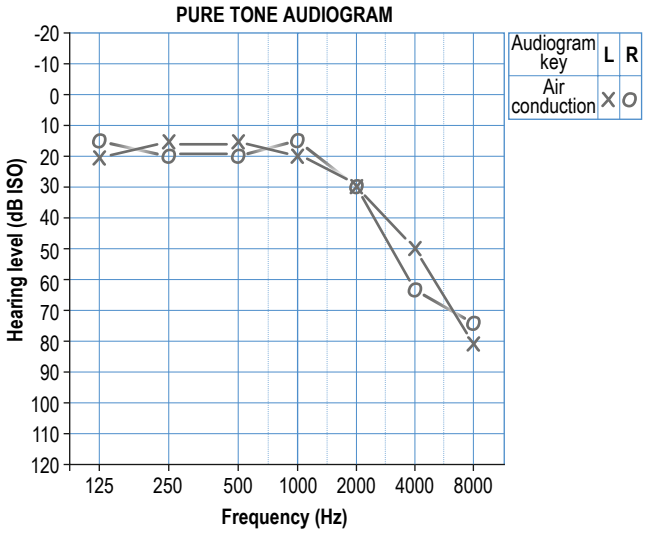


Fig. 21.6 Audiogram showing presbycusis (high-frequency loss).

Pure-tone audiometry

The patient is asked to respond when they hear sounds presented as pure tones at varying intensities and frequencies. Sounds are presented to each ear (air conduction) and each mastoid (bone conduction) in turn. An audiogram is produced by the lowest sound intensity reliably perceptible (Fig. 21.6).

Dix-Hallpike test

The patient sits on a couch with their head turned towards one ear. The head is supported by the examiner while the patient lies down so that their head is just below horizontal. This is then repeated with the patient's head turned towards the other ear. Upbeat tortional nystagmus following a latent interval of a few seconds is noted when the head is turned towards the affected ear (Fig. 21.7).

COMMON DISORDERS OF THE EAR

Common disorders of the ear are outlined in Box 21.2.

The painful ear (otalgia)

Otitis externa

When natural barriers to infection are overcome, the skin of the ear canal can become infected. Discharge and itch are initial presenting

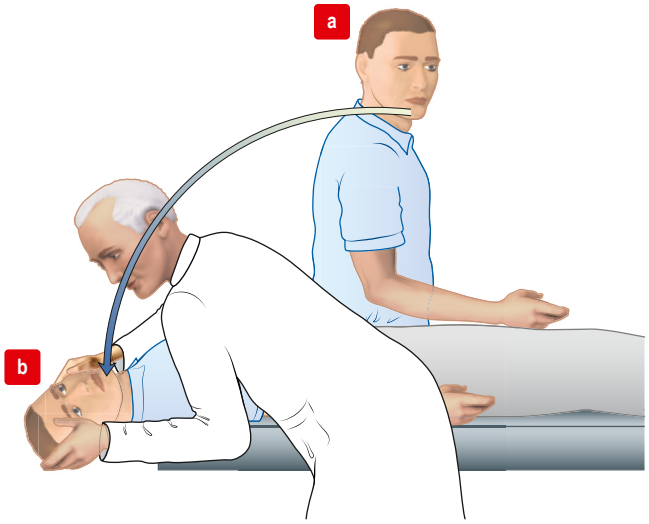


Fig. 21.7 Dix–Hallpike test. The patient sits on a couch with their head turned towards one ear. The head is supported by the examiner while the patient lies down, so their head is just below horizontal. This is then repeated with the patient's head turned towards the other ear. Geotropic (beating towards the ground) rotational nystagmus following a latent interval of a few seconds is noted when the head is turned towards the affected ear.

symptoms, followed by pain and reduction in hearing as the ear canal closes off due to swelling of the ear canal skin. Infection can spread to the pinna, causing cellulitis. Although the causative organism is most commonly bacterial (pseudomonal species > staphylococcal species), it can also be fungal. There may be swelling of the pre-auricular or post-auricular lymph nodes that can be mistaken for mastoiditis.

Examination often reveals debris in the canal, which needs to be removed by gentle mopping or preferably by suction viewed directly under a microscope. The tympanic membrane is normal, when visible. In severe cases, the canal is swollen and view of the tympanic membrane is impossible. Any foreign body should be removed by trained personnel.

Treatment includes ear plugs to prevent water ingress into the ear canal during showering or swimming. Mild infections may respond to topical ear drops containing acetic acid. These alter ear canal pH, which inhibits bacterial and fungal growth. More severe infection requires either topical antibiotic (e.g., ciprofloxacin) or combination of topical antibiotic and steroid (e.g., hydrocortisone acetate 1% and gentamicin 0.3%) if the ear canal is swollen. If symptoms do not resolve within 5 to 7 days, microsuction may be necessary.

Acute otitis media

Acute otitis media is a common infection of the middle ear seeded from the upper respiratory tract through the Eustachian tube. It

Box 21.2 Common diagnoses of the ear

Symptom	Diagnosis	Key clinical features
Otalgia	Acute otitis media	Pain, fever, hearing loss. *Complications: facial nerve palsy, mastoiditis, abscess*
	Otitis externa	Itching, discharge. *Complication: osteomyelitis*
	Temporal bone malignancy	Growth, severe pain, lymphadenopathy. *Complication: facial nerve palsy*
	Referred pain	Normal ear examination
Otorrhea	Acute otitis media	Pain, fever, hearing loss. *Complications: facial nerve palsy, mastoiditis, abscess*
	Chronic otitis media	Hearing loss, often painless. *Complication: osteomyelitis*
	Otitis externa	Itching, discharge. *Complication: osteomyelitis*
Conductive hearing loss	Acute otitis media	Pain, fever, hearing loss. *Complications: facial nerve palsy, mastoiditis, abscess*
	Chronic otitis media	Hearing loss, often painless. *Complication: osteomyelitis*
	Otitis media with effusion ('glue ear')	Usually in children, nasal obstruction. *Complication: speech delay*
	Otosclerosis	Normal tympanic membrane, family history
	Trauma	History of trauma
Sensorineural hearing loss	Sudden sensorineural hearing loss	Sudden onset
	Presbyacusis	Gradual loss in older adults
	Meniere's disease	Episodic vertigo, aural fullness, tinnitus, fluctuating loss
	Medication	Ototoxic medications e.g., cisplatin, loop diuretic, aminoglycoside
	Trauma	History of trauma
Peripheral vertigo	BPPV	Sort lived, positional, fatigable
	Vestibular neuronitis	Onset hours to days, preceding viral infection
	Meniere's disease	Episodic vertigo, aural fullness, tinnitus, fluctuating hearing loss
Central vertigo	Vestibular migraine	Episodic, history of migraine, photophobia/phonophobia
	Posterior circulation stroke	Neurological signs of stroke
	Multiple sclerosis	Presents with vertigo in 7%–10%, other neurological signs
Non-pulsatile tinnitus	Due to any cause of hearing loss	Early sign of hair cell damage which leads to hearing loss
	Eustachian tube dysfunction	Tinnitus related to respiration, autophony, may resolve on lying down
	Temporomandibular disorder	Treatment of TMD improves symptoms
Pulsatile tinnitus	Vascular disorders	Other vascular disease/risk
	Myoclonic spasm	Other neurological or ontological disease
	Middle ear disease	Features of e.g., otitis media

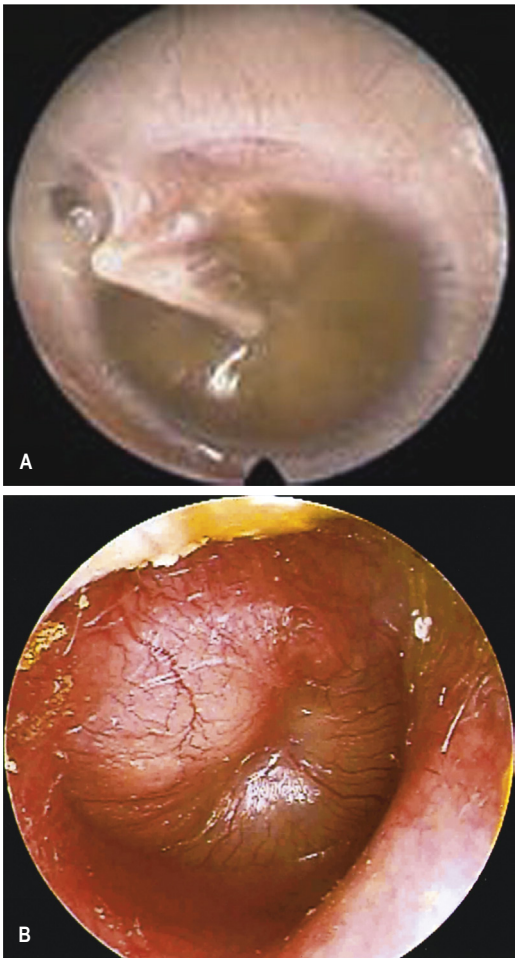


Fig. 21.8 Otitis media with effusion and acute otitis media. **(A)** The tympanic membrane is retracted and the middle ear filled with yellow serous fluid. **(B)** Acute otitis media. Note the red bulging tympanic membrane with pus in the middle ear. *([A] From Woodbury et al., 2011; [B] From Kliegman et al., 2016.)*

may also occur on a background of glue ear (otitis media with effusion) where the stagnant middle ear fluid gets secondarily infected (Fig. 21.8A). Common pathogens are those that also cause upper respiratory tract infections: respiratory syncytial virus, *Streptococcus pneumoniae*, *Haemophilus influenzae*, and *Moraxella catarrhalis*. Acute otitis media most commonly affects children under 10 years. Infection causes inflammation of the middle ear mucosa and inflammatory exudate in the middle ear space. It can lead to tympanic membrane perforation and discharge. Rare complications include mastoiditis as inflammatory fluid escapes from the middle ear into the mastoid, and meningitis due to spread through the tegmen into the intracranial cavity. Examination shows a healthy ear canal but an erythematous and bulging tympanic membrane (Fig. 21.8B). Initial treatment is with non-steroidal anti-inflammatory drugs. If symptoms do not improve within 72 hours or the patient becomes systemically unwell, they should be reassessed for antibiotics. Initial treatment with antibiotics is advised in children less than 3 months old, those less than 6 months old with temperature greater than 39°C, systemic infection, complications such as mastoiditis, or risk of complications (e.g., immunosuppression). Patients presenting with otorrhea and children under 2 with bilateral symptoms are also likely to benefit from antibiotic use. Oral amoxicillin (clarithromycin in penicillin allergy) can be used in uncomplicated cases. If symptoms recur or do not resolve, a beta-lactamase inhibitor (e.g., co-amoxiclav) may be required to cover beta-lactamase-producing bacteria (e.g., *H. influenzae*). When antibiotics are indicated, treatment should be systemic rather than topical, as any perforation may heal quickly limiting access of topical therapy to the middle ear. Topical ear drops containing anaesthetic or analgesics for pain may be used (providing there is no ear drum perforation).

Chronic otitis media is a longer-term inflammatory disorder of the middle ear of two main types:

- Mucosal: tympanic membrane perforation
- Squamous: tympanic membrane retraction or cholesteatoma

Active mucosal disease is characterized by infection. In inactive mucosal disease the perforation is dry and uninfected. Squamous disease is inactive if there is a retraction only and active when there if uncontrolled growth of squamous keratinized epithelium (cholesteatoma).

All patients with a perforation of the tympanic membrane are managed with water precautions, as water entry into the middle ear through a perforated tympanic membrane leads to infection. If otorrhea is present, topical antibiotics are used with or without adjunctive aural toilet (manual cleaning of the ear). Topical antibiotics are likely more effective than oral antibiotics due to limited systemic drug penetration through the relatively avascular mucosa of the middle ear and mastoid. Retraction pockets may be monitored to check whether they are stable or progressing and whether they may require surgery. Surgery for retraction pockets may be performed in order to prevent hearing loss if the retraction is draped onto, and at risk of eroding the ossicular chain or to prevent further retraction, infections, and cholesteatoma formation. Cholesteatoma requires surgical removal to create a dry ear, repair the tympanic membrane, eradicate infection, and prevent complications and recurrence (see below).

Referred otalgia

Pain may be referred from:

- the teeth and temporomandibular joint via the auriculotemporal nerve, a branch of the mandibular nerve;
- the cervical spine from C1 to C3;
- the tonsil and tongue base via the glossopharyngeal (Jacobson's) nerve;
- the larynx and pharynx via the vagus (Arnold's) nerve; and
- Ramsay Hunt syndrome, causing vesicles along the distribution of the facial nerve.

The innervation of the pinna is from the auriculotemporal branch of the trigeminal and first two cervical nerves (Fig. 21.9). Therefore dental pain, temporomandibular joint dysfunction, and upper cervical osteoarthritis can all present as otalgia. The ear canal is innervated by the above-mentioned nerves, in addition to facial and vagus nerves (see Fig. 21.9). Hence Ramsay Hunt syndrome (varicella reactivation along the sensory division of the facial nerve) causes otalgia in association with ear canal vesicles, and laryngeal and pharyngeal cancers occasionally present with otalgia due to referred pain along the vagus nerve. The middle ear is innervated by the glossopharyngeal nerve and therefore infections of the pharynx are associated with otalgia.

The discharging ear (otorrhoea)

Discharge from the ear is usually due to infection of the outer or middle ear. The most common cause is otitis externa, followed by otitis media with a perforated tympanic membrane (see above).

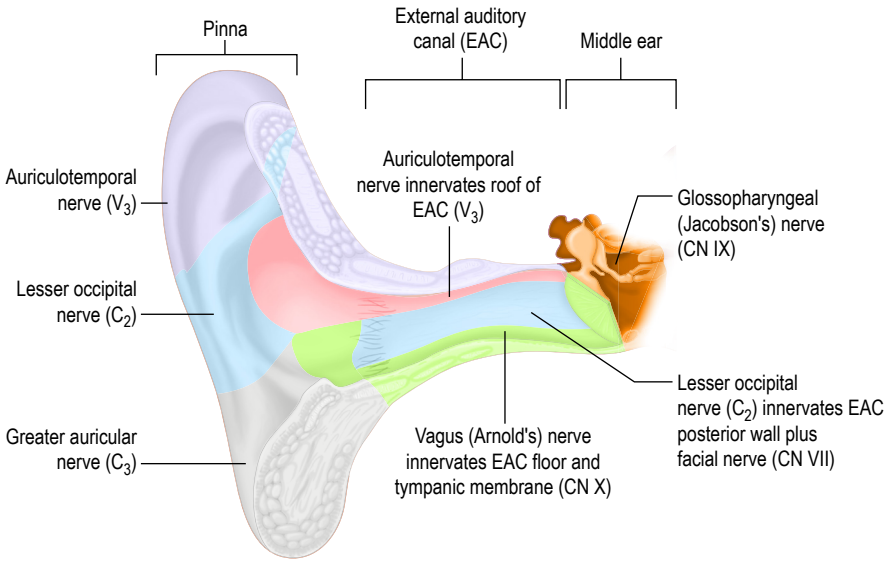


Fig. 21.9 Sensory innervation of the ear. Due to the sophisticated embryology of ear structures, the sensory innervation is complicated. Therefore otalgia can be referred from many sites, and knowledge of aural sensory innervation is a good guide for a comprehensive examination of idiopathic ear pain. The auriculotemporal nerve (a branch of the mandibular division of the trigeminal nerve) supplies innervation to the superolateral pinna and roof of the external auditory canal (EAC). Pathology related to the upper teeth, temporomandibular joint, and parotid gland can cause referred otalgia through this nerve. The lesser occipital nerve (a branch from the second cervical nerve) supplies most of the medial surface of the pinna and the posterior wall of the EAC. Cervical osteoarthritis can cause referred otalgia through this nerve. The greater auricular nerve (a branch from the third cervical nerve) supplies the inferior portion of the pinna, including the lobule. Cervical osteoarthritis can cause referred otalgia through this nerve. Arnold's nerve (a branch of the vagus cranial nerve) supplies the EAC floor and lateral surface of the tympanic membrane. Tumours of the larynx and pharynx can cause referred otalgia through this nerve. Jacobson's nerve (a branch of the glossopharyngeal nerve) supplies innervation to the mucosa of the middle ear, including the medial surface of the tympanic membrane. Tonsillitis and pharyngitis can cause referred otalgia through this nerve. *CN,* Cranial nerve.

Cholesteatoma

Although cholesteatoma is a rare cause of otorrhoea, it has significant implications if missed and should be considered in non-resolving or recurrent otorrhoea. Cholesteatoma is a subtype of chronic otitis media (active squamous disease) defined as keratinizing squamous epithelium within the middle ear cleft. It can present with foul-smelling otorrhoea (which is usually painless) as the keratinizing squamous epithelium acts as a nidus for infection. Examination shows a defect in the tympanic membrane, most commonly in the superior portion of the drum (attic), full of white, cheesy material or a brown attic crust (Fig. 21.10). Mastoid surgery is required to remove this sac of squamous debris, as it can erode local structures including the ossicles or facial nerve, and extend intracranially, leading to meningitis or an intracranial abscess.

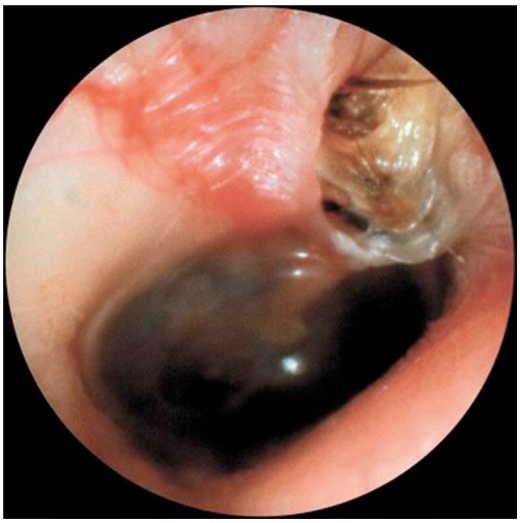


Fig. 21.10 Cholesteatoma of the pars flaccida (attic) of the tympanic membrane. *(From Self-Assessment in Otolaryngology.)*

Hearing loss

Deafness can be conductive or sensorineural, and these can be differentiated at the bedside by the Rinne and the Weber tests (Box 21.3) or with pure-tone audiometry (see investigations above).

Conductive hearing loss has many causes (Box 21.4), but occlusive wax is the most common.

Conductive hearing loss

Otitis media with effusion ('glue ear')

This is common in children when Eustachian tube dysfunction leads to poorly ventilated middle ears. The vacuum created by poor ventilation leads to a non-inflammatory effusion, which resolves in most cases but can persist or recur, causing hearing loss that impacts speech and language development. It can therefore present as hearing loss or speech/educational delay. Otalgia is not a usual feature.

Examination shows a dull tympanic membrane with loss of light reflex (Fig. 21.8A). Occasionally fluid air bubbles are visible in the

Box 21.3 Testing hearing loss

Type	Defect	Rinne test	Weber test
Conductive	Outer or middle ear	Negative. Bone conduction louder than air conduction	Sound heard louder on the affected side
Sensorineural	Inner ear or more centrally	Positive. Air conduction louder than bone conduction	Sound heard louder on unaffected side

Box 21.4 Causes of deafness

Conductive

External meatus

Wax

Foreign body

Otitis externa

Tympanic membrane

Chronis otitis media (mucosal or squamous), e.g., perforation

Middle ear

Otosclerosis

Ossicular discontinuity or fixation

Chronic otitis media (mucosal or squamous), e.g., retraction pocket with ossicular erosion or cholesteatoma

Sensorineural

Congenital

Usher syndrome (deafness and retinitis pigmentosa)

Pendred syndrome (deafness and goitre)

Jervell and Lange-Nielsen (deafness and long QT syndrome)

Vestibular nerve lesions

Vestibular schwannoma

Head injury

Inflammatory, e.g., tuberculous meningitis, neurosyphilis

Brainstem lesions (rare)

Multiple sclerosis

cerebral haemorrhage or infarct (stroke)

Other

Presbyacusis (age related)

Occupational acoustic trauma

Drugs (e.g., aminoglycoside, furosemide)

Labyrinthitis

middle ear. Children frequently have adenoidal hypertrophy and nasal blockage.

If symptoms are persistent and impact hearing, treatment options include insertion of a grommet (tympanostomy tube) through the tympanic membrane or provision of a hearing aid. If otorrhea occurs in the presence of a grommet, combined topical antibiotic-glucocorticoid ear drops are more effective than oral antibiotics. Adenoidectomy may be required if there is complete nasal blockage or recurrent upper respiratory tract infection. Grommets are extruded from the tympanic membrane as it heals (over 6 months to 2 years). In most children, the middle third of the face grows between 7 and 14 years of age and Eustachian tube dysfunction is less common after this, although in some patients it persists to adulthood.

Otosclerosis

This is usually a hereditary disorder, in which new bony deposits occur within the stapes footplate and the cochlea. Characteristically seen in the second and third decades, it is more common in females and can worsen during pregnancy. The hearing loss may be mixed, and management includes a hearing aid or replacement of the fixed stapes with a prosthesis (stapedectomy). The choice of treatment is individualized: surgery has very good success rates in the hands of a regular stapedectomist but confers a small risk of complete hearing loss. Hearing aids, while safe, require adherence for benefit.

Sensorineural hearing loss

Presbycusis

This is the most common cause of hearing loss. It is a degenerative disorder of the cochlea, typically in older age. It can occur due to loss of outer hair cells (sensory), loss of ganglion cells (neural), or strial atrophy (metabolic), or aetiology can be mixed. Environmental noise toxicity is a major factor. Onset is gradual, and higher frequencies are most often affected (**Fig. 21.6**). Speech has two components: low frequencies (vowels) and high frequencies (consonants). When high-frequency consonants are lost, speech loses intelligibility, and increasing low-frequency volume leads to the characteristic response, 'Do not shout. I am not deaf'. Management is with a high-frequency-specific hearing aid. Patients with severe to profound hearing loss, in whom adequate amplification by conventional hearing aids is not possible, may be candidates for cochlear implants and referral to a cochlear implant centre is required.

Noise trauma

Cochlear damage due to noise trauma (e.g., shooting without ear protectors, industrial noise) characteristically leads to high-frequency loss. Similar to other types of sensorineural hearing loss, management is with a conventional hearing aid.

Vestibular schwannoma

This is a slow-growing, benign tumour of the vestibular nerve that can present with progressive sensorineural hearing loss. Imaging (e.g., magnetic resonance imaging [MRI]) is required in any patient with unilateral or sudden-onset sensorineural hearing loss. Treatment is determined according to the size and growth rate of the tumour and includes surveillance with repeat MRI imaging, stereotactic radiosurgery (for growing tumours below 2.5 cm), and surgery. Unilateral hearing loss caused by this pathology may be managed with a contralateral routing of sound hearing aid, or a bone-anchored hearing aid that utilizes conduction to direct sound to the cochlear of the unaffected ear.

Vertigo

See also Chapter 11. There are many causes of dizziness, many of which are not related to the ear and vestibular system. Determining aetiology requires a detailed history. Conditions causing dizziness rather than true rotatory vertigo include:

- Postural hypotension. Ask about pre-syncopal symptoms on standing, antihypertensive medication, and history of cardiovascular disease/diabetes.
- Peripheral neuropathy, for example, secondary to diabetes or chemotherapy.
- Joint pathology, e.g., knee arthritis affecting proprioception which affects proprioception.
- Persistent postural perceptual dizziness. This is a chronic, long-lasting form of dizziness without true rotatory vertigo, worsened

by sitting upright, standing, walking, and visually complex stimuli. It is associated with anxiety and depression.

Clues as to peripheral or vestibular origin include the presence of true rotatory vertigo, or concurrent symptoms of ear disease: otalgia, otorrhoea, tinnitus, or hearing loss. Central causes of vertigo are associated with other neurological symptoms (ataxia, weakness, numbness, hoarseness) or signs (facial palsy or nystagmus [direction changing or vertical types]).

Vertigo can be classified according to the duration of vertigo:

- Seconds (<1 min): benign paroxysmal positional vertigo (BPPV)
- Minutes to hours: Ménière's disease or vestibular migraine
- Hours to days: labyrinthine or central pathology, for example, vestibular neuronitis, labyrinthitis, posterior fossa mass, cerebellar haemorrhage, multiple sclerosis

Vertigo may also be divided into peripheral and central causes.

Peripheral causes of vertigo

Benign paroxysmal positional vertigo

Benign paroxysmal positional vertigo is thought to occur when otoconia (tiny crystals of calcium carbonate) are dislodged from the utricle into the semicircular canals, commonly the posterior canal. Positional vertigo is precipitated by head movements, usually to a particular position, and often occurs when turning in bed or upon sitting up. The onset is typically sudden and distressing. The vertigo lasts seconds (<1 min), and the phenomenon becomes less severe on repeated movements (fatigue). There is no serious underlying cause, but it sometimes follows vestibular neuronitis (see later), head injury, or ear infection. Diagnosis is made based on history and by the Dix–Hallpike manoeuvre (**Fig. 21.7**). A positive Dix–Hallpike test, defined as geotropic nystagmus, which habituates on repeated testing, confirms BPPV. The differential diagnosis of a cerebellar mass can be distinguished by positional nystagmus (and vertigo) that is immediately apparent, with no latent interval, and does not fatigue, and urgent neuroimaging should be used to confirm this. BPPV, on the other hand, is treated in over 90% of cases by the Epley manoeuvre. This involves gentle but specific manipulation and rotation of the patient's head to shift loose otoliths from the semicircular canals into the utricle.

Ménière's disease

This is a rare condition. It is a clinical diagnosis characterized by episodic, rotatory vertigo lasting 30 minutes to a few hours, recurring over months or years. Classically, it is associated with a low-frequency sensorineural hearing loss, a feeling of fullness in the affected ear, loss of balance, tinnitus, and vomiting. It occurs due to a build-up of endolymphatic fluid in the inner ear, the aetiology of which is unknown. MRI imaging (with delayed contrast administration) can confirm endolymphatic hydrops, though this finding is not specific to Ménière's disease.

Treatment is with vestibular sedatives, such as cinnarizine, in the acute phase. Preventative measures include a low-salt diet and avoidance of caffeine, alcohol, and any identified triggers (e.g., nicotine, stress). Betahistine is commonly used despite a paucity of evidence that it is beneficial. The diagnosis of Ménière's should be made by ENT specialists and requires audiometric documentation of a low-frequency hearing loss. If symptoms are poorly controlled, intratympanic steroids or vestibulo-ablative intra-tympanic gentamicin can be considered. Gentamicin is more vestibulotoxic than ototoxic. It destroys the vestibular epithelium, giving the patient severe vertigo for around 2 weeks until able to compensate for the lack of vestibular input on that side. Occasional mild vertigo when the balance system is challenged is often preferable to the unpredictable, severe, and disabling vertigo of Ménière's disease. However, there is a risk of sensorineural hearing loss and complete vestibular failure if Ménière's recurs on the previously unaffected side. Surgical options include endolymphatic sac clipping to relieve endolymphatic hydrops, vestibular nerve section, or labyrinthectomy, although these are rarely performed.

Vestibular neuritis and labyrinthitis

This is due to a virus or postviral inflammation affecting the vestibular portion of the 8th cranial nerve (CN). It causes a rapid onset of vertigo, nausea, and vomiting. Examination shows spontaneous nystagmus that is unidirectional, horizontal, and not direction changing. The fast phase beats away from the side of the lesion. The patient will have a positive head thrust test and be unable to maintain visual fixation, resulting in a catch-up saccade. They may have some gait instability but will still be able to ambulate and will have no neurological symptoms. The Head Impulse, Nystagmus, Test of Skew examination (**Box 21.5**) is a useful way of distinguishing a central or peripheral cause of acute vertigo syndrome and determining who needs urgent neuroimaging. If hearing is simultaneously affected, the condition is termed labyrinthitis.

An acute vertigo syndrome where a central lesion has been excluded either clinically or on imaging is managed with steroids and vestibular sedatives in the acute phase. A patient with a deficit of vestibular function due to viral labyrinthitis or neuronitis should be able to cease vestibular sedatives within 2 weeks; long-term use can give parkinsonian side-effects, delay central compensation, and thus prolong vertigo. Vestibular rehabilitation by a physiotherapist or audiological scientist can speed up compensation, although most patients will be able to compensate themselves with time.

Central causes of vertigo

Posterior circulation stroke

Acute-onset, sustained vertigo may be a presentation of lesions in the cerebellum and brainstem including posterior circulation stroke or posterior fossa mass. Direction changing or vertical nystagmus and other neurological symptoms and signs are usually present. The Head Impulse, Nystagmus, Test of Skew examination can be used to determine whether an acute vertigo episode is central and therefore whether urgent neuroimaging is required (**Box 21.5**).

Box 21.5 The Head Impulse, Nystagmus, and Test of Skew examination[a]

	Peripheral vertigo	Central vertigo
Head Impulse Test	Abnormal, corrective saccade occurs	Normal, no corrective saccade
Nystagmus	None or unidirectional	Direction changing Vertical Torsional
Test of Skew	No vertical skew	Vertical skew present

[a]This test is performed on patients with acute vertigo syndrome to distinguish a peripheral from central vertigo. Acute central vertigo requires urgent neuroimaging. (See also: https://www.youtube.com/watch?v=1q-VTKPweuk).

Vestibular migraine

Vestibular migraine is a subtype of migraine that involves a combination of headache symptoms (severe, throbbing, usually unilateral headache, nausea/vomiting, sensitivity to light/smell/noise) in addition to vestibular symptoms including vertigo (which can last from 5 mins to 72 hours), unsteadiness, and sensitivity to motion. The migraine headaches and vertigo symptoms do not have to occur together. Tinnitus and aural fullness may also be present. Triggers and treatments are the same as for other migraine headaches (see Chapter 28).

Tinnitus

This is a sensation of a sound when there is no auditory stimulus. It can occur without hearing loss and results from heightened awareness of neural activity in the auditory pathways. It is usually benign. It is rarely due to vascular malformations (e.g., aneurysms) or vascular tumours (tinnitus may be pulsatile and commonly unilateral in these cases). It is a difficult symptom to treat. A tinnitus masker (a mechanically produced, continuous, soft sound) can help, or hearing aids can be used if there is associated audiological deficit. Cognitive behavioural therapy through audiological services may be helpful.

Further reading

Kattah JC. Use of HINTS in the acute vestibular syndrome. An Overview. *Stroke Vasc Neurol* 2018; 3.
Lempert T, Olesen J, Furman J et al. Vestibular migraine: Diagnostic criteria1. *J Vestib Res* 2022; 32(1):1–6.
Popkirov S, Staab JP, Stone J. Persistent postural-perceptual dizziness (PPPD): a common, characteristic and treatable cause of chronic dizziness. *Pract Neurol* 2018;18:5–13.
Video of the HINTS exam https://www.youtube.com/watch?v=1q-VTKPweuk.

THE NOSE

Anatomy and physiology

The function of the nose is to facilitate smell and respiration:

- Smell is a sensation conveyed by the olfactory epithelium in the roof of the nose via the olfactory (1st cranial) nerve.
- The nose filters, moistens, and warms inspired air and, in doing so, assists normal respiration.

The external portion of the nose consists of two nasal bones attached to the facial skeleton and upper and lower lateral cartilages. The internal nose is divided by a midline septum that comprises both cartilage and bone, separating the external nostril from the posterior choanae. The posterior choanae are in continuity with the nasopharynx posteriorly. The paranasal sinuses open into the lateral wall of the nose and form a system of aerated chambers within the facial skeleton (Fig. 21.11). The blood supply of the nose is derived from the internal and external carotid arteries. The internal carotid artery supplies the upper nose via the anterior and posterior ethmoidal arteries. The external carotid artery supplies the posterior and inferior portions of the nose via the superior labial, greater palatine, and sphenopalatine arteries. On the anterior nasal septum there is an area of confluence of these vessels (Little's area; Fig. 21.12A), and nasal packing can be used to provide direct compression to these vessels and help to arrest nasal bleeding Fig. 21.12B).

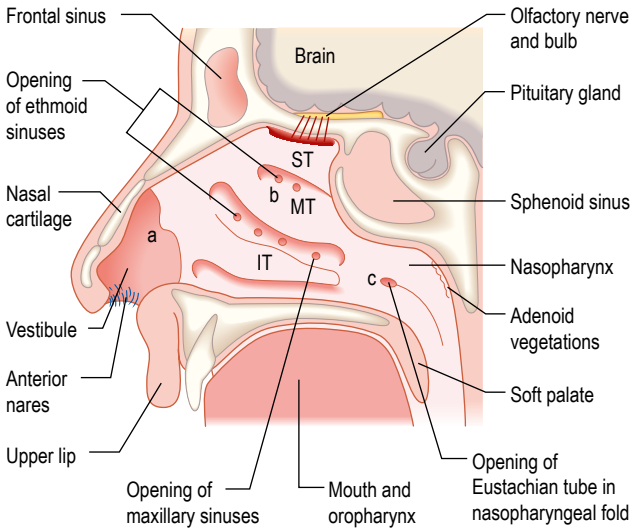


Fig. 21.11 The nose and paranasal sinus ostia. The paranasal sinuses all drain into distinct areas of the nasal cavity.

CLINICAL APPROACH TO THE PATIENT WITH A DISORDER OF THE NOSE

Examination

Inspection of the nose includes an assessment of facial proportions and symmetry. Nasal airflow is assessed by fogging on a cold spatula. A Thudicum speculum is used to examine the anterior nares for evidence of septal deviation and for prominent vessels in Little's area. The posterior nasal cavity and nasopharynx are examined using a rigid or flexible endoscope. Assessment of olfaction can be performed using 'Sniffin' sticks' or the University of Pennsylvania Smell Identification Test.

COMMON DISORDERS OF THE NOSE

Common disorders of the nose are outlined in Box 21.6.

Nasal obstruction

Nasal obstruction can significantly affect quality of life. Causes include:

- ***Allergic rhinitis.*** This is most commonly triggered by allergens, leading to erythema of the nasal mucosa and hypertrophy of inferior turbinates. Symptoms may be seasonal (e.g., grass pollen) or perennial (e.g., house dust mite). If an allergen is identified in the history and confirmed (e.g., via skin prick or radioallergosorbent [RAST] testing), then allergen avoidance is the mainstay of treatment. Oral antihistamines, topical steroids, or a combination nasal preparation may be tried. If rhinitis is severe, referral to an allergy clinic for immunotherapy may be warranted. Short-term benefit may be gained in severe nasal blockage by surgically reducing the inferior turbinate.
- ***Non-allergic rhinitis.*** Vasomotor rhinitis is the most common form, occurring in response to temperature changes, strong odours, or spicy food. Treatment is with an anticholinergic nasal spray (e.g., ipratropium bromide).

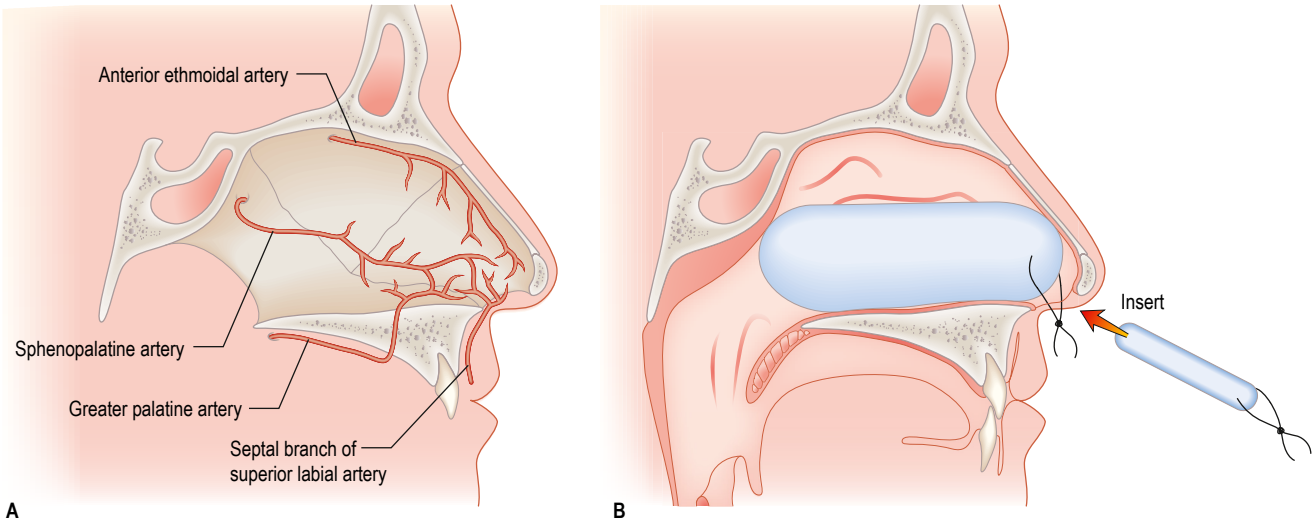


Fig. 21.12 **(A and B) Little's area and nasal packing.** **(A)** Blood supply of Little's area on the septum of the nose. **(B)** Nasal pack.

Box 21.6 Common diagnoses of the nose

Symptom	Diagnosis	Key clinical features
Obstruction	Rhinitis	Allergic symptoms
	Septal deviation	Congenital or traumatic, unilateral blockage
	Rhinosinusitis	Rhinorrhoea, facial pain, altered smell. *Complications: periorbital cellulitis, visual loss*
	Foreign body	Offensive discharge, may be bloody, aspiration risk
	Malignancy	Unilateral symptoms, red flag symptoms (see Box 21.1)
Rhinorrhoea	Rhinitis	Allergic symptoms
	Rhinosinusitis	Rhinorrhoea, facial pain, altered smell. *Complications: periorbital cellulitis, visual loss*
Anosmia	Rhinosinusitis	Rhinorrhoea, facial pain, altered smell. *Complications: periorbital cellulitis, visual loss*
	Neurological disease	Other features of neurological disease, e.g., Parkinson's, Alzheimer's
	Medication/drugs	e.g., beta-blockers, ACE-inhibitors, cocaine, smoking
	Head injury	Occurs following trauma
	Viral infection	e.g., SARS-CoV-2
	Malignancy	Unilateral, mass on endoscopy, associated epistaxis
Epistaxis	Local irritation	e.g., digital trauma, nasal oxygen, nasal corticosteroids
	Nose/face trauma	History of injury
	Foreign body	Offensive discharge, may be bloody, aspiration risk
	Juvenile nasal angiofibroma	Adolescent males, avoid biopsy due to bleeding risk
	Bleeding disorders	Other bleeding, easy bruising
	Medication	e.g., antiplatelets, anticoagulants
	Hereditary haemorrhagic telangiectasia	Telangiectasia around nose and mouth, haemoptysis, melaena

- ***Septal deviation.*** The septum may be deviated to one side of the nose, either congenitally or due to trauma, causing unilateral obstruction. Surgical correction can be undertaken.
- ***Foreign bodies.*** These are usually seen in children presenting with unilateral nasal discharge which may be offensive. Clinical examination of the nose with a light source often reveals the foreign body, which requires removal.
- ***Sinonasal tumours.*** These may be benign (e.g., inverted papilloma, see Fig. 21.13) or malignant (e.g., squamous cell carcinomas, adenocarcinomas, adenoid cystic carcinoma, olfactory neuroblastomas, lymphoma, melanoma, sinonasal undifferentiated carcinoma). Inverted papillomas are associated with squamous cell carcinoma in 5% of cases. Malignancy should be considered in the context of unusual, unilateral symptoms including nasal obstruction, epistaxis, pain, epiphora, paraesthesia of the cheek, unilateral serous otitis media, or proptosis of the orbit (Box 21.1). A unilateral polyp may be seen on examination and confirmed on imaging.
- ***Rhinosinusitis*** (see below).

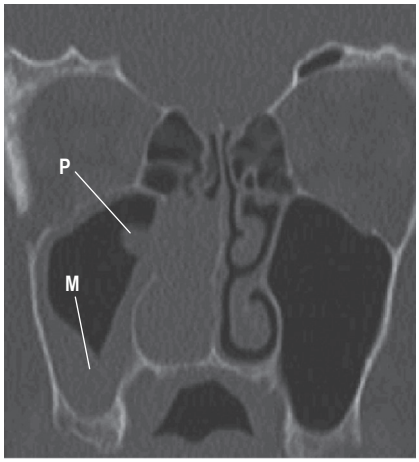


Fig. 21.13 CT of the sinuses. An inverted papilloma (P) can be seen obstructing the right nostril, and there is mucosal thickening (M) of the right maxillary sinus.

Box 21.7 Types of rhinosinusitis

Type	Duration of symptoms
Acute	Less than 12 weeks
Recurrent acute	>4 episodes of acute sinusitis per year with symptom-free intervals
Chronic	>12 weeks

Rhinorrhoea (runny nose)

Rhinosinusitis

Rhinosinusitis is inflammation of the nose and paranasal sinuses characterized by two or more symptoms, one of which should be nasal obstruction/congestion or nasal discharge, with or without change in smell, facial pain, or pressure. In addition, there must be either endoscopic signs of nasal polyps, mucopurulent discharge, or CT changes. Symptoms may be acute (<12 weeks) or chronic (≥12 weeks) (Box 21.7). Rhinosinusitis symptoms can be confused with a variety of other conditions, including migraine, trigeminal neuralgia, and temporal arteritis.

Acute rhinosinusitis is commonly associated with upper respiratory tract infection and can occur in asthma. Symptoms include frontal headache, purulent rhinorrhoea, facial pain with tenderness, and fever. Treatment includes saline spray and short courses of nasal decongestants (e.g., xylometazoline). Topical or short courses of oral corticosteroids are used to reduce mucosal swelling and can lead to symptom improvement. Signs of a superimposed bacterial infection (mainly *S. pneumoniae* and *H. influenzae*) in which antibiotic treatment is indicated include double sickening (getting worse after an initial period of recovery), fever (>38°C), unilateral symptoms, severe pain, and raised inflammatory markers.

Chronic rhinosinusitis (CRS) is classified as primary when disease originates in the nose and sinuses rather than secondary to other systemic disease, for example, cystic fibrosis. Causes of primary CRS can be determined initially by their distribution: localized to a single sinus, for example, maxillary, or diffuse and affecting all paranasal sinuses. Inflammatory subtypes are type 2 and non–type 2. Type 2 inflammation occurs due to immune system activation comparable to that of atopic dermatitis and asthma and is associated with nasal polyp formation. Non–type 2 inflammation is caused by recurrent acute bacterial infection, for example, following upper respiratory tract infection, which leads to tissue remodelling, bacterial colonization, and chronic symptoms (Box 21.8). Treatment is with daily intranasal saline irrigations and intranasal steroids. If symptoms persist or recur, nasoendoscopy can confirm the diagnosis. Endoscopic sinus surgery may be necessary to improve access to the sinuses for delivery of topical steroid therapy. CT imaging of the paranasal sinuses is undertaken prior to sinus surgery to define bony landmarks and anatomy. Patients with severe type 2 inflammatory disease (requiring multiple endoscopic sinus surgery procedures and oral steroids courses) may benefit from immune-modulating treatments including monoclonal antibodies targeting inflammatory cytokines involved in these pathways.

Periorbital cellulitis is an acute complication of acute or chronic rhinosinusitis. An urgent ENT opinion is indicated due to the potential for complications including visual loss, cavernous sinus thrombosis, or intracranial abscess, which can develop rapidly.

Fungal rhinosinusitis can be acute or chronic, invasive or non-invasive. Allergic fungal sinusitis is a chronic, non-invasive fungal sinusitis, associated with atopy, and treated similarly to other forms of chronic sinusitis with nasal steroids. Functional endoscopic sinus surgery is used as an adjunct to open the sinuses ostia, remove fungal disease, and provide access for topical therapies to the sinonasal mucosa. Acute, invasive fungal rhinosinusitis is an aggressive, life-threatening disease that occurs in immunosuppressed patients requiring aggressive surgical debridement and intravenous antifungal treatment. Prognosis is poor.

Further reading

Fokkens WJ et al. Executive summary of EPOS 2020 including integrated care pathways. *Rhinology* 2020; 58(2):82–111.

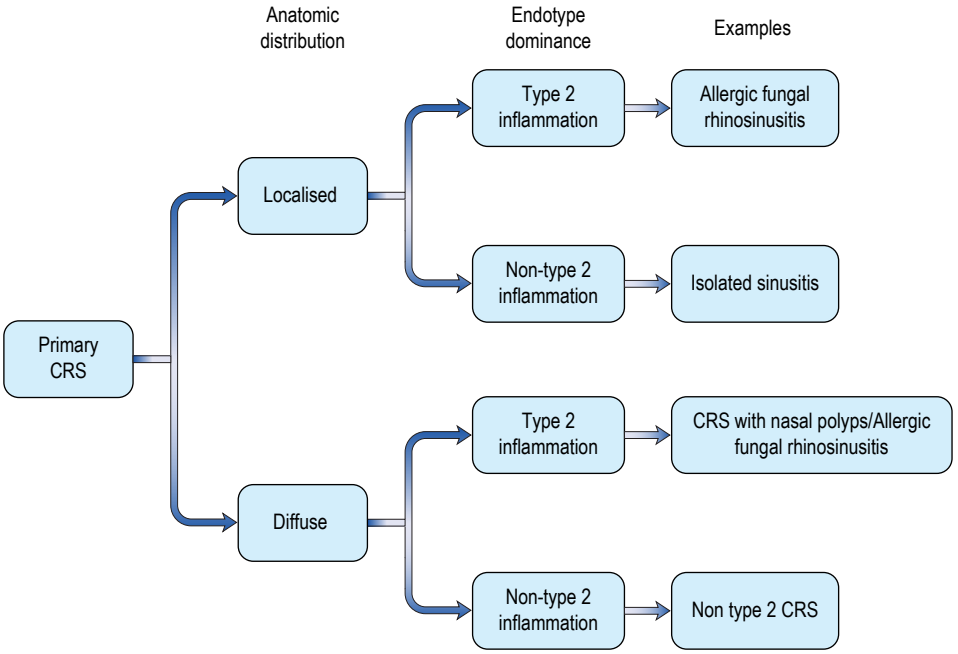
Anosmia

Anosmia is a complete loss of the sense of smell, and hyposmia is a decreased sense of smell. Olfaction is mainly transmitted by the CN I, although irritant and unpleasant nasal sensations are also stimulated by CNs V, IX, and X. A conductive deficit of smell occurs if odorant molecules do not reach the olfactory epithelium high in the nose. Sensorineural loss of smell results from affected neural transmission. Some conditions predispose to a mixed (conductive and sensorineural) loss of smell. The most common cause of a loss of smell is nasal obstruction due to upper respiratory infection or nasal polyps. In many, anosmia is idiopathic, but this is a diagnosis of exclusion, and assessment to exclude an intranasal tumour or intracranial mass should be undertaken, initially with nasendoscopy, followed by MRI in the absence of obvious nasal pathology. Other causes include sinonasal disease, infection, older age, drugs (beta-blockers, calcium channel blockers, angiotensin-converting enzyme inhibitors, tobacco, cocaine), and head injury/trauma. It can be difficult to predict the speed and extent of recovery.

Epistaxis

Nose bleeds vary in severity from minor to life-threatening. Common causes of epistaxis include nose picking, nasal foreign body, mucosal irritation from nasal oxygen cannula or intranasal steroids, and nose/facial trauma.

Box 21.8 Classification of primary chronic rhinosinusitis



(From European position paper on Rhinosinusitis and nasal polyps guidelines, 2020.)

Box 21.9 Aetiology of epistaxis

Local
Idiopathic
Trauma: foreign body, nose-picking, nasal fractures
Iatrogenic: surgery, intranasal steroids
Neoplasm: nasal, paranasal sinus, nasopharyngeal tumours, e.g., juvenile angiofibroma, squamous cell carcinoma

General
Coagulopathy: anticoagulants/antiplatelet agents, coagulation disorders
Severe hypertension
Hereditary haemorrhagic telangiectasia

Little's area is a frequent site of nasal haemorrhage (Fig. 21.12A). Immediate first aid measures include external digital compression of the anterior lower cartilaginous portion of the external nose, ice pack, and leaning forward. The patient should try to avoid swallowing blood running posteriorly as this causes nausea. If profuse bleeding continues, resuscitation in the form of intravenous access and volume/blood product replacement should be administered. Topical adrenaline and intranasal cautery with silver nitrate or bipolar cautery can be used when the bleeding vessel can be identified. If bleeding persists but the vessel cannot be directly identified, a combination of tranexamic acid, absorbable haemostatic agents, or non-absorbable intranasal packing can help to control bleeding. If the above treatments fail, then surgical ligation of the sphenopalatine artery can be undertaken endoscopically, or interventional arterial embolization can be performed. In addition to direct control of epistaxis, the underlying cause should be sought and treated (Box 21.9).

Trauma

People with a fractured nose present with epistaxis, bruising of the eyes, and nasal bridge swelling. It can be difficult to assess whether bones are deviated, particularly if there is significant swelling. The patient should be examined for a head injury and the nose checked for septal haematoma (Fig. 21.14). This is painful, can cause nasal obstruction, is fluctuant to touch on the nasal septum, and requires urgent drainage to prevent destruction of the septal cartilage (which may otherwise result in a saddle nose deformity). Reduction of the fracture should be undertaken in the first 2 to 3 weeks after injury and can be achieved by manipulation. If the fracture sets after this time, rhinoplasty may have to be undertaken at a later stage.

Fig. 21.14 Septal haematoma.

Hereditary haemorrhagic telangiectasia

Hereditary haemorrhagic telangiectasia is a genetic condition where arteriovenous malformations (AVM) occur on the nasal mucosa, as well as in the liver, GI tract, and brain. Epistaxis is a common presenting feature. Treatment includes saline spray, moisturizing creams to avoid drying of the nasal mucosa, and avoidance of mucosal trauma. Potassium titanyl phosphate (KTP) laser treatment is used to treat the mucosal AVM in the nose in order to reduce the frequency of epistaxis.

Juvenile nasal angiofibroma

Juvenile nasal angiofibroma is a rare vascular nasal tumour that presents predominantly in adolescent males and can cause severe

epistaxis. Although benign, they can invade local structures and be challenging to completely excise. The diagnosis is confirmed on CT or MRI imaging. Biopsy should be avoided due to the risk of life-threatening bleeding.

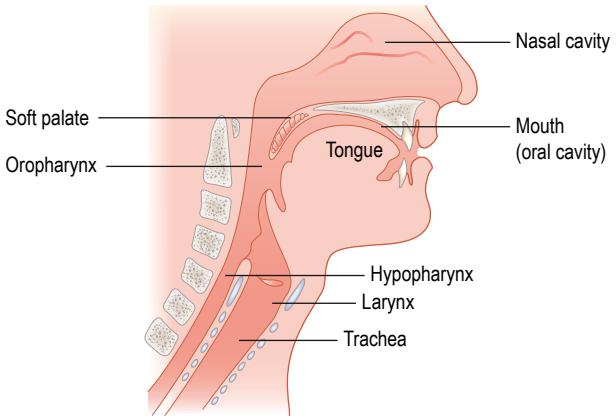


Fig. 21.15 **Normal anatomy of the throat.**

THE THROAT

Anatomy and physiology

The throat extends from the lips to the tonsils and consists of the oral cavity, pharynx, and larynx (Fig. 21.15). The pharynx is innervated from the pharyngeal plexus and is divided into three areas:

- Nasopharynx: posterior nasal openings to soft palate
- Oropharynx: soft palate to tip of epiglottis
- Hypopharynx: tip of the epiglottis to just below the level of cricoid cartilage

Lying within the hypopharynx is the larynx. This consists of cartilaginous, ligamentous, and muscular tissue with the primary function of protecting the distal airway. Two vocal cords abduct (open) during inspiration and adduct (close) to protect the airway and produce the voice (phonation). The main nerve supply of the vocal cords comes from recurrent laryngeal nerves which arise in the neck as branches of the vagus nerve. **The course of the recurrent laryngeal nerve is longer on the left than on the right** (Fig. 21.16). In phonation, normal vocal cords vibrate between 90 (male) and 180 (female) times per second, giving the voice its

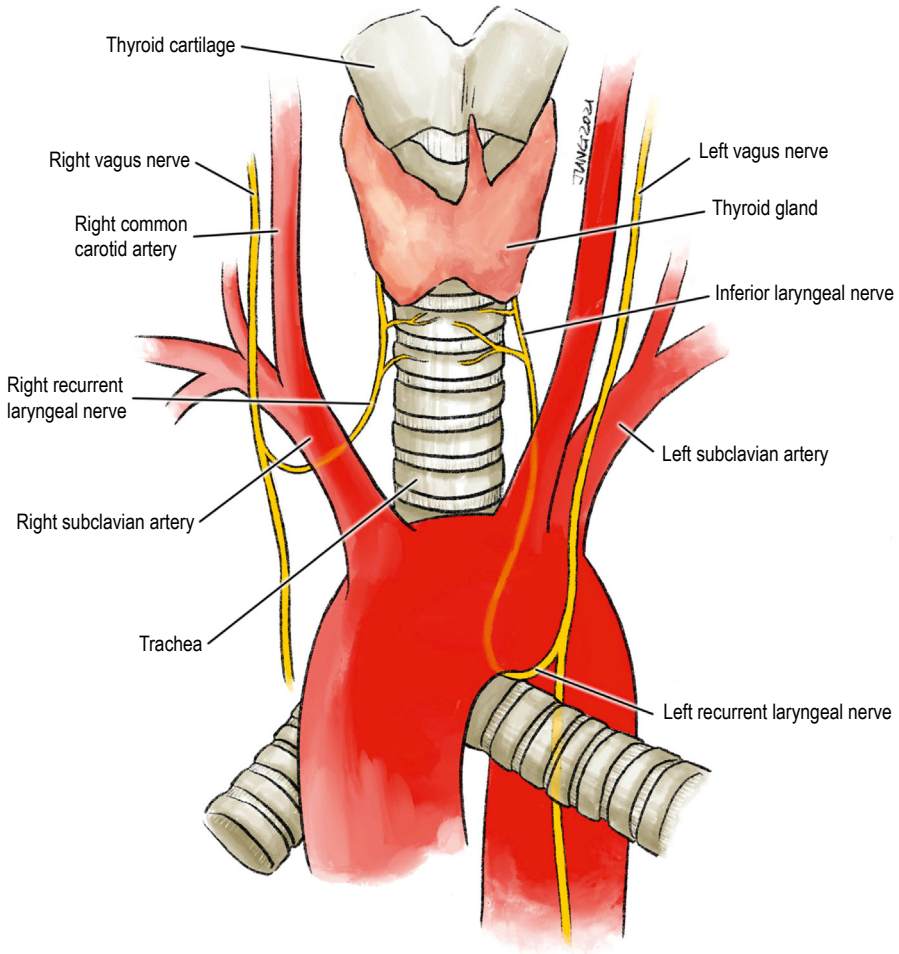


Fig. 21.16 **Anatomy of the recurrent laryngeal nerves differs on the right and left sides of the neck.** The course of the left recurrent laryngeal nerve is longer, passing down around the aortic arch before ascending in the tracheo-oesophageal groove to the larynx, meaning left vocal cord paralysis is more common than right. A patient with a paralysed vocal cord must have a CT of the neck and chest (from skull base to diaphragm) to identify pathology along the course of the vagus or recurrent laryngeal nerves.

pitch or frequency. A healthy voice requires full closure of the vocal cords with a smooth, regular pattern of vibration. Any pathology that prevents full closure results in air escaping during phonation and a 'breathy' voice.

CLINICAL APPROACH TO THE PATIENT WITH A DISORDER OF THE THROAT

Examination

Good illumination is essential. Inspect teeth, gums, tongue, floor of mouth, and oral cavity. The tonsils, soft palate, and uvula are easily visualized. A gag reflex should be present. The remainder of the pharynx and larynx can be inspected with a flexible laryngopharyngoscope. The neck should be examined for lymph nodes and other masses.

COMMON DISORDERS OF THE THROAT

Common disorders of the throat are outlined in **Box 21.10**.

Odonophagia (pain on swallowing)

Tonsillitis/pharyngitis

Most infections are of viral origin and self-limiting, settling with time and analgesia. If pyrexia, dysphagia, or lymphadenopathy is present, then bacterial infection is more likely, with β-haemolytic streptococcus being the most common organism. Infectious mononucleosis caused by Epstein–Barr virus (see Chapter 42) can present as tonsillitis with a confluent white tonsillar exudate and associated petechial rash on the soft palate and lymphadenopathy. The fever pain score is a tool used in primary care to predict the likelihood of bacterial throat infection and guide antibiotic use (**Box 21.11**). Oropharyngeal candidiasis is uncommon and warrants exclusion of immunosuppression including diabetes mellitus and HIV.

Indications for a tonsillectomy are shown in **Box 21.12**. Current surgical techniques include cold steel, diathermy dissection, and coblation (which uses an ultrasonic dissecting probe under general anaesthesia). There are strong advocates for each technique, and much depends on the individual surgeon's preference. Most tonsillectomy is performed as a day-case procedure, as most reactionary bleeding occurs in the first 8 postoperative hours.

Box 21.10 Common diagnoses of the throat

Symptom	Diagnosis	Key clinical features
Odynophagia	Tonsilitis/pharyngitis	Fever, pain on swallowing, pus on tonsils, lymphadenopathy
	Quinsy	Fever, pain on swallowing, trismus, swelling above tonsil, uvula deviation
	Parapharyngeal abscess	Fever, pain on swallowing, pain with neck movement, breathlessness. *Complications: airway compromise, mediastinitis*
	Malignancy	Pain (persistent, unilateral), mass, lymphadenopathy
Dysphagia	Foreign body	Sudden onset, pain, drooling, airway compromise
	Pharyngeal pouch	Dysphagia to solids, regurgitation, halitosis, compressible neck mass, aspiration pneumonia
	Globus pharyngeus	Worse with saliva, better with eating and drinking, normal endoscopy
	Malignancy	Progressive dysphagia, pain (unilateral), lymphadenopathy, smoking, airway compromise
Hoarseness	Vocal cord nodules	Professional voice user (e.g., teacher, singer), smoking, reflux
	Vocal cord polyps	Can follow episode of haemorrhage e.g., screaming, coughing
	Recurrent respiratory papillomatosis	Usually in children after HPV acquired at the time of delivery, airway compromise
	Reinke's oedema	Husky voice, smoking
	Vocal cord palsy	Breathy voice, aspiration pneumonia
	Laryngeal cancer	Stridor in advanced disease
Stridor	Malignancy	Hoarseness, dysphagia, odynophagia, smoking, alcohol
	Recurrent respiratory papillomatosis	Usually in children after HPV acquired at the time of delivery, airway compromise
	Supraglottitis	Odynophagia, fever, stridor. Normal oropharynx despite severe symptoms
	Parapharyngeal abscess	Fever, pain on swallowing, pain with neck movement, breathlessness. *Complications: airway compromise, mediastinitis*
	Foreign body	Sudden onset, usually in a child
	Anaphylaxis	Tongue/upper airway swelling over minutes to hours
	Angioedema	Swelling of lips, tongue, posterior pharynx, and larynx
	Trauma	Following traumatic injury
	Smoke inhalation	History of inhalation, soot around mouth. Consider early intubation

Box 21.11 Fever pain score for tonsillitis

Clinical feature		Score 1 if present
Fever during last 24 hours		
Purulence (pus) on tonsils		
Presentation within 3 days of symptoms		
Tonsils severely inflamed		
No cough or coryza		
Fever pain score		**Total (max 5)**
Fever pain score	**Likelihood of streptococcus**	**Antibiotics**
0–1	13%–18% (similar to asymptomatic carrier rate)	Not recommended
2–3	34%–40%	Delayed 24–48 hours
4–5	62%–65%	Recommended

Box 21.12 Indications for tonsillectomy

Suspected malignancy (e.g., asymmetrical tonsils, or for diagnostic workup in the case of an unknown primary)
Obstructive sleep apnoea due to tonsillar hypertrophy
Recurrent tonsillitis: three episodes per year for more than 3 years, five episodes per year for 2 years, or seven episodes in 1 year (Paradise criteria)
Quinsy
Specific conditions where tonsillitis may contribute to disease activity, e.g., periodic fever, aphthous stomatitis, pharyngitis, and cervical adenitis syndrome.

Quinsy (peritonsillar abscess)

Quinsy is a collection of pus outside the capsule of the tonsil, usually located adjacent to its superior pole. Trismus is a key feature and can make examination difficult. The area around the tonsil is hyperaemic and can push the uvula across the midline. Awake transoral needle aspiration of the abscess by an ENT surgeon is required. If untreated, infection can spread through the superior constrictor muscles of the pharynx, forming a parapharyngeal abscess. From there, pus may spread rapidly via deep fascial planes of the neck and chest, with the potential airway compromise, mediastinitis, and sepsis to occur.

Oropharyngeal squamous cell carcinoma

Unilateral throat pain with an ulcerated or asymmetrically enlarged tonsil is concerning for malignancy and urgent biopsy is required. Human papilloma virus (HPV) infection accounts for 50% to 70% of oropharyngeal malignancy. HPV-related tumours have distinct tumour characteristics and slightly more favourable outcomes compared to HPV-negative oropharyngeal squamous cell carcinomas (which are associated with the traditional risk factors of smoking and alcohol excess).

Dysphagia

See also Chapter 22. Dysphagia can occur because of disease anywhere between the throat and stomach. The pharyngeal pouch and foreign body in the pharynx are managed by ENT specialists. Aetiologies arising further down the oesophagus are managed by gastroenterology.

Pharyngeal pouch

A pharyngeal pouch is a herniation of mucosa through the fibres of the inferior pharyngeal constrictor muscle (cricopharyngeus)

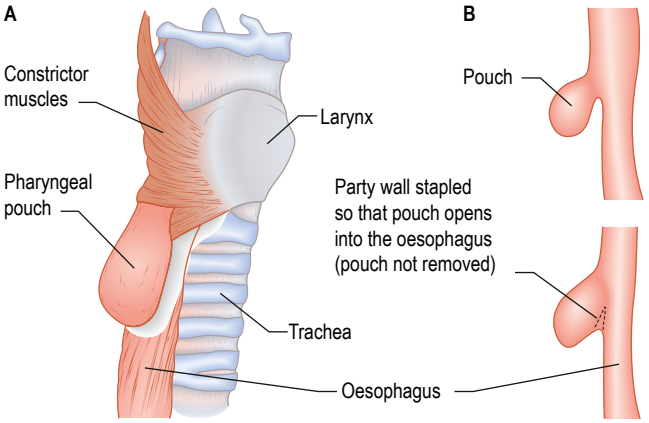


Fig. 21.17 (A and B) Pharyngeal pouch. **(A)** Position of pharyngeal pouch. **(B)** Stapling of the party wall between oesophagus and pouch.

(Fig. 21.17A). An area of weakness known as Killian's dehiscence allows a pulsion diverticulum to form. Patients present with neck swelling following a failed swallow attempt. They can occasionally compress the swelling, which pushes food particles back into the oesophagus. A gurgling sound may be heard in the neck following a swallow as liquid and food collect in the pouch. Occasionally, patients present with recurrent pneumonia following aspiration of food into the trachea. Diagnosis is made on barium swallow imaging (Fig. 21.18), and treatment is surgical, either via an external approach through the neck or, more commonly, via endoscopy with stapling of the party wall (see Fig. 21.17B).

Foreign body

Foreign bodies in the pharynx can be classified as one of four main types: soft food bolus, coins, button batteries (smooth with a halo/double ring/narrow rim on plain radiograph imaging, Fig. 21.19A and B), and bones. Soft food boluses can be managed conservatively with muscle relaxants for 24 hours. Impacted coins should be removed at the earliest opportunity. Button batteries and sharp objects require emergency removal to avoid perforation of the muscle wall. If the patient perceives the foreign body to be on one side and located above the manubrium, then the foreign body is usually found above the cricopharyngeus and can be located with flexible laryngopharyngoscopic examination; common areas are the tonsillar fossae, base of tongue, posterior pharyngeal wall, and valleculae. Lateral and anterior-posterior radiographs can identify and locate a foreign body not visible on examination, as these frequently lodge at the level of the cricopharyngeus. A coin or sharp foreign body that has passed this junction will often pass into the stomach and no further treatment may be required as it will exit naturally. While not all fish bones are radio-opaque, soft tissue changes resulting from the impacted foreign body may be seen. CT imaging may provide additional information or reassurance where the history is equivocal. Direct laryngopharyngooesophagoscopy is the best diagnostic tool.

Globus pharyngeus

This is an intermittent sensation of a lump in the throat, without true dysphagia. The sensation is perceived in the midline at the level of the cricoid cartilage, is worse when swallowing saliva, and is better or disappears when ingesting food/liquid. Examination including flexible laryngopharyngoscopy is normal. Patients with features or risk factors for head and neck cancers should undergo panendoscopy. Treatment is with explanation

and reassurance. Gastric acid suppression therapy can be tried if symptoms persist.

Malignancy of upper aerodigestive tract

Dysphagia may be caused by a malignancy of the upper aerodigestive tract, usually in the presence of red flag features, including odonophagia, hoarseness, dyspnoea, neck lymphadenopathy, and weight loss. Risk factors include tobacco and alcohol excess. Visualization of the upper aerodigestive tract with a flexible laryngopharyngoscope may reveal an irregular mass or an area of salivary pooling. MRI imaging of the head and neck and cross-sectional chest imaging are required prior to biopsy. Ultrasound localization and fine needle aspiration of neck lymphadenopathy may be performed in a one-stop diagnostic clinic and provide a same-day diagnosis. Investigation and management are undertaken with multidisciplinary discussion and consensus.

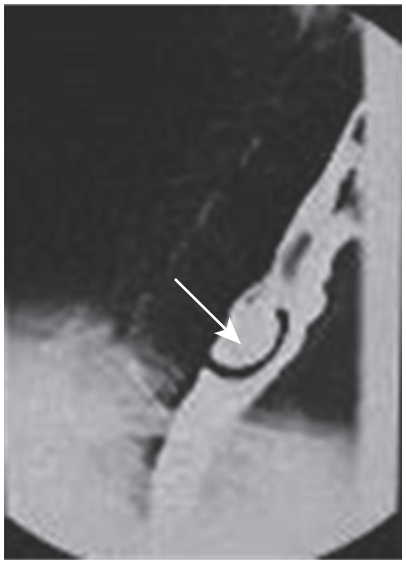


Fig. 21.18 **Pharyngeal pouch shown on barium swallow.**

Hoarseness (dysphonia)

There are three essential components for voice production: an air source (lungs), a vibratory source (vocal cords), and a resonating chamber (pharynx, nasal and oral cavities). Although chest and nasal disorders can affect voice, the majority of hoarseness is due to vocal cord pathology.

Nodules/polyps

Nodules (usually bilateral, more common in females) and polyps (usually unilateral, more common in men) (Fig. 21.20A) occur secondary to chronic vocal fold irritation: smoking, laryngopharyngeal reflux, and muscle tension dysphonia. The presence of lesions on the free edge of the vocal cord prevents full closure and results in hoarseness. They are commonly found in professionals who rely on their voice for their livelihood such as teachers, singers, and lawyers. They are usually related to the technique of voice production and can be cured with speech therapy. If surgery is needed, great care must be taken to remain in the superficial layers of the vocal cord to prevent deep scarring, which leaves the voice permanently hoarse.

Reinke's oedema

This is a collection of tissue fluid in the subepithelial layer of the vocal cord (Fig. 21.20B). The vocal cord has poor lymphatic drainage, predisposing it to oedema. Reinke's oedema is associated with irritation of the vocal cords by smoking, voice abuse, laryngopharyngeal reflux, and, rarely, hypothyroidism. Treatment is to remove the irritant cause in most cases, but surgery to incise the cords and reduce the oedema may also allow the voice to return to normal pitch.

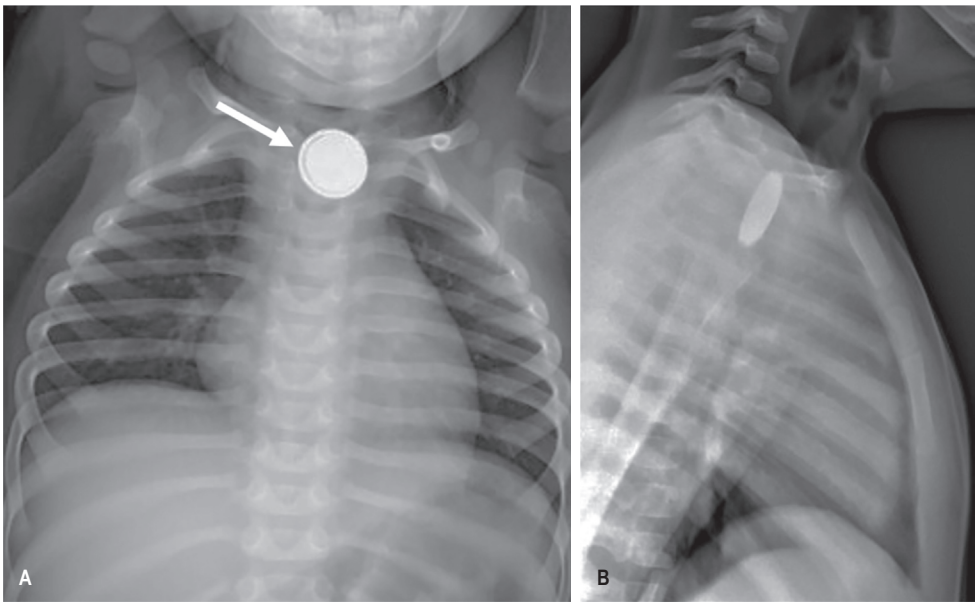


Fig. 21.19 **Button battery.** **(A)** Frontal chest radiograph in a 1-year-old shows a rounded metallic density foreign body projected over the upper oesophagus. A sub-peripheral lucency producing a 'double-ring' appearance *(arrow)* is seen to be suggestive of a button battery. **(B)** Lateral radiograph in this patient shows the battery in profile as would be expected for a foreign body in the oesophagus.

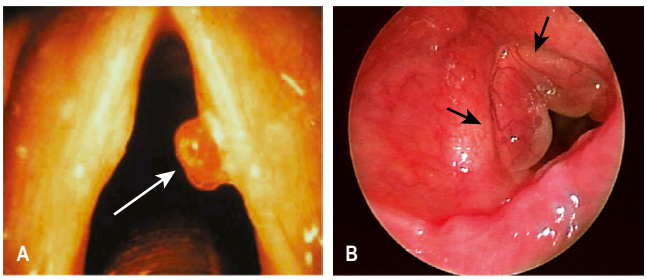


Fig. 21.20 (A and B) Disorders of the vocal cords. (A) Vocal cord polyp. (B) Reinke's oedema of the vocal cords.

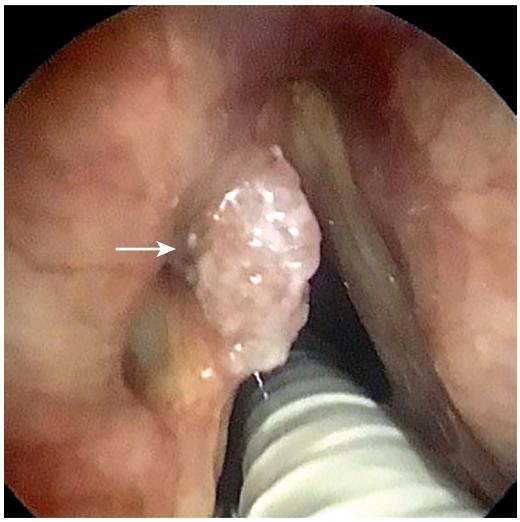


Fig. 21.21 Laryngeal carcinoma of the right vocal cord.

Laryngeal or bronchial carcinoma

Hoarseness in a person who smokes should be considered a sign of malignancy until proven otherwise. Any patient with a hoarse voice for over 6 weeks should be referred urgently for examination of the vocal cords using flexible nasopharyngolaryngoscopy. The voice may be deep, harsh, and breathy, indicating a mass on the vocal cord (Fig. 21.21), or it can be weak, suggesting a paralysed left vocal cord secondary to mediastinal disease, such as bronchial carcinoma. A patient with a paralysed vocal cord must have a CT of the neck and chest to identify pathology along the course of the vagus or recurrent laryngeal nerve (Fig. 21.16). If there is airway compromise, immediate surgical debulking of the laryngeal mass or tracheostomy to prevent life-threatening airway obstruction may be indicated.

Early squamous cell carcinoma of the larynx has a good prognosis due to limited lymphatic drainage of the vocal cords, with 5-year disease-free survival for stages I and II disease of 90% and 80%, respectively. Treatment is with a single modality of treatment, either transoral laser resection or radiotherapy. Larger tumours that spread to surrounding lymph nodes may require dual modality treatment (e.g., surgery with postoperative radiotherapy or chemoradiotherapy). Non–organ-preserving surgery (laryngectomy and neck dissection of affected lymph nodes) is reserved for large tumours that have invaded through the laryngeal cartilage, where safe swallow is already impaired, or where chemoradiotherapy has failed to cure disease. Patients with a paralysed vocal cord (e.g., from a bronchial carcinoma) can have vocal cord medialization surgery to improve their voice and help with laryngeal competence. This can be done under local or general anaesthesia.

Box 21.13 Indications for tracheostomy

Emergency

Upper airway obstruction in patients who have failed an endotracheal tube or in patients where an endotracheal tube cannot be placed, e.g., large obstructing tumour

In patients with select fractures or trauma to the oropharynx, face, or neck including penetrating laryngeal trauma

Elective

Patients who require or are likely to require long-term mechanical ventilation and have difficulty weaning off the ventilator

Patients with poor airway protection who need ventilation or secretion control (e.g., neuromuscular disorders, recurrent aspiration)

Severe obstructive sleep apnoea, vocal cord paralysis, and subglottic stenosis who have failed other therapies

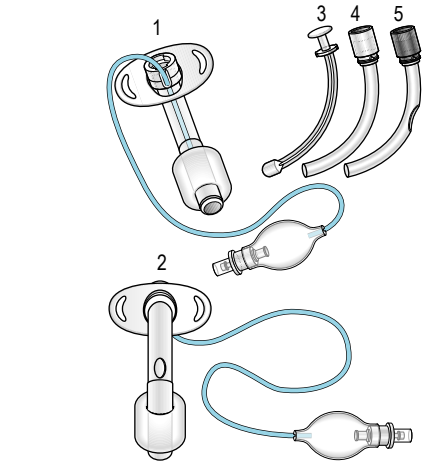


Fig. 21.22 Different types of adult tracheostomy tubes.

Stridor

Stridor, or noisy breathing, can be divided into the following types:

- ***Inspiratory:*** obstruction is at the level of the vocal cords or above.
- ***Mixed*** (inspiratory and expiratory): obstruction is in the subglottis or extrathoracic trachea.
- ***Expiratory:*** obstruction is in the intrathoracic trachea or distal airways.

All patients with stridor are at risk of asphyxiation and should be investigated. Volume of stridor is not a marker of severity if it is causing less laryngeal airflow. Stridor with significant respiratory distress may be an indication for intubation or a tracheostomy (Box 21.13). Types of tracheostomy tubes (Fig. 21.22) include:

- ***Cuffed or uncuffed.*** A high-volume, low-pressure cuff is used to prevent aspiration and to allow positive-pressure ventilation.
- ***Fenestrated or unfenestrated.*** A fenestrated cuff has a small hole on the greater curvature of the tube (both outer and inner), allowing air to escape upwards to the vocal cords; the patient can therefore speak. This tube is often used with a valve that allows air to enter from the stoma but closes on expiration, directing air through the fenestration.

Most adult long-term tracheostomy tubes have an inner and an outer tube, while paediatric tubes are usually single lumen to maximize internal diameter for ventilation. In adult tracheotomy

tubes the inner tube fits inside the outer tube and projects beyond its lower end. A major problem with a tracheostomy tube is the crusting of its distal end with dried secretions, and this arrangement allows the inner tube to be removed, cleaned, and replaced as frequently as required without disrupting the outer tube. In children with a single lumen tube frequent suctioning and humification are required to prevent tube obstruction with dried secretions. Decannulation (removal of a tracheostomy tube) requires laryngeal competence; assessment by ENT and speech therapists is needed. A tracheostomy tube can impair swallowing due to compression of the oesophagus (particularly cuffed tubes) and prevention of the larynx from rising during normal swallowing.

Stertor (snoring)

Snoring is caused by high-pressure airflow, resulting in vibration of soft tissue above the level of the larynx. It is a common symptom (50% of 50-year-old males). Obstruction to airflow can occur in three sites: nose, palate, and/or hypopharynx. There is a strong association between snoring and sleep-disordered breathing, including obstructive sleep apnoea. The Epworth Sleepiness Scale is a subjective measure of a patient's sleepiness and can prompt further investigation (see **Box 20.21**). Most patients will benefit from lifestyle changes, including weight loss and avoiding alcohol. Continuous positive airway pressure is used for obstructive sleep apnoea, which affects quality of life or daytime activities (see Chapter 20). Patients unable to tolerate standard continuous positive airway pressure can try changing from a nasal to orofacial mask, heated humification, or the use of a mandibular advancement splint if good dentition. Sleep nasendoscopy assesses a sedated, snoring patient using a flexible nasendoscope to identify the location of reduced airway tone and collapse. Nasal pathology such as polyps or enlarged tonsils can be removed surgically with good results.

Further reading

NICE 2021. *Obstructive sleep apnoea/hypopnoea syndrome and obesity hypoventilation syndrome in over 16*. https://www.nice.org.uk/guidance/ng202.

Questions for this chapter can be found on Elsevier eBooks+.

22 Gastroenterology

Noor Jawad and Charlotte I. Skinner

Clinical skills 602
History and examination 602
Visualizing the bowel 602
Background 604
Anatomy and physiology of the gastrointestinal tract 604
Mouth, oropharynx, and oesophagus 604
Stomach and duodenum 604
Small intestine anatomy 605
Colon and rectal anatomy 611
Clinical features of gastrointestinal disease 612
Dysphagia 612
Dyspepsia 612
Abdominal pain 612
Nausea and vomiting 613
Diarrhoea 613
Constipation 615
Investigation of gastrointestinal disease 617
Functional gastrointestinal disorders 619
Mouth 620
Neoplasia (squamous cell carcinoma) 621
Non-neoplastic lesions 621
Tongue 621
Gums 621
Teeth 621
Salivary glands 621
Pharynx and oesophagus 621
Gastro-oesophageal reflux disease 621
Complications of gastro-oesophageal reflux disease 625
Motility disorders 625
Other oesophageal disorders 628
Oesophageal tumours 630
Functional oesophageal disorders 632
Stomach and duodenum 632
Gastritis and gastropathy 632
Helicobacter pylori infection 632
Gastric tumours 637
Acute and chronic gastrointestinal bleeding 640
Small intestine 642
Irritable bowel syndrome 642
Malabsorption 644
Miscellaneous intestinal diseases 650
Tumours of the small intestine 651
Inflammatory bowel disease 653
Colon and rectum 665
Miscellaneous colonic conditions 665
Diverticular disease 666
Anorectal disorders 667
Colonic tumours 669
The acute abdomen 674
Peritoneum 677
Disorders affecting the peritoneum 677

CORE SKILLS AND KNOWLEDGE

Gastrointestinal disease is common throughout the world. The WHO estimates that diarrhoeal illnesses alone caused 1.4 million deaths in 2016. Gastrointestinal disease accounts for approximately 10% of primary care appointments, and colorectal cancer is the third commonest cancer globally.

Gastroenterologists care for patients in inpatient settings (e.g., for those with acute flares of inflammatory bowel disease or gastrointestinal bleeding), outpatient clinics (for long-term management of chronic gastrointestinal conditions, or diagnosis of newly presenting patients), and they provide diagnostic, therapeutic, and screening endoscopy services. Gastroenterology research is a fertile area with many exciting new therapies and technologies appearing in recent years, and gastroenterologists may also be active in medical education or hospital management.

Key skills in gastroenterology include:

- Managing and investigating acute and chronic gastrointestinal bleeding
- Understanding different treatment strategies for inflammatory bowel disease, including the role for biological agents and surgical treatment
- Developing an understanding of functional gastrointestinal disease and how to help patients suffering with problems such as irritable bowel syndrome.

There are many opportunities to gain the knowledge and skills to successfully manage patients with gastrointestinal diseases, including reviewing patients on general gastroenterology wards, attending general and specialist outpatient clinics, observing endoscopic procedures, and attending multidisciplinary team (MDT) meetings. MDTs commonly exist for upper and lower gastrointestinal cancers and inflammatory bowel disease, and allow the discussion of patients by gastroenterologists, radiologists, histopathologists, surgeons, oncologists, and nurse specialists.

CLINICAL SKILLS

Box 22.1 History-taking in gastroenterology

Symptoms of upper gastrointestinal disease

- **Dysphagia**—distinguish between a likely ***neurological cause*** (often high in mouth or pharynx, worse with swallowing liquids, often speech involvement), and a likely ***anatomical cause*** (lower down in chest or upper abdomen, worse with larger and more solid food boluses).
- **Dyspepsia**—often a non-specific symptom, but establish the relationship with food ingestion posture—symptoms of reflux disease are worsened by both.
- **Early satiety**—may be a symptom of stomach cancer
- **Nausea and vomiting**—these may be mediated by central causes, but a close relationship with food (e.g., projectile vomiting after eating) suggests a pyloric stricture.
- **Haematemesis**—vomiting of fresh red blood should be treated as a medical emergency. Vomiting of altered blood (sometimes described as 'coffee grounds') is more difficult to evaluate, but should still be taken seriously. Haematemesis after a prolonged bout of vomiting may suggest a Mallory–Weiss tear.
- **Melaena**—this describes the passage of thick, black, 'tarry' altered blood from the rectum and indicates upper gastrointestinal bleeding—a medical emergency.

Symptoms of lower gastrointestinal disease

- **Change in bowel habit**—while constipation is often a symptom of inadequate diet or physical exercise, unexplained loosening of stool, particularly if persistent, may be a symptom of colonic diseases, including cancer and colitis.
- **Rectal bleeding**—fresh blood dripping after passing stool, or present on wiping, suggests anorectal disease including haemorrhoids. Blood mixed with stool may indicate colitis or colonic malignancy.
- **Tenesmus**—a feeling of incomplete rectal emptying after defecation suggests anorectal pathology such as a tumour.

'Red flag' symptoms suggestive of malignancy

- Progressive weight loss.
- Anorexia.
- Symptoms of anaemia from chronic blood loss.

Past medical history

- Gastrointestinal disease.
- Autoimmune disease.
- Abdominal surgery.

Drug history

- All prescription drugs.
- Non-steroidal anti-inflammatories, over-the-counter drugs, and herbal remedies.
- Recent courses of antibiotics.

Social history

- Smoking.
- Alcohol consumption.

History and examination

Gastrointestinal disease can present with a variety of different symptoms and signs and an experienced clinician needs to be able to elicit these in a focused manner, and also ask about relevant risk factors (**Boxes 22.1 and 22.2**). The figure opposite illustrates signs of common gastrointestinal diseases to be found on examination.

Visualizing the bowel

The different types of endoscopy are discussed fully in the 'Endoscopy' section. **Box 22.3** illustrates the ability of different techniques of endoscopy (including contrast techniques in ERCP) to visualize different parts of the gastrointestinal and biliary tract. Most are readily available in the majority of units, although capsule endoscopy and balloon enteroscopy remain more specialist procedures. All except capsule endoscopy allow biopsies to be taken and a range of therapeutic procedures to be performed.

Box 22.2 Core knowledge in gastroenterology

The 'top ten' presentations

- Acute upper GI bleed
- Dysphagia
- Dyspepsia
- Abdominal pain
- Malabsorption
- Diarrhoea
- Iron-deficiency anaemia
- Rectal bleeding
- Constipation
- Tenesmus

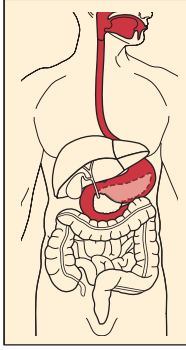
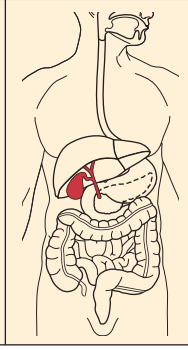
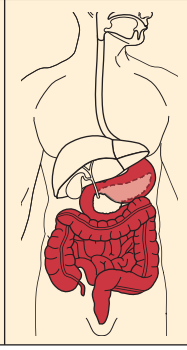
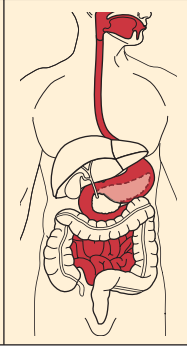
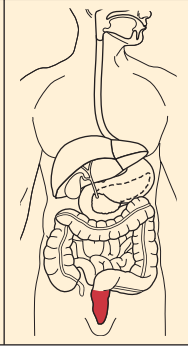
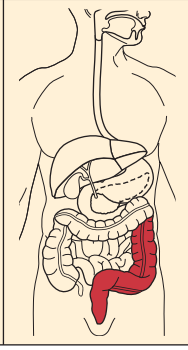
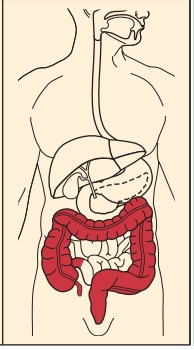
The 'top ten' conditions

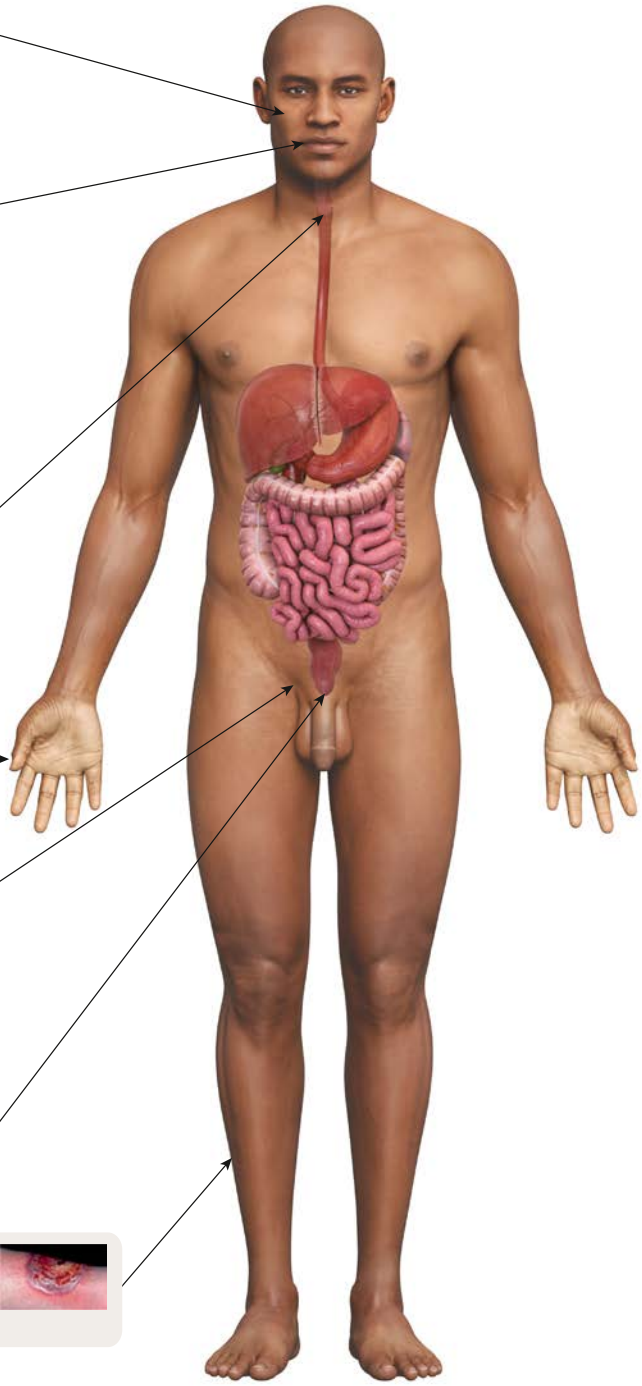
- Functional bowel disorders
- Peptic ulcer disease
- Gastro-oesophageal reflux disease
- Oesophageal cancer
- Gastric cancer
- Coeliac disease
- Inflammatory bowel disease
- Infectious diarrhoea
- Colorectal cancer
- Diverticulosis

The 'top ten' medications

- Proton pump inhibitors and H_2 receptor antagonists
- *Helicobacter pylori* eradication regimens
- Antiemetics
- 5-aminosalicylic acids
- Oral and intravenous steroids
- Azathioprine and methotrexate
- Anti-tumour necrosis factor antibodies
- Oral iron preparations
- Antispasmodics
- Antibiotics for infectious diarrhoea, and when to use them

Box 22.3 Endoscopy techniques

Oesophagogastroduodenoscopy (OGD)	Endoscopic retrograde cholangiopancreatography (ERCP)	Capsule endoscopy	Balloon enteroscopy	Proctoscopy	Flexible sigmoidoscopy	Colonoscopy
						



Face
Conjunctival pallor

Mouth

Angular stomatitis (malnutrition)
Oral ulcers (Crohn's)
Glossitis (inflammation of the tongue) can be due to iron-deficiency anaemia, B12 deficiency or malabsorption

Neck
Cervical and clavicular lymphadenopathy
Virchow's node (enlarged left supraclavicular lymph node) which is a sentinel lymph node for gastric cancer

Hands
Koilonychia (spoon-shaped nails, associated with iron-deficiency anaemia)
Clubbing

Hernial orifices
Signs of incarcerated hernia if intestinal obstruction is suspected

Digital examination of the rectum
Haemorrhoids
Anorectal cancer
Empty rectum (intestinal obstruction)
Full rectum (constipation)
Blood on finger (malignancy or colitis)

Legs

Pyoderma gangrenosum or erythema nodosum (both seen in inflammatory bowel disease)

General features
Cachexia
Jaundice
Pallor

Observation chart
Tachycardia or shock (in acute bleeding)

Abdomen
Inspection
- scars
- abdominal distension

Abdomen
Soft touch palpation
- abdominal tenderness
- masses

Abdomen
Deeper palpation
- masses
- hepatomegaly
- splenomegaly

Abdomen
Percussion
- shifting dullness (ascites)
- suprapubic dullness (distended bladder)
- hepatomegaly
- splenomegaly

Abdomen
Auscultation
- generally not of great value
- absent bowel sounds in peritonism
- high pitched or 'tinkling' in intestinal obstruction

BACKGROUND

The gastrointestinal tract has many functions, such as digestion, absorption, and excretion, as well as the synthesis of hormones, growth factors, and cytokines. In addition, a complex enteric nervous system (ENS) controls its function and communicates with the central and peripheral nervous systems. Finally, as the gastrointestinal tract is the part of the body containing the largest source of foreign antigens, it has very well-developed arms of both the innate and the acquired immune systems. The intestinal microbiota exerts a range of beneficial functions aiding nutrition, absorption, and the development of a mature immune system. Disruption of the composition and diversity of the microbiota is associated with the development of gastrointestinal and systemic disease.

ANATOMY AND PHYSIOLOGY OF THE GASTROINTESTINAL TRACT

Mouth, oropharynx, and oesophagus

The oral cavity extends from the lips to the pharynx and contains the tongue, teeth, and gums. Its primary functions are mastication, swallowing, and speech. The ***pharynx*** consists of the nasal, oral, and laryngeal sections. The latter is the part of the throat that connects to the oesophagus.

The ***oesophagus*** is a muscular tube approximately 20 cm long that connects the pharynx to the stomach just below the diaphragm. Its only function is to transport food from the mouth to the stomach. In the upper portion of the oesophagus, both the outer longitudinal layer and inner circular muscle layers are striated. In the lower two-thirds of the oesophagus, including the thoracic and abdominal parts containing the lower oesophageal sphincter (LOS), both layers are composed of smooth muscle.

The oesophagus is lined by stratified squamous epithelium, which extends distally to the squamocolumnar junction where the oesophagus joins the stomach, recognized endoscopically by a zigzag ('Z') line, just above the most proximal gastric folds.

The oesophagus is separated from the pharynx by the ***upper oesophageal sphincter*** (UOS), which is normally closed due to tonic activity of the nerves supplying the cricopharyngeus. The ***lower oesophageal sphincter*** consists of a 2- to 4-cm zone in the distal end of the oesophagus that has a high resting tone and, assisted by the diaphragmatic sphincter, is largely responsible for the prevention of gastric reflux.

Swallowing

During swallowing, the bolus of food is voluntarily moved from the mouth to the pharynx. This process is mediated by a complex reflex involving a swallowing centre in the dorsal motor nucleus of the vagus in the brainstem. Once activated, the swallowing centre neurons send pre-programmed discharges of inhibition followed by excitation to the motor nuclei of the cranial nerves. This results in initial relaxation, followed by distally progressive activation of neurons to the oesophageal smooth muscle and LOS. Pharyngeal and oesophageal peristalsis mediated by this swallowing reflex causes ***primary peristalsis***. ***Secondary peristalsis*** arises as a result of stimulation by a food bolus in the lumen, mediated by a local intra-oesophageal reflex. ***Tertiary contractions*** indicate pathological non-propulsive contractions resulting from aberrant activation of local reflexes within the myenteric plexus.

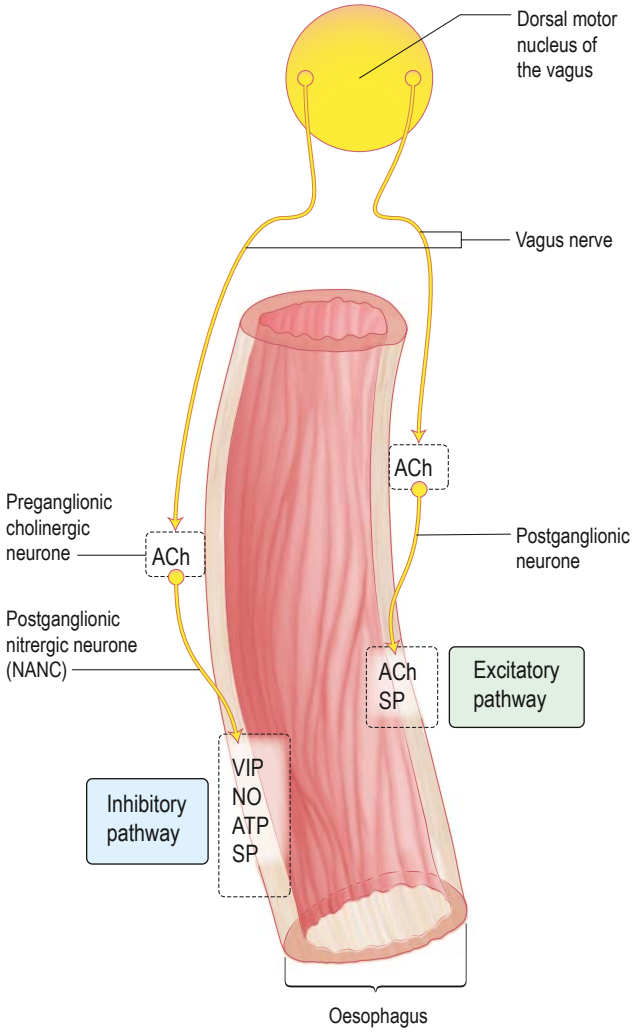


Fig. 22.1 Innervation of the oesophagus. The excitatory pathway consists of vagal preganglionic neurons releasing acetylcholine (ACh), connecting to postganglionic neurons that release ACh and substance P (SP). The inhibitory pathway consists of vagal preganglionic neurons releasing ACh, connecting to postganglionic neurons that release nitric oxide (NO), vasoactive intestinal peptide (VIP), adenosine triphosphate (ATP), and SP.

The smooth muscle of the thoracic oesophagus and LOS is supplied by vagal autonomic motor nerves consisting of extrinsic preganglionic fibres and intramural postganglionic neurons in the myenteric plexus (Fig. 22.1). There are parallel excitatory and inhibitory pathways.

Stomach and duodenum

The stomach occupies a small area immediately distal to the oesophagus (the cardia), the upper region (the fundus, under the left diaphragm), the mid-region or body, and the antrum, which extends to the pylorus (see Fig. 22.13). It serves as a reservoir where food can be retained and broken up before being actively expelled into the proximal small intestine.

The smooth muscle of the wall of the stomach has three layers: outer longitudinal, inner circular, and innermost oblique layers. There are two sphincters: the gastro-oesophageal sphincter and the pyloric sphincter. The latter is largely made up of a thickening of the

circular muscle layer and controls the exit of gastric contents into the duodenum.

The duodenum has outer longitudinal and inner circular smooth muscle layers. It is C-shaped and the pancreas sits in the concavity. It terminates at the duodenojejunal flexure, where it joins the jejunum.

- The ***mucosal lining*** of the stomach can stretch in size with feeding. The greater curvature of the undistended stomach has thick folds or rugae. The mucosa of the upper two-thirds of the stomach contains ***parietal cells***, which secrete hydrochloric acid, and ***chief cells***, which secrete pepsinogen (which initiates proteolysis). There is often a colour change at the junction between the body and the antrum of the stomach, which can be seen macroscopically and confirmed by measuring surface pH.
- The ***antral mucosa*** secretes bicarbonate and contains mucus-secreting cells and ***G cells***, which secrete gastrin, stimulating acid production. There are two major forms of gastrin, G17 and G34, depending on the number of amino-acid residues. G17 is the major form found in the antrum. Somatostatin, a suppressant of acid secretion, is also produced by specialized antral cells ***(D cells)***.
- ***Mucus-secreting cells*** are present throughout the stomach and secrete mucus and bicarbonate. The mucus is made of glycoproteins called mucins.
- The ***'mucosal barrier'***, made up of the plasma membranes of mucosal cells and the mucus layer, protects the gastric epithelium from damage by acid and, for example, alcohol, aspirin, non-steroidal anti-inflammatory drugs (NSAIDs), and bile salts. Prostaglandins stimulate secretion of mucus, and their synthesis is inhibited by aspirin and NSAIDs, which inhibit cyclo-oxygenase (see 'NSAIDs, Helicobacter and ulcers' section).
- The ***duodenal mucosa*** has villi like the rest of the small bowel, and also contains Brunner's glands, which secrete alkaline mucus. This, along with the pancreatic and biliary secretions, helps to neutralize the acid secretion from the stomach when it reaches the duodenum.

Physiology

Acid secretion is central to the functionality of the stomach; factors controlling acid secretion are shown in Fig. 22.2. Acid is not essential for digestion but does prevent some food-borne infections. It is under neural and hormonal control, and both stimulate acid secretion through the direct action of histamine on the parietal cell. Acetylcholine and gastrin also release histamine via the enterochromaffin cells. Somatostatin inhibits both histamine and gastrin release, and therefore acid secretion.

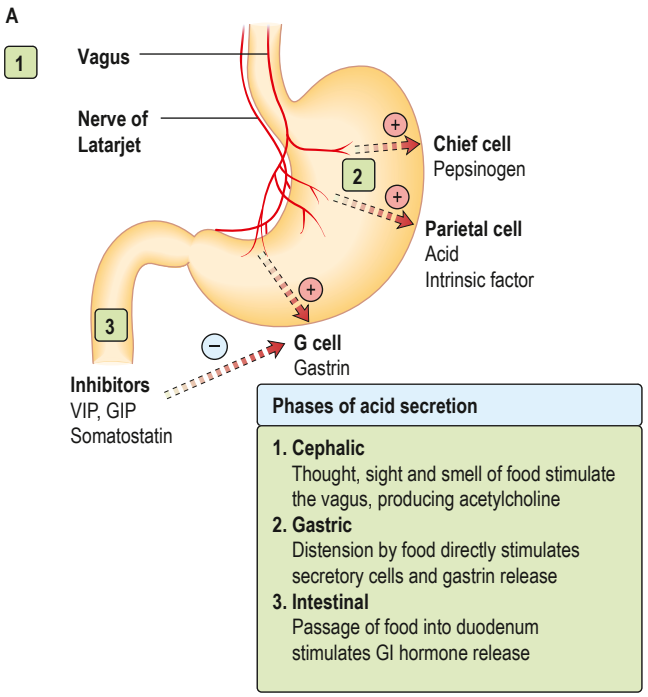
Other major gastric functions are:

- reservoir for food
- emulsification of fat and mixing of gastric contents
- secretion of intrinsic factor
- absorption (of only minimal importance).

Gastric emptying depends on many factors. There are osmoreceptors in the duodenal mucosa, which control gastric emptying by local reflexes and the release of gut hormones. In particular, intraduodenal fat delays gastric emptying by negative feedback through duodenal receptors.

Small intestine anatomy

The small intestine extends from the duodenum to the ileocaecal valve. It is 3 to 6 m in length, and 300 m^2 in surface area. The upper



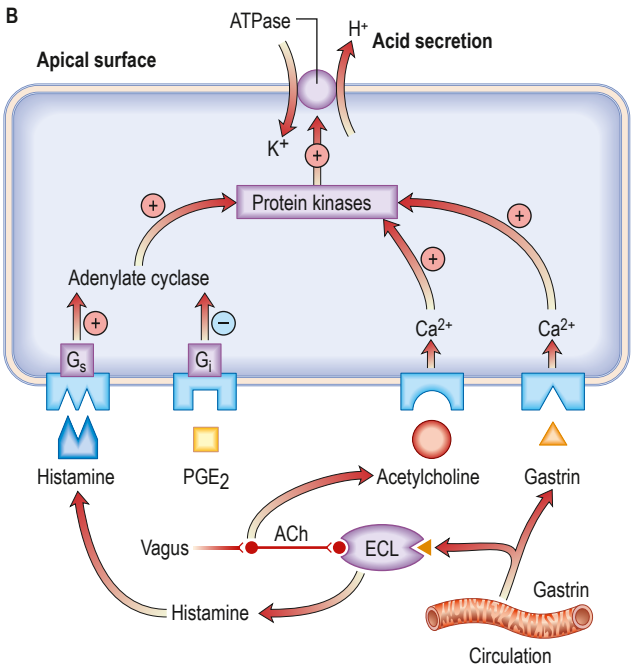


Fig. 22.2 Pathophysiology of acid secretion. **(A)** Control of acid secretion. **(B)** Mechanisms involved in acid secretion. Histamine stimulates the G_s protein via the H_2 receptors and acts via cyclic adenosine monophosphate (cAMP). Prostaglandin E_2 (PGE_2) activates the G_i protein and *inhibits* acid secretion. Acetylcholine (ACh) acts via the vagus M_3 receptors. ACh also acts via the enterochromaffin cell (ECL). Gastrin acts via the cholecystokinin B (CCKBB)–gastrin receptor, increasing the intracellular free calcium, and also via the ECL cell, stimulating histamine. *GIP,* Gastric inhibitory polypeptide; *Gs and Gi,* stimulating and inhibiting G-proteins; *VIP,* vasoactive intestinal polypeptide.

40% is the duodenum and jejunum; the remainder is the ileum. Its surface area is enormously increased by circumferential mucosal folds that bear multiple finger-like projections called villi. On the villi, the surface area is further increased by microvilli on the luminal side of the epithelial cells (enterocytes) (Fig. 22.3).

Each villus consists of a core that contains blood vessels, lacteals (lymphatics), and cells (e.g., plasma cells and lymphocytes). The lamina propria contains plasma cells, lymphocytes, macrophages, eosinophils, and mast cells. The crypts of Lieberkühn are the spaces between the bases of the villi.

Enterocytes are formed at the bottom of the crypts and migrate towards the tops of the villi, where they are shed. This process takes 3 to 4 days. On its luminal side, the enterocyte is covered by microvilli and a gelatinous layer called the glycocalyx. Scattered between the epithelial cells are mucin-secreting goblet cells and occasional intraepithelial lymphocytes and Paneth cells. Most of the blood supply to the small intestine is via branches of the superior mesenteric artery. The terminal branches are end arteries; there are no local anastomotic connections.

Enteric nervous system

The enteric nervous system (ENS) controls the functioning of the small bowel; it is an independent system that coordinates absorption, secretion, blood flow, and motility. It is estimated to contain 10^8 neurons (as many as the spinal cord), organized in two major ganglionated plexuses: the *myenteric plexus* between the muscular layers of the intestinal wall, and the *submucosal plexuses* associated with the mucosa. The ENS communicates with the central nervous system (CNS) via autonomic afferent and efferent pathways but can operate autonomously.

Coordination of small intestinal function involves a complex and only partly understood interplay between many neuroactive mediators and their receptors, ion channels, gastrointestinal hormones, nitric oxide, and other transmitters. Acetylcholine, adrenaline (epinephrine), adenosine triphosphate (ATP), vasoactive intestinal peptide (VIP), and other hormones and opioids have been shown to have actions in the small bowel but the exact role of each is not yet clear.

Gut motility

The contractile patterns of the small intestinal muscular layers are primarily determined by the ENS. The CNS and gut hormones also have a modulatory role in motility. The interstitial cells of Cajal, which lie within the smooth muscle, appear to govern rhythmic contractions.

During fasting, a distally migrating sequence of motor events, termed the migrating motor complex (MMC), occurs in a cyclical fashion. The MMC consists of:

- a period of motor quiescence *(phase I)*
- a period of irregular contractile activity *(phase II)*
- a short (5–10-min) burst of regular phasic contractions *(phase III).*

Each MMC cycle lasts for approximately 90 minutes. In the duodenum, phase III is associated with increased gastric, pancreatic, and biliary secretions. The role of the MMC is unclear, but the strong phase III contractions propel secretions, residual food, and desquamated cells towards the colon. It is named the 'intestinal housekeeper'.

After a meal, the MMC pattern is disrupted and replaced by irregular contractions. This seemingly chaotic pattern lasts typically for 2 to 5 hours after feeding, depending on the size and nutrient content of the meal. The irregular contractions of the fed pattern have a mixing function, moving intraluminal contents to and fro and aiding the digestive process.

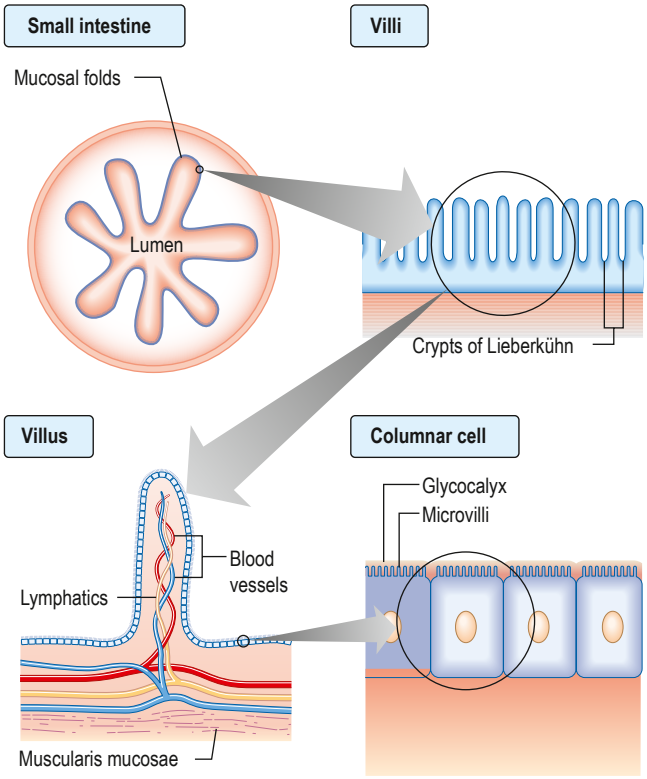


Fig. 22.3 Structure of the small intestine.

Neuroendocrine peptide production

The hormone-producing cells of the gut are scattered diffusely throughout its length and also occur in the pancreas.

Gut hormones play a part in the regulation and integration of the functions of the small bowel and other metabolic activities, including appetite. Their actions are complex and interactive, both with each other and with the ENS (Box 22.4).

Small intestine physiology

In the small bowel, digestion and absorption of nutrients and ions take place, as does the regulation of fluid absorption and secretion. The epithelial cells of the small bowel form a physical barrier that is selectively permeable to ions, small molecules, and macromolecules. Digestive enzymes, such as proteases and disaccharidases, are produced by intestinal cells and expressed on the surface of microvilli; others, such as lipases produced by the pancreas, are associated with the glycocalyx. Some nutrients are absorbed most actively in specific parts of the small intestine: iron and folate in the duodenum and jejunum, and vitamin B_{12} and bile salts in the terminal ileum, where they have specific receptors.

General principles of absorption

Simple diffusion

This process is non-specific, requires no carrier molecule or energy, and takes place if there is a concentration gradient from the intestinal lumen (high concentration) to the bloodstream (low concentration). Vitamin B_{12} can be absorbed from the jejunum by this means.

Facilitated diffusion

Absorption takes place down a concentration gradient but a membrane carrier protein is involved, conferring specificity on the process. Fructose transport is an example.

Box 22.4 Gut regulatory peptides

Peptide	Localization	Main actions
Gastrin/cholecystokinin family		
Cholecystokinin (CCK): multiple forms from CCK8 (8 amino acids) to CCK83; 8, 33 and 58 are predominant Terminal 5 amino acids same as gastrin	Duodenum and jejunum (I cells) Enteric nerves CNS	Causes gall bladder contraction and sphincter of Oddi relaxation. Trophic effects on duodenum and pancreas. Pancreatic secretion (minor role) Role in satiety—acting on CNS
Gastrin	G cells in gastric antrum and duodenum	Stimulates acid secretion. Trophic to mucosa
Secretin-glucagon family		
Secretin	Duodenum and jejunum (S cells)	Stimulates pancreatic bicarbonate secretion
Glucagon	Alpha cells of pancreas	Opposes insulin in blood glucose control
Vasoactive intestinal polypeptide (VIP)	Enteric nerves	Intestinal secretion of water and electrolytes. Neurotransmitter. Splanchnic vasodilation, stimulates insulin release
Glucose-dependent insulinotropic peptide (GIP)	Duodenum (K cells) Gastric antrum Ileum	Release by intraduodenal glucose causes greater insulin release by islets than i.v. glucose (incretin effect)
Glucagon-like peptide-1 (GLP-1)	Ileum and colon (L cells)	Incretin. Stimulates insulin synthesis. Trophic to islet cells. Inhibits glucagon secretion and gastric emptying Stimulates growth of enterocytes
Glicentin	L cells, A cells	Stimulates insulin secretion and gut growth, inhibits gastric secretion
Growth hormone-releasing factor (GHRF)	Small intestine	Unclear
Pancreatic polypeptide family		
Pancreatic polypeptide (PP)	Pancreas (PP cells)	Inhibits pancreatic and biliary secretion
Peptide YY (PYY)	Ileum and colon (L cells)	Inhibits pancreatic exocrine secretion. Slows gastric and small bowel transit ('ileal brake'). Reduces food intake and appetite
Neuropeptide Y (NPY)	Enteric nerves	Stimulates feeding. Regulates intestinal blood flow
Other		
Motilin	Whole gut	Increases gastric emptying and small bowel contraction
Ghrelin	Stomach	Stimulates appetite, increases gastric emptying
Obestatin	Stomach and small intestine	Opposes ghrelin
Oxyntomodulin	Colon	Inhibits appetite
Gastrin releasing-polypeptide (bombesin)	Whole gut and pancreas	Stimulates pancreatic exocrine secretion and gastric acid secretion
Somatostatin	Stomach and pancreas (D cells) Small and large intestine	Inhibits secretion and action of most hormones
Substance P	Enteric nerves	Enhances gastric acid secretion, smooth muscle contraction
Neurotensin	Ileum	Affects gut motility. Increases jejunal and ileal fluid secretion
Insulin	Pancreatic β cells	Increases glucose utilization
Chromogranins	Neuroendocrine cells	Precursor for other regulatory peptides that inhibit neuroendocrine secretion

Active transport

Absorption occurs via a specific carrier protein, powered by cellular energy, allowing a substance to be transported against a concentration gradient. Many carrier proteins are powered by ion gradients across the enterocyte wall. For example, glucose crosses the enterocyte microvillous membrane from the lumen into the cell against a concentration gradient by using a ***co-transporter*** carrier molecule. This is the sodium/glucose co-transporter, SGLT1 (Fig. 22.4). The process is powered by the energy derived from the flow of Na^+ ions from a high concentration outside the cell to a low concentration inside. The sodium gradient across the cell wall is maintained by a separate ATP-consuming Na^+/K^+ exchanger in the basolateral membrane. Glucose leaves the cell on the serosal side by facilitated diffusion via a sodium-independent carrier (GLUT-2) in the basolateral membrane.

Another active transport mechanism operates for Na^+ absorption in the ileum using an Na^+/H^+ ***exchange*** mechanism, powered by the outwardly directed gradient of H^+ across the cell membrane.

Absorption of nutrients in the small intestine

Carbohydrate

Dietary carbohydrate consists mainly of starch, with some sucrose and a small amount of lactose. Starch is a polysaccharide made up of numerous glucose units. In order to have a nutrient value,

starch must be digested into smaller oligo-, di- and finally monosaccharides, which may then be absorbed. Polysaccharide hydrolysis begins in the mouth and is catalyzed by salivary amylase, though the majority takes place under the action of pancreatic amylase in the upper intestine. The breakdown products of starch digestion are maltose and maltotriose, together with sucrose and lactose. These are further hydrolyzed on the microvillous membrane by specific oligo- and disaccharidases to form glucose, galactose, and fructose. These monosaccharides are then able to be transported across the enterocytes into the blood (see **Fig. 22.4**).

Protein

Dietary protein is digested by pancreatic proteolytic enzymes to amino acids and peptides prior to absorption. These enzymes are secreted by the pancreas as pro-enzymes and transformed to active forms in the lumen. Protein in the duodenal lumen stimulates the enzymatic conversion of trypsinogen to trypsin, and this, in turn, activates the other pro-enzymes, chymotrypsin, and elastase.

These enzymes break down protein into oligopeptides. Some di- and tripeptides are absorbed intact by carrier-mediated processes, while the remainder are broken down into free amino acids by peptidases on the microvillous membranes of the enterocytes, prior to absorption into the cell by a variety of amino acid and peptide carrier systems.

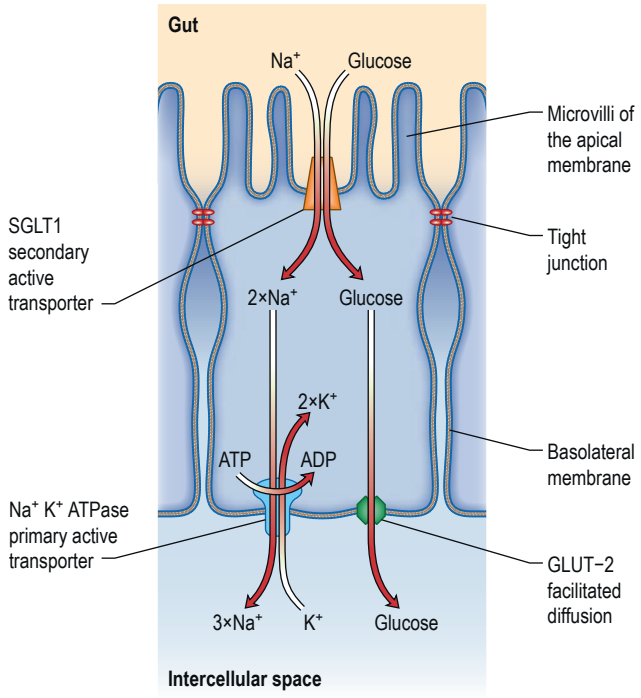


Fig. 22.4 Transcellular uptake of glucose across the intestinal epithelia. Glucose is co-transported across the apical membrane with sodium ions by the sodium-dependent glucose transporter (SGLT). This is secondary active transport, as the sodium is travelling down its electrochemical gradient. The sodium gradient is maintained by the primary active transport of sodium across the basolateral membrane by the Na^+/K^+ ATPase (thus intracellular Na^+ is kept low). Transcellular transport of glucose is achieved by facilitated diffusion across the basolateral membrane as glucose is moved down its concentration gradient by GLUT-2.

Fat

Dietary fat consists mainly of triglycerides with some cholesterol and fat-soluble vitamins. Fat is emulsified by mechanical action in the stomach. Bile containing the amphipathic detergents, bile acids, and phospholipids enters the duodenum following gall bladder contraction. These substances act to solubilize fat and promote hydrolysis of triglycerides in the duodenum by pancreatic lipase to yield fatty acids and monoglycerides. Bile acids, phospholipids, and the products of fat digestion cluster together with their hydrophilic ends on the outside to form aggregations called mixed micelles. Trapped in the centre of the micelles are the hydrophobic monoglycerides, fatty acids, and cholesterol. At the cell membrane, the lipid contents of the micelles are absorbed, while the bile salts remain in the lumen. Inside the cell, the monoglycerides and fatty acids are re-esterified to triglycerides. The triglycerides and other fat-soluble molecules (e.g., cholesterol, phospholipids) are then incorporated into chylomicrons to be transported into the lymph.

Any unabsorbed lipids that reach the ileum delay gastric emptying via peptide YY, which is secreted by the ileum (called the ***'ileal brake'***). This delay allows more time for absorption of lipids in the small intestine.

Medium-chain triglycerides (MCTs, fatty acids of chain length 6–12) are transported via the portal vein with a small amount of long-chain fatty acid. Patients with pancreatic exocrine or bile salt insufficiency can therefore supplement their fat absorption with MCTs.

Bile salts are not absorbed in the jejunum, so the intraluminal concentration in the upper gut is high. They pass down the intestine to be absorbed in the terminal ileum and are transported back to the liver. This enterohepatic circulation prevents excess loss of bile salts (see Chapter 23).

The pathophysiology of fat absorption is shown in **Fig. 22.5**. Interference with absorption can occur at all stages, as indicated, giving rise to steatorrhoea (>17 mmol or 6 g of faecal fat per day).

Water and electrolytes

Large amounts of water and electrolytes, partly dietary but mainly from intestinal secretions, are absorbed, coupled with absorption of monosaccharides, amino acids, and bicarbonate in the upper jejunum. Water and electrolytes are also absorbed paracellularly (between the enterocytes) down electrochemical and osmotic gradients. Additional water and electrolytes are absorbed in the ileum and colon, where active sodium transport is not coupled to solute absorption. Secretion of fluid and electrolytes occurs together to maintain the normal functioning of the gut. Secretory diarrhoea (see 'Secretory diarrhoea' section) can occur because of defects in intestinal secretory mechanisms.

Water-soluble vitamins, essential metals, and trace elements

These are all absorbed in the small intestine. Vitamin B_{12} (see Chapter 38) and bile salts are absorbed by specific transport mechanisms in the terminal ileum; malabsorption of both these substances often occurs following ileal resection.

Calcium

Calcium absorption is discussed in Chapter 33.

Iron

Iron absorption is discussed in Chapter 38.

Response of the small bowel to antigens and pathogens

The small bowel has a number of mechanisms to prevent colonization and invasion by pathogens while simultaneously preventing

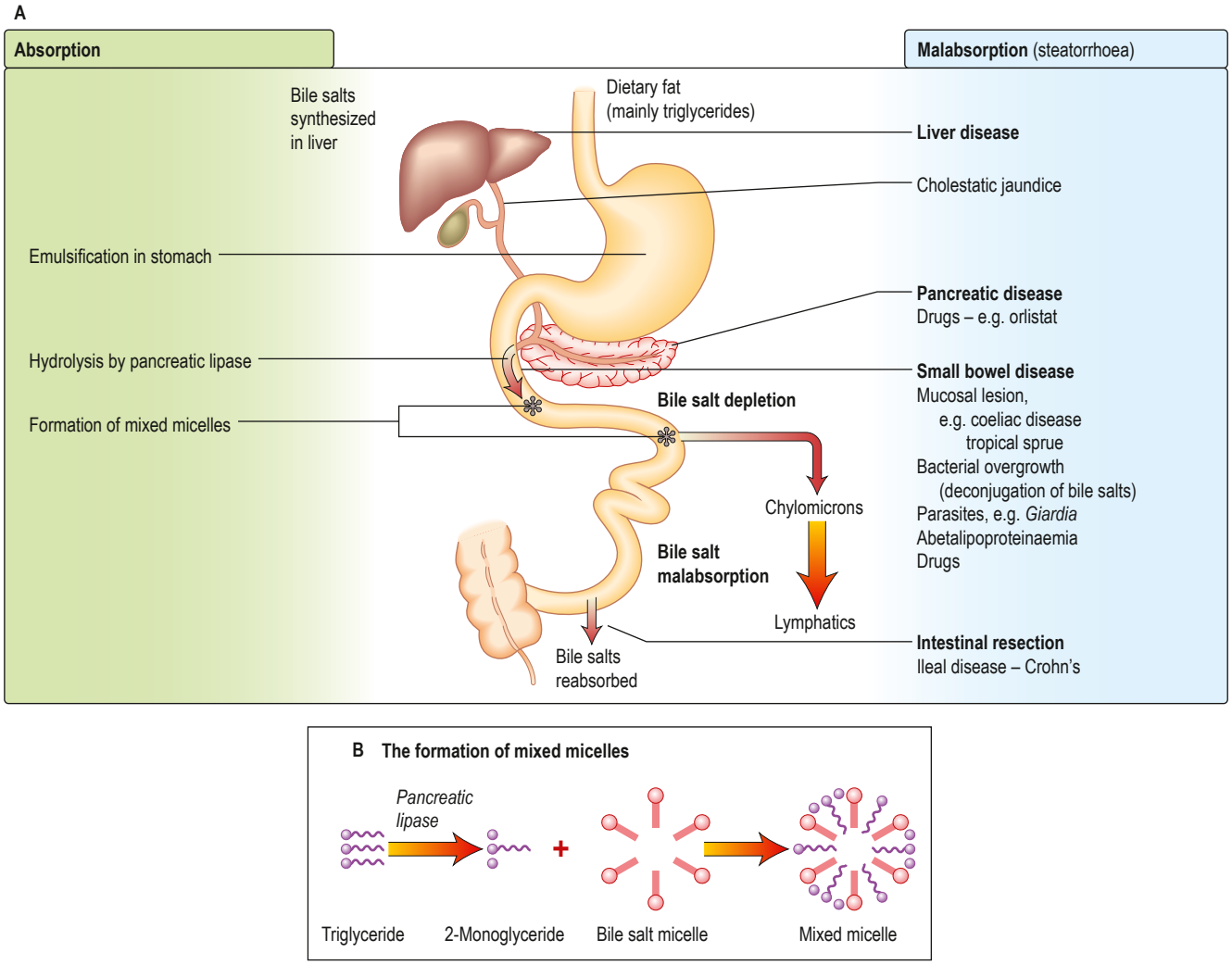


Fig. 22.5 Fat absorption in the small intestine. **(A)** The pathophysiology of fat absorption. Diseases impairing fat absorption, that may therefore cause steatorrhoea, are shown on the right of the diagram. **(B)** The formation of mixed micelles.

inappropriate responses to foreign antigens or the indigenous bacterial population. At the same time, commensal bacteria maintain the integrity of the small bowel and play a major role in host physiology.

Mechanisms

Physical defence

- The mucus layer.
- Continuous shedding of surface epithelial cells.
- The physical movement of the luminal contents.
- Colonization resistance—the ability of the indigenous microbiota to outcompete pathogens for a survival niche in the gut.

Innate chemical defence

- ***Enzymes*** such as lysozyme and phospholipase A_2, secreted by Paneth cells at the base of the crypts, help ensure an infection-free environment in the gut, even in the presence of commensal bacteria.
- ***Antimicrobial peptides*** are secreted from enterocytes and Paneth cells in response to pathogenic bacteria. These include ***defensins***, which are 15 to 20 amino-acid peptides with potent activity against a broad range of pathogens, including Gram-positive and Gram-negative bacteria, fungi, and viruses.
- ***Trefoil peptides*** are a family of small proteins secreted by goblet cells. They consist of a three-loop structure with intra-chain disulphide bonds, which makes the molecules highly resistant to digestion. Their actions include stabilization of mucus, promotion of cell migration to injured areas, and promotion of repair. Three trefoil factors (TFFs) are found in humans (TFF1, TFF2, and TFF3), all of which have been implicated in the response to gastrointestinal injury in experimental models. Their molecular mode of action is not yet known.

Innate immunological defence

- ***Humoral defence.*** IgA is the principal mucosal antibody. It mediates mucosal immunity by agglutinating and neutralizing pathogens in the lumen and preventing colonization of the epithelial surface (Fig. 22.6). IgA is secreted from immunocytes in the lamina propria as dimers joined by a protein called the 'joining chain' (J-chain); in this form, it is known as polymeric IgA (pIgA). This pIgA is internalized by endocytosis at the basolateral membrane of enterocytes. It crosses the cell as a complex of pIgA/pIgAR and is secreted on to the mucosal surface.

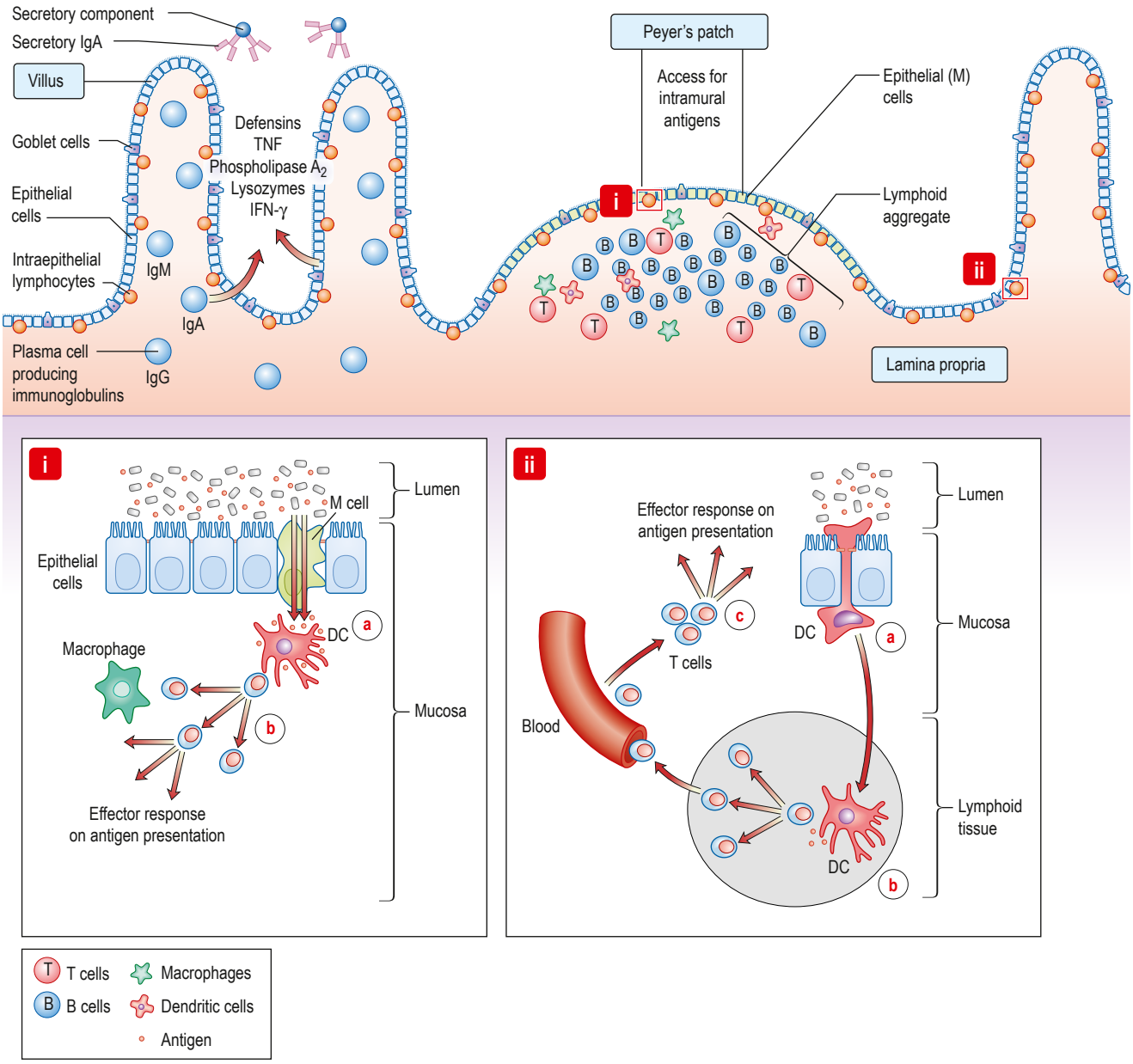


Fig. 22.6 Small intestinal mucosa with a Peyer's patch, showing the gut-associated lymphoid tissue (GALT). Inset i: **(A)** Specialized 'M' cells within a lymphoid follicle pass antigen to underlying antigen-presenting cells (APCs), such as dendritic cells (DCs). **(B)** DCs present antigen to T cells, resulting in activation and effector responses. Inset ii: **(A)** DCs pass dendrites through epithelial tight junctions to sample luminal antigen. **(B)** DCs traffic to the mesenteric lymph node and present antigen to naïve T cells, which differentiate into effector or regulatory subtypes and are imprinted with a gut-homing integrin. **(C)** On subsequent antigen presentation, activated T cells home to lamina propria and exert a regulatory or inflammatory effect, depending on their phenotype. *IFN-γ,* Interferon-gamma; *Ig,* immunoglobulin; *TNF,* tumour necrosis factor.

- ***B-cell sensitization.*** Antigens from the lumen of the bowel are transported by M cells and dendritic cells in the follicle-associated epithelium (FAE). This covers Peyer's patches in the 'dome' region that contain abundant virgin B cells, helper T cells, and antigen-presenting cells. Activated B cells then produce IgA locally and are programmed to home back to the lamina propria. They travel through mesenteric lymph nodes and then via the thoracic duct to the blood and back to the small bowel and other mucosal surfaces (such as the airways), where they undergo terminal differentiation into plasma cells. Homing back to the gut is facilitated by the α4β7-integrin on gut-derived lymphocytes binding to MAdCAM-1, uniquely expressed on blood vessels in the gut.
- ***Cellular defence.*** T lymphocytes also provide host defence and initiate, activate, and regulate adaptive immune responses. ***Intestinal T lymphocytes*** occur principally in three major compartments:
  - *Organized gut-associated lymphoid tissue (GALT)*, such as Peyer's patches, where mucosal T-cell responses are generated, and after which cells leave the organized lymphoid tissue and home back to the mucosa.

- *The lamina propria*, containing mostly CD4 cells.
- The *surface epithelium*, where these lymphocytes are known as intraepithelial lymphocytes (IELs) and are mostly CD8 cells. T cells are sensitized to antigen in the Peyer's patch lymphoid tissue in a similar fashion to B cells, and pass through mesenteric lymph nodes into the thoracic duct and into the circulation, homing back to the small bowel to end up in the lamina propria or the epithelium. It is probable that IELs are cytotoxic cells, capable of killing virally or bacterially infected epithelial cells. CD4 cells in the lamina propria of healthy individuals are highly activated cells, probably protecting against low-grade infections, since loss of these cells, as in HIV infection, leads to colonization of the gut by protozoa such as cryptosporidia.

The gut microbiome

The relationship between the hundred thousand billion microbes in the human gut and the host is only beginning to be appreciated. New molecular sequencing techniques have allowed the identification and classification of the gut microbiota. The use of metabonomics has aided the assessment of its functional output. There is increasing evidence that a reduction in the diversity of the gut microbiota is associated with a range of conditions, including inflammatory bowel disease (IBD) and metabolic syndrome. Germ-free mice have essentially no mucosal immune system, showing that the abundant and activated immune system seen in healthy individuals is driven by the flora, without adverse effects. Bacteria also release chemical signals, such as lipopolysaccharide (LPS) and lipoteichoic acid, which are recognized by Toll-like receptors (TLRs) (see Chapter 6) present on a variety of intestinal cells, priming repair processes and enhancing the ability of the epithelium to respond to injury.

Oral tolerance

The immune system must guard against pathogens and toxins while avoiding an excessive response to the multiplicity of food antigens and commensal bacteria. The mechanisms by which tolerance occurs are multiple, including maintenance of barrier function to prevent excess antigen uptake, active inhibition via regulatory T cells, and promotion of tolerogenic rather than immunogenic T-cell responses by dendritic cells. All of these are likely to play a role in diseases such as coeliac disease, caused by an excessive T-cell response to gluten, or Crohn's disease, where tolerance to the indigenous bacterial population is defective.

Colon and rectal anatomy

The large intestine starts at the caecum, on the posterior medial wall of which is the appendix.

The ***colon*** is made up of ascending, transverse, descending, and sigmoid parts, which join the rectum at the rectosigmoid junction (Fig. 22.7). The muscle wall consists of an inner circular layer and an outer longitudinal layer. The outer layer is incomplete, coming together to form the taeniae coli, which produce the haustral pattern seen in the normal colon.

The mucosa of the colon is lined with epithelial cells and goblet cells. There are crypts but no villi, so that the surface is flat. A variety of cells, mainly lymphocytes and macrophages, are found in the lamina propria.

The blood supply to the colon is from the superior and inferior mesenteric vessels. The colon is innervated mainly by the ENS with input from the parasympathetic and sympathetic pathways. Spinal afferent neurons from the dorsal root ganglia innervate the entire colon.

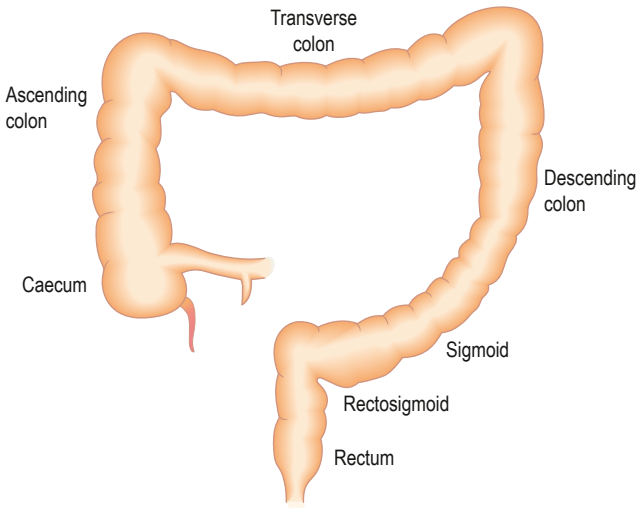


Fig. 22.7 Structure of the large intestine.

Box 22.5 Input and output of water and electrolytes in the gastrointestinal tract over 24 hours

	Water (mL)	Sodium (mmol)	Potassium (mmol)
Input			
Diet	1500	150	80
Gastrointestinal secretions	7500	1000	40
Totals	9000	1150	120
Output			
Faeces	150	5	12
Ileostomy (adapted)	500–1000	60–120	4

The ***rectum*** is about 12 cm long. Its interior is divided by three crescentic circular muscles that produce shelf-like folds. These are the rectal valves, which can be seen at sigmoidoscopy. The anal canal has an internal and an external sphincter.

Physiology of the colon

The main roles of the colon are the absorption of water and electrolytes (Box 22.5) and the propulsion of contents from the caecum to the anorectal region. Approximately 1.5 to 2 L of fluid pass the ileocaecal valve each day. Absorption is stimulated by short-chain fatty acids, which are produced predominantly in the right colon by the anaerobic metabolism of dietary fibre by bacterial polysaccharidase enzyme systems. Colonic contents are mixed, aiding absorption by non-propagative segmenting muscular contractions. High-amplitude propagative colonic contractions cause propulsion. Peristalsis is induced by the release of serotonin (5-HT) from neuroendocrine cells in response to luminal distension. Serotonin activates the HT_4 receptors, which, in turn, results in the activation of sensory (calcitonin gene-related peptide, CGRP) neurons. Normal colonic transit time is 24 to 48 hours with normal stool weights of up to 250 g/day.

Physiology of defecation

The role of the rectum and anus in defecation is complex. The rectum is normally empty. Stool is propelled into the rectum by propagated colonic contractions. A sensation of fullness, a desire to

defecate, and urgency to defecate are experienced with increasing volumes of rectal content (threshold 100 mL). The sensations are associated with rectal contraction and a relaxation of the internal anal sphincter, both of which serve to push the stool down into the proximal anal canal. This increases the defecatory urge, which can only be suppressed by vigorous contraction of the external sphincter and puborectalis. If conditions are appropriate for defecation, the subject sits or squats, contracts the diaphragm and abdominal muscles, and relaxes the pelvic floor muscles, including puborectalis, and the anal sphincter muscles, with the result that stool is expelled.

CLINICAL FEATURES OF GASTROINTESTINAL DISEASE

Dysphagia

Dysphagia refers to a difficulty in swallowing. There are a multitude of different causes (see **Box 22.6**) and can be a difficulty with swallowing liquids, solids, or both. It can be intermittent or progressive depending on the cause.

As described in the previous section, swallowing is a complex process that involves physical structures from the mouth down through the pharynx into the oesophagus and finally into the stomach. There is neurological involvement from the brainstem, and cranial and peripheral nerves. Dysfunction or disruption to any of these structures or processes can result in dysphagia. Broadly, dysphagia can be categorized into oropharyngeal or oesophageal.

Patients often describe a sensation of food getting stuck and the commonest place for this sensation to occur is the GOJ. Sometimes dysphagia can be associated with pain on swallowing—odynophagia—but this is not always the case. In the event that a patient's swallowing is impaired to the extent that they cannot maintain sufficient calorie intake, then they may experience sustained weight loss as well.

Investigation usually involves upper GI endoscopy if an oesophageal cause is suspected. If this does not identify a cause, then further investigations may include imaging, such as a CT or a barium swallow, or functional investigations such as oesophageal manometry. Treatment depends on the underlying cause.

Dyspepsia

'Indigestion' is common: 80% of the population will suffer from this symptom at some time. Dyspepsia is an inexact term used to describe a number of upper abdominal symptoms such as heartburn, acidity, pain or discomfort, nausea, wind, fullness, or belching. Patients who use the term 'indigestion' may also be describing lower gastrointestinal symptoms such as constipation or the presence of undigested vegetable material in the stool, so obtaining a precise history is necessary.

Features of dyspepsia that are suggestive of serious diseases such as cancer are known as ***'alarm' symptoms***. They include:

- dysphagia
- weight loss
- vomiting
- anorexia
- haematemesis or melaena.

Patients aged 55 years or over who demonstrate any of these features have a higher possibility of significant gastrointestinal pathology and should be investigated on an urgent basis. However, most patients with dyspepsia can be managed in primary care without need for referral to a gastroenterologist. In the UK, the National Institute of Clinical Excellence (NICE) has published a guideline algorithm for managing a patient with dyspepsia (https://www.nice.org.uk/guidance/cg184). A significant number of patients with these symptoms will have functional dyspepsia (see 'Functional dyspepsia' section).

? Box 22.6 Differential Diagnosis

Oesophageal	• Oesophagitis • Benign reflux stricture • Candidiasis • Eosinophilic oesophagitis • Oesophageal carcinoma • Schatzki ring • Pharyngeal pouch
Neuromuscular	• Achalasia • Scleroderma • Diffuse oesophageal spasm • Myasthenia gravis • Bulbar/pseudobulbar palsy • Myotonic dystrophy • Stroke
Extrinsic compression	• Retrosternal goitre • Mediastinal mass

Abdominal pain

Abdominal pain is an extremely common symptom and will be experienced by almost everyone in the population at some point or other during their lives. It can be acute or chronic, mild through to severe, and continuous or intermittent. Pain in the abdomen can be due to a wide range of causes, many of them not related to the gastrointestinal tract, for example, biliary and pancreatic problems, gynaecological problems, urinary tract problems including kidney stones, surgical issues including herniae, vascular problems such as abdominal aortic aneurysm. Abdominal pain is also one of the main presenting features of an acute abdomen (see 'The acute abdomen/Pain' section).

Gastrointestinal causes of abdominal pain can include gastro-oesophageal reflux (see 'Gastro-oesophageal reflux disease' section), dyspepsia (see earlier), acute infection, IBD, irritable bowel syndrome (IBS), and colorectal cancer. The key to diagnosis is to take a thorough history in order to establish:

- The location of the pain—is this constant or does it shift or radiate? Does it fit with referred pain (i.e., umbilical pain of appendicitis) or is it more generalized?
- The timeframe of the pain—acute versus chronic, intermittent versus continuous.
- Severity at its best and worst—a score out of 10 can be helpful.
- Nature of the pain—reflux is often a burning pain in the epigastrium, infectious diarrhoea, or colitis usually gives a crampy lower abdominal pain. However, pain can sometimes be difficult to characterize and patients can describe pain that fits multiple causes.
- Precipitants—for example, food (can suggest reflux, IBS, biliary issues), movement (acute abdomen, musculoskeletal problems), constipation.
- Relieving factors—analgesia (remember that chronic opioid use can lead to and worsen abdominal pain), defaecation (the crampy nature of colitis pain may be temporarily relieved by passing stool).
- Any associated symptoms—for example, nausea, vomiting, diarrhoea, fever can help narrow down or exclude a diagnosis.

Box 22.7 Commonly used antiemetic drugs

Antiemetic class	Drugs	When to use	Side-effects
Antihistamines	Cyclizine Promethazine	Motion sickness Vestibular nausea	i.v. cyclizine can be addictive
Dopamine-R antagonists	Prochlorperazine Chlorpromazine (phenothiazines)	Post-chemotherapy	Extra-pyramidal reactions—avoid in Parkinson's
	Haloperidol	Palliative	QT prolongation
	Metoclopramide	Gastroenteritis Migraine Postoperative	Increased gastric emptying Tardive dyskinesia with prolonged use
Serotonin receptor antagonists (5-HT_3)	Ondansetron	Chemotherapy-induced (1st-line)	Constipation Headache Dizziness
Glucocorticoids	Dexamethasone	Chemotherapy-induced (in combo with 5-HT_3 RA)	Steroid side-effect profile

This is an amended version of the SOCRATES mnemonic which can be a helpful tool to structure a history around.

Investigation and treatment will depend on the suspected differential diagnosis that is gained from the history.

Nausea and vomiting

Nausea is the sensation of being about to vomit, which can occur with or without vomiting. It can be more unpleasant and disabling for patients than vomiting alone. There are two central areas that are responsible for triggering nausea and vomiting. The nucleus tractus solitarius in the medulla, which acts as a central pattern generator for vomiting, and the area postrema in the floor of the fourth ventricle. This contains the 'chemoreceptor trigger zone' (CTZ), which has five different receptors:

- 5-HT receptors
- histamine H1 receptors
- muscarinic receptors
- dopamine D2 receptors
- substance P (neurokinin-1-neuropeptide).

Substances such as opioids, hormones, ketoacids, and urea can all stimulate the CTZ directly as they are carried in the bloodstream.

Nausea and vomiting can be caused by a huge number of varied conditions, physiological states (e.g., pregnancy), infections, medications, and toxins (see Box 11.38). The choice of antiemetic depends on the cause of the vomiting and the side-effect profile of the antiemetic (**Box 22.7**).

Diarrhoea

The word 'diarrhoea' comes from the Greek meaning 'to flow through'. It is a very common manifestation of GI disease with many causes (**Box 22.8**; see also Box 11.39 for common causes in the acute medical unit). There are many definitions but commonly it is accepted as being more than three loose stools in 24 hours.

Pathophysiology

Osmotic diarrhoea

The gut mucosa acts as a semipermeable membrane and fluid enters the bowel if there are large quantities of non-absorbed hypertonic substances in the lumen. This occurs because the patient:

- has ingested a non-absorbable substance (e.g., a purgative, such as magnesium sulphate or magnesium-containing antacid)
- has generalized malabsorption, so that high concentrations of solute (e.g., glucose) remain in the lumen
- has a specific absorptive defect (e.g., disaccharidase deficiency or glucose-galactose malabsorption).

Box 22.8 Causes of diarrhoea

Non-infective causes

- Inflammatory bowel disease
- Radiation proctitis or colitis
- Behçet's disease
- Diverticular disease
- Ischaemic colitis
- Gastrointestinal lymphoma
- Carcinoma of the colon (change in bowel habit)
- Malabsorption
- Gut resection
- Bile acid malabsorption
- Drugs—many, including:
  - Laxatives
  - Metformin
  - Anticancer drugs
  - Statins
  - Proton pump inhibitors
- Faecal impaction with overflow
- Irritable bowel syndrome and functional diarrhoea
- Endocrine causes:
  - Zollinger–Ellison syndrome
  - VIPoma
  - Somatostatinoma
  - Glucagonoma
  - Carcinoid syndrome
  - Thyrotoxicosis
  - Medullary carcinoma of thyroid
  - Diabetic autonomic neuropathy
- Factitious diarrhoea:
  - Purgative abuse
  - Dilutional diarrhoea

Infective causes

- Bacterial, for example,
  - *Campylobacter jejuni*
  - *Salmonella* spp.
  - *Shigella*
  - *Escherichia coli* (see Chapter 42)
  - *Staphylococcal enterocolitis*
  - *Bacillus cereus*
  - *Clostridium perfringens, C. botulinum, C. difficile*
  - Gastrointestinal tuberculosis
- Viral, for example,
  - Rotavirus
- Fungal, for example,
  - Histoplasmosis
- Parasitic, for example,
  - Amoebic dysentery *(Entamoeba histolytica)*
  - Schistosomiasis
  - *Giardia intestinalis*

The volume of diarrhoea produced by these mechanisms is reduced by the absorption of fluid by the ileum and colon. The diarrhoea stops when the patient stops eating or the malabsorptive substance is discontinued.

Secretory diarrhoea

In this disorder, there is both active intestinal secretion of fluid and electrolytes, and decreased absorption. The mechanism of intestinal secretion is shown in **Fig. 22.8A**.

Common causes of secretory diarrhoea are:

- enterotoxins (e.g., cholera, *E. coli* thermolabile or thermostable toxin, *Clostridium difficile* toxin)
- hormones (e.g., VIP in Verner–Morrison syndrome; see Chapter 24)
- bile salts (in the colon) following ileal resection

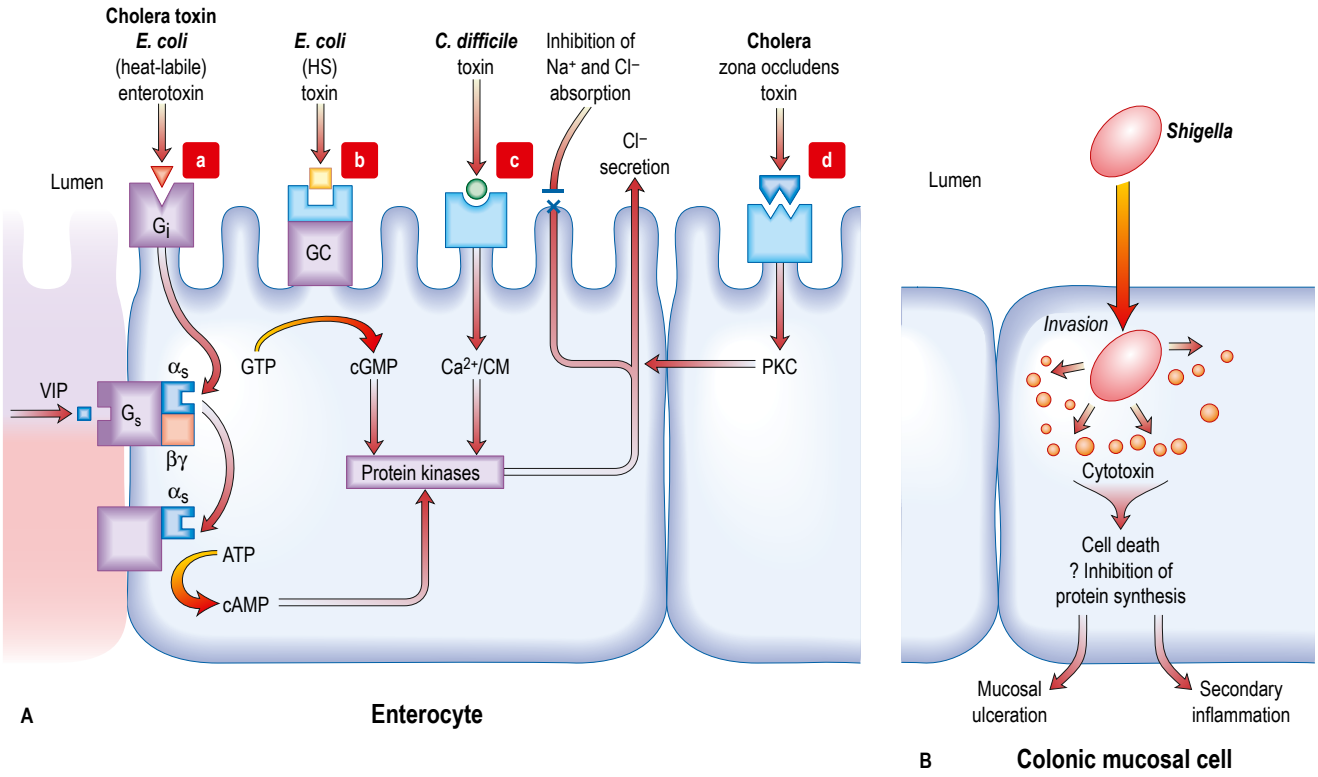


Fig. 22.8 Mechanisms of diarrhoea. (A) Small intestinal secretion of water and electrolytes. (a) *Cholera toxin* binds to its receptor (monosialoganglioside G_i) via fimbria (toxin co-regulated pili) on its β-subunit. This activates the α_s subunit (of the G_s protein), which in turn dissociates and activates cyclic AMP (cAMP). The increase in cAMP activates intermediates (e.g., protein kinase and Ca^{2+}), which then act on the apical membrane, causing Cl^- secretion (with water) and inhibition of Na^+ and Cl^- absorption. *Heat-labile E. coli enterotoxin* shares a receptor with cholera toxin. (b) *Heat-stable (HS) E. coli toxin* binds to its own receptor and activates guanylate cyclase (cGMP), producing the same effect on secretion. (c) *Clostridium difficile* activates the protein kinases via Ca^{2+}/calmodulin (Ca^{2+}/CM). (d) *Zona occludens toxin* is the product of the *ZOT* gene, which loosens tight junctions and is required for function of the cholera toxin. It has enterotoxic activity, producing secretion. Cholera and *E. coli* cause these effects without invasion of the cell. **(B)** Colonic mucosal cell. This demonstrates one of the mechanisms by which an invasive pathogen (e.g., *Shigella*) acts. Following penetration, the pathogens generate cytotoxins, which lead to mucosal ulceration and cell death. *ATP,* Adenosine triphosphate; *G,* G protein consisting of subunits α, β, γ; *GC,* guanyl cyclase; *GMP,* guanosine monophosphate; *GTP,* guanosine triphosphate; *i,* inhibitory; *PKC,* protein kinase C; *s,* stimulatory; *VIP,* vasoactive intestinal polypeptide.

- fatty acids (in the colon) following ileal resection
- some laxatives (e.g., docusate sodium).

Inflammatory diarrhoea (mucosal destruction)

Diarrhoea occurs because of damage to the intestinal mucosal cell so that there is a loss of fluid and blood (Fig. 22.8B). In addition, there is defective absorption of fluid and electrolytes. Causes are usually infective (e.g., dysentery due to *Shigella*) or inflammatory (IBD).

Abnormal motility

Diabetic, post-vagotomy, and hyperthyroid diarrhoea are all due to abnormal motility of the upper gut. Symptoms may be exacerbated by small bowel bacterial overgrowth.

Acute diarrhoea

Acute diarrhoea is defined as lasting less than 14 days. The cause is usually infectious, most commonly viral in children but sometimes bacterial (usually more severe cases). Common infectious causes include:

- Viral infection, for example, norovirus or rotavirus
- *E. coli* food poisoning (causes a secretory diarrhoea)
- Food poisoning from *Salmonella*, *Campylobacter*, or *Staphylococcus* bacteria
- *C. difficile* infection (usually due to recent course of antibiotics)
- Contaminated food or water ('traveller's diarrhoea'), often *Giardia* or *Entamoeba histolytica.*

Clinical features associated with acute diarrhoea include fever, abdominal pain, and vomiting. Stools are usually loose and watery but can also be bloody. If the diarrhoea is severe then dehydration can occur, especially in very young or elderly patients.

The key investigation is to send a stool sample to try to isolate an organism. ***Stool microscopy, culture, and sensitivities (MCS)*** should be sent in all patients. Samples for ***C. difficile toxin*** should be sent in at-risk patients (elderly, nursing home residents, recent antibiotic use). Stool for ova, cysts, and parasites (OCP) should be sent in the following cases:

- persistent diarrhoea (>14 days <30 days)
- men who have sex with men
- immunocompromised patients
- recent travel abroad

- if suspecting giardiasis/amoebiasis/cryptosporidium.

Management of acute diarrhoea focuses on volume repletion, usually with oral rehydration solutions that contain water, salt, and sugar. In severe cases of dehydration, patients may need to be admitted to hospital for intravenous fluid resuscitation. Antibiotics are not usually empirically given unless it is a particularly severe case or complications develop (**Box 22.9** lists common antibiotic regimens). Antimotility agents (e.g., codeine phosphate or loperamide) should be avoided if possible.

Indications for antibiotic treatment include:

- severe/prolonged symptoms (>5 days)
- systemic signs of infection
- extremes of age
- immunocompromise
- presence of complications
- bloody or mucoid stools.

Chronic diarrhoea

This is defined as diarrhoea present for longer than 30 days. The causes are listed in **Box 22.10**.

A thorough history is important including the onset, timeframe, and severity of symptoms. Important points to address are any associated features such as bloody stools, abdominal pain, weight loss, steatorrhoea, recent changes in medication, food/other triggers, and a full past medical history. A full examination is needed.

Initial investigations should include:

- full blood count, serum creatinine and electrolytes, thyroid function
- coeliac screen (tissue transglutaminase [tTG] or endomysial [EMA] antibodies)
- faecal calprotectin (see 'Stool tests' section for more information)
- stool MCS, *C. difficile* toxin.

If the above do not reveal a cause then patients should undergo a flexible sigmoidoscopy (FS). In the following circumstances, however, patients should undergo a full colonoscopy:

- iron deficiency anaemia
- abnormal faecal calprotectin with a suspicion of IBD
- older patients where screening for polyps and colorectal cancer would be advantageous.

Further investigations include bowel imaging (CT abdomen, MRI small bowel, pancreatic CT), video capsule endoscopy, SeHCAT scan, and lactose hydrogen breath tests; the choice will depend on the individual patient's symptoms. Management of chronic diarrhoea involves treating the underlying cause and this will be discussed in more detail later. See **Box 22.11** for a summary of take-home messages for the patient.

Box 22.9 Common antibiotic regimens for acute diarrhoea

Infection	Antibiotic
Empirical therapy	Ciprofloxacin p.o. 500 mg twice daily for 5 days
Salmonella/Shigella	Ciprofloxacin p.o. 500 mg twice daily for 5 days
Campylobacter	Clarithromycin p.o. 500 mg twice daily for 5 days
Clostridium difficile	Metronidazole p.o. 400 mg three times daily for 10–14 days OR Vancomycin p.o. 125–250 mg four times daily for 10–14 days

Box 22.10 Causes of chronic diarrhoea

Common causes

- Irritable bowel syndrome
- Microscopic colitis
- Inflammatory bowel disease
- Colonic cancer
- Medications
- Diet
- Bile acid diarrhoea
- Pancreatic insufficiency
- Coeliac disease
- Overflow diarrhoea

Infrequent causes

- Radiation enteritis
- Lymphoma
- Diabetes
- Faecal incontinence
- Small bowel overgrowth
- Hyperthyroidism
- Small bowel resection
- Mesenteric ischaemia
- Chronic pancreatitis

Rare causes

- VIPoma
- Carcinoid
- Gastrinoma
- Whipple's disease
- Amyloidosis
- Tropical sprue
- Addison's disease

Constipation

'Constipation' is a very common symptom, particularly in women and the elderly. A consensus definition used in research (the Rome IV criteria) defines constipation as having two or more of the following for at least 3 months:

- infrequent passage of stools (<3/week)
- straining >25% of time
- passage of hard stools (Bristol stool chart form 1–2) in >25% defecations
- incomplete evacuation and sensation of anorectal blockage in >25% defecations
- manual manoeuvres to facilitate >25% defecations (digital evacuation, support of the pelvic floor).

Other symptoms include abdominal bloating and/or discomfort (undistinguishable from irritable bowel syndrome), as well as local and perianal pain. The causes of constipation are shown in **Box 22.12**.

Diagnosis

This relies on the history. When there has been a recent change in bowel habit in association with other significant symptoms (e.g., rectal bleeding), a colonoscopy or CT pneumocolon is indicated. By these means, gastrointestinal causes, such as colorectal cancer and narrowed segments due to diverticular disease, can be excluded.

Constipation can be classified into three broad categories but there is much overlap:

- normal transit through the colon (59%)
- defecatory disorders (25%)
- slow transit (13%).

Defecatory disorders with slow transit can occur together (3%).

Normal-transit constipation

In normal-transit constipation, stool traverses the colon at a normal rate, the stool frequency is normal, and yet patients believe they are constipated. This is likely to be due to perceived difficulties of evacuation or the passage of hard stools. Patients may complain of abdominal pain or bloating. Normal-transit constipation can be distinguished from slow-transit constipation by undertaking marker studies of colonic transit. Capsules containing 20 radio-opaque shapes are swallowed on days 1, 2, and 3 and an abdominal X-ray is obtained 120 hours after ingestion of the first capsule. Each capsule contains shapes of different configuration and the presence of

Box 22.11 Take-home messages for patients presenting with diarrhoea

	Acute diarrhoea	Chronic diarrhoea
Length	<14 days	>30 days
Common causes	Unknown Viral illness Bacterial if severe	Irritable bowel syndrome, inflammatory bowel disease, medications, overflow diarrhoea, diet, coeliac
Key investigations	Stool MCS OCP if risk factors	Coeliac screen Faecal calprotectin Flexible sigmoidoscopy Imaging
Management	Volume repletion Antibiotics only if severe/complications	Treat cause

MCS, Microscopy culture and sensitivities; *OCP,* ova, cysts, and parasites.

Box 22.12 Causes of constipation

General
- Inadequate fibre intake
- Immobility

Metabolic/endocrine
- Diabetes mellitus
- Hypercalcaemia
- Hypothyroidism
- Porphyria

Functional
- Irritable bowel syndrome
- Idiopathic slow transit

Drugs
- Opiates
- Antimuscarinics
- Calcium-channel blockers, for example, verapamil
- Antidepressants, for example, tricyclics
- Iron

Neurological
- Spinal cord lesions
- Parkinson's disease

Psychological
- Depression
- Anorexia nervosa
- Repressed urge to defecate

Gastrointestinal disease
- Intestinal obstruction and pseudo-obstruction
- Colonic disease, for example, carcinoma, diverticular disease
- Aganglionosis, for example, Hirschsprung's disease, Chagas' disease
- Painful anal conditions, for example, anal fissure

Defecatory disorders
- Rectal prolapse, mucosal prolapse intussusception, and solitary rectal ulcer syndrome
- Large rectocele
- Pelvic floor dyssynergia/anismus
- Megarectum

Fig. 22.9 Slow-transit constipation. Frontal abdominal X-ray taken on day 6 after ingestion of capsules, each containing 20 radio-opaque shapes, which were administered daily on days 1, 2, and 3. All the markers are retained, confirming the diagnosis of severe slow-transit constipation.

more than 4 shapes from the first capsule, 6 from the second and 12 from the third denotes moderate to severe slow transit (Fig. 22.9).

Defecatory disorders

A 'paradoxical' contraction, rather than the normal relaxation of puborectalis and the external anal sphincter and associated muscles during straining, may prevent evacuation (pelvic floor dyssynergia, anismus). These are mainly due to dysfunction of the anal sphincter and pelvic floor. An anterior rectocele is a common problem where there is a weakness of the rectovaginal septum, resulting in protuberance of the anterior wall of the rectum with trapping of stool if the diameter is more than 3 cm. In some patients, the mucosa of the anterior rectal wall prolapses downwards during straining (see 'Rectal prolapse, intussusception, and solitary rectal ulcer syndrome' section), impeding the passage of stool, while in others there may be a higher mucosal intussusception.

In some patients, the rectum can become unduly sensitive to the presence of small volumes of stool, resulting in the urge to pass frequent amounts of small-volume stool and the sensation of incomplete evacuation.

The defecatory disorders can often be characterized by performing evacuation proctography and tests of anorectal physiology.

Slow-transit constipation

Slow-transit constipation occurs predominantly in young women who have infrequent bowel movements (usually less than once a week). The condition often starts at puberty and the symptoms include an infrequent urge to defecate, bloating, abdominal pain, and discomfort. Some patients with severe slow-transit constipation have delayed emptying of the proximal colon and others a failure of 'meal-stimulated' colonic motility. Histopathological abnormalities have been demonstrated in the colons of some patients with severe slow-transit constipation, and some patients have coexisting disorders of small intestinal motility, consistent with a diagnosis of chronic idiopathic intestinal pseudo-obstruction (see 'Acute colonic pseudo-obstruction' section).

Management

Any underlying cause should be treated. In patients with normal and slow-transit constipation, the main focus should be directed towards increasing the fibre content of the diet in conjunction with increasing fluid intake.

The use of laxatives should be restricted to cases where symptoms impact on the patient's quality of life. The types of laxatives available are listed in **Box 22.13**. *Osmotic laxatives* act by increasing the colonic inflow of fluid and electrolytes; this acts not only to soften the stool but also to stimulate colonic contractility. The *polyethylene glycols (macrogols)* have the advantage over the synthetic disaccharide lactulose in that they are not fermented anaerobically in the colon to gas that can distend the colon and cause pain. The osmotic laxatives are preferred to the stimulatory laxatives, which act by stimulating colonic contractility and by causing intestinal secretion. *Prucalopride* is a high-affinity 5-HT_4 agonist that increases colonic transit and is an effective therapy for refractory constipation.

Linaclotide, a minimally absorbed peptide agonist of guanylate cyclase-C receptor, works by activating the cystic fibrosis transmembrane conductance regulator to stimulate chloride secretion and increase gastrointestinal fluid secretion. Significant benefits over placebo have been shown in clinical trials. *Lubiprostone* is an orally active agonist for type 2 chloride channels and therefore also increases gastrointestinal fluid secretion.

Box 22.13 Laxatives and enemas

Bulk-forming laxatives
- Dietary fibre
- Wheat bran
- Methylcellulose
- Mucilaginous gums—sterculia
- Mucilaginous seeds and seed coats, for example, ispaghula husk

Stimulant laxatives
(Stimulate motility and intestinal secretion)
- Phenolphthalein
- Bisacodyl
- Anthraquinones—senna and dantron (only for the terminally ill)
- Docusate sodium
- Methylnaltrexone (for opioid-induced constipation)
- Lubiprostone
- Prucalopride
- Linaclotide
- Sodium picosulfate

Osmotic laxatives
- Magnesium sulphate
- Lactulose
- Macrogols

Suppositories
- Bisacodyl
- Glycerol

Enemas
- Arachis oil
- Docusate sodium
- Hypertonic phosphate
- Sodium citrate

Patients with defecatory disorders should be referred to a specialist centre, as surgery may be indicated: for example, for anterior rectocele or internal anal mucosal intussusception. Anterior mucosal prolapse can be treated by injection, and those with pelvic floor dyssynergia (anismus) can benefit from biofeedback therapy.

INVESTIGATION OF GASTROINTESTINAL DISEASE

Routine haematology and biochemistry, followed by endoscopy and radiology, are the principal investigations.

Blood tests

All patients undergoing investigation for gastrointestinal disease should have routine blood tests. These include a full blood count, serum creatinine and electrolytes, liver function tests, and clotting. This basic screen checks for anaemia, elevated/depleted white cell or platelet counts, electrolyte disturbances, and clotting abnormalities, which can be present in several gastrointestinal diseases.

Stool tests

Stool tests are vital investigation for GI disease. Stool should always be sent for microscopy, culture, and sensitivities (MCS) in order to exclude infection in patients presenting with diarrhoea. For patients who have travelled abroad recently it can be useful to also look for stool ova, cysts, and parasites to check for infections such as giardiasis, amoebiasis, or tapeworm. **Faecal calprotectin** is a protein found in stool during intestinal inflammation (e.g., IBD or infection). It is commonly used in outpatients both diagnostically (it is 93% sensitive and 96% specific for IBD in adults) and for monitoring disease activity. Below 100 μg/g is within normal limits and a level higher than 200 μg/g is suggestive of IBD or an infective colitis. It should not be used in older patients where malignancy is suspected.

Endoscopy

Video endoscopes relay colour images to a high-definition television monitor. The tip of the endoscope can be angulated in all directions, and channels in the instrument are used for air insufflation, water injection, suction, and the passage of accessories such as biopsy forceps or brushes for obtaining tissue, snares for polypectomy and needles for injection therapies.

- ***Oesophagogastroduodenoscopy*** (***OGD***, 'gastroscopy') is the investigation of choice for upper gastrointestinal disorders with the possibility of therapy and mucosal biopsy. Findings include reflux oesophagitis, gastritis, ulcers, and cancer. Therapeutic OGD is used to treat upper gastrointestinal haemorrhage and both benign and malignant obstruction. The mortality for diagnostic endoscopy is 0.001% with significant complications in 1:10000, usually when performed as an emergency (e.g., gastrointestinal haemorrhage).
- ***Colonoscopy*** allows good visualization of the whole colon and terminal ileum. Biopsies can be obtained and polyps removed. Benign strictures can be dilated and malignant strictures stented. Cancer, polyps, and diverticular disease are the most common significant findings. Perforation occurs in 1:1000 examinations but this is higher (up to 2%) after polypectomy and endoscopic mucosal resection (**Box 22.14**).
- ***Proctoscopy*** (**Box 22.15**) is performed in all patients with a history of bright red rectal bleeding to look for anorectal pathology such

Box 22.14 Gastroscopy and colonoscopy

- Explain the procedure to the patient, including benefits and risks.
- Discuss the need for sedation.
- Obtain written informed consent.

Gastroscopy

- The patient should be fasted for at least 4 h.
- Give oxygen and monitor oxygen saturation with an oximeter.
- Give lidocaine throat spray or sedation (midazolam ± opiate if required).
- Pass the gastroscope to the duodenum under direct vision.
- Examine during insertion and withdrawal.
- Gastroscopy takes 5–15 min, depending on the indication and findings.
- Withhold fluid and food until local anaesthetic/sedation wears off.
- Complications are rare but beware of over-sedation, perforation, and aspiration.
- The patient must be accompanied home if sedation has been given.

Colonoscopy

- Stop oral iron a week before the procedure.
- Restrict the diet to low-residue foods for 48 h; clear fluids only for 24 h.
- Use a local bowel-cleansing regime, usually starting 24 h beforehand (e.g., two sachets of sodium picosulfate with magnesium citrate and 2–4 bisacodyl tablets, or macrogols 2–4 L, or local alternative; more if constipated).
- Give oxygen and monitor O_2 levels.
- Give sedation (midazolam ± opiate) if required by the patient.
- Pass the colonoscope to the caecum or ileum under direct vision.
- Examine in detail during withdrawal.
- Colonoscopy takes 15–30 min, depending on the colon anatomy, indication, and findings.
- Withhold fluid and food until sedation wears off.
- Observe the patient for at least an hour after sedation has been given.
- Complications are rare but beware of over-sedation, perforation, and aspiration.
- The patient must be accompanied home if sedation has been given.

For patients on antiplatelet therapy and anticoagulants (including antithrombotic agents), see Veitch AM, Vanbiervliet G, Gershlick AH et al. Endoscopy in patients on antiplatelet or anticoagulant therapy including direct oral anticoagulants. British Society of Gastroenterology (BSG) and European Society of Gastrointestinal Endoscopy (ESGE) guidelines. Gut *2016; 65:374–389.*

Box 22.15 Proctoscopy

- The proctoscope is passed into the anus and the obturator is removed.
- The patient strains down as the proctoscope is removed.
- Haemorrhoids are seen as purplish veins in the left lateral, right posterior, or right anterior position.
- Fissures may also be seen, but pain often prevents the procedure from being performed.

as haemorrhoids; a rigid sigmoidoscope is too narrow and long to enable adequate examination of the anal canal.

- ***Sigmoidoscopy***, either flexible or rigid, is part of the routine hospital examination in cases of diarrhoea and in patients with lower abdominal symptoms such as a change in bowel habit or rectal bleeding. ***Rigid sigmoidoscopy*** can visualize the distal 20 to 25 cm of large bowel, whereas ***flexible sigmoidoscopy*** can reach up to the splenic flexure (60 cm), and is typically performed in the endoscopy unit after evacuation of the distal colon using an enema or suppository.
- ***Balloon enteroscopy***, either double- or single-balloon, can examine the small bowel from the duodenum to the ileum, adopting both cranial and caudal approaches, and using specialized enteroscopes in expert centres.
- ***Capsule endoscopy*** is used for the evaluation of obscure gastrointestinal bleeding (after negative gastroscopy and colonoscopy) and for the detection of small bowel tumours and occult IBD. It should be avoided if strictures are suspected.
- ***pH studies, impedance monitoring, and manometry*** are all discussed in more detail in the upper gastrointestinal disease section.

Imaging

Plain X-rays

Plain X-rays of the chest and abdomen are chiefly used in the investigation of an acute abdomen. Interpretation depends on analysis of gas shadows inside and outside the bowel. Plain films are particularly useful where obstruction or perforation is suspected, to exclude toxic megacolon in colitis, and to assess faecal loading in constipation. Calcification may be seen with gall bladder stones and in chronic pancreatitis, though computed tomography is more sensitive for both.

Ultrasound

Ultrasonography involves no radiation and is the first-line investigation for abdominal distension: for example, in ascites, a mass or suspected inflammatory conditions. It can show dilated, fluid-filled loops of bowel in obstruction, and thickening of the bowel wall. It can be used to guide biopsies or percutaneous drainage. In an acute abdomen, ultrasound can diagnose cholecystitis, appendicitis, enlarged mesenteric glands, and other inflammatory conditions.

Computed tomography

Computed tomography (CT) involves a significant dose of radiation (approximately 10 millisieverts). Modern multislice fast scanners and techniques, involving intraluminal and intravenous contrast, enhance diagnostic capability. Intraluminal contrast may be positive (Gastrografin or Omnipaque) or negative (usually water). The bowel wall and mesentery are well seen after intravenous contrast, especially with negative intraluminal contrast. Clinically unsuspected diseases of other abdominal organs are quite often also revealed (Fig. 22.10A).

CT is widely used as a first-line investigation for the acute abdomen. It is sensitive for small volumes of gas from a perforated viscus, as well as leakage of contrast from the gut lumen. Inflammatory conditions, such as abscesses, appendicitis, diverticulitis, Crohn's disease and its complications, are well demonstrated. In high-grade bowel obstruction, CT is usually diagnostic of both the presence and the cause of the obstruction.

CT is widely used in cancer staging and as guidance for biopsy of tumour or lymph nodes.

CT pneumocolon/CT colonography (virtual colonoscopy) after CO_2 insufflation into a previously cleansed colon provides an alternative to colonoscopy for diagnosis of colonic mass lesions (see Fig. 22.10B). It is being evaluated as a screening test for colon pathology with sensitivities of over 90% for polyps larger than 10 mm.

Magnetic resonance imaging

Magnetic resonance imaging (MRI) uses no radiation and is particularly useful in the evaluation of rectal cancers and of abscesses and fistulae in the perianal region. It is also useful in small bowel disease (MR enteroclysis) and in hepatobiliary and pancreatic disease.

Positron emission tomography

Positron emission tomography (PET) relies on detection of the metabolism of fluorodeoxyglucose. It is used for staging oesophageal, gastric, and colorectal cancer and in detecting metastatic and recurrent disease. PET/CT adds additional anatomical information.

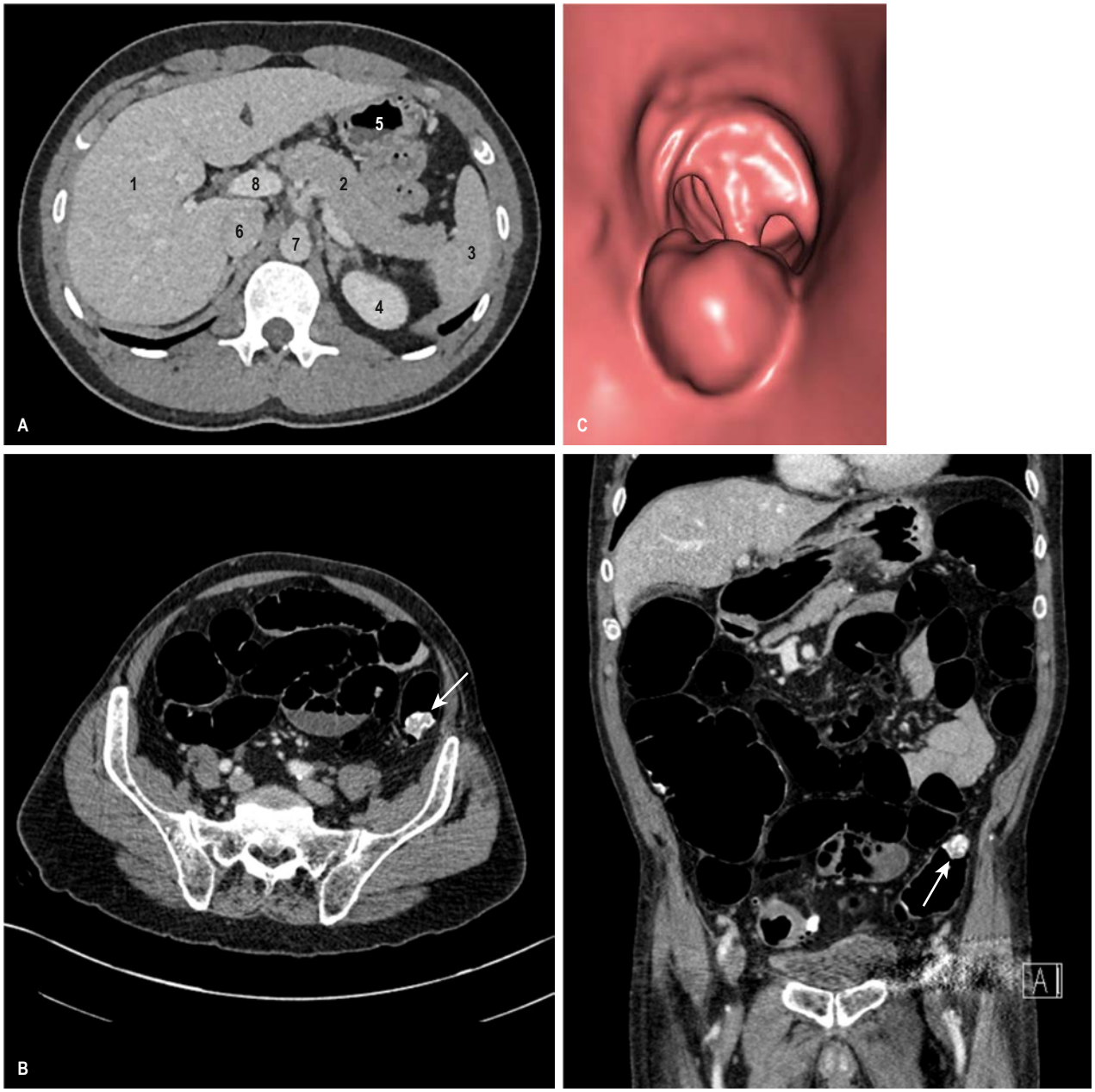


Fig. 22.10 Computed tomography in gastrointestinal disease. (A) Annotated axial CT image of the abdomen. (1) Liver; (2) pancreas; (3) spleen; (4) left kidney; (5) stomach; (6) inferior vena cava; (7) aorta; (8) portal vein. **(B)** Axial and coronal CT colonography images demonstrate a stalked polyp in the distal descending colon *(arrows)*. **(C)** 3D view of the polyp demonstrated in (B).

Contrast studies

These include barium swallow examination, double-contrast barium meal, and small bowel follow-through or meal. ***Absorbable water-soluble (Gastrografin or Omnipaque) contrast agents*** should be used in preference to barium when perforation is suspected anywhere in the gut. These are useful examinations for the upper GI tract and small bowel but do rely on good technique. They are gradually being superseded by other modalities of imaging, particularly MRI.

Radioisotopes

Radionuclides are used to a varying degree, depending on availability and expertise. Some of the more common indications are measuring the rate of gastric emptying, demonstration of a Meckel's diverticulum, evaluation of neuroendocrine tumours, assessment of bile salt malabsorption (SeHCAT scan), or detection of bacterial overgrowth. These will be explored in more detail in the relevant sections.

FUNCTIONAL GASTROINTESTINAL DISORDERS

There is a large group of gastrointestinal disorders that are termed 'functional' because symptoms occur in the absence of any demonstrable abnormalities in the digestion and absorption of nutrients,

fluid, and electrolytes, and no structural abnormality can be identified in the gastrointestinal tract, although there may be discernible abnormalities in neuromuscular function, such as dysmotility and visceral hypersensitivity, which are not routinely investigated.

Pathophysiology and brain–gut interactions

People with functional gastrointestinal disorders (FGIDs) are characterized by having a greater gastrointestinal motility response to life events than normal subjects. There is, however, a poor association between measured gastrointestinal motility changes and symptoms in many of the FGIDs. Patients with FGID have been shown to have abnormalities in visceral sensation and to have a lower pain threshold when tested with balloon distension (visceral hyperalgesia). Visceral hypersensitivity possibly relates to:

- altered receptor sensitivity at the viscus itself
- increased excitability of the spinal cord dorsal horn neurons
- altered central modulation of sensations.

A systematic review of published studies suggests that 10% of patients who experience an acute infective gastroenteritis develop a degree of FGID. It is not clear whether this is caused by post-infectious bile salt malabsorption, alterations in the mucosal immune system, or the use of antibiotics to treat the index infection.

The brain–gut axis describes a combination of intestinal motor, sensory, and activities (Fig. 22.11). Thus, extrinsic (e.g., vision, smell) and intrinsic (e.g., emotion, thought) information can affect gastrointestinal sensation because of the neural connections from higher centres. Conversely, viscerotropic events can affect central pain perception, mood, and behaviour.

Psychological stress can exacerbate gastrointestinal symptoms, and psychological disturbances are more common in patients with FGIDs. These disturbances alter attitude to illness, promote healthcare seeking, and often lead to a poor clinical outcome. They have psychosocial consequences with poor quality of life at home and work. Early in life, genetic and environmental influences (e.g., family attitudes towards bowel training, verbal or sexual abuse, exposure to an infection) may affect psychosocial development (susceptibility to life stress, psychological state, coping skills, development of social support) or the development of gut dysfunction (abnormal motility or visceral hypersensitivity). Therefore, FGID should be regarded as a dysregulation of brain–gut function and is often referred to as disorder of gut–brain interaction (DGBI).

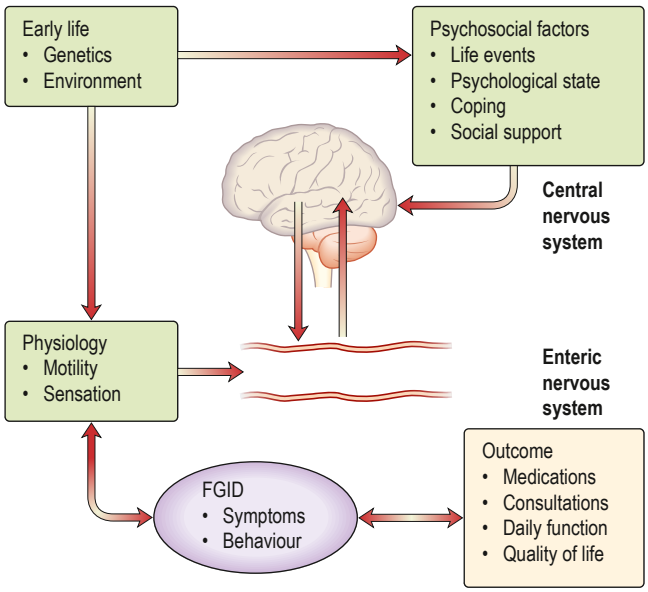


Fig. 22.11 A biopsychosocial conceptualization of the pathogenesis and clinical expression of functional gastrointestinal disorders (FGIDs). Genetic, environmental, and psychological factors may interact to cause dysregulation of brain–gut function. *(Modified from Drossman DA. The functional gastrointestinal disorders and the Rome III process.* Gastroenterology *2006; 130:1377–1390, with permission from Elsevier.)*

MOUTH

Recurrent aphthous ulceration

Idiopathic aphthous ulceration is common and affects up to 25% of the population. Recurrent painful round or ovoid mouth ulcers are seen with inflammatory halos. They are more common in females and non-smokers, usually appear first in childhood, and tend to reduce in number and frequency before the age of 40. Other family members may be affected. There is no sign of systemic disease.

- ***Minor aphthous ulcers*** are the most common; they are less than 10 mm diameter, have a grey/white centre with a thin erythematous halo, and heal within 14 days without scarring. They rarely affect the dorsum of the tongue or hard palate.
- ***Major aphthous ulcers*** (>10 mm diameter) often persist for weeks or months and heal with scarring.

The cause is not known. Deficiencies of iron, folic acid, or vitamin B_{12} (with or without gastrointestinal disorders) are sometimes found but are not causally linked. Secondary causes, such as Crohn's disease, should be excluded (Box 22.16).

There are no specific effective therapies. Sufferers should avoid oral trauma, and acidic foods or drinks that cause pain. Topical (1% triamcinolone) or systemic corticosteroids may lessen the duration and severity of the attacks. Chlorhexidine gluconate or tetracycline mouthwash, dapsone, colchicine, thalidomide, and azathioprine have all been used with variable effect.

? Box 22.16 Causes of mouth ulcers

- Idiopathic aphthous ulceration (most common)
- Gastrointestinal disease:
  - Inflammatory bowel disease
  - Coeliac disease
- Infection:
  - Viral—herpes simplex virus (HSV), human immunodeficiency virus (HIV), Coxsackie
  - Fungal—candidiasis
  - Bacterial—syphilis, tuberculosis
- Systemic disease:
  - Reactive arthritis (see Chapter 32)
  - Behçet's syndrome
  - Systemic lupus erythematosus
- Trauma:
  - For example, dentures
- Neoplasia:
  - For example, squamous cell carcinoma
- Drugs:
  - In erythema multiforme major, toxic epidermal necrolysis
  - Chemotherapy, antimalarials
- Skin disease:
  - Pemphigoid
  - Pemphigus
  - Lichen planus

Neoplasia (squamous cell carcinoma)

Malignant tumours of the mouth account for 1% of all malignant tumours in the UK. The majority develop on the floor of the mouth or lateral borders of the tongue. Early lesions may be painless, but advanced tumours are easily recognizable as hard, indurated ulcers with raised and rolled edges. Aetiological agents include tobacco, heavy alcohol consumption, and the areca nut. Human papillomavirus 16 causes some oral cancers. Pre-malignant lesions include leucoplakia (single adherent white patch), lichen planus, submucous fibrosis, and erythroplakia (a red patch).

Management

Treatment is by surgical excision, which may require extensive neck dissection to remove involved lymph nodes and/or radiotherapy. Ablative treatment with photodynamic therapy is being pioneered for early lesions.

Pigmented neoplastic lesions

These include melanotic naevi on the hard palate and buccal mucosa. These are rarer in the mouth than on the skin. Malignant melanomas are rare, but more common in males, and occur mainly on the upper jaw. The 5-year survival is only 5%.

Non-neoplastic lesions

Oral white patches

Transient white patches either are due to *Candida* infection or are very occasionally found in systemic lupus erythematosus. Local causes include mechanical, irritative, or chemical trauma from drugs (e.g., ill-fitting dentures or aspirin). Oral candidiasis in adults is seen following therapy with broad-spectrum antibiotics or inhaled steroids, and in people with diabetes, patients who are seriously ill or those who are immunocompromised.

Persistent white patches can be due to ***leucoplakia***, which is associated with alcohol and (particularly) smoking; it is pre-malignant. A biopsy should always be taken; histology shows alteration in the keratinization and dysplasia of the epithelium. Treatment is unsatisfactory. Isotretinoin possibly reduces disease progression. ***Oral lichen planus*** presents as white striae, which can rarely extend into the oesophagus.

Oral pigmented lesions

Racial pigmentation is scattered and symmetrically distributed. Amalgam tattoo is the most common form of localized oral pigmentation and consists of blue–black macules involving the gingivae; it results from dental amalgam sequestering into the tissues. Diseases causing pigmentation include Peutz–Jeghers syndrome and Addison's disease. Heavy metals (e.g., lead, bismuth, and mercury) and drugs (e.g., phenothiazines and antimalarials) all cause pigmentation of the gums.

Tongue

The tongue may be affected by inflammatory or malignant processes, with similar lesions to those described above.

Glossitis

Glossitis is a red, smooth, sore tongue associated with vitamin B_{12}, folate, iron, riboflavin, and nicotinic acid deficiency. It is also seen in infections due to *Candida*.

Geographic tongue

A geographic tongue is an idiopathic condition occurring in 1% to 2% of the population and may be familial. There are erythematous areas surrounded by well-defined, slightly raised irregular margins (Fig. 22.12). The lesions are usually painless and the patient should be reassured.

Gums

The gums (gingivae) are the mucous membranes covering the alveolar processes of the mandible and the maxilla. Diseases of the gum are shown in Box 22.17.

Teeth

Dental caries occur as a result of bacterial damage to tooth structures leading to tooth decay and 'cavities'. The main cause in humans is *Streptococcus mutans*, which is cariogenic only in the presence of dietary sugar. Dental caries can progress to pulpitis and pulp necrosis, and spreading infection can cause dentoalveolar abscesses. If there is soft tissue swelling, antibiotics (e.g., amoxicillin or metronidazole) should be prescribed prior to dental intervention.

Erosion of the teeth can result from exposure to acid (e.g., in bulimia nervosa) or, very occasionally, from severe gastro-oesophageal reflux disease.

SALIVARY GLANDS

Diseases of the salivary glands are shown in Box 22.18.

PHARYNX AND OESOPHAGUS

Gastro-oesophageal reflux disease

Epidemiology and pathophysiology

Gastro-oesophageal reflux disease (GORD) is common with a prevalence of 10% to 20% in the developed world. It is more common in the Western world than in Asia. There are a number of lifestyle risk factors as well as possible genetic factors, and the cardinal clinical feature is heartburn. Psychosocial factors are often determinants of symptom severity, with heartburn increased during times of stress. Factors associated with GORD are shown in Box 22.19.

Fig. 22.12 Geographic tongue. A florid example, demonstrating well-delineated areas of erythema partially surrounded by white serpiginous borders. *(Courtesy, Carl M. Allen, MD, and Charles Camisa, MD. From* Dermatology Essentials, *Second Edition.)*

Box 22.17 Gum diseases

	Cause	Management
Chronic gingivitis	Bacterial plaque	Plaque removal
Acute (necrotizing) ulcerative gingivitis (Vincent's angina)	Spirochaete and fusiform bacteria (poor hygiene and smoking)	Oral metronidazole 400 mg 3 times a day for 3 days Chlorhexidine mouthwash
Desquamative gingivitis (smooth red atrophic gingivae)	Lichen planus or mucous membrane pemphigoid	Confirmation with biopsy
Gingival swelling	Inflammation (e.g., pregnancy, scurvy), infiltrations (leukaemia), or fibrous hyperplasia (e.g., drug-induced—phenytoin, ciclosporin, nifedipine)	

Box 22.18 Diseases of the salivary glands

- Dry mouth (xerostomia):
  - Sjögren's syndrome
  - Drugs (e.g., antimuscarinic, anti-parkinsonian, lithium, tricyclics, monoamine oxidase inhibitors (MAOIs))
  - Radiotherapy
  - Psychogenic
  - Dehydration
- Acute sialadenitis:
  - Viral (e.g., mumps)
  - Bacterial (e.g., *Staphylococcus*, *Streptococcus*)
- Salivary duct obstruction:
  - Calculus
- Sarcoidosis (see Chapter 20)
- IgG4 disease (see Chapter 6)
- Neoplasms (3% of all tumours worldwide):
  - Pleomorphic adenoma (15% become malignant and can cause VII lower motor neuron signs)

Box 22.19 Factors associated with gastro-oesophageal reflux

- Pregnancy or obesity
- Fat, chocolate, coffee, or alcohol ingestion
- Large meals
- Cigarette smoking
- Drugs—antimuscarinic, calcium-channel blockers, nitrates
- Systemic sclerosis
- Treatment for achalasia
- Hiatus hernia

The underlying cause of GORD is reflux of gastric acid, pepsin, bile, and duodenal contents back into the oesophagus. This can be influenced by many factors that overcome the innate defence mechanisms, primarily the LOS. Between swallows, the muscles of the oesophagus are relaxed, except for those of the two sphincters. The LOS in the distal oesophagus remains closed because of the unique property of the muscle, and relaxes when swallowing is initiated (see 'Swallowing' section). ***Transient lower (o)esophageal sphincter relaxations*** (TLESRs) are part of normal physiology, but occur more frequently in patients with GORD, allowing gastric acid to flow back into the oesophagus (Fig. 22.13). Increased abdominal pressure (pregnancy) and low LOS pressure also predispose to GORD.

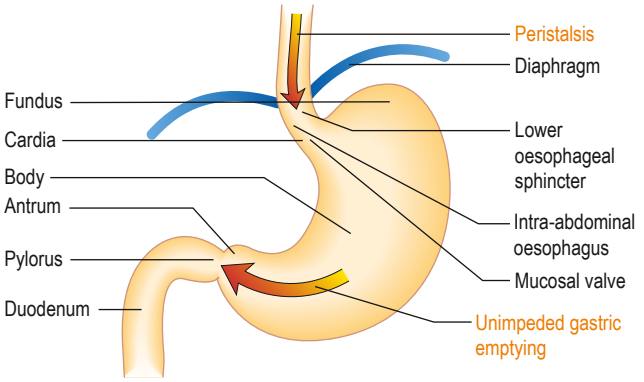


Fig. 22.13 The main anti-reflux mechanisms. These are shown on the right.

Box 22.20 Hiatus hernia

Sliding hiatus hernia

The oesophageal–gastric junction and part of the stomach 'slide' through the hiatus so that it lies above the diaphragm.

- Present in 30% of people over 50 years
- Produces no symptoms—any symptoms are due to reflux

Rolling or para-oesophageal hernia

Part of the fundus of the stomach prolapses through the hiatus alongside the oesophagus.

- The lower oesophageal sphincter remains below the diaphragm and remains competent
- Occasionally, severe pain occurs due to volvulus or strangulation

Other anti-reflux mechanisms involve the intra-abdominal segment of the oesophagus, which acts as a flap valve. In addition, the mucosal rosette, formed by folds of the gastric mucosa, and the crural diaphragm at the LOS, which contracts and acts like a pinchcock, prevent acid reflux. A hiatus hernia can impair this mechanism (Box 22.20). Normally, the oesophagus is also 'cleared' rapidly of refluxate by secondary peristalsis, gravity, and salivary bicarbonate.

The clinical features of reflux occur when the anti-reflux mechanisms fail, allowing acidic gastric contents to make prolonged contact with the lower oesophageal mucosa. The sphincter relaxes transiently, independently of a swallow, after meals, and this is the cause of almost all reflux in normal subjects and in about two-thirds of GORD patients.

The role of the ***acid pocket*** in GORD has gained more credence in the last 10 years. Food and fluid generally increase the gastric pH, but the acid pocket is an area of unbuffered gastric acid that accumulates postprandially in the proximal stomach. It serves as a reservoir for acid reflux, particularly in individuals with a hiatus hernia where the distal oesophagus is supradiaphragmatic, especially on lying flat. This can be specifically targeted with an antacid–alginate combination.

Clinical features

Heartburn is the major feature. This is a burning chest pain that is aggravated by bending, stooping, and lying down, all of which promote acid exposure. The patient typically complains of pain on eating spicy food or drinking hot liquids or alcohol.

Box 22.21 Classic features of the pain of gastro-oesophageal reflux and cardiac ischaemia

Reflux pain: burning, worse on bending, stooping, or lying down
- Seldom radiates to the arms
- Worse with spicy food, hot drinks, or alcohol
- Relieved by antacids

Cardiac ischaemic pain
- Gripping or crushing
- Radiates to neck or left arm
- Worse with exercise
- Accompanied by dyspnoea

The correlation between heartburn and oesophagitis is poor. Some patients have mild oesophagitis but severe heartburn, while others have severe oesophagitis in the absence of symptoms, and can present with haematemesis or iron deficiency anaemia from chronic blood loss.

Differentiation of cardiac and oesophageal pain can be difficult; 20% of cases seen in emergency departments have GORD (Box 22.21). A trial of antacid therapy is often useful, resolving reflux-induced pain, before progressing to 24-hour pH studies only if symptoms persist.

Regurgitation of food and acid into the mouth occurs, particularly on bending or lying flat. This can lead to excess salivation in the mouth, commonly known as water-brash. Aspiration pneumonia is unusual without an accompanying obstruction, but cough and asthma can occur and respond slowly (1–4 months) to a proton pump inhibitor (PPI).

Laryngopharyngeal reflux disease (LPRD) is the transport of gastric contents into the larynx and pharynx, usually, although not exclusively, in the context of GORD. Heartburn may be one of the associated symptoms but cough, hoarse voice, postnasal drip, and asthma are more frequently seen, but are not common.

Diagnosis and investigations

The ***clinical*** diagnosis can usually be made without investigation. Unless there are alarm symptoms, especially dysphagia (see 'Dysphagia' section), patients under the age of 45 years can safely be treated initially without investigations. If investigation is required, there are two aims:

- ***Assess oesophagitis and hiatal hernia by endoscopy.*** If there is oesophagitis (Fig. 22.14) or Barrett's oesophagus (see 'Barrett's oesophagus' section), reflux is confirmed.
- ***Intraluminal monitoring*** (Fig. 22.15). Intraluminal pH monitoring or impedance combined with manometry is helpful if there is no response to PPIs and should always be performed to confirm reflux before surgery. Ambulatory 24-hour pH monitoring consists of passing a nasogastric tube with a pH probe at its tip which is placed at the level of the LOS. This allows measurements of acid exposure and is worn for 24 hours. Some centres now offer a wireless probe (known as a Bravo capsule) that is clipped to the lower oesophagus without the need for an NG tube. This then gradually detaches and is passed in the stool. A composite score of the various parameters measured is then calculated, with a score higher than 14.72 indicating reflux. This is known as a DeMeester score. There should also be a good correlation between reflux (pH < 4.0) and symptoms. Impedance testing is when a catheter is used to measure the resistance to flow of 'alternating current' in the contents of the oesophagus. It can also be helpful to assess oesophageal dysmotility as a potential cause of the symptoms, leading to impaired oesophageal clearance.

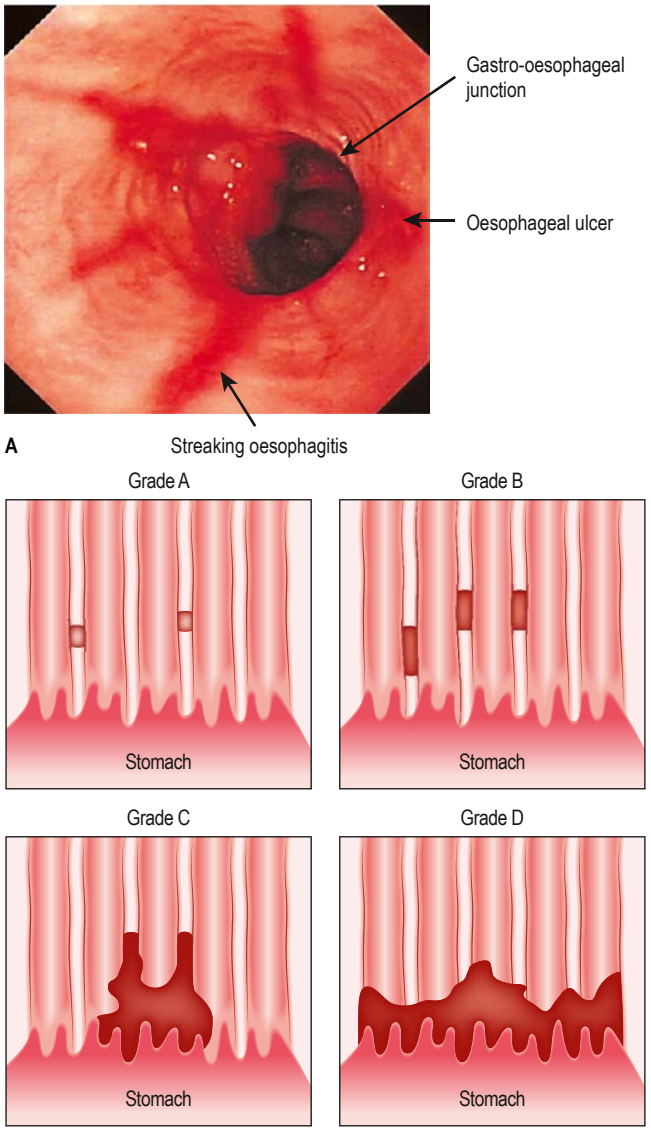


Fig. 22.14 Oesophagitis. (A) Oesophagitis seen on endoscopy. (B) Los Angeles classification of oesophagitis. **Grade A** = mucosal breaks confined to the mucosal folds, each no longer than 5 mm. **Grade B** = at least one mucosal break longer than 5 mm confined to the mucosal fold but not continuous between two folds. **Grade C** = mucosal breaks that are continuous between the tops of mucosal folds but not circumferential. **Grade D** = extensive mucosal breaks engaging at least 75% of the oesophageal circumference.

Management

Approximately half of patients with reflux symptoms in primary care can be treated successfully with simple antacids, loss of weight, and raising the head of the bed at night. Precipitating factors should be avoided, with dietary measures, reduction in alcohol and caffeine consumption, and cessation of smoking. These measures are simple to say but difficult to carry out, though they are useful in mild disease in compliant patients.

Drugs

Alginate-containing antacids

Alginate-containing antacids (10 mL three times daily) are the most frequently used 'over-the-counter' agents for GORD. They form a

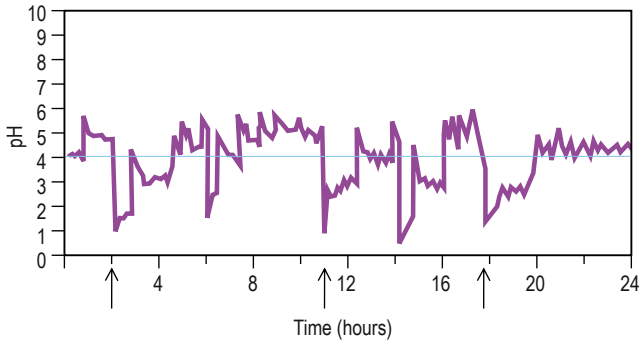


Fig. 22.15 Twenty-four-hour intraluminal pH monitoring. Five reflux episodes (pH < 4) occurred but only three produced symptoms *(arrowed)*.

gel or 'foam raft' with gastric contents to reduce reflux. Magnesium-containing antacids tend to cause diarrhoea while aluminium-containing compounds may cause constipation.

Proton pump inhibitors

Proton pump inhibitors (PPIs; omeprazole, rabeprazole, lansoprazole, pantoprazole, esomeprazole) inhibit gastric hydrogen/potassium-adenosine triphosphatase. PPIs reduce gastric acid secretion by up to 90% and are the drugs of choice for all but mild cases. Most patients with GORD will respond well, with approximately 60% symptom-free after 4 weeks of a once-daily PPI. Patients with severe symptoms may need twice-daily PPIs and prolonged treatment, often for years. Once oesophageal sensitivity has normalized, a lower dose, for example, omeprazole 20 mg, may be sufficient for maintenance. Long-term PPI prescription is not uncommon, and although some recent data have suggested side-effects such as osteoporosis, and an increase in gastrointestinal infections such as *C. difficile*, these are uncommon and tend to occur in at-risk patients.

Patients who do not respond to a PPI and have continuing symptoms with a normal endoscopy are described as having ***non-erosive reflux disease (NERD)*** (Fig. 22.16). These patients are usually female and often the symptoms are functional, although a small group have a 'hypersensitive' oesophagus, giving discomfort with only slight changes in pH. Isomers of some of the original PPIs (e.g., dexlansoprazole) have the benefit of more effective gastric acid inhibition over a longer time period, as their metabolism of the active metabolite is slower.

H_2-receptor antagonists

H_2-receptor antagonists (e.g., cimetidine, ranitidine, famotidine, and nizatidine) are frequently used for acid suppression if antacids fail, as they can easily be obtained. They can be used in conjunction with PPIs for patients with more severe GORD.

Dopamine antagonist prokinetic agents

The dopamine antagonist prokinetic agents metoclopramide and domperidone are occasionally helpful, as they enhance peristalsis and speed gastric emptying, but there are few data to substantiate this. The role of domperidone has been limited still further following reports of serious cardiac side-effects.

Endoluminal gastroplication

In this endoscopic procedure, multiple plications or pleats are made below the gastro-oesophageal junction. Randomized controlled trials have shown benefit with reduction in heartburn, acid reflux episodes and PPI usage, but not sustained improvements in the oesophageal pH measurements.

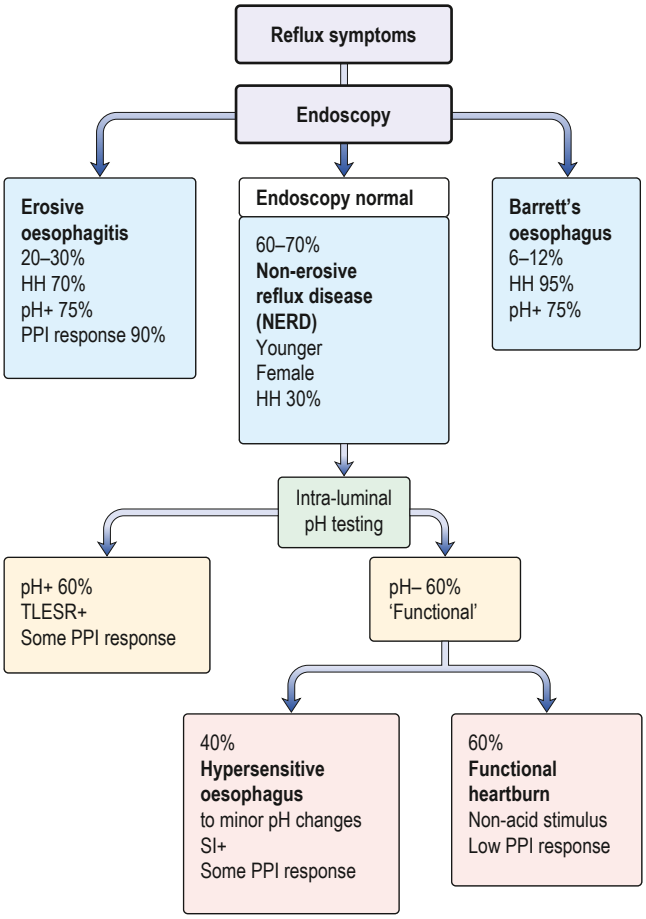


Fig. 22.16 Outcome of patients with reflux symptoms. *HH,* Hiatus hernia; *pH⁺,* meets standard criteria for excessive acid reflux on intraluminal testing; *pH⁻,* does not meet standard criteria for excessive acid reflux; *PPI,* proton pump inhibitor; *SI+,* symptom index (proportion of pH drops resulting in symptoms, >0.5 is significant); *TLESR,* transient lower (o)esophageal sphincter relaxation.

Surgery

Surgery should never be performed for a hiatus hernia alone. The best predictors of a good surgical result are typical reflux symptoms with documented acid reflux, which correlates with symptoms and response to a PPI. With such highly selected cases in experienced hands, the laparoscopic Nissen fundoplication has a satisfaction rate of over 90% at 5 years, and available 10-year data show satisfaction rates remain high at 88%. Current surgical techniques return the oesophagogastric junction to the abdominal cavity, mobilize the gastric fundus, close the diaphragmatic crura snugly, and involve a short, tension-free fundoplication.

The ***Linx reflux management system*** consists of a row of magnets that increase LOS closure pressure, allowing food passage during swallowing; it is inserted laparoscopically. Improvement in quality of life with no PPI therapy has been shown but further studies are necessary. Patients cannot have an MRI study with the device *in situ*.

Indications for operation are not clear-cut but include intolerance to medication, the desire for freedom from medication, the expense of therapy, and concern about long-term side-effects.

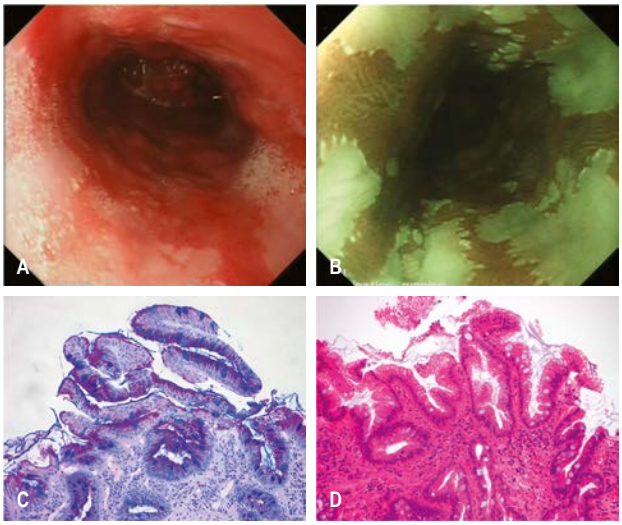


Fig. 22.17 Barrett's oesophagus. **(A)** Endoscopic picture of Barrett's oesophagus. **(B)** Endoscopic narrow-band imaging of Barrett's mucosa. **(C)** Alcian blue staining of Barrett's mucosa highlighting the blue-staining goblet cells. **(D)** Haematoxylin and eosin staining of Barrett's columnar mucosa.

The most common cause of mechanical fundoplication failure is recurrent hiatus hernia.

Patients with oesophageal dysmotility unrelated to acid reflux, those with no response to PPIs and those with underlying functional bowel disease should rarely have surgery.

Complications of gastro-oesophageal reflux disease

Peptic stricture

Since the advent of PPIs, peptic strictures have become far less common. They usually occur in patients over the age of 60 and present with intermittent dysphagia for solids, which worsens gradually over a long period. Mild cases may respond to PPIs alone. More severe cases need endoscopic dilation and long-term PPI therapy. Surgery is required if medical treatment fails.

Barrett's oesophagus

Barrett's oesophagus (Fig. 22.17) is a condition in which part of the normal oesophageal squamous epithelium is replaced by metaplastic columnar mucosa to form a segment of 'columnar-lined oesophagus' (CLO). It is a complication of GORD and there is almost always a hiatus hernia present.

Diagnosis and classification

The diagnosis is made if endoscopy shows proximal displacement of the squamocolumnar mucosal junction and biopsy demonstrates columnar lining above the proximal gastric folds (>1 cm); intestinal metaplasia is no longer a requirement of the British Society of Gastroenterology definition but is central to the American College of Gastroenterology guidelines. Barrett's oesophagus may be seen as a continual circumferential sheet, as finger-like projections extending upwards from the squamocolumnar junction or as islands of columnar mucosa interspersed with areas of squamous mucosa. The Prague classification (Fig. 22.18) is used for recording the endoscopic distribution, stating both the length of circumferential CLO (C measurement) and the maximum length (M measurement), the distance from the top of the gastric folds to the most proximal tongue of the columnar mucosa.

Central obesity increases the risk of Barrett's by 4.3 times. Long-segment (>3 cm) and short-segment (<3 cm) Barrett's is found, respectively, in 5% and 15% of patients undergoing endoscopy for reflux symptoms. Barrett's is also often found incidentally in endoscoped patients without reflux symptoms. It is most common in middle-aged obese males. The major concern is that approximately 0.12% to 0.5% of Barrett's patients develop oesophageal adenocarcinoma per year; the majority, probably, through a gradual transformation from intestinal metaplasia to low-grade then high-grade dysplasia, before invasive adenocarcinoma. Early studies suggested that Barrett's increases the chance of developing oesophageal adenocarcinoma 30- to 50-fold but recent studies have shown the risk to be much lower and closer to a 1% lifetime risk in the typical patient.

Screening for Barrett's oesophagus is not routinely recommended because of the poor correlation between symptoms and the presence of Barrett's. Once detected however, ongoing surveillance is carried out at regular intervals. The interval between endoscopies is determined by the length of the CLO and the degree of cellular disturbance within it (Fig. 22.19).

Surveillance involves inspection of the oesophageal mucosa with a high-definition gastroscope, and the taking of targeted biopsies of any focally abnormal tissue in addition to random biopsies from all four quadrants (every 1–2 cm) of the CLO. High-grade dysplasia (HGD) is usually associated with endoscopically visible nodules or ulceration, optimally visualized with a high-definition endoscope. Chromo-endoscopy (the topical application of stains or pigments via the endoscope), narrow band and autofluorescence imaging may aid the diagnosis of dysplasia and carcinoma. Endoscopic technology has improved the detection of pre-malignant lesions, enabling their removal by either endoscopic mucosal resection (EMR) or endoscopic submucosal dissection, therefore preventing surgical oesophagectomy.

If ***low-grade dysplasia*** is found on endoscopic surveillance, a repeat endoscopy with quadrantic biopsies every 1 cm is usually performed within 6 months, while on high-dose proton pump inhibition. Recent guidelines now suggest that if low-grade dysplasia persists then these patients should be offered endoscopic ablation therapy or 6-monthly surveillance.

If ***high-grade dysplasia*** is found, it is usually in the context of an endoscopically visible lesion, which, if nodular, is removed by endoscopic mucosal resection for more accurate histological staging. If high-grade dysplasia is detected in the absence of any endoscopically visible lesion, high-dose proton pump inhibition is started and repeat biopsies taken within 3 months. Endoscopic ultrasound is frequently used to stage this patient group more accurately, in order to exclude cancer and associated significant lymphadenopathy.

Radiofrequency ablation (RFA) has superseded photodynamic therapy as the technique of choice for endoscopic treatment of dysplasia within Barrett's segments following removal of any nodular lesions, returning the oesophagus to squamous lining. The benefit of RFA in low-grade dysplasia is currently under evaluation.

Motility disorders

Achalasia

Achalasia is characterized by oesophageal aperistalsis and impaired relaxation of the LOS. The lower oesophageal pressure is elevated in more than half of patients.

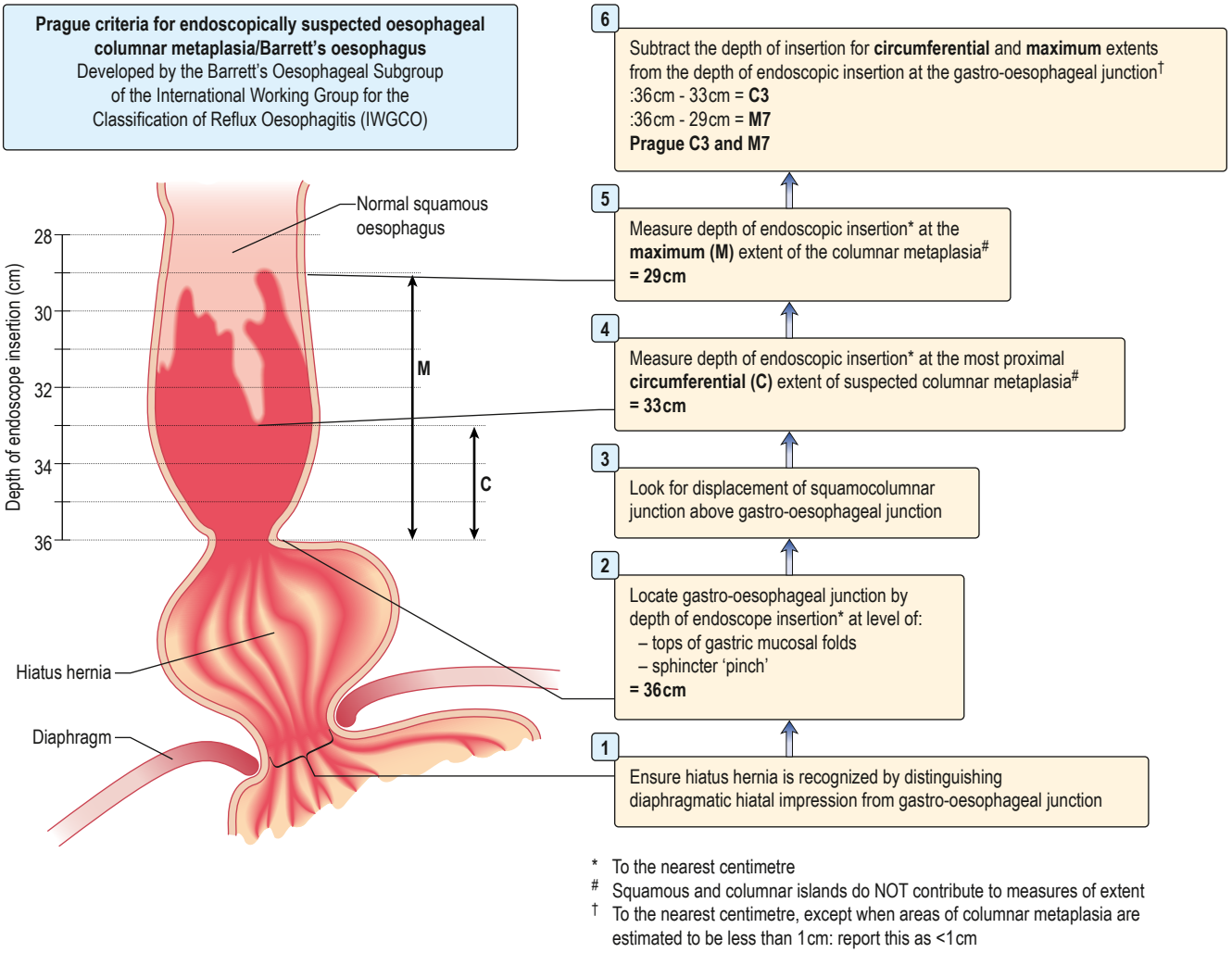


Fig. 22.18 The Prague criteria for endoscopically suspected oesophageal columnar metaplasia/Barrett's oesophagus.

Clinical features

The incidence of achalasia is 1:100000 and is equal in males and females. It occurs at all ages but is rare in childhood. Patients usually have a long history of intermittent dysphagia, characteristically for both liquids and solids from the onset. Regurgitation of food from the dilated oesophagus occurs, particularly at night, and aspiration pneumonia is a complication. Spontaneous chest pain occurs and is said to be due to oesophageal 'spasm'; it may be misdiagnosed as cardiac pain. Dysphagia may be mild and accepted by the patient as normal; weight loss is usually minimal.

Pathogenesis

The aetiology of achalasia is unknown, with autoimmune, viral, and neurodegenerative aetiologies all being postulated. A similar clinical picture is seen in chronic Chagas' disease (American trypanosomiasis; see Chapter 42), where there is damage to the neural plexus of the gut.

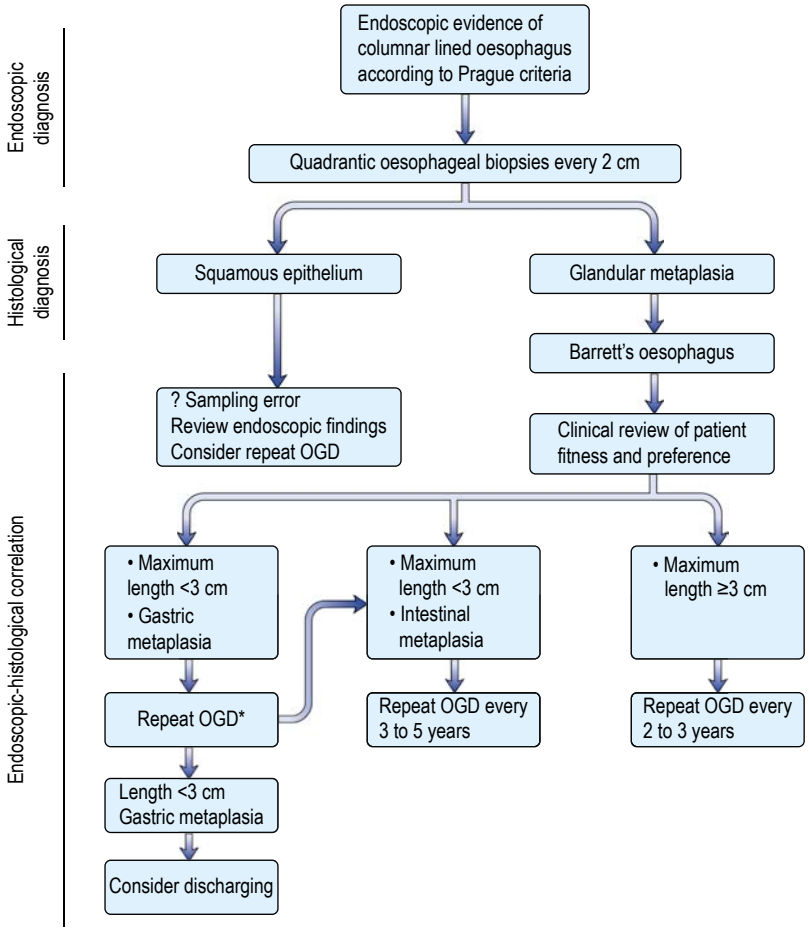
Histopathology shows inflammation of the myenteric plexus of the oesophagus with reduction of ganglion cell numbers. Cholinergic innervation appears to be preserved. Reduction in nitric oxide synthase-containing neurons has been shown by immunohistochemical staining. Pharmacological studies in patients with achalasia support the selective loss of ***inhibitory***, nitrergic neurons.

Differential diagnosis

The differential diagnosis of achalasia worldwide includes genetic syndromes, infectious diseases, neoplasms, and chronic inflammatory conditions.

Investigations

- ***Chest X-ray*** shows a dilated oesophagus, sometimes with a fluid level seen behind the heart. The fundal gas shadow is absent.
- ***Barium swallow*** shows lack of peristalsis and often synchronous contractions in the body of the oesophagus, sometimes with dilation. The lower end shows a 'bird's beak' due to failure of the sphincter to relax (Fig. 22.20).
- ***Oesophagogastroduodenoscopy*** is performed to exclude a carcinoma at the lower end of the oesophagus, as this can produce a similar X-ray appearance. When there is marked dilation, a 24-h liquid-only diet and a washout, prior to endoscopy, are useful to remove food debris. In true achalasia, the endoscope passes through the LOS with little resistance.



* Interval depends on the degree of clinical confidence about diagnosis (accuracy of endoscopic report and number of biopsies)

Fig. 22.19 British Society of Gastroenterology (BSG) Guidelines for surveillance.

- ***CT scan*** excludes distal oesophageal cancer.
- ***Manometry*** shows aperistalsis of the oesophagus and failure of relaxation of the LOS (Fig. 22.21).

There are three different subtypes (I, II, and III) based on findings of the high-resolution manometry. Type I is sometimes referred to as classic achalasia and is characterized by minimal oesophageal contractility. Type II features intermittent periods of oesophageal contraction and type III shows spastic or premature contractions of the distal oesophagus. All three types feature impaired relaxation of the LOS.

Management

All current forms of treatment for achalasia are ***palliative***. Drug therapy rarely produces satisfactory or durable relief; nitrates or sildenafil can be tried initially.

Endoscopic and surgical therapies are equally effective. Endoscopic dilation of the LOS, using a pneumatic balloon under X-ray control, weakens the sphincter and is initially successful in 80% of cases. About 50% of patients require a second or third dilation in the first 5 years. There is a low but significant risk of perforation. Intrasphincteric injection of botulinum toxin A produces satisfactory initial results but the effects wear off within months. Further injections can be given. It is safer and simpler than dilation, so may be valuable in patients at risk of death if a perforation occurs. Neither pneumatic dilation nor botulinum toxin works as well in younger patients.

Surgical division of the LOS, Heller's operation, usually performed laparoscopically, is the surgical treatment of choice. Per oral endoscopic myotomy (POEM) is a relatively novel technique, which is a division of the LOS using a gastroscope. This can be very effective method of treatment in all subtypes of achalasia.

Reflux oesophagitis complicates all procedures and aperistalsis of the oesophagus remains.

Complications

There is a slight increase in the incidence of squamous carcinoma of the oesophagus in both treated and untreated cases (7% after 25 years).

Systemic sclerosis

The oesophagus is involved in almost all patients with this disease. Diminished peristalsis and oesophageal clearance, detected manometrically (see Fig. 22.21) or by barium swallow, is due to replacement of the smooth muscle by fibrous tissue. LOS pressure is decreased, allowing reflux with consequent mucosal damage. Strictures may develop. Initially, there are no symptoms, but dysphagia and heartburn occur as the oesophagus becomes more severely involved. Similar motility abnormalities may be found in other autoimmune

rheumatic disorders, particularly if Raynaud's phenomenon is present. Treatment is as for reflux disease (see 'Gastro-oesophageal reflux disease/Management' section) and benign stricture.

Diffuse oesophageal spasm

This is a severe form of oesophageal dysmotility that can sometimes produce retrosternal chest pain and dysphagia. It can accompany GORD. Swallowing is accompanied by bizarre and marked contractions of the oesophagus without normal peristalsis (see **Fig. 22.21**). On barium swallow, the appearance may be that of a 'corkscrew' oesophagus. However, asymptomatic oesophageal 'dysmotility' is not infrequent, particularly in patients over the age of 60 years.

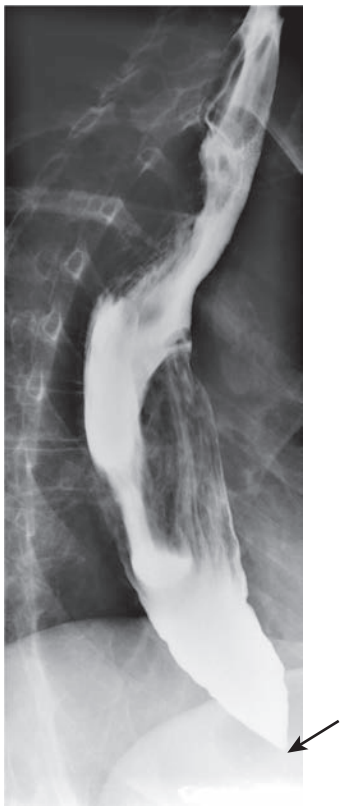


Fig. 22.20 Barium swallow demonstrating achalasia. The oesophagus is dilated and tapers inferiorly with a 'bird-beak' *(arrow)* appearance at the gastro-oesophageal junction.

A variant of diffuse oesophageal spasm is the 'nutcracker' oesophagus, which is characterized by very high-amplitude peristalsis (pressures >200 mmHg) within the oesophagus. Chest pain is more common than dysphagia.

Management

True oesophageal spasm producing severe symptoms is uncommon and treatment is often difficult. PPIs may be successful if reflux is a factor. Antispasmodics, nitrates, calcium-channel blockers, and γ-aminobutyric acid (GABA) receptor agonists (e.g., baclofen) are used. Occasionally, balloon dilation or even longitudinal oesophageal myotomy is necessary.

Miscellaneous motility disorders

Abnormalities of motility that occasionally produce dysphagia are found in the elderly, and in diabetes mellitus, myotonic dystrophy, oculopharyngeal muscular dystrophy, and myasthenia gravis, as well as neurological disorders involving the brainstem.

Other oesophageal disorders

Oesophageal diverticulum

Diverticula occur:

- immediately ***above*** the UOS (pharyngeal pouch—Zenker's diverticulum; see Chapter 21)
- near the ***middle*** of the oesophagus (traction diverticulum due to inflammation, or associated with diffuse oesophageal spasm or mediastinal fibrosis)
- just above the ***lower*** oesophageal sphincter (epiphrenic diverticulum—associated with achalasia).

Usually detected incidentally on a barium swallow performed for other reasons, these are often asymptomatic. Dysphagia and regurgitation can occur with a pharyngeal pouch, becoming more problematic as the diverticula increase in size. These can be treated surgically and endoscopically.

Rings and webs

An oesophageal web is a thin, membranous tissue flap covered with squamous epithelium. Most acquired webs are located anteriorly in the postcricoid region of the cervical oesophagus and are well seen on barium swallow. They may produce dysphagia. In the ***Plummer–Vinson syndrome*** (or ***Paterson–Brown–Kelly syndrome***), a web is associated with chronic iron deficiency anaemia, glossitis, and angular stomatitis. This rare syndrome affects mainly females and its aetiology is not understood. The web may be difficult to see at

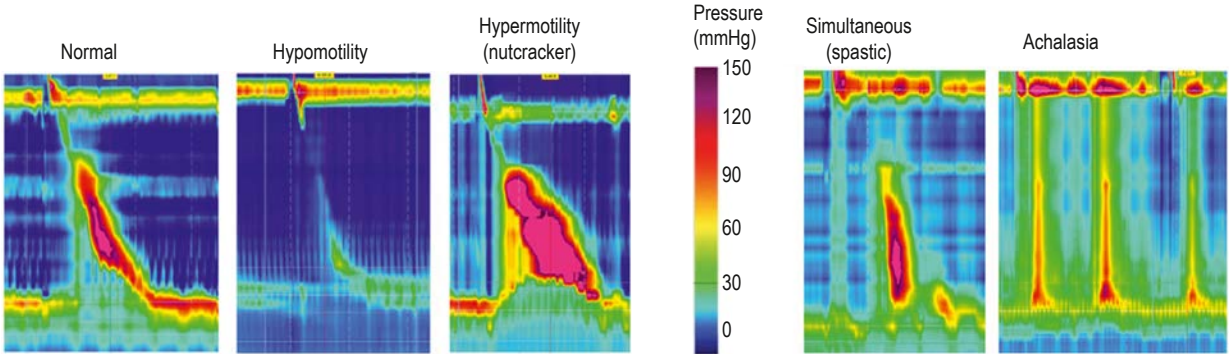


Fig. 22.21 High-resolution manometry of the oesophagus. Results are shown in the normal subject and in patients with hypomotility, hypermotility, simultaneous (spastic) contractions, and achalasia. Changes in colour reflect changes in pressure.

endoscopy and is often ruptured unintentionally by the passage of the endoscope. Dilation of the web is rarely necessary. Iron is given for the iron deficiency.

Lower oesophageal rings

Lower oesophageal rings are of two types:

- ***Mucosal (Schatzki's ring, also called B ring)*** is located at the squamocolumnar mucosal junction and is common (Fig. 22.22). It is associated with a characteristic history of intermittent bolus obstruction.
- ***Muscular (A ring)*** is located proximal to the mucosal ring and is uncommon. It is covered by squamous epithelium and may cause dysphagia.

Management for these rings is usually with reassurance and dietary advice, but dilation is occasionally necessary. After a single dilation, 68% of patients with Schatzki's rings are symptom-free at 1 year and 35% remain symptom-free after 2 years, but only 11% are symptom-free at 3 years. Many also respond to oral PPI, either alone or with dilation.

Benign oesophageal stricture

Peptic stricture (Fig. 22.23) secondary to reflux is the most common cause of benign strictures (for treatment, see 'Peptic stricture' section). They also occur after the ingestion of corrosives, radiotherapy, sclerotherapy of varices, and prolonged nasogastric intubation. Dysphagia is usually treated by endoscopic dilation. Surgery is sometimes required.

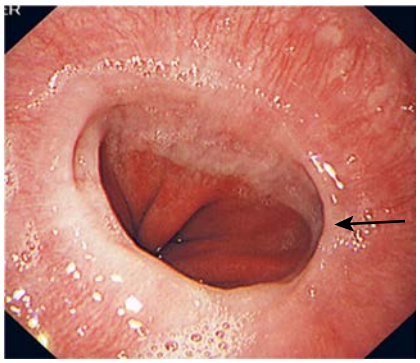


Fig. 22.22 Schatzki's ring. Endoscopic view.

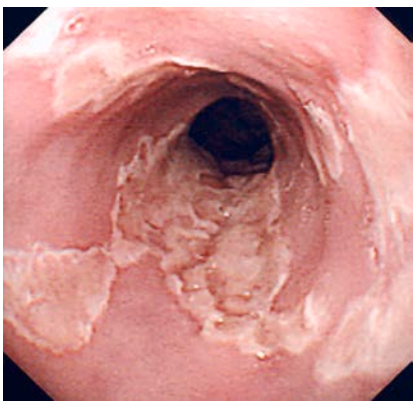


Fig. 22.23 Benign oesophageal reflux stricture. Endoscopic view.

Oesophageal infections

Infection is a cause of painful swallowing and is seen particularly in immunosuppressed (e.g., on chemotherapy) and debilitated patients, and in those with acquired immunodeficiency syndrome (AIDS). Infection can occur with:

- Candida
- herpes simplex
- cytomegalovirus
- *Mycobacterium tuberculosis.*

It is occasionally difficult to distinguish between these disorders on oesophagoscopy, as only widespread ulceration is seen. In candidiasis, the characteristic white plaques are frequently found; oral candidiasis is not always present. The diagnosis of *Candida* infection can be confirmed by examining a direct smear taken at endoscopy, but often infections are mixed and cultures and biopsies must be performed. Tuberculosis causes deep ulceration with associated mediastinal lymphadenopathy.

Management

Most patients on large doses of immunosuppressive agents are treated prophylactically for candidiasis with nystatin, fluconazole, or amphotericin. Antifungal or antiviral treatment is prescribed appropriately for other infections.

Mallory–Weiss syndrome

This is described in the 'Mallory–Weiss tear' section.

Eosinophilic oesophagitis

Eosinophils can be seen in the oesophageal mucosa (which is usually devoid of eosinophils microscopically) due to a variety of causes, such as eosinophilic (or allergic) oesophagitis and GORD.

Eosinophilic oesophagitis (Fig. 22.24) is increasingly recognized but its pathogenesis is unknown. There may be a personal or family history of allergic disorders, such as food allergy, eczema, or asthma.

Patients present with a long history of dysphagia, food impaction, 'heartburn', and oesophageal pain caused by the eosinophil-induced oesophageal inflammation. Usually, the patient is male and white, and has an average age at diagnosis of 35, but eosinophilic oesophagitis is becoming more common in children.

Typical endoscopic abnormalities include mucosal furrowing, loss of vascular pattern due to a thickened mucosa, plaques of eosinophilic surface exudate, and prominent circular folds, but the oesophagus may appear macroscopically normal. Reflux oesophagitis and Schatzki's rings may coexist. Endoscopic forceps biopsies

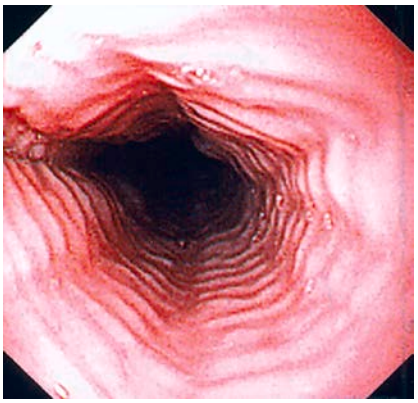


Fig. 22.24 Eosinophilic oesophagitis.

should be taken throughout the oesophagus for histology and eosinophil numbers calculated.

The eosinophilic infiltration of the oesophagus due to ***reflux disease*** tends to have a different microscopic appearance and fewer eosinophils.

Management

First-line treatment is with topical steroids, such as swallowing fluticasone spray or budesonide syrup. If this is not effective, systemic steroids or empirical elimination diets may also be used (dietary treatment is more beneficial in children). A cohort of patients respond to PPIs in the absence of GORD. Dilation is sometimes necessary, with a risk of perforation of 2%.

Oesophageal perforation

Oesophageal perforation most commonly occurs at the time of endoscopic dilation and, rarely, following insertion of a nasogastric tube, gastroscope, or transoesophageal echoprobe. Patients with malignant, corrosive, or post-radiotherapy strictures are more likely to perforate than those with a benign peptic stricture.

Management

This normally involves placement of an expanding covered oesophageal stent, which usually seals the hole. A water-soluble contrast X-ray is performed after 2 to 3 days to check the perforation has been sealed.

Oesophageal rupture

'Spontaneous' oesophageal rupture occurs with violent vomiting (Boerhaave's syndrome), producing severe chest pain and collapse in typical cases. Diagnosis can be difficult because classic symptoms are absent in about one-third of cases, and delays in presentation for medical care are common. Rupture may follow alcohol ingestion. A chest X-ray shows a hydropneumothorax. The diagnosis is made with a water-soluble contrast swallow or on CT. The mortality rate is approximately 35%, making it the most lethal perforation of the gastrointestinal tract. The best outcomes are associated with early diagnosis and definitive surgical management within 12 hours of rupture. If intervention is delayed longer than 24 hours, the mortality rate (even after surgery) rises to above 50%, and to nearly 90% after 48 hours.

Oesophageal tumours

Cancer of the oesophagus

This is the sixth most common cancer worldwide. Squamous cancers occurring in the middle third account for 40% of tumours, and for 15% in the upper third. Adenocarcinomas occur in the lower third of the oesophagus and at the cardia, and represent approximately 45% of tumours. Primary small cell cancer of the oesophagus is extremely rare.

Epidemiology and aetiology

Squamous cell carcinoma

The geographic variation in incidence is greater than for any other carcinoma—often in regions very close to one another. Squamous cell carcinoma (SCC) is common in Ethiopia, China, South and East Africa, and the Caspian regions of Iran. By contrast, North, Central, and West Africa have low rates.

In the UK, the incidence is 5 to 10 per 100 000 and SCC represents 2.2% of all malignant disease. The incidence is decreasing, in contrast to that of adenocarcinoma. SCC of the oesophagus is more common in men (2:1). Risk factors are shown in **Box 22.22**.

Box 22.22 Risk factors for cancer of the oesophagus

Squamous cell carcinoma
- Tobacco smoking
- High alcohol intake
- Plummer–Vinson syndrome
- Achalasia
- Corrosive strictures
- Coeliac disease
- Breast cancer treated with radiotherapy
- Tylosis[a]

Adenocarcinoma
- Longstanding heartburn
- Barrett's oesophagus
- Tobacco smoking
- Obesity
- Breast cancer treated with radiotherapy
- Older age

[a]Tylosis is a rare autosomal dominant condition with hyperkeratosis of the palms and soles.

High levels of alcohol consumption increase the risk of squamous cell cancer of the oesophagus, while tobacco use is associated with an increased incidence of both squamous cell and adenocarcinomas of the oesophagus. Smoking, obesity, and low fruit and vegetable consumption are implicated in approximately 9 in 10 squamous cell cancers of the oesophagus.

Diets rich in fibre, carotenoids, folate, vitamin C, and non-starchy vegetables probably decrease the risk of oesophageal cancer, whereas diets high in saturated fat, cholesterol, and refined cereals have been associated with an increased risk. Red and processed meat intake has been associated with an increased risk of both oesophageal SCC and adenocarcinoma. Conversely, fish and white meat consumption have been inversely associated with risk of oesophageal SCC in case–control studies from Italy, Switzerland, and Uruguay.

Adenocarcinoma

These tumours primarily arise in columnar-lined epithelium in the lower oesophagus. The incidence of this tumour is increasing in Western industrialized countries. It currently accounts for more than 70% of all new oesophageal cancer diagnoses. Extension of an adenocarcinoma of the gastric cardia into the oesophagus can present with the same symptoms. Previous reflux symptoms increase the risk up to eightfold and the risk is proportional to their severity.

Clinical features

Carcinoma of the oesophagus occurs mainly in those aged 60 to 70 years. Dysphagia is progressive and unrelenting. Initially, there is difficulty in swallowing solids but, typically, dysphagia for liquids follows within weeks. Impaction of food causes pain but more persistent pain implies infiltration of adjacent structures.

The lesion may be ulcerative, proliferative, or scirrhous, extending variably around the wall of the oesophagus to produce a stricture. Direct invasion of the surrounding structures and metastases to lymph nodes are more common than disseminated metastases. Weight loss, due to the dysphagia as well as to anorexia, is frequent. Oesophageal obstruction eventually causes difficulty in swallowing saliva, coughing, and aspiration into the lungs.

Weight loss, anorexia, and lymphadenopathy are the most common physical signs.

Investigations

Diagnosis

- ***Endoscopy*** provides histological proof of the carcinoma (Fig. 22.25A).
- ***Barium swallow*** can be useful where the differential diagnosis of dysphagia includes a motility disorder such as achalasia (see Fig. 22.25B).

Staging

The TNM staging system is used (see 'Gastric tumours/Staging' section); it is similar to the one used for gastric cancer. Tumour invasion of the wall of the oesophagus (T), presence of tumour in lymph nodes (N), and metastases (M) are combined into stage categories. Tumours arising in the cervical, thoracic, or abdominal oesophagus, including those that arise from within 5 cm from the gastro-oesophageal junction, share the same TNM staging criteria, but recent reclassification has differentiated between ***squamous*** and ***adenocarcinoma*** cancers.

- ***CT scan*** of the thorax and upper abdomen shows the volume of the tumour, local invasion, peritumoral and coeliac lymph node involvement, and metastases in the lung and elsewhere.
- ***MRI*** is equivalent to CT in local staging but not as good for pulmonary metastases.
- ***Endoscopic ultrasound*** has an accuracy rate of nearly 90% for assessing depth of tumour and infiltration, and of 80% for staging lymph node involvement. It is useful if CT has not already demonstrated that a cancer is too advanced for surgery. A fine-needle aspiration (FNA) of lymph nodes improves staging accuracy. Accurate T staging is necessary, as cancers confined to the superficial mucosa can be removed endoscopically (see Fig. 22.25C).
- ***Laparoscopy*** is useful if the tumour is at the cardia, to look for peritoneal and node metastases.
- ***PET*** after fluorodeoxyglucose is used principally to confirm distant metastases suspected on CT.

Management of oesophageal cancer is discussed in Chapter 41.

Other oesophageal tumours

Most other tumours are rare. ***Gastrointestinal stromal tumours*** (see Chapter 41) and ***leiomyomas*** (both submucosal tumours) are found usually by chance; 10% cause dysphagia or bleeding. Surgical removal is performed for symptomatic lesions or those over 3 cm, which are more likely to harbour malignancy. Small, benign tumours are relatively common and often do not require treatment.

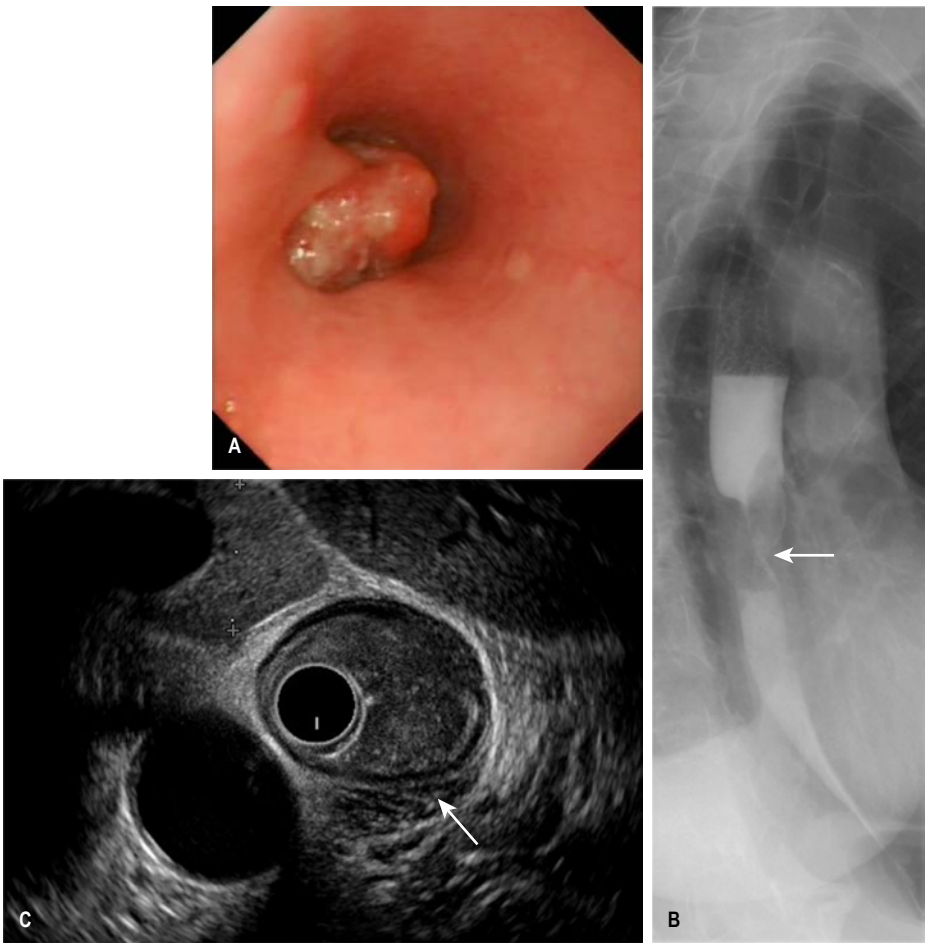


Fig. 22.25 Carcinoma of the oesophagus. **(A)** Endoscopic image. **(B)** Barium swallow demonstrating a tight stricture in the mid-oesophagus *(arrow)* due to oesophageal cancer. **(C)** Endoscopic ultrasound. The central concentric circles are the probe. The *arrow* points to a break in the muscle layer and the soft tissue mass of the carcinoma.

Kaposi's sarcoma is found in the oesophagus as well as the mouth (see Chapter 44) and hypopharynx in patients with AIDS.

Functional oesophageal disorders

The criteria for diagnosis rest mainly on compatible symptoms. However, pathological gastro-oesophageal reflux and eosinophilic oesophagitis may need to be excluded (see 'Eosinophilic oesophagitis' section).

Globus

This presents as:

- persistent or intermittent sensation of a lump or foreign body in the throat
- occurrence of the sensation between meals
- absence of dysphagia and pain on swallowing (odynophagia).

Management is with explanation and reassurance, and a trial of anti-reflux therapy. Antidepressants may be tried.

Functional chest pain of presumed oesophageal origin

This is characterized by episodes of mainly midline chest pain, not burning in nature, that are potentially of oesophageal origin and occur in the absence of a cardiological cause, gastro-oesophageal reflux, and achalasia.

More than half of patients will respond to high-dose acid-suppression therapy in the first week; some will respond to nitrates and calcium-channel blockers.

Antidepressant therapy, such as amitriptyline or the selective serotonin reuptake inhibitor (SSRI) citalopram, has been shown to be effective.

Further reading

Dhar A, Haboubi HN, Attwood SE et al. BSG and BSPGHAN joint consensus guidelines on the diagnosis and management of eosinophilic oesophagitis in children and adults. *Gut* 2022; 71:1459–1487.
Katz PO, Dunbar KB et al. ACG clinical guideline for the diagnosis and management of gastro-oesophageal reflux disease. *Am J Gastroenterol* 2022; 117(1):27–56.
NICE. Gastro-oesophageal reflux disease and dyspepsia in adults: investigation and management. *Clinical Guideline* 2014; 184. (updated 2019).
Pophali P, Halland M. Barrett's oesophagus: diagnosis and management. *BMJ* 2016; 353:i2373.
Tuason J, Inoue H. Current status of achalasia management: a review on diagnosis and treatment. *J Gastroenterol* 2017; 52:401–406.
Werner VB, Hakanson B, Martinek et al. Endoscopic or surgical myotomy in patients with idiopathic achalasia. *N Engl J Med* 2019; 381:2219–2229.
https://www.nice.org.uk/guidance/CG184.

STOMACH AND DUODENUM

Gastritis and gastropathy

'Gastritis' indicates inflammation associated with mucosal injury (although the term is often used loosely by endoscopists to describe 'redness'), and ***'gastropathy'*** indicates epithelial cell damage and regeneration without inflammation.

Gastritis

Several classifications of gastritis (e.g., Sydney classification) have been proposed but are controversial due to lack of correlation between endoscopic and histological findings. *H. pylori* infection is the most common cause of gastritis, with autoimmune gastritis being seen in 5% of cases; the remaining causes include viruses (e.g., cytomegalovirus and herpes simplex), duodenogastric reflux and specific causes, for example, Crohn's, more common in children than adults. Chronic inflammation, particularly if induced by *H. pylori*, can lead to gastric intestinal metaplasia, a precursor to gastric cancer. The role of surveillance in these patients is unclear.

Autoimmune gastritis

This affects the fundus and body of the stomach (pangastritis), leading to atrophic gastritis and loss of parietal cells, with achlorhydria and intrinsic factor deficiency causing the clinical syndrome of 'pernicious anaemia' (see Chapter 38). Metaplasia, usually of the intestinal type, is almost always in the context of atrophic gastritis. Serum autoantibodies to gastric parietal cells are common and non-specific; antibodies to intrinsic factor are rarer and more significant.

Gastropathy

Gastropathy is usually caused by irritants (drugs, NSAIDs and alcohol), bile reflux, and chronic congestion. Acute erosive/haemorrhagic gastropathy can also be seen after severe stress (stress ulcers); secondary to burns (Curling ulcers), trauma, shock, or renal failure; and in portal hypertension (called portal gastropathy). The underlying mechanism for these ulcers is unknown but may be related to an alteration in mucosal blood flow.

Helicobacter pylori infection

Helicobacter pylori is a slow-growing, spiral, Gram-negative, flagellate, urease-producing bacterium (**Fig. 22.26**), which plays a major role in gastritis and peptic ulcer disease. It colonizes the mucous layer in the gastric antrum, but is also found in the duodenum in areas of gastric metaplasia. *H. pylori* is found in greatest numbers under the mucous layer in gastric pits, where it adheres specifically to gastric epithelial cells. It is protected from gastric acid by the juxtamucosal mucous layer, which traps bicarbonate secreted by antral cells, and ammonia produced by bacterial urease.

Epidemiology

The prevalence of *H. pylori* is high in low and middle income countries (80%–90% of the population) and much lower (20%–50%) in high income countries. Infection rates are highest in lower-income groups. Infection is usually acquired in childhood; although the exact route is uncertain, it may be faecal–oral or oral–oral. The incidence increases with age, probably due to acquisition in childhood when hygiene was poorer (cohort effect) rather than infection in adult life, which is most likely far less than 1% per year in high income countries.

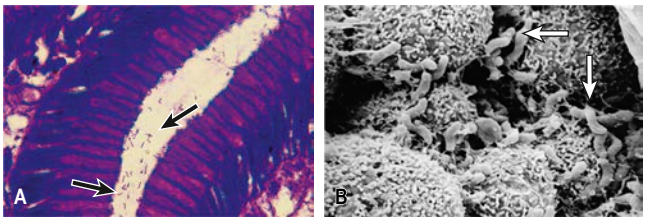


Fig. 22.26 ***Helicobacter pylori.*** **(A)** Organisms *(arrowed)* are shown on the gastric mucosa (cresyl fast violet (modified Giemsa) stain). **(B)** Scanning electron microscopy, showing the spiral-shaped bacterium *(arrowed)*. (*A, Courtesy of Dr Alan Phillips, Department of Paediatric Gastroenterology, Royal Free Hospital.*)

Pathogenesis

The pathogenetic mechanisms are not fully understood, with the majority of the colonized population remaining asymptomatic throughout their life. *H. pylori* is highly adapted to the stomach environment, exclusively colonizing gastric epithelium and inhabiting the mucous layer, or just beneath. It adheres by a number of adhesion molecules including BabA, which binds to the Lewis antigen expressed on the surface of gastric mucosal cells and causes gastritis in all infected subjects. Damage to the gastric epithelial cell is caused by the release of enzymes and the induction of apoptosis through binding to class II major histocompatibility complex (MHC) molecules. The production of urease enables the conversion of urea to ammonium and chloride, which are directly cytotoxic. Ulcers are most common when the infecting strain expresses *CagA* (cytotoxic-associated protein) and *VacA* (vacuolating toxin) genes secondary to a more pronounced inflammatory and immune response. Expression of *CagA* and *VacA* is associated with greater induction of interleukin 8 (IL-8), a potent mediator of gastric inflammation. Genetic variations in the host are also thought to be involved; for example, polymorphisms leading to increased levels of IL-1β are associated with atrophic gastritis and cancer.

Results of *H. pylori* infection

- Inflammation (antral gastritis and gastric intestinal metaplasia).
- Peptic ulcers (duodenal and gastric).
- Gastric cancer (see 'Adenocarcinoma' section).

Antral gastritis

Antral gastritis is the usual effect of *H. pylori* infection. It is normally asymptomatic, although patients without ulcers do sometimes experience relief of dyspeptic symptoms after *Helicobacter* eradication. Chronic antral gastritis causes hypergastrinaemia due to gastrin release from antral G cells. The subsequent increase in acid output is usually asymptomatic but can lead to duodenal ulceration.

Duodenal ulcer

The prevalence of *H. pylori* infection in patients with duodenal ulcers (DUs; Fig. 22.27A) is falling and in the developed world is now between 50% and 75%, whereas duodenal ulceration was once rare in the absence of *H. pylori* infection. This has been attributed to a decrease in prevalence of the bacterium and an increase in NSAID use. Eradication of the infection improves ulcer healing and decreases the incidence of recurrence.

The precise mechanism of duodenal ulceration is unclear, as only 15% of patients infected with *H. pylori* (50%–60% of the adult population worldwide) develop duodenal ulcers. Factors that have been implicated include increased gastrin secretion, smoking, bacterial virulence, and genetic susceptibility.

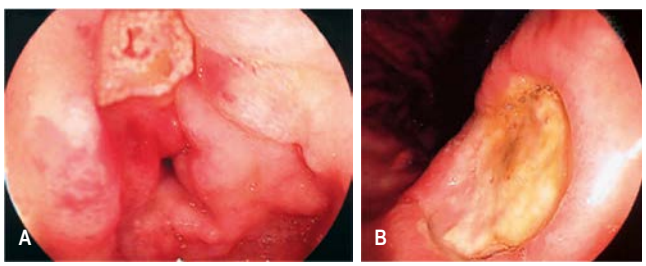


Fig. 22.27 Endoscopic views in *Helicobacter pylori* infection. (A) Duodenal ulcer with inflamed duodenal folds. **(B)** Benign gastric ulcer.

Gastric ulcer

Gastric ulcers (GUs; see Fig. 22.27B) are associated with a gastritis affecting the body as well as the antrum of the stomach (pangastritis), causing parietal cell loss and reduced acid production. The ulcers are thought to occur because of a reduction of gastric mucosal resistance due to cytokine production caused by the infection, or perhaps because of alterations in gastric mucus.

Peptic ulcer disease

A ***peptic ulcer*** consists of a break in the superficial epithelial cells penetrating down to the muscularis mucosa of either the stomach or the duodenum; there is a fibrous base and an increase in inflammatory cells. ***Erosions***, by contrast, are superficial breaks in the mucosa alone. DUs are most commonly found in the duodenal cap; the surrounding mucosa appears inflamed, haemorrhagic, or friable (duodenitis). GUs are most commonly seen on the lesser curve near the incisura, but can be found in any part of the stomach.

Epidemiology of peptic ulcer disease

DUs affect approximately 10% of the adult population and are two to three times more common than GUs.

Ulcer rates are declining rapidly for younger men and increasing for older individuals, particularly females. Both DUs and GUs are common in the elderly. There is considerable geographical variation, with peptic ulcer disease being more prevalent in low and middle income countries related to the high *H. pylori* infection. In the developed world, the percentage of NSAID-induced peptic ulcers is increasing as the prevalence of *H. pylori* declines.

Clinical features of peptic ulcer disease

The characteristic feature of peptic ulcer is recurrent, burning epigastric pain. It has been shown that if a patient points with a single finger to the epigastrium as the site of the pain, this is strongly suggestive of peptic ulcer disease. The relationship of the pain to food is variable and, on the whole, not helpful in diagnosis. The pain of a DU classically occurs at night (as well as during the day) and is worse when the patient is hungry, but this is not reliable. The pain of both GUs and DUs may be relieved by antacids.

Nausea may accompany the pain; vomiting is infrequent but can relieve the pain. Anorexia and weight loss may occur, particularly with GUs. Persistent and severe pain suggests complications, such as penetration into other organs. Back pain suggests a penetrating posterior ulcer. Severe ulceration can occasionally be symptomless, as many who present with acute ulcer bleeding or perforation have no preceding ulcer symptoms.

Untreated, the symptoms of a DU relapse and remit spontaneously. The natural history is for the disease to remit over many years due to the onset of atrophic gastritis and a decrease in acid secretion.

Epigastric tenderness is common in both ulcer and non-ulcer dyspepsia.

Diagnosis of Helicobacter pylori infection

Diagnosis of *H. pylori* is necessary if the clinician plans to treat a positive result. This is usually in the context of active peptic ulcer disease, previous peptic ulcer disease, or mucosa-associated lymphoid tissue (MALT) lymphoma, or to 'test and treat' patients with dyspepsia under the age of 55 with no alarm symptoms (i.e., weight loss, anaemia, dysphagia, vomiting or family history of gastrointestinal cancer). ***Examination*** is usually unhelpful.

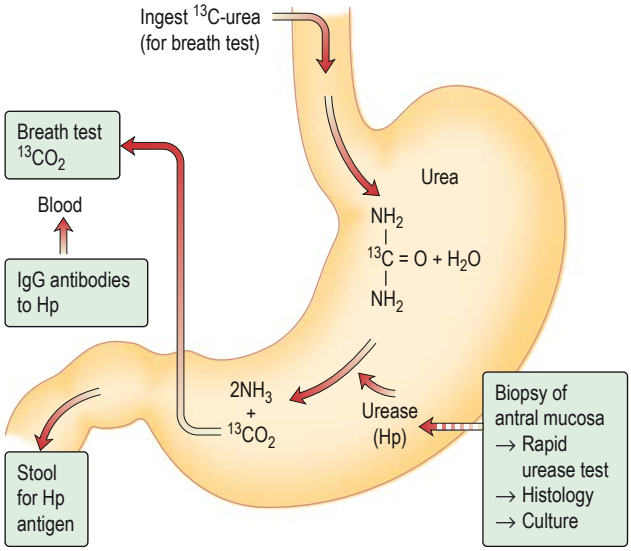


Fig. 22.28 Metabolism of urea by *Helicobacter pylori* (Hp). The different tests available for the detection of *H. pylori* are shown.

Non-invasive methods

- ***Serological tests*** detect immunoglobulin G (IgG) antibodies and are reasonably sensitive (90%) and specific (83%). They have been used in diagnosis and in epidemiological studies. IgG titres may take up to 1 year to fall by 50% after eradication therapy and therefore are not useful for confirming eradication or the presence of a current infection. Antibodies can also be found in the saliva but tests are not as sensitive or specific as serology.
- ***^{13}C-Urea breath test*** (Fig. 22.28) is a reliable test for *H. pylori* and can be used as a screening test. The measurement of $^{13}CO_2$ in the breath after ingestion of ^{13}C-urea requires a mass spectrometer. The test is sensitive (90%) and specific (96%), but the sensitivity can be improved by ensuring the patient has not taken antibiotics in the 4 weeks before the test and PPIs in the previous 2 weeks.
- ***Stool antigen test*** has superseded breath testing as the method used to determine *H. pylori* status. A specific immunoassay, using monoclonal antibodies for the qualitative detection of *H. pylori* antigen, is now widely available. The overall sensitivity is 97.6% and specificity is 96%. The test is useful in the diagnosis of *H. pylori* infection and for monitoring efficacy of eradication therapy. Patients should be off PPIs for 2 weeks but can continue with H_2-blockers. Newer stool antigen tests are being developed that can be performed in the clinic setting, although at present the sensitivity and specificity are not as good as for those performed in the laboratory.

Invasive methods (endoscopy)

- ***Biopsy urease test (Campylobacter-like organism—CLO test).*** Gastric biopsies, usually antral unless additional material is needed to exclude proximal migration, are added to a substrate containing urea and phenol red. If *H. pylori* is present, the urease enzyme that the bacteria produce splits the urea to release ammonia, which raises the pH of the solution and causes a rapid colour change (yellow to red). This enables patients' *H. pylori* status to be determined before they leave the endoscopy suite. The test may be falsely negative if patients are taking PPIs or antibiotics at the time.
- ***Histology.*** *H. pylori* can be detected histologically on routine (Giemsa) stained sections of gastric mucosa obtained at endoscopy. The sensitivity is reduced if a patient is on PPIs, but less so than with the urease test. Sensitivity can be improved with immunohistochemical staining using an anti-*H. pylori* antibody.
- ***Culture.*** Biopsies obtained can be cultured on a special medium, and *in vitro* sensitivities to antibiotics can be tested. This technique is typically used for patients with refractory *H. pylori* infection to identify the appropriate antibiotic regimen; routine culture is rare.

Investigation of suspected peptic ulcer disease

- ***Patients under 55 years of age***, with typical symptoms of peptic ulcer disease who test positive for *H. pylori*, can start eradication therapy without further investigation.
- ***Older patients*** require endoscopic diagnosis (see Fig. 22.27) and exclusion of cancer. All gastric ulcers must be biopsied to exclude an underlying malignancy and should be followed up endoscopically until healing has taken place.
- All patients with 'alarm symptoms' should undergo endoscopy:
  - iron deficiency anaemia
  - weight loss
  - anorexia
  - haematemesis/melaena
  - persistent vomiting
  - epigastric mass.

Management

Eradication therapy

Current recommendations are that all patients with duodenal and gastric ulcers should have *H. pylori* eradication therapy if the bacteria are present. Many patients have incidental *H. pylori* infection with no GU or DU. On balance, whether all such patients should have eradication therapy is controversial.

The increase in the prevalence of GORD and adenocarcinoma of the lower oesophagus in the last few years is currently unexplained, but has been postulated to be linked to eradication of *H. pylori*; this seems unlikely but is not disproven.

Depending on local antibacterial resistance patterns, standard eradication therapies in the developed world are successful in approximately 90% of patients. Re-infection is very uncommon (1%) in high income countries. In low and middle income countries, re-infection is more common, compliance with treatment may be poor, and metronidazole resistance is high (>50%, as it is frequently used for parasitic infections), so failure of eradication is common.

There are many regimens for eradication, but all must take into account that:

- Good compliance is essential.
- There is a high incidence of resistance to metronidazole and clarithromycin, particularly in some populations. Clarithromycin resistance has doubled in Europe in the last decade.
- Oral metronidazole has frequent side-effects.
- Bismuth chelate is unpleasant to take, even as tablets.

Metronidazole, clarithromycin, amoxicillin, tetracycline, and bismuth are the most widely used agents. Resistance to amoxicillin (1%–2%) and tetracycline (<1%) is low, except in countries where they are available without prescription, where resistance may exceed 50%. Quinolones (such as ciprofloxacin), furazolidone, and rifabutin are also used when standard regimens have failed ('rescue therapy'). None of these drugs is effective alone; eradication regimens therefore usually comprise two antibiotics, given with powerful acid suppression in the form of a PPI.

Bismuth-containing quadruple therapy is increasingly advocated as first-line treatment because of increasing clarithromycin resistance; the standard clarithromycin-based triple therapy has been replaced as the treatment of choice in areas where resistance is high.

Example regimens

- Omeprazole 20 mg + clarithromycin 500 mg and amoxicillin 1 g—all twice daily.
- Omeprazole 20 mg + metronidazole 400 mg and clarithromycin 500 mg—all twice daily.

These should be given for 7 or 14 days. Two-week treatments increase the eradication rates but increased side-effects may reduce compliance.

In ***eradication failures*** and in areas of clarithromycin resistance, bismuth chelate (120 mg 4 times daily), metronidazole (400 mg 3 times daily), tetracycline (500 mg 4 times daily), and a PPI (20–40 mg twice daily) for 14 days is used. Sequential courses of therapy are also used in such cases (5 days of PPI and amoxicillin, followed by a 5-day period of PPI with clarithromycin and tinidazole). With the increase in clarithromycin resistance, many are using this quadruple therapy for initial treatment.

Prolonged therapy with a PPI after a course of PPI-based 7-day triple therapy is not necessary for ulcer healing in most *H. pylori*-infected patients. The effectiveness of treatment for uncomplicated duodenal ulcer should be assessed symptomatically. If symptoms persist, breath or stool testing should be performed to check eradication (off PPI therapy).

Patients with a risk of bleeding or those with complications, such as haemorrhage or perforation, should always have a ^{13}C-urea breath test or stool test for *H. pylori* 6 weeks after the end of treatment to be sure that eradication has been successful. Long-term PPIs may be necessary if a rebleed would be likely to be fatal.

General measures

Stopping smoking should be strongly encouraged, as smoking slows mucosal healing.

Patients with gastric ulcers should be routinely re-endoscoped at 6 weeks to confirm mucosal healing and exclude an underlying gastric cancer. Repeat biopsies may be necessary.

Complications of peptic ulcer disease

Haemorrhage

See Chapter 11.

Perforation

The frequency of perforation of peptic ulceration is decreasing, partly because of medical therapy. DUs perforate more commonly than GUs, usually into the peritoneal cavity; perforation into the lesser sac also occurs. Detailed management of perforation is described in the 'Acute peritonitis/Management' section. Laparoscopic surgery is usually performed to close the perforation and drain the abdomen. Conservative management using nasogastric suction, intravenous fluids, and antibiotics is occasionally used in elderly and very sick patients.

Gastric outlet obstruction

The obstruction may be pre-pyloric, pyloric, or duodenal. The obstruction occurs either because there is an active ulcer with surrounding oedema or because the healing of an ulcer has been followed by scarring. However, obstruction due to peptic ulcer disease and gastric malignancy are now uncommon; Crohn's disease or external compression from a pancreatic carcinoma is a more common cause. Adult hypertrophic pyloric stenosis is a very rare cause.

After gastric outlet obstruction, the stomach becomes full of gastric juice and ingested fluid and food, giving rise to the main symptom of vomiting, usually without pain, as the characteristic ulcer pain has abated owing to healing.

Vomiting is infrequent, projectile, and large in volume; the vomitus contains particles of previous meals. On examination of the abdomen, there may be a succussion splash (**Fig. 22.29**). The diagnosis is made by endoscopy but can be suspected from the nature of the vomiting; by contrast, psychogenic vomiting is frequent, small volume, and usually noisy.

Severe or persistent vomiting causes loss of acid from the stomach and a hypokalaemic hypochloraemic metabolic alkalosis (see Chapter 27). Vomiting will often settle with intravenous fluid (0.9% NaCl solutions preferred in the presence of hypochloraemic alkalosis) and electrolyte replacement, gastric drainage via a nasogastric tube, and potent acid suppression therapy. Endoscopic dilation of the pyloric region is useful, as is luminal stenting, and overall, 70% of patients can be managed without surgery.

Surgical treatment and its long-term consequences

Once the mainstay of treatment, surgery is now used in peptic ulcer disease only for complications including:

- recurrent uncontrolled haemorrhage
- perforation, which is oversewn.

No other procedure, such as gastrectomy or vagotomy, is required.

In the past, two types of operation were performed: a partial gastrectomy or a vagotomy. In the latter, either a truncal vagotomy with a pyloroplasty or gastro-jejunostomy was performed, or highly selective vagotomy or proximal gastric vagotomy, which did not require a bypass procedure.

Long-term complications of surgery, which are still seen occasionally, include:

Fig. 22.29 Succussion splash. The patient is rocked gently while auscultating over the epigastrium; the presence of a 'sloshing' noise indicates gastric outlet obstruction. (*From* Clinical Examination, *Fourth Edition. B9780723434542500131/f07-62-9780723434542.*)

- ***Recurrent ulcer*.** If this occurs, check for *H. pylori*; rule out Zollinger–Ellison syndrome (see Chapter 24). Malignancy needs to be excluded in all cases.
- ***Dumping*.** This term describes a number of upper abdominal symptoms (e.g., nausea and distension associated with sweating, faintness, and palpitations) that occur in patients following gastrectomy or gastroenterostomy. It is due to 'dumping' of food into the jejunum, causing rapid fluid shifts from plasma to dilute the high osmotic load with reduction of blood volume. The symptoms are usually mild and patients adapt to them. It is rare for it to be a long-term problem, and if so, the symptoms usually have a functional element. Hypoglycaemia can also occur.
- ***Diarrhoea*.** This was chiefly seen after vagotomy. Recurrent severe episodes occurred in about 1% of patients. Antidiarrhoeals are the usual treatment.
- ***Nutritional complications*.** In the long term, almost any gastric surgery, but particularly gastrectomy, may be followed by:
  - iron deficiency, due to poor absorption
  - folate deficiency, usually due to poor intake
  - vitamin B_{12} deficiency, due to intrinsic factor deficiency
  - weight loss, usually due to reduced intake.

Other *H. pylori*-associated diseases

- ***Gastric adenocarcinoma*.** The incidence of distal (but not proximal) gastric cancer parallels that of *H. pylori* infection in countries with a high incidence of gastric cancer. Serological studies show that people infected with *H. pylori* have a higher incidence of distal gastric carcinoma (see 'Adenocarcinoma' section).
- ***Gastric B-cell lymphoma*.** Over 70% of patients with gastric B-cell lymphomas (MALT) have *H. pylori. H. pylori* gastritis has been shown to contain the clonal B cell that eventually gives rise to the MALT lymphoma (see 'Primary gastric lymphoma' section).

NSAIDs, Helicobacter and ulcers

Aspirin and other NSAIDs deplete mucosal prostaglandins by inhibiting the cyclo-oxygenase (COX) pathway, which leads to mucosal damage. Cyclo-oxygenase occurs in two main forms: COX-1, the constitutive enzyme; and COX-2, inducible by cytokine stimulation in areas of inflammation. COX-2-specific inhibitors have less effect on the COX-1 enzyme in the gastric mucosa; they still produce gastric mucosal damage but less than with other conventional NSAIDs. Their use is limited by concern regarding cardiovascular side-effects.

Some 50% of patients taking regular NSAIDs will develop gastric mucosal damage and approximately 30% will have ulcers on endoscopy. Only a small proportion of patients have symptoms (about 5%) and only 1% to 2% have a major problem: that is, gastrointestinal bleed or perforation. Because of the large number of patients on NSAIDs, including low-dose aspirin for vascular prophylaxis, this is a significant problem, particularly in the elderly.

H. pylori and NSAIDs are independent and synergistic risk factors for the development of ulcers. In a meta-analysis, the odds ratio (OR) for the incidence of peptic ulcer was 61.1 in patients infected with *H. pylori* and also taking NSAIDs, compared with uninfected controls not taking NSAIDs.

Management

- Stop the ingestion of NSAIDs.
- Give a PPI.
- Start *H. pylori* eradication therapy if the patient is *H. pylori*-positive.

In many people with severe arthritis, stopping NSAIDs may not be possible. Therefore use:

- ***An NSAID*** with low GI side-effects at the lowest dose possible. If there is low cardiovascular risk, a COX-2 NSAID can be considered (see Chapter 32).
- ***Prophylactic cytoprotective therapy***, such as PPI or misoprostol (a synthetic analogue of prostaglandin E_1 800 μg/day) for all high-risk patients, that is over 65 years, those with a peptic ulcer history, particularly with complications, and those on corticosteroid or anticoagulant therapy. PPIs reduce the risk of endoscopic duodenal and gastric ulcers and are better tolerated than misoprostol, which causes diarrhoea.

Gastroparesis

Gastroparesis results when there is delayed gastric emptying. There are a number of known causes such as:

- diabetes
- amyloidosis
- scleroderma
- parkinsonism
- multiple sclerosis
- stress
- medications (tricyclic antidepressants, dopamine agonists, calcium-channel blockers, octreotide, cyclosporine, glucagon-like peptide agonists, phenothiazines)
- post-surgical
- idiopathic.

Symptoms include nausea, vomiting, early satiety, fullness, bloating, and upper abdominal pain.

Investigation should exclude mechanical obstruction thus all patients with suspected delayed gastric emptying should undergo an OGD. Further imaging may then be needed with either CT/MRI or a barium meal. Once obstruction has been ruled out, the diagnosis of delayed gastric emptying is confirmed by scintigraphy. This is a nuclear medicine scan that determines the rate of gastric emptying. The patient must fast overnight before eating an isotope-labelled meal. The residual content of the stomach is then measured at 4 hours with more than 10% considered abnormal. Gastroduodenal manometry can help to differentiate between myopathic (amyloid, scleroderma) and neuropathic (diabetes) causes.

Management is directed at relief of symptoms and correction of nutritional deficiencies. Improvement of gastric emptying can sometimes be achieved by medical or surgical treatments, such as pro-motility agents (metoclopramide, domperidone, or erythromycin) or gastric pacemaker insertion. A diet of frequent, small, low-fibre meals can help symptoms in a number of patients. In diabetics, glycaemic control is helpful.

Functional dyspepsia

This is a disorder of gut–brain interaction (DGBI) and it is important to recognize that the majority of patients who present with dyspepsia symptoms are likely to have functional dyspepsia (FD). The main symptoms are those of epigastric pain/burning, early satiety, and postprandial fullness but can include belching, nausea/vomiting, and bloating. The Rome IV criteria state that symptoms need to have occurred for at least 3 months with onset at least 6 months prior to diagnosis. This syndrome overlaps with gastroparesis as one of the most common DGBIs and the two are often difficult to differentiate. It is likely that both disorders are part of the same spectrum with underlying pathological gastric neuromuscular dysfunction. However, there is also a complex interaction between psychopathology, gut barrier function, gut microbiome, immune system, and visceral

sensitivity to physical and chemical stimuli, which underlies these disorders.

Gastric-emptying studies (GES) can show delayed gastric emptying in a subset of patients but the correlation with symptoms is poor. Rapid gastric emptying and failure of proximal stomach accommodation can also be seen on GES. A detailed history can be enough for a positive diagnosis and this should be emphasized to patients. Upper GI endoscopy (± imaging) should be performed in the presence of red-flag symptoms but does not need to be undertaken for a diagnosis to be made.

Central to the management of FD and other DGBIs is clear, empathic communication with the patient. It is important to emphasize the chronic nature of many DGBIs but also that symptom improvement is attainable and patients can take ownership of this aspect of their care. The BSG guidelines for management of FD are very comprehensive and there is an algorithm to aid this (https://gut.bmj.com/content/71/9/1697). Briefly, this entails testing for and eradicating *H. pylori* if present, giving clear lifestyle advice regarding dietary triggers and the importance of exercise, acid suppression or prokinetics if helpful, trial of low-dose neuromodulators (e.g., tricyclic antidepressants such as amitriptyline) and referral for psychological therapies. There are a number of other foregut DGBIs which are listed in **Box 22.23** along with their common features.

Cyclical vomiting syndrome

This syndrome is characterized by typical bouts of intense vomiting lasting for hours to days, separated by periods with no symptoms. It can occur in children or adults. There is a link to migraine headaches but the cause is not known. In some cases, provocative triggers can be identified; for example, there is a specific syndrome associated with frequent cannabis use—cannabinoid hyperemesis syndrome (CHS).

Box 22.23 Foregut disorders of gut–brain interaction

Foregut DGBI	Features and management
Rumination syndrome	Regular vomiting preceded by high-pressure gastric contractions. Short-term bridging NJ feeding to other therapies may be indicated if malnourished.
Cyclical vomiting syndrome	Intermittent bouts of severe nausea and vomiting with periods of normality. Often associated with history of migraines. May need short hospital admission for parenteral fluids and antiemetics.
Centrally mediated abdominal pain and narcotic bowel syndrome	Continuous chronic pain despite escalating doses of narcotics (usually opioids). Optimize nutrition (avoid NG/NJ feeding if possible), reduce opiate doses. May need psychological support.
Somatoform disorder	Multiple overlapping functional symptoms. Needs psychiatric as well as gastroenterological input.
Avoidant Restrictive Food Intake Disorder (ARFID)	Severely restrictive food intake that is **not** driven by body image (i.e., disorded eating rather than an eating disorder) but rather by fear/anxiety/beliefs. Needs psychological and psychiatric input. Enteral tube feeding may be needed in the short term if malnourished.

Diagnosis of cyclical vomiting syndrome is by recognition of the pattern of symptoms and ruling out of other causes. Patients with CHS often report taking lengthy hot showers or baths, which provide short-term symptomatic relief.

Management

Treatment is with antiemetics and intravenous rehydration and electrolyte replacement during bouts, with avoidance of triggers where these are known. Prophylactic treatment of migraine, when it coexists, may help.

Gastric tumours

Adenocarcinoma

Gastric cancer is currently the fourth most common cancer found worldwide and the second leading cause of cancer-related mortality. The incidence increases with age (peak incidence 50–70 years), and it is rare under the age of 30 years. The highest incidence of the disease is found in Eastern Asia, Eastern Europe, and South America. The incidence in men is twice that in women and varies throughout the world, being high in Japan (M: 53/100000, F: 21.3/100000) and Chile, and relatively low in the USA (M: 7/100000, F: 2.9/100000). In the UK, carcinoma of the stomach (**Fig. 22.30**) is the eighth most common cancer (M: 16/100000, F: 9/100000). The overall worldwide incidence of gastric carcinoma is falling, even in Japan, probably due to reductions in the incidence of *Helicobacter* and, before this, improvements in food storage. However, the incidence of proximal gastric cancers is increasing in the West and they have very similar demographic and pathological features to Barrett's-associated oesophageal adenocarcinoma.

Epidemiology and pathogenesis

- ***H. pylori infection*** and distal gastric cancer are strongly linked. *H. pylori* is recognized by the International Agency for Research in Cancer (IARC) as a group 1 (definite) gastric carcinogen. *H. pylori* infection causes chronic gastritis, which eventually leads to atrophic gastritis and pre-malignant intestinal metaplasia (**Fig. 22.31**). Much of the earlier epidemiological data (i.e., the increase of cancer in lower socioeconomic groups) can be explained by the intrafamilial spread of *H. pylori.* Epstein–Barr virus is detected in 2% to 16% of gastric cancers worldwide, but its role in aetiology is not well understood.
- ***Dietary factors*** may also be involved (as both initiators and promoters) and have separate roles in carcinogenesis. Diets high in salt probably increase the risk. Dietary nitrates can be converted

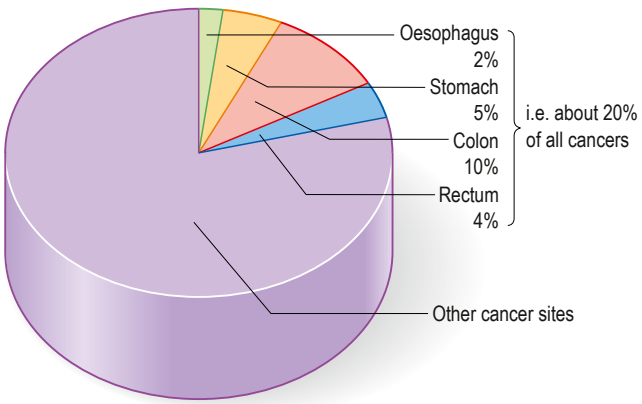


Fig. 22.30 Relative contribution of gastrointestinal sites to all cancers.

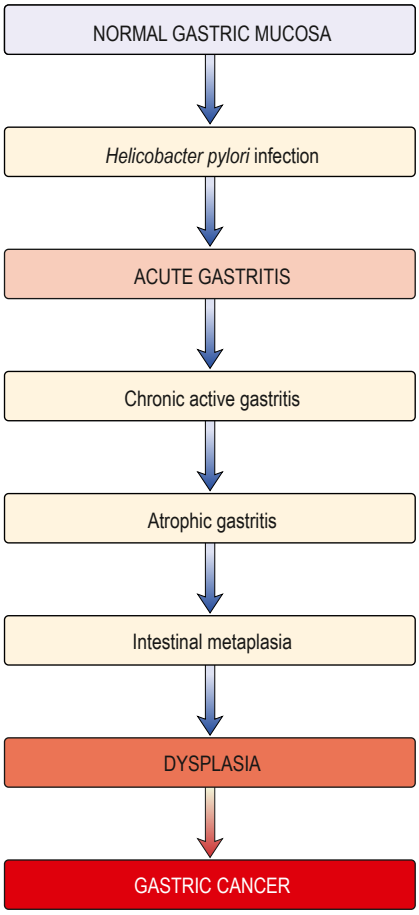


Fig. 22.31 Flow diagram showing the development of gastric cancer associated with *H. pylori* infection.

into nitrosamines by bacteria at neutral pH; nitrosamines are known to be carcinogenic in animals but the evidence in human carcinogenesis is limited. Nitrosamines are also present in the stomach of patients with achlorhydria, who have an increased cancer risk.

- ***Tobacco smoking*** is associated with an increased incidence of stomach cancer.
- ***Genetic abnormality*** is also a factor. The most common abnormality is a loss of heterozygosity (LOH) of tumour suppressor genes such as *p53* (in 50% of cancers, as well as in pre-cancerous states) and the gene encoding adenomatous polyposis coli *(APC)* (in over one-third of gastric cancers). These abnormalities are similar to those found in colorectal cancers. Some rare families with diffuse gastric cancer have been shown to have mutations in the E-cadherin gene *(CDH-1).* There is a higher incidence of gastric cancer in blood group A patients.
- ***First-degree relatives*** of patients with gastric cancer have a two- to threefold increased relative risk of developing the disease, but this may be environmental rather than inherited.
- ***Pernicious anaemia*** carries a small increased risk of gastric carcinoma due to the accompanying atrophic gastritis.
- ***Partial gastrectomy*** (postoperative stomach) carries an increased risk of gastric cancer, whether performed for a GU or DU; this is probably due to untreated *H. pylori* infection.

Screening

Earlier diagnosis has been advocated in an attempt to improve the poor prognosis of gastric cancer. Although the incidence of gastric cancer is falling in Japan, where aggressive screening by barium studies is followed by endoscopy if there is doubt, there is no evidence that screening has had an effect on overall mortality. Similarly, early investigation of dyspepsia has had little effect on mortality, possibly because of the relatively low prevalence of cancer.

Early gastric cancer

Early gastric cancer is defined as a carcinoma that is confined to the mucosa or submucosa, regardless of the presence of lymph node metastases. It is associated with 5-year survival rates of approximately 90%, but many of these patients would have survived 5 years without treatment. In Japan, mass screening with mobile units has increased the proportion of early gastric cancers (EGC) diagnosed. In a large series of patients from the UK with gastric cancer, only 0.7% were identified as having EGC. They are usually detected by chance, as although EGC exists in Western populations, endoscopists do not readily recognize it at present.

Pathology

There are two major types of gastric cancer:

- ***Intestinal (type 1)*** with well-formed glandular structures (differentiated). The tumours are polypoid or ulcerating lesions with heaped-up, rolled edges. Intestinal metaplasia is seen in the surrounding mucosa, often with *H. pylori.* This type is more likely to involve the distal stomach and occur in patients with atrophic gastritis. It has a strong environmental association.
- ***Diffuse (type 2)*** with poorly cohesive cells (undifferentiated) that tend to infiltrate the gastric wall. It may involve any part of the stomach, especially the cardia, and has a worse prognosis than the intestinal type. Loss of expression of the cell adhesion molecule E-cadherin is the key event in the carcinogenesis of diffuse gastric cancers. Unlike type 1 gastric cancers, type 2 cancers have similar frequencies in all geographic areas and occur in a younger population.

Some 50% of gastric cancers in Western countries occur in the proximal stomach.

Clinical features

Symptoms

Around 50% of patients with EGC discovered at screening have no symptoms. Most patients with carcinoma of the stomach have advanced disease at the time of presentation. The most common symptom of advanced disease is epigastric pain, indistinguishable from the pain of peptic ulcer disease; it may be relieved by food and antacids. The pain can vary in intensity but may be constant and severe, and there may also be nausea, anorexia, and weight loss. Vomiting is frequent and can be severe if the tumour encroaches on the pylorus. Dysphagia can occur with tumours involving the fundus. Gross haematemesis is unusual but anaemia from occult blood loss is frequent. No pattern of symptoms is suggestive of EGC.

Widely spreading submucosal gastric cancer causes diffuse thickening and rigidity of the stomach wall and is called ***'linitis plastica'***.

Patients can present at a late stage with malignant ascites or jaundice due to liver involvement. Metastases also occur in bone, brain and lung, producing appropriate symptoms.

Signs

Weight loss is often the dominant feature. Nearly 50% of patients have a palpable epigastric mass with abdominal tenderness. A palpable lymph node is sometimes found in the supraclavicular fossa (Virchow's node, usually on the left side), and metastases are

present in up to one-third of patients at presentation. This cancer is the most frequently associated with dermatomyositis (see Chapter 32) and acanthosis nigricans.

Diagnosis

Gastroscopy (Fig. 22.32) allows biopsies to be taken for histological assessment. Positive biopsies can be obtained in almost all cases of obvious carcinoma, but a negative biopsy does not necessarily rule out the diagnosis. For this reason, 8 to 10 biopsies should be taken from suspicious lesions. Diffuse-type gastric cancer infiltrates the submucosa and muscularis propria and can be undetected on endoscopy; multiple deep biopsies help.

Staging

- ***CT scan of the chest and abdomen*** with a gastric water load can demonstrate gastric wall thickening, lymphadenopathy, and lung and liver secondaries, but has limited ability to determine the depth of local tumour invasion.
- ***Endoscopic ultrasound*** is useful for local staging to demonstrate the depth of penetration of the cancer through the gastric wall and extension into local lymph nodes. It complements CT and ultrasound but is most relevant to confirm that a cancer is confined to the superficial mucosa, before endoscopic resection.
- ***Laparoscopy*** is useful in patients being considered for surgery to exclude serosal disease.
- ***PET and CT/PET*** can be helpful in further delineation of the cancer.

The TNM classification is used. The tumour grade (T) indicates depth of tumour invasion, N denotes the presence or absence of lymph nodes, and M indicates presence or absence of metastases. TNM classification is then combined into stage categories 0 to 4. At presentation, two-thirds of patients are at stage 3 or 4: that is, advanced disease (Box 22.24). The histological grade of the tumour also determines survival.

Management of gastric cancer is discussed in Chapter 41.

Gastrointestinal stromal tumours

Gastrointestinal stromal tumours (GISTs) are a subset of gastrointestinal mesenchymal tumours of varying differentiation. They are usually asymptomatic and found by chance but occasionally they can ulcerate and bleed. There are 200 to 900 new cases each year in the UK. GISTs mostly affect people between 55 and 65 years of age.

These tumours were previously classified as gastrointestinal leiomyomas, leiomyosarcomas, leiomyoblastomas, or schwannomas. Truly benign leiomyomas do occur, mainly in the oesophagus, but GISTs are now recognized as a distinct group of mesenchymal tumours and comprise about 80% of gastrointestinal mesenchymal tumours. They are of stromal origin and are thought to share a common ancestry with the interstitial cells of Cajal. They have varying differentiation, with mutations occurring in the cellular proto-oncogene *KIT* (which leads to activation and cell-surface expression of the tyrosine kinase KIT [CD 117]) in 80%, and also in platelet-derived growth factor receptor-α (PDGFRA) in up to 10% of patients.

Management

Treatment is surgical as far as possible. These tumours generally grow slowly but may be malignant. Imatinib, a tyrosine kinase inhibitor (see Chapter 41), is chosen for unresectable or metastatic disease, and is now used as adjunctive therapy after surgical removal of the primary in the absence of metastatic disease. Some patients are resistant to this; sunitinib can be used as an alternative agent over a short time period.

Primary gastric lymphoma

Mucosa-associated lymphatic tissue (MALT) lymphomas are indolent B-cell marginal zone lymphomas that primarily involve sites other than lymph nodes (gastrointestinal tract, thyroid, breast, or skin). They constitute about 10% of all types of non-Hodgkin's lymphoma (NHL).

Aetiology

About 90% of cases are due to *H. pylori* infection. Chromosome abnormalities t(1;14)(p22; q32) and t(11;18)(q21; q21) have also been noted in this form of NHL.

Box 22.24 Gastric cancer: staging and 5-year survival rates

Stage	TNM[a] stage	5-year survival (%)
1	T1N0M0, T1N1M0 or T2N0M0	88
2	T1N2M0, T2N1M0 or T3N0M0	65
3a	T2N2M0, T3N1M0 or T4N0M0	35
3b	T3N2M0	35
4	T4N1–3M0, TxN3M0 or TxNxM1[b]	5

[a]T, tumour; N, nodes; M, metastases.
[b]Tx indicates any T stage; Nx, any N stage.

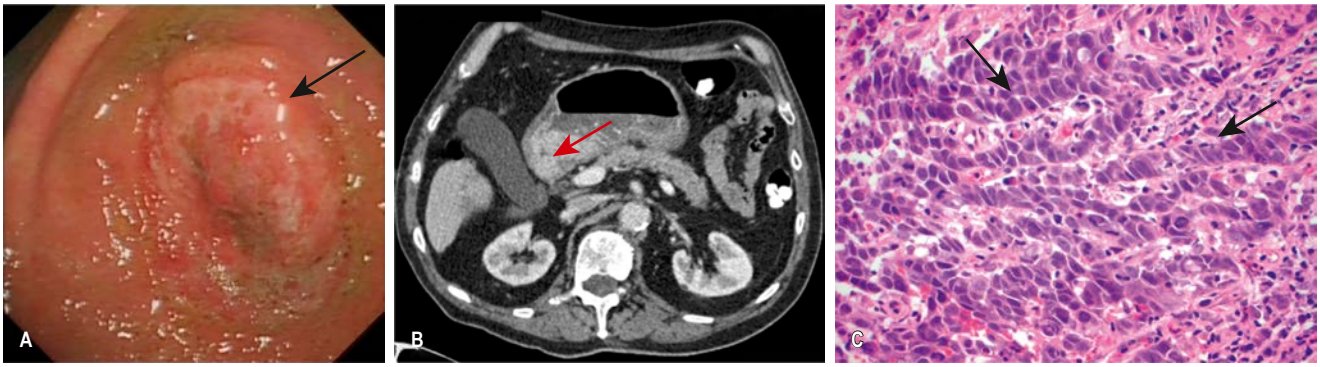


Fig. 22.32 Carcinoma of the stomach. **(A)** Endoscopic picture showing a large, irregular ulcer with rolled edges *(arrowed)*. **(B)** Axial CT scan *(ulcer arrowed)*. **(C)** Histopathology showing an adenocarcinoma *(arrowed)*.

Clinical features

Most patients are diagnosed in their 60s with stage I or stage II disease outside the lymph nodes. Patients have stomach pain, ulcers, or other localized symptoms, but rarely have systemic complaints such as fatigue or fever.

Management

Eradication of *H. pylori* infection may resolve cases of local gastric involvement. After standard eradication regimens, 50% of patients show resolution at 3 months. Other patients may resolve after 12 to 18 months of observation. Stage III or IV disease is treated with surgery or chemotherapy with or without radiation. The prognosis is good, with an estimated 90% 5-year survival.

Gastric polyps

Gastric polyps are found in about 1% of endoscopies, usually by chance. They rarely produce symptoms, but larger lesions can result in anaemia or haematemesis.

Endoscopic biopsy is the usual approach to diagnosis and treatment is possible polypectomy based on histological finding. Occasional large or multiple polyps may require surgery.

- ***Hyperplastic polyps*** are by far the most common type. Most are smaller than 2 cm. The polyps are rarely pre-malignant, but may be accompanied by pre-malignant atrophic gastritis.
- ***Adenomatous polyps*** are usually solitary lesions in the antrum. Approximately 3% progress to gastric cancer, especially if greater than 2 cm in diameter, but they are not a common cause of gastric cancer (compare this with colorectal cancer).
- ***Cystic gland polyps*** contain microcysts that are lined by fundic-type parietal and chief cells. They are located in the fundus and body of the stomach. They are found in otherwise normal subjects, but are especially common in familial polyposis syndromes and patients on PPIs. Their malignant potential is negligible, although low-grade dysplasia is seen in the absence of familial adenomatous polyposis coli (FAP), and high-grade dysplasia exclusively in its presence.
- ***Inflammatory fibroid polyps*** are benign spindle cell tumours infiltrated by eosinophils. Excision of these polyps is indicated because of their propensity to enlarge and cause obstruction.

Further reading

Crowe SE. *Helicobacter pylori* infection. *N Engl J Med* 2019; 380:1158–1165.
Kelly CM, Gutierrez Sainz L, Chi P. The management of metastatic GIST: current standard and investigational therapeutics. *J Hematol Oncol* 2021; 14:2.

ACUTE AND CHRONIC GASTROINTESTINAL BLEEDING

Acute upper gastrointestinal bleeding

The differential diagnosis, clinical approach, and emergency management of acute upper gastrointestinal bleeding are discussed in Chapter 11.

Endoscopic management of acute gastrointestinal bleeds

Endoscopy can detect the cause of haemorrhage in 80% or more of cases. These causes include:

- Oesophageal and gastric varices
- Ulcers
- Severe oesophagitis
- Dieulafoy lesion (vascular malformation)
- Malignancy.

It also enables risk stratification and allows therapy to be performed. Current guidance advises that endoscopy should be performed within 24 hours of presentation to hospital. If a patient is acutely unwell and haemodynamically unstable with a large UGI bleed they may need an emergency endoscopy performed under general anaesthetic rather than under sedation in the endoscopy department. Patients with pre-endoscopy Rockall scores of 0 or 1 may be candidates for immediate discharge and early outpatient endoscopy, depending on local policy (see Box 11.41).

The method of endoscopic therapy will depend on the cause of bleeding.

- Oesophageal varices are treated with band ligation if they are large, have high-risk stigmata (cherry red spots, red wale markings), or are actively bleeding.
- Gastric varices cannot be banded and need to be injected with cyanoacrylate glue if they are large, actively bleeding, or have high-risk stigmata.
- Ulcers that are actively bleeding or show stigmata of recent bleeding (visible vessel or adherent clot) should be treated with at least two, and sometimes three, haemostatic methods. These include injection with adrenaline (epinephrine), endoscopic clipping, and thermal coagulation (with heater probe, bipolar probe, or argon plasma coagulation). Dual and triple therapy is more effective than monotherapy in reducing re-bleeding.
- Malignancy-related bleeds are often difficult to control. The surrounding tissue may be very friable and not amenable to endoscopic clipping or thermal coagulation. Haemostatic powders that are applied widely to the bleeding area through a catheter at endoscopy can be very useful here and in challenging ulcers.
- In the case of uncontrollable portal hypertensive bleeding (oesophageal or gastric varices) where endoscopic management is not possible, a Sengstaken-Blakemore tube should be inserted under endoscopic vision. A Danis stent can be used in cases of oesophageal variceal bleeding. These can remain *in situ* for up to a week and do not require a patient to be intubated but are slightly more complex to insert and remove.
- All patients with a gastric ulcer should be rescoped 8 weeks later to check for healing and to rule out a malignancy.
- Antral biopsies for *H. pylori* should be taken although in reality many patients with duodenal ulcers will be commenced on empirical eradication therapy.

Following endoscopy, calculation of the post-endoscopy Rockall score gives an indication of the risk of rebleeding (see Box 11.42). In patients with peptic ulcers, stigmata of recent bleed indicate a higher risk of rebleeding.

Specific conditions associated with acute gastrointestinal bleeding

Oesophageal varices

These are discussed in Chapter 23.

Mallory–Weiss tear

This is a linear mucosal tear occurring at the oesophagogastric junction and produced by a sudden increase in intra-abdominal pressure. It often follows a bout of coughing or retching, and is classically seen after alcoholic ‘dry heaves’. There may, however, be no antecedent history of retching. Most bleeds are minor and discharge is usual within 24 hours. The haemorrhage may be large

but most patients stop spontaneously. Early endoscopy confirms diagnosis and allows therapy such as clipping if necessary. Surgery with oversewing of the tear is rarely needed.

Chronic peptic ulcer

Eradication of *H. pylori* is started as soon as possible (see 'Eradication therapy' section). A PPI is continued for 4 weeks to ensure ulcer healing. Eradication of *H. pylori* should always be checked in a patient who has bled, and long-term acid suppression is given if *H. pylori* eradication cannot be achieved. If bleeding is not controlled, the patient should either undergo angiography and embolization or be referred directly for surgery.

Gastric carcinoma

Most of these patients do not have large bleeds but surgery is occasionally necessary for uncontrolled or repeat bleeding. Usually, surgery can be delayed until the patient has been fully evaluated (see 'Primary gastric lymphoma' section). Oozing from gastric cancer is very difficult to control endoscopically. Radiotherapy can occasionally be successful but its effects are not immediate.

Bleeding after percutaneous coronary intervention

In the era of ever more aggressive percutaneous coronary intervention (PCI), the list of antithrombotic medication grows longer: glycoprotein IIb/IIIa inhibitors, unfractionated heparin, low-molecular-weight heparin, fondaparinux and platelet inhibitors (e.g., clopidogrel, prasugrel, and ticagrelor). Taken in addition to the oral anticoagulants that this group of patients are often taking, these give rise to a gastrointestinal bleeding rate of approximately 2% of patients undergoing PCI (who are on antiplatelet therapy, e.g., clopidogrel), and there is a high mortality of 5% to 10%. It has become increasingly evident in this patient group that gastroscopy should be performed on an urgent basis and not deferred for days or weeks. A bolus of intravenous PPI is administered, followed by an infusion; platelet infusion is given to counter the effect of clopidogrel. Management is difficult, as cessation of antiplatelet therapy has a high risk of acute stent thrombosis and also an associated high mortality. Using a risk assessment score (e.g., Blatchford), a reasonable approach is to temporarily pause antiplatelet therapy in high-risk patients but continue it in low-risk ones. This can be restarted once endoscopic haemostasis has been achieved and the patient is stable. Co-prescribed proton pump inhibition does not decrease the antiplatelet effect of clopidogrel, as was first thought. These patients should be under the combined care of a cardiologist and a gastroenterologist.

Prognosis

The mortality from gastrointestinal haemorrhage has not changed from 5% to 12% over the years, despite many changes in management, mainly because of a demographic shift to more elderly patients with co-morbidity. The lowest mortality rates are achieved in dedicated medical/surgical gastrointestinal units.

Acute lower gastrointestinal bleeding

Massive bleeding from the lower gastrointestinal tract is rare and is usually due to diverticular disease or ischaemic colitis. Common causes of small bleeds are haemorrhoids and anal fissures (**Fig. 22.33**).

Management

Most acute lower gastrointestinal bleeds start and stop spontaneously. The few patients who continue bleeding and are haemodynamically unstable need resuscitation using the same principles as for upper gastrointestinal bleeding (see Chapter 11).

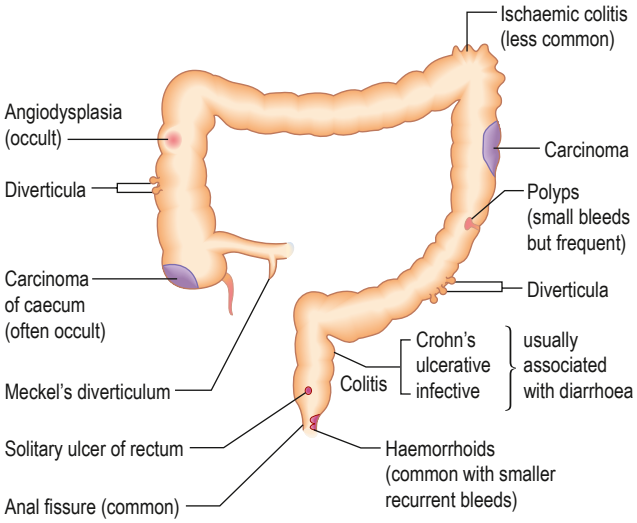


Fig. 22.33 Causes of lower gastrointestinal bleeding. The sites shown are illustrative—many of the lesions can be seen in other parts of the colon.

The diagnosis and management of lower gastrointestinal bleeding can present more of a challenge compared with upper gastrointestinal bleeding—nearly a quarter of patients who present to hospital in the UK with lower gastrointestinal bleeding are discharged home without a definitive diagnosis.

Current guidelines advise that patients who are clinically suspected to have an active bleed or are haemodynamically unstable should undergo an urgent CT angiogram. If this detects an active bleed, this should be treated endoscopically or using interventional radiology techniques, such as embolization. If this is unsuccessful then salvage surgery may be needed in those patients fit to undergo an emergency operation.

Stable patients who present with a lower gastrointestinal bleed should be categorized as a major or minor bleed using a risk assessment tool (e.g., Oakland score which uses age, gender, heart rate, blood pressure, and haemoglobin). Patients who are low risk can be discharged with a plan for outpatient investigation (usually colonoscopy as first line). Patients with a major bleed should be admitted to hospital for lower gastrointestinal endoscopy. If no lesion is found and the patient continues to have bleeding then further investigations with CT angiography and/or video capsule endoscopy (**Fig. 22.34**) can be considered.

In all patients, consideration should be taken with regard to holding or reversing anticoagulants or antiplatelet therapy.

Chronic gastrointestinal bleeding

Patients with chronic bleeding usually present with iron deficiency anaemia (see Chapter 38).

Chronic blood loss producing iron deficiency anaemia in males, and in females after the menopause, is always due to bleeding from the gastrointestinal tract. The primary concern is to exclude cancer, particularly of the stomach or right colon, and coeliac disease. Stool tests detecting occult blood (e.g., faecal immunohistochemical test, FIT) are insufficiently sensitive in this setting.

Diagnosis

Chronic blood loss can occur with any lesion of the gastrointestinal tract that produces acute bleeding (see Figs. 11.24 and 22.33). However, oesophageal varices usually bleed overtly and rarely present as chronic blood loss. Although uncommon in high income

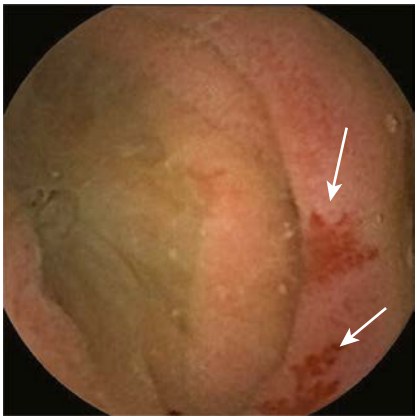


Fig. 22.34 **Angiodysplastic lesions *(arrows)* seen on video capsule endoscopy.**

countries, hookworm is the most common worldwide cause of chronic gastrointestinal blood loss.

History and examination may indicate the most likely site of the bleeding, but if no clue is available, it is usual to investigate both the upper and lower gastrointestinal tract endoscopically at the same session ('top and tail'), especially in males and postmenopausal females:

- ***Upper gastrointestinal endoscopy*** is usually performed first. Duodenal biopsies should always be taken to diagnose coeliac disease, even if coeliac serology has been performed.
- ***Colonoscopy*** follows and any lesion should be biopsied or removed, though it is unsafe to assume that colonic polyps are the cause of chronic blood loss.
- ***Unprepared CT*** scanning is a reasonable test to look for colon cancer in frail patients.
- ***CT colonography*** can be used as an alternative to colonoscopy.

If gastroscopy, colonoscopy, and duodenal biopsy have not revealed the cause, investigation of the small bowel is necessary. ***Capsule endoscopy*** is the diagnostic investigation of choice but currently has no therapeutic ability. Positive diagnostic yield varies from 60% to 85%, depending on series. Bleeding lesions can be identified and later treated with balloon-assisted enteroscopy.

Occasionally, intravenous technetium-labelled colloid may be used to demonstrate a potential bleeding site in a Meckel's diverticulum.

Management

The cause of the bleeding should be dealt with, if found. Oral iron is given to treat anaemia (see Chapter 38), although intravenous infusions are occasionally required. Some patients will require maintenance with regular transfusion as a last resort.

Further reading

Laine L. Timing of endoscopy in patients hospitalized with upper gastrointestinal bleeding. *N Engl J Med* 2020; 382:1361–1363.
Pennazio M, Spada C, Eliakim R et al. Small-bowel capsule endoscopy and device-assisted enteroscopy for diagnosis and treatment of small-bowel disorders: European Society of Gastrointestinal Endoscopy (ESGE) Clinical Guideline. *Endoscopy* 2015; 47:352–376.
https://www.bsg.org.uk/resource/bsge-acute-upper-gi-bleed-care-bundle.html. *British Society of Gastroenterology acute upper GI bleed care bundle*.

SMALL INTESTINE

Irritable bowel syndrome

Irritable bowel syndrome (IBS) is the most common functional gastrointestinal disorder. In Western populations, up to 1 in 5 people report symptoms consistent with IBS. Approximately 50% will consult their doctors and, of these, up to 30% will be referred by their doctor to a hospital specialist. Up to 40% of all patients seen in specialist gastroenterology clinics will have IBS. Estimates in the UK put the annual cost of IBS to healthcare resources as £45.6 million; in the USA, the cost is higher at $8 billion. In the UK, approximately one-quarter of IBS patients take time off work for periods ranging from 7 to 13 days each year.

The factors that determine whether an IBS sufferer in the community seeks medical advice include higher illness attitude scores and higher anxiety and depression scores than non-consulters. Consulters perceive that their symptoms are more severe than those of non-consulters, and consulting behaviour may be determined by the number of presenting symptoms. Female consulters outnumber male consulters by a factor of 2–3 to 1.

Box 22.25 Non-gastrointestinal features of irritable bowel syndrome

Gynaecological symptoms
- Painful periods (dysmenorrhoea)
- Pain following sexual intercourse (dyspareunia)

Urinary symptoms
- Frequency
- Urgency
- Passing urine at night (nocturia)
- Incomplete emptying of bladder

Other symptoms
- Joint hypermobility
- Back pain
- Headaches
- Bad breath, unpleasant taste in the mouth
- Poor sleeping
- Fatigue

Box 22.26 Some factors that can trigger onset of irritable bowel symptoms

- Affective disorders, for example, depression, anxiety
- Psychological stress and trauma
- Gastrointestinal infection
- Antibiotic therapy
- Sexual, physical, or verbal abuse
- Pelvic surgery
- Eating disorders

IBS—a multisystem disorder

IBS patients suffer from a number of non-intestinal symptoms (**Box 22.25**), which may be more intrusive than the classical features. IBS coexists with chronic fatigue syndrome (see Chapter 30), fibromyalgia (see Chapter 32), and temporomandibular joint dysfunction.

The biopsychosocial conceptualization of the pathogenesis and clinical expression of FGIDs is particularly relevant to IBS (see **Fig. 22.11**), and **Box 22.26** lists some common factors that have been shown to trigger IBS symptoms. Infectious diarrhoea precedes the onset of IBS symptoms in 7% to 30% of patients. Whether this is a factor for all patients or just a small subgroup remains controversial. Risk factors in these patients have been shown to include female sex, severity and duration of diarrhoea, pre-existing adverse life events, and high hypochondriacal anxiety and neurotic scores at the time of the initial illness. Symptoms of anxiety and depression are more common in IBS patients, and stress or adverse life events often precede the onset of chronic bowel symptoms.

Box 22.27 Subtyping irritable bowel syndrome by predominant stool pattern

Type	Description
IBS with constipation (IBS-C)	Hard lumpy stools >25% and loose (mushy) or watery stools <25% of bowel movements
IBS with diarrhoea (IBS-D)	Loose (mushy) or watery stools >25% and hard or lumpy stools <25% of bowel movements
Mixed IBS (IBS-M)	Hard or lumpy stools >25% and loose (mushy) or watery stools >25% of bowel movements
Unsubtyped IBS	Insufficient abnormality of stool consistency to meet criteria for IBS-C, D, or M

Diagnosis

Diagnostic criteria (Rome IV 2016) state that, in the preceding 3 months, there should be at least 1 day per week in the last 3 months of recurrent abdominal pain associated with two or more of the following:

1. Related to defecation.
2. Onset associated with a change in frequency of stool.
3. Onset associated with a change in form (appearance) of stool.

These are useful for comparative studies. Subgroups of IBS patients can be identified according to the criteria listed in **Box 22.27**.

The decision as to whether to investigate and the choice of investigations should be based on clinical judgement. Pointers to the need for thorough investigation are the presence of the above symptoms in association with rectal bleeding, nocturnal pain, fever and weight loss, and a clinical suspicion of organic diarrhoea. A raised stool calprotectin or lactoferrin would suggest inflammation needing further investigation.

Management

Current strategies for the treatment of IBS include therapies which target central and end-organ pathways (**Box 22.28**); these are not mutually exclusive.

Patients with IBS are often worried that their symptoms are due to a serious disease such as cancer. Alarm symptoms much be investigated when present. However, a positive diagnosis of IBS with an explanation of the symptoms and reassurance is often helpful and may require no further investigation or treatment. It is helpful to establish and maintain a good clinician–patient relationship from the beginning. **Box 22.28** shows the overall management strategies used in IBS for those with severe and longstanding problems. Lifestyle and dietary modifications should be the mainstay of management for patients with mild–moderate symptoms with pharmacologic therapy used adjunctively for those who fail to respond or who have more severe symptoms.

Pain/gas/bloat syndrome/midgut dysmotility

Disordered motility and visceral sensation that predominantly affects the small intestine or midgut results in symptoms of pain and bloating without altered defecation. Other symptoms include postprandial fullness, nausea and, on occasions, anorexia and weight loss.

Management of patients with pain/gas/bloat syndrome is not easy, and in some, pain can be chronic and severe. The main approaches are summarized below:

- Narcotics should always be avoided.

Box 22.28 Approaches to the management of the irritable bowel syndrome

Treatment modality	Action
End-organ treatment	
Exploration of dietary triggers	Refer to dietician
High-fibre diet ± fibre supplements for constipation, low FODMAP diet for bloating	Refer to dietician
Alteration of microbiota	Rifaximin has shown short-term benefit in IBS patients without constipation (target I and II trials) Pro- and prebiotics
Antidiarrhoeal drugs for bowel frequency	Loperamide Codeine phosphate Co-phenotrope [a]Eluxadoline
Constipation	5-HT_4 receptor agonist, e.g., prucalopride
Smooth muscle relaxants for pain	Mebeverine hydrochloride Dicycloverine hydrochloride Peppermint oil
Central treatment	
Explanation of physiology and symptoms	At consultation (leaflets with diagrams help)
Psychotherapy	Refer to clinical psychologist (see Chapter 30)
Hypnotherapy	
Cognitive behavioural therapy	Refer to psychiatrist
Antidepressants	Functional diarrhoea—clomipramine Diarrhoea-predominant IBS—tricyclic group, e.g., amitriptyline Constipation-predominant IBS—SSRI, e.g., paroxetine

FODMAP, Fermentable oligo-, di- and monosaccharides and polyols; *SSRI*, selective serotonin reuptake inhibitor.
[a]Eluxadoline—a μ- and κ-opioid receptor agonist and δ-opioid receptor antagonist.

- Central and end-organ-targeted treatment approaches should be combined: for example, the SSRI paroxetine combined with a prokinetic agent, such as domperidone, or a smooth muscle relaxant, such as ***mebeverine***.
- Small bowel bacterial overgrowth can be a contributory feature that can be treated with non-absorbed antibiotics such as ***rifaximin***, however this is expensive and cheaper antibiotics such as doxycycline can also offer symptomatic benefit.
- Research has highlighted the benefit of altering the fermentable components of the diet. A diet with reduced fermentable oligo-, di- and monosaccharides and polyols (low ***FODMAP***) will exclude a range of food types, including garlic, onions, specific beans, fructose-containing fruit, wheat-containing products, and certain natural and synthetic sweeteners. Small sham-diet-controlled clinical trials have shown significant benefits in patients with functional symptoms but long-term randomized controlled studies are necessary.
- The role of **probiotics** for management of IBS patients is not yet clear. A recent meta-analysis found some evidence for the efficacy of various probiotics in helping with symptoms such as pain, bloating, and flatulence but with varying methodological limitations in most of the studies. Two studies found

that *Bifidobacterium infantis* versus placebo led to symptomatic improvement in patients with IBS. It is likely that future work will focus on subgroups of patients, particularly those with diarrhoea-predominant IBS, and aim to correct the microbiome on an individual level according to the particular alterations of the compositional flora rather than give a generic probiotic.

Some patients with pain/gas/bloat syndrome have particularly severe and chronic symptoms, which may be nocturnal. A small subgroup of these has been shown to have manometric features consistent with a diagnosis of ***chronic idiopathic intestinal pseudo-obstruction (CIIP)***, and specifically of an enteric neuropathy. Full-thickness small intestinal biopsies confirm this diagnosis by showing a deficiency of α actin staining in the inner circular layer of smooth muscle. More appropriately, these patients should be considered as having a gastrointestinal neuromuscular disorder of the gut. About 10% of these individuals are subsequently found to have an underlying autoimmune overlap disorder (see Chapter 32).

Management of patients with neuromuscular disorders of the gut requires a multidisciplinary approach, with an emphasis on the management of pain, psychological state, and nutrition. Patients with underlying autoimmune rheumatic disorders may benefit from primary treatment of these. Patients with intestinal failure as a result of CIIP need long-term parenteral nutrition.

Functional diarrhoea

In this form of functional bowel disease, symptoms occur in the absence of abdominal pain. They commonly include:

- the passage of several stools in rapid succession, usually first thing in the morning; no further bowel action may occur that day, or defecation takes place only after meals
- a first stool of the day that is usually formed, the later ones being mushy, looser, or watery
- urgency of defecation
- anxiety, and uncertainty about bowel function with restriction of movement (e.g., travelling)
- exhaustion after defecation.

Chronic diarrhoea without pain is caused by many diseases that are indistinguishable by history from functional diarrhoea. Features that are atypical of a functional disorder (e.g., large-volume stools, rectal bleeding, nutritional deficiency, and weight loss) call for more extensive investigations.

Treatment of functional diarrhoea is with **loperamide**, often combined with a tricyclic antidepressant prescribed at night (e.g., clomipramine 10–30 mg).

Purgative abuse

This is most commonly seen in females who surreptitiously take high-dose purgatives and are often extensively investigated for chronic diarrhoea. The diarrhoea is usually of high volume (>1 L daily) and patients may have a low serum potassium. Biochemical analysis of the stool may help diagnose laxative abuse. Management is difficult, as most patients deny purgative ingestion. Purgative abuse often occurs in association with eating disorders and patients may need psychiatric help.

Malabsorption

In many small bowel diseases, malabsorption of specific substances occurs, but these deficiencies do not usually dominate the clinical picture. An example is Crohn's disease, in which malabsorption of vitamin B_{12} can be demonstrated, but this is not usually the major problem; diarrhoea and general ill-health are the major features.

The major disorders that cause malabsorption are shown in **Box 22.29**. These can broadly be split into three major categories: gastric causes, small intestine causes, and hepato-pancreato-biliary (HPB) causes (plus a number of other miscellaneous causes).

Assessment of malabsorption should focus on the following points, which will help to categorize the cause and determine the severity:

- Is there chronic diarrhoea?
- Are stools pale, bulky, foul smelling, and difficult to flush? If so, this indicates steatorrhoea (the presence of excessive fat within the faeces).
- Is there weight loss?
- Are there abnormalities in the menstrual cycle?
- Does the patient have a condition that predisposes them to malabsorption, such as coeliac disease, chronic pancreatitis, or small bowel Crohn's disease.
- Any unexplained nutrient deficiencies?
- Is there any evidence of anaemia? If so, this suggests global malabsorption.

Investigations should start with blood tests, which should include the following: FBC, U+Es, vitamin B_{12}, folate, ferritin, calcium, serum albumin, serum iron, iron saturation, magnesium, anti-tissue transglutaminase (tTG; see 'Coeliac disease' section), immunoglobulins, 25-hydroxyvitamin D and prothrombin time/INR (loss of fat-soluble vitamins, A, D, E, K).

Other investigations can include the following:

- Stool analysis—faecal fat to confirm steatorrhoea.
- Faecal elastase if suspect pancreatic insufficiency.

? Box 22.29 Disorders of the small intestine causing malabsorption

Gastric causes	Small intestinal causes	HPB causes	Other
Atrophic gastritis	Coeliac disease	Chronic pancreatitis	Lactase deficiency
Pernicious anaemia	Tropical sprue	Bile acid malabsorption	Lymphoma
Zollinger-Ellison syndrome	Bacterial overgrowth	Liver cirrhosis	Diabetic enteropathy
Post-Billroth II surgery	Whipple's disease	Chronic cholestasis	Thyrotoxicosis
Eosinophilic gastroenteritis	Intestinal resection		
	Radiation enteropathy		
	Small bowel Crohn's disease		
	Parasite infection (e.g., *Giardia intestinalis*)		
	Intestinal lymphangiectasia		

Box 22.30 Malabsorption—helpful hints when assessing patients

- Remember malabsorption can present in a very subtle way, not necessarily with weight loss and chronic diarrhoea.
- Sometimes the only findings may be anaemia or mildly raised LFTs (e.g., in coeliac disease).
- Look for nutrient deficiencies.
- It is uncommon for patients with malabsorption to present with abdominal pain—exceptions are Crohn's disease, intestinal pseudo-obstruction, and chronic pancreatitis.
- Remember to ask if there is any history of previous abdominal surgery.
- Check if there is history of recurrent peptic ulcer disease.
- Do not forget to ask about alcohol history.
- Is there a family history of coeliac disease or gluten sensitivity? (1 in 100 people have coeliac disease).

- If tTG positive—proceed to Upper GI Endoscopy with multiple biopsies from second and third parts of the duodenum (coeliac disease).
- Consider CT pancreas or MRCP (especially if history of alcohol excess or chronic pancreatitis).
- If history of adhesions or strictures consider hydrogen breath testing (small intestine bacterial overgrowth).
- Consider MRI small bowel if has anaemia and abdominal pain.
- If history of previous cholecystectomy—perform SeHCAT scan (see 'Bile acid malabsorption' section) to confirm bile salt malabsorption.

It is important to remember that malabsorption does not always present with drastic weight loss and diarrhoea; it can be more subtle and indolent. See **Box 22.30** for a list of helpful hints when assessing a patient with potential malabsorption.

Coeliac disease (gluten-sensitive enteropathy)

Coeliac disease is an autoimmune condition in which there is inflammation of the mucosa of the upper small bowel that improves when gluten is withdrawn from the diet and relapses when gluten is reintroduced. Gluten is the entire protein content of the cereals wheat, barley, and rye. Up to 1% of many populations are affected, though most have clinically silent disease.

Aetiology

The aetiology of coeliac disease is multifactorial with both genetic and environmental factors needed for the disease to manifest. Prolamins (gliadin from wheat, hordeins from barley, secalins from rye) are the damaging factors in these patients. These proteins are resistant to digestion by pepsin and chymotrypsin because of their high glutamine and proline content, and remain in the intestinal lumen, triggering the immune responses that result in coeliac disease. Gliadin peptides pass through the epithelium (para- and/or intracellularly) and are deaminated by tTG, which increases their immunogenicity. Gliadin peptides then bind to antigen-presenting cells, which interact with $CD4^+$ T cells in the lamina propria via HLA class II molecules DQ2 or DQ8. These T cells produce pro-inflammatory cytokines, particularly interferon-γ. $CD4^+$ T cells also interact with B cells to produce endomysial and tTG antibodies. Gliadin peptides also cause release of IL-15 from enterocytes, activating intraepithelial lymphocytes with a natural killer (NK) cell marker. This inflammatory cascade releases metalloproteinases and other mediators, which contribute to the villous atrophy and crypt hyperplasia that are typical of the disease.

The mucosa of the proximal small bowel is predominantly affected, and the mucosal damage decreasing in severity towards the ileum as gluten is digested into smaller 'non-toxic' fragments.

Genetic factors

There is an increased incidence of coeliac disease within families but the exact mode of inheritance is unknown; 10% to 15% of first-degree relatives will have the condition, although it may be asymptomatic. The concordance rate in identical twins is about 70%.

HLA-DQ2 (DQAI*0501, DQBI*0201) and HLA-DQ8 (DQAI*0301, DQBI*0302) are associated with coeliac disease. Over 90% of patients will have HLA-DQ2, compared with 20% to 30% of the general population. Studies in twins and siblings indicate that HLA genes are responsible for less than 50% of the genetic cause of the disease. Many unaffected people also carry these genes, so other factors must also be involved. Non-HLA genes may also contribute to coeliac disease: for example, chromosome regions 19p13.1, 11q, 5q31–33, and 6q21–22. The *CD28/CTLA4/ILO5* gene cluster has also shown linkage with coeliac disease.

Environmental factors

Breast-feeding and the age of introduction of gluten into the diet are significant. Rotavirus infection in infancy also increases the risk; adenovirus-12, which has sequence homology with α-gliadin, was suspected as a causative agent but this is now thought to be unlikely.

Clinical features

Coeliac disease can present at any age. In infancy, it sometimes appears after weaning on to gluten-containing foods. The peak period for diagnosis in adults is in the fifth decade, with a female preponderance. Many patients are asymptomatic (silent) and come to attention because of routine blood tests: for example, a raised MCV, or iron deficiency in pregnancy. The symptoms are very variable and often non-specific; they include tiredness and malaise, often associated with anaemia.

Gastrointestinal symptoms may be absent or mild. Coeliac disease should be tested for in all patients with symptoms suggestive of irritable bowel syndrome. Diarrhoea or steatorrhoea, abdominal pain, and weight loss suggest more severe disease. Mouth ulcers and angular stomatitis are frequent and can be intermittent. Infertility and neuropsychiatric symptoms of anxiety and depression can occur.

Rare complications include tetany, osteomalacia (both manifestations of severe vitamin D deficiency), or gross malnutrition with peripheral oedema. Long-term problems include osteoporosis, which occurs even in patients on long-term gluten-free diets. Neurological symptoms, such as paraesthesia, ataxia (due to cerebellar calcification), muscle weakness, or a polyneuropathy occur; the prognosis for these symptoms is variable. There is an increased incidence of atopy and autoimmune disease, including thyroid disease, type 1 diabetes, and Sjögren's syndrome. Other associated diseases include:

- IBD
- primary biliary cholangitis
- chronic liver disease
- interstitial lung disease
- epilepsy
- IgA deficiency is more common than in the general population.

Physical signs are usually few and non-specific, and are related to anaemia and malnutrition.

Diagnosis

Small bowel biopsy is still considered to be the 'gold standard' for positive diagnosis and is therefore desirable in all but the most clear-cut cases, because treatment involves a life-long diet that is both expensive and socially limiting. However, with the increasing accuracy of serological tests, it is no longer necessary to take duodenal biopsies for suspected coeliac disease in patients without antibodies. For example, in patients undergoing endoscopy for iron deficiency anaemia with negative coeliac serology, the pretest value of small bowel histology is less than 0.03%.

If biopsies are to be taken, four to six forceps biopsies should be taken from the second part of the duodenum and the bulb because the disease is sometimes patchy and it can be difficult to orientate endoscopic biopsies for histological section. Endoscopic signs, including absence of mucosal folds, mosaic pattern of the surface, and scalloping of mucosal folds, are often present; however, their absence is not conclusive because they are markers of relatively severe disease. Patients must be consuming gluten at the time of the endoscopy for the test to be diagnostic.

Histology

Histological changes are of variable severity and are graded using the Marsh–Oberhuber classification (**Box 22.31**) with types 2 and 3 being diagnostically supportive of coeliac disease. Villous atrophy can be caused by other conditions (e.g., tropical sprue, common variable immunodeficiency, non-coeliac enteropathy), but coeliac disease is the most common cause of subtotal villous atrophy.

Serology

Indications for testing include fatigue, persistent diarrhoea, folate or iron deficiency, unexplained abnormal liver biochemistry, a family history of coeliac disease and associated autoimmune disease.

The most sensitive tests are for endomysial (EMA) and tTG antibodies (95% sensitivity). Titres of either of these correlate with the severity of mucosal damage and so they can be used for dietary monitoring. Standard tests use IgA class antibodies. Selective IgA deficiency occurs in 2.5% of coeliac disease patients but only 0.25% of normals. In a few cases, this may render these tests falsely negative. In situations in which CD is strongly suspected in a patient negative for EMA antibodies, IgA levels should be measured; if low, IgG-based tests should then be used (e.g., deaminated gliadin peptide (DGP) antibody).

HLA typing

HLA-DQ2 is present in 90% to 95% of coeliac disease patients and HLA-DQ8 in about 8%—most of the rest. The *absence* of both alleles has a high negative predictive value for coeliac disease. HLA typing can be useful for risk assessment: for example, in patients already on a gluten-free diet, in whom serology would be negative.

Box 22.31 Marsh–Oberhuber classification for histological changes in coeliac disease

Marsh grade	Histological appearance
0	Normal mucosa
1	Increased intraepithelial lymphocytes
2	Crypt hyperplasia
3a	Partial villous atrophy
3b	Subtotal villous atrophy
3c	Total villous atrophy
4	Hypoplasia of small bowel architecture

Other investigations

- ***Haematology*.** Mild or moderate anaemia is present in 50% of cases. Folate deficiency is common, often causing macrocytosis. Vitamin B_{12} deficiency is rare. Iron deficiency, due to malabsorption of iron and increased loss of desquamated cells, is common. A blood film may therefore show microcytes and macrocytes (i.e., a dimorphic picture), as well as hypersegmented polymorphonuclear leucocytes and Howell–Jolly bodies (see Chapter 38) due to splenic atrophy.
- ***Biochemistry*.** *In severe cases*, biochemical evidence of osteomalacia may be seen (low calcium and high phosphate) and there is hypoalbuminaemia.
- ***Imaging*.** A small bowel barium follow-through or MRI enteroclysis may show dilation of the small bowel with slow transit. Folds become thicker, and in severe disease, total effacement is seen. Imaging is mainly used when a complication, such as lymphoma, is suspected.
- ***Bone densitometry*.** Dual-energy X-ray absorptiometry (DXA) should be performed on all patients because of the risk of osteoporosis.
- ***Capsule endoscopy*** (see 'Endoscopy' section). This is used to look for gut abnormalities when a complication is suspected.

Management

Replacement minerals and vitamins, such as iron, folic acid, calcium, and vitamin D, may be needed initially to replace body stores.

Management is with a gluten-free diet for life. Dietary elimination of wheat, barley, and rye usually produces a clinical improvement within days or weeks. Morphological improvement often takes months, especially in adults. Oats are tolerated by most people with CD but must not be contaminated with flour during their production. Meat, dairy products, fruits, and vegetables are naturally gluten-free and are all safe.

Gluten-free products can be expensive, unless subsidized by national health services; however, this is slowly improving as gluten-free diets are becoming more popular among the general population in the industrialized world. Patient support organizations, such as the Coeliac Society (UK), are valuable as information sources and for advice about diet, recipes, and gluten-free processed foods. Referral to a dietician for help with what to eat can be very beneficial for some patients. Despite advice, many patients do not keep to a strict diet but maintain good health. The long-term effects of this low gluten intake are uncertain but osteoporosis can occur, even in treated cases.

The usual cause of ***failure to respond*** to the diet is poor compliance. Dietary adherence can be monitored by serial tests for EMA and tTG. If clinical progress is suboptimal, then a repeat intestinal biopsy should be taken. If the diagnosis is equivocal on the diagnostic mucosal biopsy, or if the patient has already started on a gluten-free diet, then a gluten challenge, that is re-introduction of oral gluten, with evidence of jejunal morphological change, can confirm the diagnosis.

Patients should have ***pneumococcal vaccinations*** (because of splenic atrophy) once every 5 years (see Chapter 38).

Complications

A few patients do not improve on a strict diet and are said to have ***non-responsive coeliac disease***. Many of these patients are still ingesting gluten. A few of the others may have concomitant problems, such as microscopic colitis, IBD, small bowel bacterial overgrowth, or lactase deficiency.

A very small percentage will have the rare complication of ***refractory coeliac disease*** (RCD):

- Type 1 RCD—the lymphocytes are normal and the T-cell receptors are polyclonal; 5-year survival is 93%.
- Type 2—there are abnormal clonal lymphocytes with loss of CD8 and CD3 surface markers. The 5-year survival rate is 40% to 60%.

Very rarely, enteropathy-associated ***T-cell lymphoma (EATCL)*** (8%–20% 5-year survival) or ulcerative jejunitis can occur as part of a spectrum of neoplastic T-cell disorders.

Ulcerative jejunitis presents with fever, abdominal pain, perforation, and bleeding.

Diagnosis of these conditions is with MRI or barium studies, but laparoscopy with full-thickness small bowel biopsies is often required. Steroids and immunosuppressive agents, such as azathioprine, are used in ulcerative jejunitis.

The incidence of ***carcinoma of the oesophagus and small bowel adenocarcinoma***, as well as that of extragastrointestinal cancers, is also increased. Malignancy seems to be unrelated to the duration of the disease but the incidence is reduced by a gluten-free diet.

Dermatitis herpetiformis

This is an uncommon, blistering, subepidermal eruption of the skin associated with a gluten-sensitive enteropathy (see also Chapter 45). Rarely, gross malabsorption occurs, but usually the jejunal morphological abnormalities are not as severe as in coeliac disease. The inheritance and immunological abnormalities are the same as for coeliac disease. The skin condition responds to ***dapsone*** but a gluten-free diet improves both the enteropathy and the skin lesion, and is recommended for long-term benefit.

Non-coeliac gluten intolerance

There is a recognized group of patients who are sensitive to dietary wheat and gluten-containing foods but do not have coeliac disease, in so far as their coeliac serology is negative and duodenal biopsies are normal. These patients have a range of symptoms, including diarrhoea, bloating, and abdominal pain, which improve on avoidance of gluten. The mechanism is not yet clear.

Tropical sprue

This condition presents with chronic diarrhoea and malabsorption, and occurs in residents of or visitors to tropical areas where the disease is endemic: most of Asia, some Caribbean islands, Puerto Rico, and parts of South America. Epidemics occur, lasting up to 2 years; in some areas, repeated epidemics are seen at varying intervals of up to 10 years.

The term tropical sprue is reserved for severe malabsorption (of two or more substances) accompanied by diarrhoea and malnutrition. A mild degree of malabsorption, sometimes following an enteric infection, is quite common in the tropics; it is usually asymptomatic and is sometimes called tropical malabsorption.

Aetiology

The aetiology is unknown but is likely to be infective because the disease occurs in epidemics and patients improve on antibiotics. A number of agents have been suggested but none has been unequivocally shown to be responsible. Different agents could be involved in different parts of the world.

Clinical features

These vary in intensity and consist of diarrhoea, anorexia, abdominal distension, and weight loss. The onset is sometimes acute and occurs either a few days or many years after being in the tropics. Epidemics can break out in villages, affecting thousands of people at the same time. The onset can also be insidious, with chronic diarrhoea and evidence of nutritional deficiency. The clinical features of tropical sprue vary in different parts of the world, particularly as different criteria are used for diagnosis.

Diagnosis

Acute infective causes of diarrhoea must be excluded (**Box 22.8**), particularly *Giardia*, which can produce a syndrome very similar to tropical sprue. Malabsorption should be demonstrated, particularly of fat and vitamin B_{12}. The jejunal mucosa is abnormal, showing some villous atrophy (partial villous atrophy). In most cases, the lesion is less severe than that found in coeliac disease, although it affects the whole of the small bowel. Mild mucosal changes can be seen in asymptomatic individuals in the tropics.

Management

Many patients improve when they leave the sprue area and take folic acid (5 mg daily). Most patients also require an antibiotic to ensure a complete recovery (usually tetracycline 1 g daily for up to 6 months).

Severely ill patients require resuscitation with fluids and electrolytes for dehydration, and nutritional deficiencies should be corrected. Vitamin B_{12} (1000 μg) is also given to all acute cases.

Prognosis

The prognosis is excellent. Mortality is usually associated with water and electrolyte depletion, particularly in epidemics.

Bacterial overgrowth

The gut contains many resident bacteria in the terminal ileum and colon. Anaerobic bacteria, such as *Bacteroides*, bifidobacteria, are 100 to 1000 times more abundant than aerobic bacterial (facultative anaerobes), such as *Escherichia, Enterobacter*, and *Enterococcus*. This gut microflora has major functions, including metabolic ones, such as fermentation of non-digestible dietary residues into short-chain fatty acids as an energy source in the colon.

The microflora that influences epithelial cell proliferation is involved in the development and maintenance of the immune system and protects the gut mucosa from colonization by pathogenic bacteria. Bacteria also initiate vitamin K production.

The upper part of the small intestine is almost sterile, containing only a few organisms derived from the mouth. Gastric acid kills some ingested organisms and intestinal motility keeps bacterial counts in the jejunum low. The normal terminal ileum contains faecal-type organisms, mainly *Escherichia coli* and anaerobes, and the colon has abundant bacteria.

Bacterial overgrowth is normally found in association with a structural abnormality of the small intestine, such as a stricture or diverticulum, although it can occur occasionally in the elderly without such an abnormality. *E. coli* and/or *Bacteroides*, both in concentrations of greater than 10^6/mL, are found as part of a mixed flora. These bacteria are capable of deconjugating and dehydroxylating bile salts, so that unconjugated and dehydroxylated bile salts can be detected in small bowel aspirates.

Clinical features

The clinical features of overgrowth are chiefly diarrhoea and steatorrhoea. There may also be symptoms caused by the underlying small bowel pathology. Steatorrhoea (excess fat within the stool) occurs because of conjugated bile salt deficiency. Some bacteria

can metabolize vitamin B_{12} and interfere with its binding to intrinsic factor, leading to mild B_{12} deficiency (see Chapter 38); it is rarely severe enough to produce a neurological deficit. Some bacteria produce folic acid, giving a high serum folate. Bacterial overgrowth has only minimal effects on the absorption of other substances. Confirmation of bacterial overgrowth is with the hydrogen breath test:

- ***Hydrogen breath test*.** Bacteria are present in the oral cavity so the mouth should be rinsed out with an antiseptic mouthwash beforehand. The appearance of a breath hydrogen peak after ***oral glucose*** is used to estimate mouth-to-caecum transit time. An earlier rise in the breath hydrogen after glucose indicates bacterial breakdown in the small intestine.

  This test is simple to perform and does not involve radioisotopes. However, interpretation is often difficult, and sensitivity and specificity are low.

Management

If possible, the underlying lesion should be corrected (e.g., a stricture should be resected). If this is not possible then initial management should be with the non-absorbable antibiotic rifaximin, usually for 14 days. If using other antibiotics such as doxycycline, tetracycline, metronidazole, or amoxicillin-clavulanate, a 10-day course can be used.

Empirical treatment with a second course of antibiotics can be considered if there is a partial improvement in symptoms or symptoms recur within 3 months of initial antimicrobial treatment.

Limited data indicates that probiotics may be helpful in the management of small intestinal bacterial overgrowth, however, they are not helpful in preventing it.

Bile acid malabsorption

Bile is released from the gall bladder after eating and is essential for fat digestion. In a healthy gut, most of the bile is reabsorbed in the small intestine and recycled (enterohepatic bile acid circulation). However, this reabsorption is interrupted in some patients leading to more bile passing through into the colon. The increased concentration of bile acids in the colon leads to diarrhoea by reducing absorption of water and electrolytes as well as increasing secretion and colonic motility. Causes include:

- Crohn's disease with ileal disease
- ileal resection
- HIV infection
- idiopathic/primary bile acid diarrhoea
- post-infective gastroenteritis
- post-cholecystectomy.

Diagnosis is made by either a trial of medication or with a SeHCAT scan. This is carried out by giving oral ***SeHCAT*** (a synthetic taurine conjugate) and measuring the retention of the bile acid by whole-body counting at 7 days. Patients are managed by giving bile acid sequestrants such as colestyramine or colesevelam. These bind and inactivate the action of bile acids in the colon.

Lactose intolerance

This is a clinical syndrome where ingestion of lactose causes symptoms such as diarrhoea, abdominal pain, bloating, and flatulence. This is occasionally due to lactase deficiency but can also be due to secondary lactose malabsorption (usually due to underlying intestinal disease). There is a high incidence of lactase deficiency in many parts of the world (e.g., the Mediterranean countries and parts of Africa and Asia). There can be variability in symptoms among patients and a crossover with IBS symptoms. Symptoms occur after more than two servings of dairy a day. If lactose intolerance is suspected, the patient can be asked to eliminate lactose from their diet for 5 to 7 days. If symptoms recur on reintroduction of lactose, this is diagnostic of lactose intolerance. Symptoms usually fully resolve after lactose avoidance for 5 to 7 days.

Short bowel syndrome

This is a relatively rare malabsorptive disorder of the bowel that occurs in approximately 15% of patients undergoing an intestinal resection—it usually occurs after a massive resection or following multiple individual resections when less than 180 to 200 cm of the small bowel remains. Causes for intestinal resection include Crohn's disease, malignancy, and vascular diseases. There are causes of functional short bowel syndrome where the length of small bowel is unchanged but its function is significantly reduced. These include radiation enteritis, congenital villous atrophy, refractory sprue, chronic pseudo-obstruction syndrome, and trauma.

The GI tract carries out a number of absorptive functions along its length and removing large sections of bowel can lead to a number of consequences. The clinical presentation and its severity can vary however depending on a number of factors. These include the amount of absorptive surface remaining, any loss of site-specific transporters (e.g., B_{12} in terminal ileum), endocrine loss (e.g., GI hormones), presence of colon and/or ileo-caecal valve.

Malabsorption results in loss of fluids and nutrients leading to weight loss and symptoms of vitamin and micronutrient deficiency. Some of the symptoms can include abdominal pain, diarrhoea/high stoma output, and dehydration.

The following can also occur after ileal resection (**Fig. 22.35**):

- ***Bile-salt-induced diarrhoea*.** Bile salts and fatty acids enter the colon and cause malabsorption of water and electrolytes (see also 'Secretory diarrhoea' section).

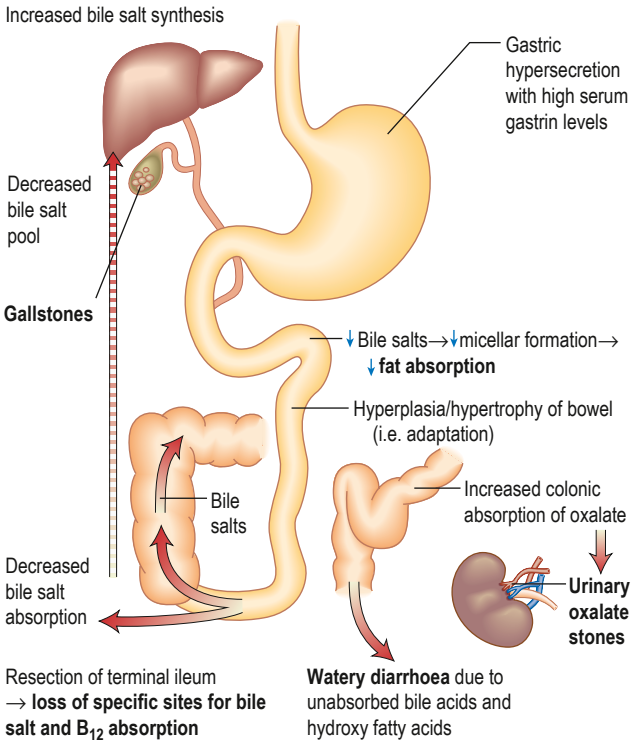


Fig. 22.35 The effects of resecting the distal small bowel.

- ***Steatorrhoea and gallstone formation*.** Increased bile salt synthesis can compensate for loss of approximately one-third of the bile salts in the faeces. Greater loss than this results in decreased micelle formation and steatorrhoea, and lithogenic bile and gallstone formation.
- ***Oxaluria and oxalate stones*.** Bile salts in the colon cause increased oxalate absorption with oxaluria, leading to urinary stone formation.
- ***B_{12} deficiency*.** Low serum B_{12}, macrocytosis, and other effects of B_{12} deficiency are seen.

Following a large intestinal resection, there is a process called intestinal adaptation, which occurs for up to 2 years afterwards and acts to increase the absorptive function of the remaining bowel. The evidence suggests that this happens either through the stimulatory effect of nutrients in the remaining section of small bowel or it is through a process of upregulating the peptide transporter (PepT1) in the colon. Intestinal adaptation is thought to have three phases:

- Phase 1 is the acute phase in the first 1 to 2 months (although can be up to 6) and is characterized by hypersecretion and excess fluid losses.
- Phase 2 sees adaptation occur gradually with improved absorptive function driven by intestinal hormones and growth factors that lead to structural and functional changes in the remaining bowel.
- Phase 3 is the maintenance phase, where maximal adaptation is reached at about 2 years.

Management of fluid and nutrient deficiencies is key and total parenteral nutrition (TPN) is often the mainstay of treatment initially. TPN prescription can be tailored to individual fluid and nutrient needs depending on the remaining section of bowel present. Other management tools include anti-secretory and antidiarrhoeal medications as well as digestive enzymes.

Intestinal failure and TPN are discussed further in Chapter 25. Broadly speaking, patients will often need complete fluid and nutritional support with TPN during phase 1. As the adaptation starts and the patient's absorptive capacity increases throughout phase 2, the PN can slowly be weaned, and once phase 3 is reached the PN can either be weaned completely, reduced in frequency or patients may only need i.v. fluid supplementation.

Whipple's disease

Whipple's disease is a rare infectious bacterial disease caused by *Tropheryma whipplei*. Some 87% of patients are males, and are most commonly white and middle-aged. Clinical features can progress over many years and can include:

- arthritis and arthralgia
- weight loss
- diarrhoea or steatorrhoea
- abdominal pain
- systemic symptoms of fever and weight loss
- peripheral lymphadenopathy and involvement of the heart, lung, joints, and brain occur, simulating many neurological conditions.

Blood tests show features of chronic inflammation and malabsorption. Endoscopy typically shows pale, shaggy duodenal mucosa with eroded, red, friable patches.

Diagnosis

Diagnosis is made by small bowel biopsy. Periodic acid–Schiff (PAS)-positive macrophages are present but are non-specific. On electron microscopy, the characteristic trilaminar cell wall of *T. whipplei* can be seen within macrophages. *T. whipplei* antibodies can be identified by immunohistochemistry. A confirmatory polymerase chain reaction (PCR)-based assay is available.

Management

Treatment is with antibiotics that cross the blood–brain barrier, such as 160 mg trimethoprim and 800 mg sulfamethoxazole (co-trimoxazole) daily for 1 year. This is preceded by a 2-week course of ceftriaxone or streptomycin and penicillin. Treatment periods of less than 1 year are associated with relapse in about 40%. If left untreated, it can be fatal.

Radiation enteritis

Radiation of more than 40 Gy will damage the intestine. The chronic effects of radiation are muscle fibre atrophy, ulcerative changes due to ischaemia, and obstruction due to radiation-induced fibrotic strictures.

Pelvic irradiation is frequently used for gynaecological and urinary tract malignancies, and so the ileum and rectum are the areas most often involved.

At the time of the irradiation, there may be nausea, vomiting, diarrhoea, and abdominal pain, usually improving within 6 weeks of completion of therapy.

Chronic radiation enteritis is diagnosed if symptoms persist for 3 months or more. The prevalence is higher than 15% of patients receiving radiotherapy that includes the abdomen. Abdominal pain due to obstruction is the main symptom. Malabsorption can be due to bacterial overgrowth in dilated segments and mucosal damage. Many patients suffer from increased bowel frequency. ***Management*** is symptomatic, although often unsuccessful in chronic radiation enteritis. Surgery should be avoided if possible, being reserved for obstruction or perforation. Hyperbaric oxygen has been shown to be beneficial but it is not available in many centres.

Acute radiation damage to the rectum produces a ***radiation proctitis*** with diarrhoea and tenesmus, with or without blood. Local steroids sometimes help initially. When the acute phase heals, mucosal telangiectases form and may cause persistent bleeding. If there is resistant anaemia, these can be treated with argon plasma coagulation or, under a light anaesthetic, by packing the rectum with a formalin-soaked swab for 2 minutes, both of which destroy the telangiectases.

Parasite infestation

- ***Giardia intestinalis*** (see Chapter 42) not only produces diarrhoea but also can produce malabsorption with steatorrhoea. Minor changes are seen in the jejunal mucosa and the organism can be found in the jejunal fluid or mucosa.
- ***Cryptosporidiosis*** (see Chapter 42) can also produce malabsorption.
- ***HIV infection***, when uncontrolled, causes patients to be particularly prone to parasitic infestation (**Box 22.32**).

Other causes of malabsorption

- ***Drugs that bind bile salts*** (e.g., colestyramine) and some antibiotics (e.g., neomycin) produce steatorrhoea.
- ***Orlistat*** is used in obesity to reduce fat absorption by inhibiting gastric and pancreatic lipase, so causing diarrhoea and steatorrhoea. A low-fat diet is also necessary, which leads to weight loss.
- ***Thyrotoxicosis*** causes diarrhoea, rarely with steatorrhoea, owing to increased gastric emptying and increased motility.
- ***Zollinger–Ellison syndrome*** is described in Chapter 24.
- ***Intestinal lymphangiectasia*** produces diarrhoea and, rarely, steatorrhoea.

Box 22.32 Gastrointestinal problems in patients with AIDS

Symptoms	Causes
Mouth/oesophagus	
Dysphagia	Herpes simplex virus (HSV)
Retrosternal discomfort	Cytomegalovirus (CMV)
Oral ulceration	Candidiasis
Small bowel/colon	
Chronic diarrhoea Steatorrhoea Weight loss	**Parasites:**
	Entamoeba histolytica
	Giardia intestinalis
	Blastocystis hominis
	Isospora belli
	Microsporidia
	Cyclospora cayetanensis
	Viruses:
	Cytomegalovirus, herpes simplex virus, adenovirus
	Bacteria:
	Salmonella
	Campylobacter
	Shigella
	Mycobacterium avium-intracellulare
	Non-infective enteropathy—cause unknown
Rectum/colon	
Bloody diarrhoea	Bacterial infection (e.g., *Shigella*)
Any site	
Weight loss Diarrhoea	**Neoplasia:**
	Kaposi's sarcoma
	Lymphoma
	Squamous carcinoma
	Infection—disseminated, e.g.,:
	Mycobacterium avium-intracellulare
	Anti-retroviral therapy (ART)

- ***Lymphoma*** that has infiltrated the small bowel mucosa causes malabsorption.
- ***Diabetes mellitus*** (see Chapter 35) causes diarrhoea, malabsorption, and steatorrhoea, sometimes due to bacterial overgrowth from autonomic neuropathy that leads to small bowel stasis.
- ***Hypogammaglobulinaemia***, which is seen in a number of conditions including lymphoid nodular hyperplasia, causes steatorrhoea due either to an abnormal jejunal mucosa or to secondary infestation with *Giardia intestinalis.*

Miscellaneous intestinal diseases

Protein-losing enteropathy

Protein-losing enteropathy refers to intestinal conditions that lead to protein loss, and usually manifest with hypoalbuminaemia. The causes include:

- Crohn's disease
- tumours
- Ménétrier's disease, a condition with giant rugal folds (Fig. 22.36)

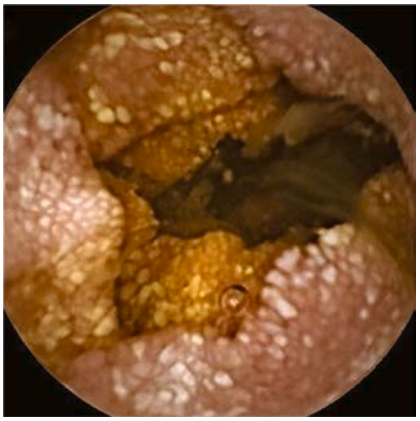


Fig. 22.36 Protein-losing enteropathy showing giant rugal folds.

- coeliac disease
- lymphatic disorders (e.g., lymphangiectasia).

Usually, protein-losing enteropathy forms a minor part of the generalized disorder but, occasionally, hepatic synthesis of albumin cannot compensate for the protein loss, and peripheral oedema dominates the clinical picture. Confirmation of the diagnosis is by measurement of α_1-antitrypsin clearance. It does not require an isotope. Alpha$_1$-antitrypsin is a large molecule (>50 000 daltons), which is resistant to proteolysis. Simultaneous measurements of serum and stool concentration (24-hour collection) are made. The treatment is that of the underlying disorder.

Meckel's diverticulum

This is the most common congenital abnormality of the gastrointestinal tract, affecting 2% to 3% of the population. The diverticulum projects from the wall of the ileum approximately 60 cm from the ileocaecal valve. It is usually symptomless. However, 50% contain gastric mucosa that secretes hydrochloric acid, and peptic ulcers can occur and may bleed or perforate.

Acute inflammation of the diverticulum also occurs and is indistinguishable clinically from acute appendicitis. Rarely, there is obstruction from an associated band.

Management is surgical removal, often laparoscopically.

Tuberculosis

Tuberculosis (TB; see also Chapter 20) can affect the intestine, as well as the peritoneum (see 'Peritoneum' section). In high income countries, most patients are from ethnic minority groups, or are immunocompromised because of HIV or drugs. Intestinal TB is due to reactivation of primary disease caused by *Mycobacterium tuberculosis*. Bovine TB occurs in areas where milk is unpasteurized and is rare in Western countries.

Clinical features

These are abdominal pain, weight loss, anaemia, fever with night sweats, obstruction, right iliac fossa pain, or a palpable mass. The ileocaecal area is most commonly affected, but the colon—and, rarely, other parts of the gastrointestinal tract—can be involved. One-third of patients present acutely with intestinal obstruction or generalized peritonitis; 50% have X-ray evidence of pulmonary TB.

Diagnosis

Differential diagnosis includes Crohn's disease and caecal carcinoma.

- ***Small bowel follow-through*** may show transverse ulceration, and diffuse narrowing of the bowel with shortening of the caecal pole.
- ***Ultrasound, MRI, or CT*** shows additional mesenteric thickening and lymph node enlargement.
- ***Histology and culture of tissue*** is desirable but not always possible. Specimens can be obtained by colonoscopy or laparoscopy but laparotomy is required in some cases. The histological findings include chronic inflammation with caseating granuloma. Acid-fast bacilli may be seen on dedicated stains. However, the histology is not always distinguishable from that of Crohn's disease.

Management

Drug treatment is similar to that for pulmonary TB (see Chapter 20). Treatment should be started if there is a high degree of suspicion.

Amyloidosis

Systemic amyloidosis (see Chapter 36) may affect any part of the gastrointestinal tract. Rectal biopsy may be diagnostic. Occasionally, amyloid deposits occur as polypoid lesions. The symptoms depend on the site of involvement; amyloidosis in the small intestine gives rise to diarrhoea.

Autoimmune rheumatic diseases

Systemic sclerosis (see Chapter 32) most commonly affects the oesophagus, although the small bowel and colon are often found to be involved if investigated. There may be no symptoms of this involvement, but diarrhoea and steatorrhoea can occur due to bacterial overgrowth caused by reduced motility, dilation, and the presence of diverticula.

Intestinal ischaemia

Intestinal ischaemia results from occlusion of arterial inflow, occlusion of venous outflow, or failure of perfusion; these factors may act alone or in combination and usually occur in the elderly.

- ***Arterial inflow occlusion*** can be caused by atheroma, thrombosis, and embolism (cardiac arrhythmia), including cholesterol emboli, aortic disease (occluding ostia of mesenteric vessels), vasculitis (including Takayasu's arteritis) or thromboangiitis.
- ***Venous outflow occlusion*** occurs in 5% to 15% of cases, usually in sick patients with circulatory failure.
- ***Infarction without occlusion*** can occur due to reduced cardiac output, hypotension, and shock, causing reduced intestinal blood flow.

Acute small intestinal ischaemia

An embolus from the heart in a patient with atrial fibrillation is the most common cause and usually occludes the superior mesenteric artery. Patients present with sudden abdominal pain and vomiting, with a distended and tender abdomen and absent bowel sounds. The patient is hypotensive and ill. Surgery is necessary to resect the gangrenous bowel. Mortality is high (up to 90%) and is related to coexisting disease, the development of multiorgan failure (MOF; see Chapter 13), and massive fluid and electrolyte losses in the postoperative period. Survivors may go on to develop nutritionally inadequate short bowel syndrome (see Chapter 25).

Ischaemic colitis

See 'Ischaemic disease of the colon (ischaemic colitis)' section.

Chronic small intestinal ischaemia

This is due to atheromatous occlusion or cholesterol emboli of the mesenteric vessels, and is found particularly in the elderly. Good collateral circulation can minimize clinical effects. The characteristic symptom is postprandial abdominal pain and weight loss. Loud bruits may be heard but, as these are heard in normal subjects, they are of doubtful significance. The diagnosis is made using angiography.

Eosinophilic gastroenteritis

In this condition of unknown aetiology there is eosinophilic infiltration and oedema of any part of the gastrointestinal mucosa. The gastric antrum and proximal small intestine are usually involved, hosting either a localized lesion (eosinophilic granuloma) or diffuse sheets of eosinophils in the serosal and submucosal layers. There is an association with asthma, eczema, and urticaria.

The condition occurs mainly in the third decade. The clinical presentation depends on the site of gut involvement. Abdominal pain, nausea and vomiting, and upper gastrointestinal bleeding occur. Peripheral eosinophilia is present in only 20% of patients. Endoscopic biopsy is useful for making the diagnosis histologically. Radiology may demonstrate mass lesions.

Treatment is with corticosteroids for the widespread infiltration, particularly if peripheral eosinophilia is present.

In some adults, the condition appears to be allergic (allergic gastroenteritis) and is associated with peripheral eosinophilia and high levels of plasma and tissue IgE. The relationship of eosinophilic oesophagitis (see 'Eosinophilic oesophagitis' section) to eosinophilic gastroenteritis is unclear.

Intestinal lymphangiectasia

Dilation of the lymphatics may be primary or secondary to lymphatic obstruction, such as occurs in malignancy or constrictive pericarditis. Hypoproteinaemia with ankle oedema is the main feature. The rare primary form may be detected incidentally as dilated lacteals on a jejunal biopsy or it can produce steatorrhoea of varying degrees. White-tipped villi are seen on capsule endoscopy. Serum immunoglobulin levels are reduced, with low circulating lymphocytes. Management is with a low-fat diet, mid-chain triglycerides, and fat-soluble vitamin supplements as required. Octreotide has a dramatic effect in a few primary cases, although the mechanism of action is unknown.

Abetalipoproteinaemia

This rare congenital disorder is due to a failure of apo B-100 synthesis in the liver and apo B-48 in the intestinal cell, so that chylomicrons are not formed. This leads to fat accumulation in the intestinal cells, lending a characteristic histological appearance to the jejunal mucosa. Clinical features include acanthocytosis (spiky red cells owing to membrane abnormalities), a form of retinitis pigmentosa, and mental and neurological abnormalities. The latter can be prevented by vitamin E injections.

Tumours of the small intestine

The small intestine is relatively resistant to the development of neoplasia and only 3% to 6% of all gastrointestinal tumours and less than 1% of all malignant lesions occur here. The reason for the rarity of tumours is unknown. Explanations include the fluidity and relative sterility of small bowel contents and the rapid transit time, reducing the time of exposure to potential

carcinogens. It is also possible that the high population of lymphoid tissue and secretion of IgA in the small intestine protect against malignancy.

Adenocarcinoma and lymphoma

Adenocarcinoma of the small intestine is rare and found most frequently in the duodenum (in the periampullary region) and in the jejunum. It is the most common tumour of the small intestine, accounting for up to 50% of primary tumours.

Lymphomas are most frequently found in the ileum. These are of the non-Hodgkin's type and must be distinguished from peripheral or nodal lymphomas involving the gut secondarily.

In high income countries, the most common type of lymphoma is the B cell type arising from MALT. These lymphomas tend to be annular or polypoid masses in the distal or terminal ileum, whereas most T-cell lymphomas are ulcerated plaques or strictures in the proximal small bowel.

A tumour similar to Burkitt's lymphoma also occurs and commonly affects the terminal ileum of children in North Africa and the Middle East.

Predisposing factors for adenocarcinoma and lymphoma

Coeliac disease

There is an increased incidence of lymphoma of the T cell type and adenocarcinoma of the small bowel, as well as an unexplained increase in all malignancies, both in the gastrointestinal tract and elsewhere. The reason for the local development of malignancy is unknown. It is now accepted that coeliac disease is a pre-malignant condition but there is no association with the length of the symptoms. Management with a gluten-free diet can reduce the risk of both lymphoma and carcinoma.

Crohn's disease

There is a small increase in the incidence of adenocarcinoma of the small bowel in Crohn's disease.

Immunoproliferative small intestinal disease

Immunoproliferative small intestinal disease (IPSID) is a rare B cell disorder in which there is proliferation of plasma cells in the lamina propria of the upper small bowel, producing truncated monoclonal heavy chains, without associated light chains. The α heavy chains are found in the gut mucosa on immunofluorescence and can also be detected in the serum. IPSID usually occurs in countries surrounding the Mediterranean, but it has also been found in low and middle income countries in South America and the Far East. It predominantly affects people in lower socioeconomic groups in areas with poor hygiene and a high incidence of bacterial and parasitic infection of the gut. IPSID presents as a malabsorptive syndrome associated with diffuse lymphoid infiltration of the small bowel and neighbouring lymph nodes, progressing in some cases to a lymphoma. The condition has also been documented in the developed world.

Clinically, patients present with abdominal pain, diarrhoea, anorexia, weight loss, and symptoms of anaemia. There may be a palpable mass, and an ultrasound followed by an MRI scan may detect a mass lesion. Endoscopic biopsy is useful where lesions are within reach. Ultrasound and CT may show bowel wall thickening and the involvement of lymph nodes, which is common with lymphoma. Wireless capsule endoscopy can be used where obstruction by the capsule is not likely, but cannot deliver histology.

Management of small intestinal tumours

Adenocarcinoma

Most patients are treated surgically with a segmental resection. The overall 5-year survival rate is 20% to 35%; this varies with the histological grade and the presence or absence of lymph node involvement. Radiotherapy and chemotherapy are used in addition.

IPSID

If there is no evidence of lymphoma, antibiotics, such as tetracycline, should be tried initially. In the presence of lymphoma, combination chemotherapy is used; in one series, the 3- to 5-year survival rate was 58%.

Lymphoma

Most patients require surgery and radiotherapy with chemotherapy for more extensive disease. The prognosis varies with the type. The 5-year survival rate for T cell lymphomas is 25% but is better for B cell lymphomas, varying from 50% to 75%, depending on the grade of lymphoma.

Carcinoid tumours

These originate from the enterochromaffin cells (APUD cells) of the intestine. They make up 10% of all small bowel neoplasms, the most common sites being the appendix and terminal ileum. It is often difficult to be certain histologically whether a particular tumour is benign or malignant. A total of 10% of carcinoid tumours in the appendix present as acute appendicitis, secondary to obstruction. Surgical resection of the tumour is usually performed.

Most carcinoids do not secrete hormones or vasoactive compounds, and may present with liver enlargement due to metastases.

Carcinoid syndrome occurs in only 5% of patients with carcinoid tumours and only when there are liver metastases. Biologically active amines and peptides secreted by carcinoid tumours are usually subject to first-pass metabolism by the liver, but may enter the systemic circulation in patients with liver metastases. Patients complain of spontaneous or induced bluish-red flushing, predominantly on the face and neck, sometimes leading to permanent changes with telangiectases.

Gastrointestinal symptoms consist of abdominal pain and recurrent watery diarrhoea. Cardiac abnormalities are found in 50% of patients and take the form of pulmonary stenosis or tricuspid incompetence. Examination of the abdomen reveals hepatomegaly. The tumours secrete a variety of biologically active amines and peptides, including serotonin (5-hydroxytryptamine, 5-HT), bradykinin, histamine, tachykinins, and prostaglandins. The diarrhoea and cardiac complications are probably caused by 5-HT itself, but the cutaneous flushing is thought to be produced by one of the kinins, such as bradykinin. This is known to cause vasodilation, bronchospasm, and increased intestinal motility.

Diagnosis of carcinoid syndrome

- ***Ultrasound examination*** confirms the presence of liver secondary deposits.
- ***Urine*** shows a high concentration of 5-hydroxyindoleacetic acid (5-HIAA), which is the major metabolite of 5-HT.
- ***Serum chromogranin A*** is raised in nearly all hindgut tumours and 80% to 90% of symptomatic foregut and midgut tumours.

Management of carcinoid syndrome

Treatment is with octreotide and lanreotide; both are octapeptide somatostatin analogues that inhibit the release of many gut hormones. They alleviate flushing and diarrhoea, and can control a carcinoid crisis. Interferon and other chemotherapeutic regimens also

Fig. 22.37 Circumoral pigmentation in Peutz-Jeghers syndrome.

occasionally reduce tumour growth, but have not been shown to increase survival.

Most patients survive for 5 to 10 years after diagnosis.

Peutz–Jeghers syndrome

This consists of mucocutaneous pigmentation (circumoral in 95% of patients, and on the hands in 70% and feet in 60%) and gastrointestinal polyps (**Fig. 22.37**). It has an autosomal dominant inheritance. The gene *STK11* (also known as *LKB1*) that is responsible for Peutz–Jeghers syndrome codes for a serine protein kinase and can be used for genetic analysis. The brown buccal pigment is characteristic of the condition. The polyps, which are hamartomas, can occur anywhere in the gastrointestinal tract but are most frequent in the small bowel. They may bleed or cause small bowel obstruction or intussusception (50% of patients).

Management

Management is by endoscopic polypectomy. Balloon enteroscopy may be necessary to reach all the small bowel polyps. Bowel resection should be avoided if possible, but may be necessary in patients presenting with gangrenous bowel due to intussusception. Follow-up is with yearly pan-endoscopy. There is an increased incidence of gastrointestinal cancers. Non-gastrointestinal cancers also occur with increased frequency, so yearly screening for uterine, ovarian, and cervical cancer should start in the teens, and breast and testicular screening by the age of 20.

Other tumours

Adenomas, lipomas, and stromal tumours (see 'Gastrointestinal stromal tumours' section) are rarely found and are usually asymptomatic and picked up incidentally. They occasionally present with iron deficiency anaemia. In familial adenomatous polyposis (FAP), duodenal adenomas form in one-third of patients and may progress to adenocarcinoma. This is the most common cause of death in FAP patients who have been treated by prophylactic colectomy.

Further reading

Catassi C, Verdu EF, Cesar Bai J et al. Coeliac disease. *Lancet* 2022; 399:2413–2426.
El-Abassi R, Soliman MY, William F et al. Whipple's disease. *J Neurol Sci* 2017; 377:197–206.
Massironi S, Cavalcoli F, Rausa E et al. Understanding short bowel syndrome: current status and future perspectives. *Dig Liver Dis* 2020; 52(3):253–261.
Penny HA, Baggus EMR, Rej A et al. Non-responsive coeliac disease: a comprehensive review. *Nutrients* 2020; 12(1):216.
Vasant DH, Paine PA, Black CJ et al. British Society of Gastroenterology guidelines on the management of irritable bowel syndrome. *Gut* 2021; 70:1214–1240.

INFLAMMATORY BOWEL DISEASE

Two major forms of IBD are recognized:

- Crohn's disease (CD), which can affect any part of the gastrointestinal tract
- Ulcerative colitis (UC), which affects only the colon.

There is a degree of overlap between these two conditions in their aetiopathogenesis, clinical features, histological and radiological abnormalities; in 10% of cases of IBD-causing colitis, a definitive diagnosis of either UC or CD is not possible and the diagnosis is termed colitis of undetermined type (indeterminate colitis). It is clinically useful to distinguish between UC and CD because of differences in their management, although, in reality, they may represent two aspects of the same disease.

Another form of colitis related to microscopic inflammation is termed microscopic colitis; this is subdivided into lymphocytic and collagenous types (see 'Microscopic colitis' section). The distinction between this and IBD is the absence of macroscopic evidence of inflammation.

Epidemiology

- The incidence of CD varies from country to country but is approximately 4 to 10 per 100 000 annually, with a prevalence of 25 to 100/100 000.
- The incidence of UC is stable at 6 to 15/100 000 annually, with a prevalence of 80 to 150/100 000.

Although both conditions have a worldwide distribution, the highest incidence rates and prevalence have been reported from Northern Europe, the UK, and North America. Both race and ethnic origin affect the incidence and prevalence of CD and UC. Thus, in North America, prevalence rates of CD are lower in Hispanic and Asian people (4.1/100 000 and 5.6/100 000, respectively) compared with white individuals (43.6/100 000). Jewish people are more prone to IBD than any other ethnic group. Prevalence rates also change after migration; thus there is an increasing incidence of CD in the UK-born children of migrants from South-east Asia. Recent studies suggest that the incidence of both CD and UC is increasing in traditional low-prevalence areas such as South-east Asia.

Approximately 25% of patients are diagnosed before their 18th birthday and there is evidence that disease commencing in youth is more extensive and more aggressive than that occurring in older patients.

Aetiology and pathogenesis

Although the aetiology of IBD is unknown, it is increasingly clear that IBD represents the interaction between several co-factors: genetic susceptibility, the environment, the intestinal microbiota, and host immune response (**Fig. 22.38**) See **Box 22.33**.

Genetic factors

CD and UC are complex polygenic diseases and having a positive family history is the largest independent risk factor for development of IBD. Up to 1 in 5 patients with CD and 1 in 6 patients with UC will have a first-degree relative with the disease. The monozygotic and dizygotic twin concordance rates for CD are 20% to 50% and 10%, respectively.

Genome-wide association studies have identified multiple susceptibility loci, and many of the underlying risk variants have been identified. The major genetic factors for CD include the *NOD2 (CARD 15)* gene (nucleotide oligomerization domain 2), the autophagy genes, and the Th17 pathway (IL-23–type 17 helper T cells). The NOD2 protein on chromosome 16 is an intracellular sensor of bacterial peptidoglycan, present in bacterial cell walls (see below). NOD2 is expressed in epithelial cells, macrophages, and endothelial

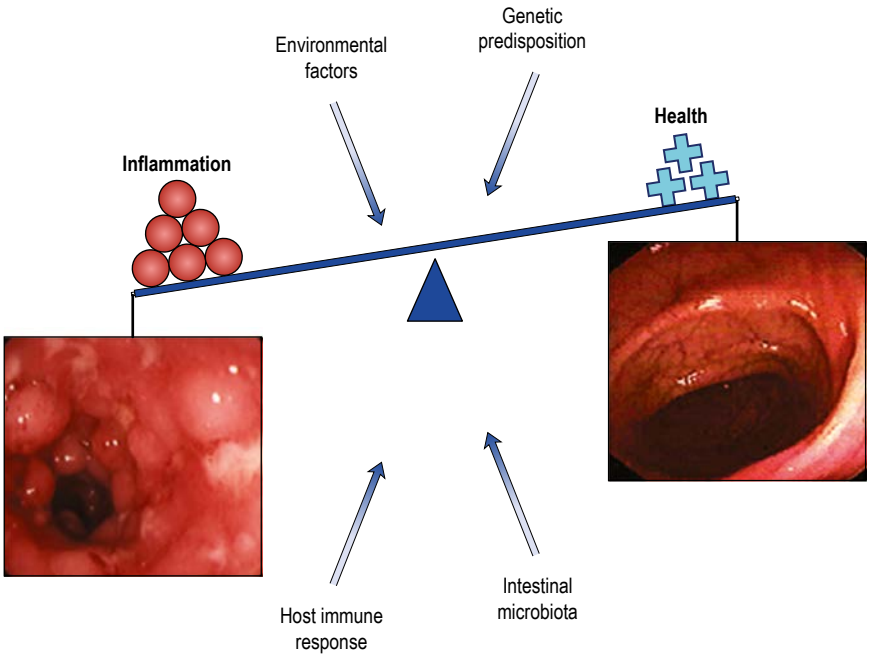


Fig. 22.38 **The aetiopathogenesis of inflammatory bowel disease.**

Box 22.33 Pathogenesis of IBD: summary

- There is overwhelming evidence that inflammatory bowel disease (IBD) is a result of dysregulated immune responses to luminal bacteria and/or their products.
- Evidence suggests that IBD patients demonstrate fundamental differences in the numbers and functionality of intestinal microbiota.
- Various genetic variations that are important in pathways regulating intestinal inflammation have been found.
- Novel agents targeting immune dysregulation, changes in intestinal microenvironment are in operational development.

cells. Individuals who are homozygote or compound heterozygote for one of several mutations in the *NOD2* gene have a significantly increased risk of developing ileal CD. Likewise, mutations in the autophagy genes *ATG16L1* and *IRGM* (immunity-related GTP-ase M-protein) and IL-23 receptor gene increase CD risk, and mutations in genes associated with the mucosal barrier increase UC risk. However, the presence of IBD-associated genes in many unaffected individuals and the failure of the approximately 71 genetic susceptibility loci identified thus far to explain more than around one-fifth of the genetic risk of CD highlight the complexity of the genetic basis of IBD. Specific genetic defects in the IL-10 receptor pathway are associated with a severe form of extremely early-onset colitis and perianal disease in children.

Apart from susceptibility, HLA genes on chromosome 6 also appear to have a role in modifying the disease. The *DRB*0103* allele is linked to a particularly aggressive course of UC and the need for surgery, as well as with colonic CD. *DRB*0103* and *MICA*010* are associated with perianal disease, and *DRB*0701* with ileal CD. For the extraintestinal disease complications and their HLA associations, see 'Extraintestinal manifestations' section.

Environmental and other factors

- ***Smoking*.** Patients with CD are more likely to be smokers, and smoking has been shown to exacerbate CD and increase the risk of disease recurrence after surgery. By contrast, there is an increased risk of UC in non- or ex-smokers and nicotine has been shown to be an effective treatment in one small clinical trial.
- ***NSAIDs*.** NSAID ingestion is associated with both the onset of IBD and flares of disease in patients with an established diagnosis.
- ***Hygiene*.** Poor and large families living in crowded conditions have a lower risk of developing CD. A 'clean' environment may not expose the intestinal immune system to pathogenic or non-pathogenic microorganisms such as helminths, which seems to alter the balance between effector and regulatory immune responses.
- ***Nutritional factors*.** Many foods and food components have been suggested as playing a role in the aetiopathogenesis of IBD (e.g., high sugar and fat intake) but, unfortunately, the results of numerous studies designed to define risk have been equivocal. However, breast-feeding may provide protection against the development of IBD in offspring.
- ***Psychological factors*.** Factors such as chronic stress and depression seem to increase relapses in patients with quiescent disease.
- ***Appendicectomy*.** This appears to be 'protective' against the development of UC, particularly if performed for appendicitis or for mesenteric lymphadenitis before the age of 20. It also influences the clinical course of UC, with a lower incidence of colectomy and reduced need for immunosuppressive therapy. By contrast, appendicectomy may increase the risk of development of CD.

Intestinal microbiota

The gut is colonized by 10 times more bacterial organisms than there are host cells, there being 300 to 400 distinct bacterial species within each host intestine. The intestinal microbiota plays a crucial role in perpetuating intestinal inflammation, in both animal models of disease and patients with IBD. The number of mucosal adherent

bacteria is increased in patients with CD compared to healthy subjects, and diversion of the bacterial component of the faecal stream induces clinical remission. However, there is also evidence of an immunoregulatory role for the commensal microbiota, which protects against intestinal inflammation and upregulates epithelial defence mechanisms in animal models of colitis.

Mechanisms by which the intestinal microbiota may relate to the aetiology of IBD include:

- ***Intestinal dysbiosis*.** There is an alteration in the bacterial flora in patients with CD. Although results vary due to differences in both the patient groups studied and the microbiological method utilized, the most consistent finding in patients with IBD is a reduced diversity of microbial species. In addition, higher concentrations of *Bacteroides* and *E. coli*, and lower concentrations of bifidobacteria and *Faecalibacterium prausnitzii* have been reported in faecal and mucosal samples from patients with CD compared to healthy controls. Lower concentrations of *F. prausnitzii* have been found in patients with active compared with quiescent disease, and low levels of this organism in CD resection specimens predict subsequent endoscopic disease recurrence.
- ***Specific pathogenic organisms*.** It has been shown that there is increased *E. coli* adherence to the ileal epithelial cells in CD, with evidence of invasion into the mucosa. *E. coli's* type 1 pili adhere to a protein called carcinoembryonic antigen-related cell adhesion molecule 6 (CEACAM6). Many authors have also suggested a link between CD and *Mycobacterium paratuberculosis* (MAP), although recent PCR-based studies have failed to confirm this and therapeutic trials of anti-*Mycobacterium tuberculosis* (MTB) therapy were not effective.
- ***Bacterial antigens*.** Bacteria exert their influence by the interaction of ligands such as peptidoglycan-polysaccharides (PG-PS) and LPS with host pattern recognition receptors such as the TLR family (cell surface) and the NOD family (intracellular).
- ***Defective chemical barrier or intestinal defensins*** (see 'Small intestine physiology [innate chemical defence]' section). Evidence suggests a decrease in human α defensin-1 (HD-1) in the mucosa of both CD and UC, and a lack of induction of HD-2, HD-3, and HD-5 in CD.
- ***Impaired mucosal barrier function*.** This may explain the presence of unusual and potentially pathogenic bacteria, such as MAP, *Listeria*, and mucosal adherent *E. coli.* However, their presence does not necessarily imply causation of the disease, and they may reflect previous disease activity.

Intestinal immune system

IBD occurs when the mucosal immune system exerts an inappropriate response to luminal antigens, such as bacteria, which may enter the mucosa via a leaky epithelium (Fig. 22.39). Bacterial ligands interact with the innate and acquired mucosal immune system via TLRs expressed on both epithelial and antigen-presenting cells. Deficiencies occur in the clearance of invading bacteria by aspects of the innate immune system, such as neutrophils, which may allow inappropriate activation of the acquired immune system. In keeping with the genetic susceptibility loci identified, these findings highlight a deficiency in patients with IBD in a component of the inflammasome (an intracellular danger sensor of the innate immune system that can trigger caspase-1-dependent processing of inflammatory mediators, such as IL-1β and IL-18). In addition, individual bacterial species have distinct immunological effects mediated by dendritic cells (DCs), which sample bacteria from the intestinal lumen and direct the subsequent functional differentiation of naïve T cells into effector or regulatory populations. IBD is associated with an imbalance in the relative numbers of intestinal homing effector (Th1 and Th17) and regulatory T-cell populations, which disturbs the normal tolerance to the luminal antigenic load.

The pro-inflammatory cytokines released by these activated effector T cells stimulate macrophages to secrete pro-inflammatory cytokines, such as tumour necrosis factor-alpha (TNF-α), IL-1, and IL-6 in large quantities. These mechanisms result in increased adhesion molecule expression on the intestinal vascular endothelium, which facilitates the recruitment of leucocytes from the circulation and the release of chemokines, all of which lead to tissue damage and also attract more inflammatory cells in a vicious circle.

Pathology

CD is a chronic inflammatory condition that may affect any part of the gastrointestinal tract from the mouth to the anus but has a particular tendency to affect the terminal ileum and ascending colon (ileocolonic disease) (Fig. 22.40). The disease can involve one small area of the gut, such as the terminal ileum, or multiple areas with relatively normal bowel in between (skip lesions). It may also involve the whole of the colon (total colitis), sometimes without macroscopic small bowel involvement. It is also associated with the development of perianal fistulae and fissures (Box 22.34).

UC can affect the rectum alone (proctitis), can extend proximally to involve the sigmoid and descending colon (left-sided colitis), or may involve the whole colon (extensive colitis) (Fig. 22.41). In a few of these patients, there is also inflammation of the distal terminal ileum (backwash ileitis).

Macroscopic changes

In ***CD***, the involved bowel is usually thickened and is often narrowed. Deep ulcers and fissures in the mucosa produce a cobblestone appearance. Intra-abdominal fistulae and abscesses may be seen, which reflect penetrating disease. An early feature is aphthoid ulceration in the colon, usually seen at colonoscopy (Fig. 22.42); later, larger and deeper ulcers appear in a patchy distribution, again producing a cobblestone appearance.

In ***UC***, the mucosa looks reddened and inflamed, and bleeds easily (friability). In severe disease, there may be extensive ulceration, with the adjacent mucosa appearing as post-inflammatory (pseudo-) polyps.

In ***fulminant colonic disease*** of either type, most of the mucosa is lost, leaving a few islands of oedematous mucosa (mucosal islands), and toxic dilation occurs. On healing, the mucosa can return to normal, although there is usually some residual scarring.

Microscopic changes

In ***CD***, the inflammation extends through all layers (transmural) of the bowel, whereas in UC superficial inflammation limited to the mucosa is seen. In CD, there is an increase in chronic inflammatory cells and lymphoid hyperplasia, and in 50% to 60% of patients, granulomas are present. These granulomas are non-caseating epithelioid cell aggregates with Langhans' giant cells.

In ***UC***, the mucosa shows a chronic inflammatory cell infiltrate in the lamina propria. Crypt abscesses and goblet cell depletion are also seen.

These two diseases can usually be differentiated not only on the basis of clinical and radiological data but also on the histological differences seen in the rectal and colonic mucosa obtained by biopsy (Box 22.35).

It is occasionally not possible to distinguish between the two disorders, particularly if biopsies are obtained in the acute phase,

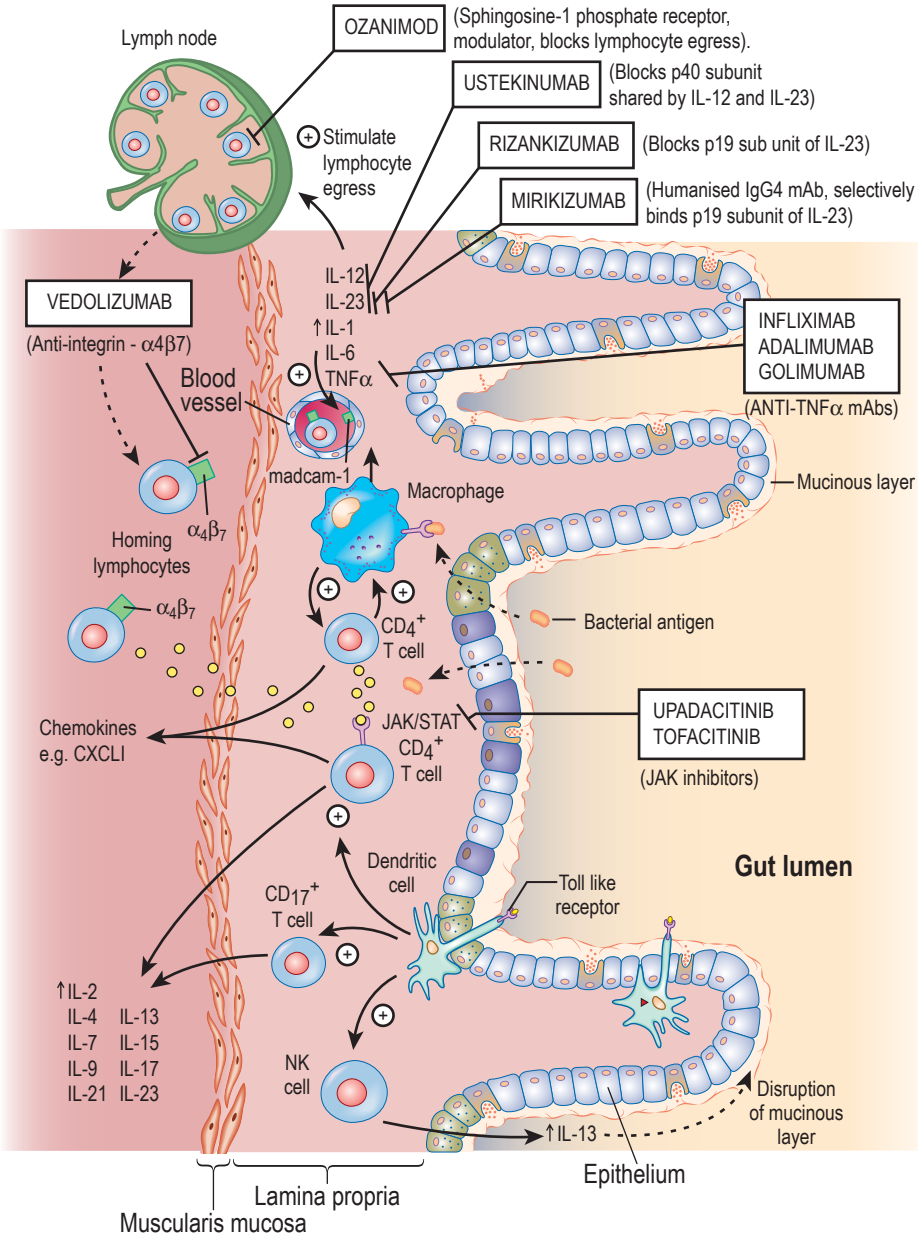


Fig. 22.39 Cellular intestinal mechanisms involved in the pathogenesis of inflammatory bowel disease. Bacterial ligands attach to the epithelium and antigen-presenting cells (e.g., dendritic cells and macrophages) via Toll-like receptors and the NOD protein. The dendritic cells migrate to the Peyer's patches and mesenteric lymph nodes. They present antigens to naïve T cells, releasing a number of cytokines that cause damage. The T-cell subsets enter the blood vessels, where the enterotropic molecules α4-β7 are induced and allow homing back to the gut, causing further damage. Labels show biologics currently approved for treatment of inflammatory bowel disease and their molecular target. *IFN-γ,* Interferon-gamma; *IL,* interleukin; *TGF-β,* transforming growth factor-beta; *TNF-α,* tumour necrosis factor-alpha.

and such patients are considered to have indeterminate colitis. Serological testing for anti-neutrophil cytoplasmic antibodies (ANCA) in UC and anti-*Saccharomyces cerevisiae* antibodies (ASCA) in CD may be of value in differentiating the two conditions (see below), although an exact diagnosis can sometimes be made only after examining a surgical colectomy specimen. Occasionally, examination of the colectomy specimen still does not lead to a diagnosis of CD or UC and the patient is labelled as having indeterminate colitis.

Extraintestinal manifestations

These occur with both diseases (Box 22.36). Joint complications are most common; the peripheral arthropathies are classified as:

- ***Type 1 (pauciarticular)*** attacks are acute and self-limiting (<10 weeks), and occur with IBD relapses; they are associated with other extraintestinal manifestations of IBD activity.
- ***Type 2 (polyarticular)*** arthropathy lasts longer (months to years), is independent of IBD activity and usually associated with uveitis.

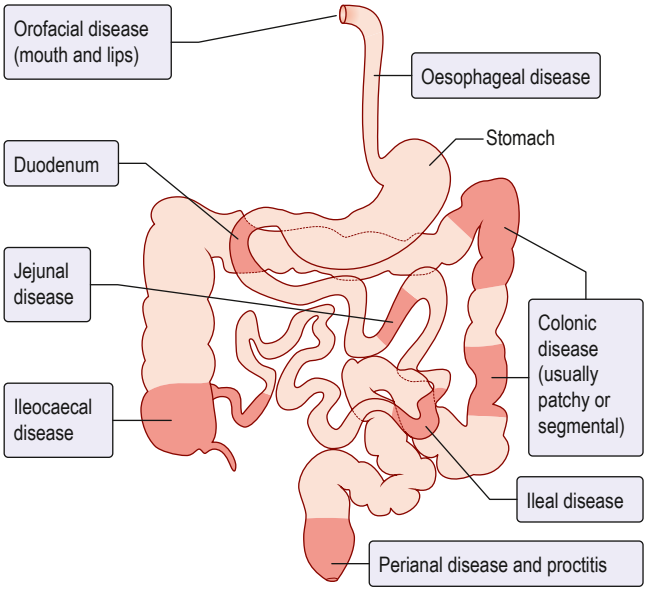


Fig. 22.40 Sites of Crohn's disease.

Box 22.34 Anal and perianal complications of Crohn's disease

- Fissure *in ano* (multiple and indolent)
- Haemorrhoids
- Skin tags
- Perianal abscess
- Ischiorectal abscess
- Fistula *in ano* (may be multiple)
- Anorectal fistulae

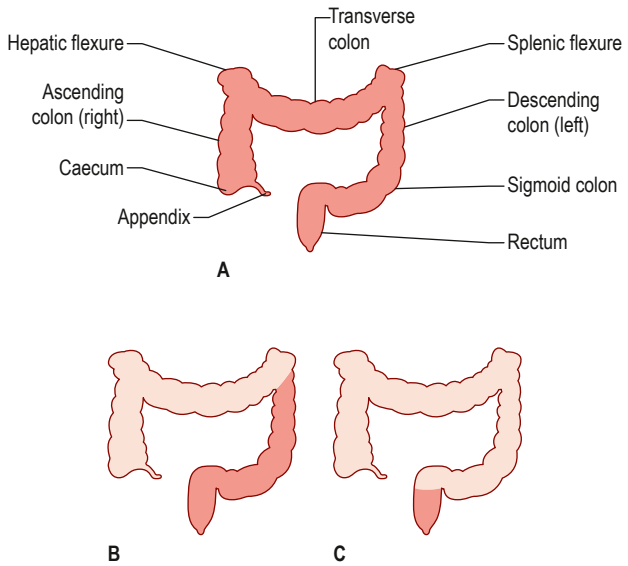


Fig. 22.41 Sites of ulcerative colitis (Montreal classification). **(A)** *Extensive colitis (E3):* pancolitis, total colitis: entire colon and rectum inflamed. **(B)** *Distal colitis (E2):* left-sided colitis: rectum, sigmoid and descending colon inflamed. **(C)** *Proctitis (E1):* rectum only inflamed.

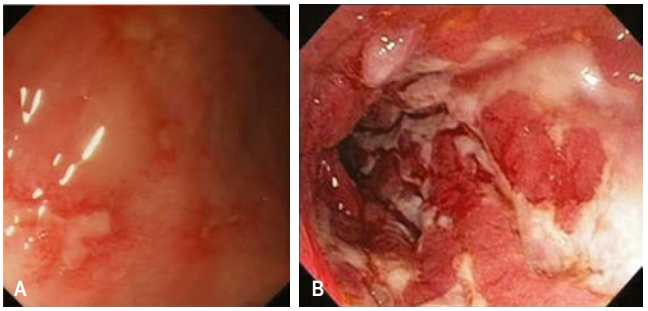


Fig. 22.42 Crohn's disease: colonoscopic appearances. (A) Aphthoid ulcers typical of Crohn's disease. **(B)** Deep, serpiginous ulceration in colonic Crohn's disease.

Box 22.35 Histological differences between Crohn's disease and ulcerative colitis

Histological feature	Crohn's disease	Ulcerative colitis
Inflammation	Deep (transmural) patchy	Mucosal continuous
Granulomas	++	Rare
Goblet cells	Present	Depleted
Crypt abscesses	+	++

Box 22.36 Extragastrointestinal manifestations of inflammatory bowel disease

- Eyes:
  - Uveitis
  - Episcleritis, conjunctivitis
- Joints:
  - Type I (pauciarticular) arthropathy
  - Type II (polyarticular) arthropathy
  - Arthralgia
  - Ankylosing spondylitis
  - Inflammatory back pain
- Skin:
  - Erythema nodosum
  - Pyoderma gangrenosum
- Liver and biliary tree:
  - Sclerosing cholangitis
  - Fatty liver
  - Chronic hepatitis
  - Cirrhosis
  - Gallstones
- Nephrolithiasis
- Venous thrombosis

Joint and other extragastrointestinal manifestations of IBD are shown in **Box 22.36**. There is an association of HLA *DRB1*0103* with pauciarticular large-joint arthritis in UC and CD, and HLA-B44 with small-joint symmetrical arthritis. HLA-B27 is associated with sacroiliitis.

Differential diagnosis

Alternative causes of diarrhoea should be excluded (see **Box 22.8**) and stool cultures (including *C. difficile* toxin assay) must always be performed. However, symptoms persisting beyond 5 days are unlikely to be caused by infective gastroenteritis. Stool microscopy for parasitic diseases such as amoebiasis should be carried out in patients with a relevant travel history. CD should be considered in all individuals with evidence of vitamin malabsorption (e.g., megaloblastic anaemia) or malnourishment, as well as in children with reduced growth velocity. Ileocolonic tuberculosis (see 'Tuberculosis' section) is common in low and middle income countries, such as India, which makes a diagnosis of CD difficult. Microscopy and culture for TB of any available tissue is

essential in these countries. A therapeutic trial of anti-TB therapy may be required. Lymphomas can occasionally involve the ileum and caecum, although they are rare in the patient population at risk from IBD.

Crohn's disease

Clinical features

The major symptoms are diarrhoea, abdominal pain, and weight loss. Constitutional symptoms of malaise, lethargy, anorexia, nausea, vomiting, and low-grade fever may be present and in 15% of these patients there are no gastrointestinal symptoms. Reduced growth velocity and delayed puberty may be the main presenting features in children. Despite the recurrent nature of this condition, some patients have an almost normal lifestyle. However, patients with extensive disease have frequent recurrences and progress from inflammatory to stricturing and penetrating disease. Approximately 50% of patients will require an intestinal resection within 5 years of diagnosis.

Clinical features are very variable and depend partly on the region of the bowel that is affected. The disease may present insidiously or acutely. Abdominal pain can be colicky, suggesting obstruction, but it usually has no special characteristics and sometimes in colonic disease only minimal discomfort is present. Diarrhoea occurs in 80% of all cases and in colonic disease it usually contains blood, making it difficult to differentiate from UC. Steatorrhoea can be present in small bowel disease. Diarrhoea can also be due to bile acid malabsorption, occurring as a consequence of ileal resection or ileal disease.

CD can also present as an emergency with acute right iliac fossa pain mimicking appendicitis. If laparotomy is undertaken, an oedematous, reddened terminal ileum is found. Other causes of an acute ileitis include infections such as *Yersinia* and TB.

CD is complicated by anal and perianal disease, and this is the presenting feature in 25% of cases, often preceding colonic and small intestinal symptoms (see **Box 22.34**). Enteric fistulae—for example, to bladder, vagina, or abdominal wall—occur in 20% to 40% of cases.

Examination

Physical signs are few, apart from loss of weight and signs of malnutrition. Aphthous ulceration of the mouth is often seen. Abdominal examination may be normal, although tenderness and/or a right iliac fossa mass are occasionally found. The mass is due either to inflamed loops of bowel that are matted together or to an abscess, which may also cause psoas muscle irritation. The anus should always be examined to look for oedematous anal tags, fissures, or perianal abscesses.

The presence of extraintestinal features of IBD should be assessed (see **Box 22.36**).

Investigations

Blood tests

- ***Anaemia*** is common and may be the normocytic, normochromic anaemia of chronic disease. However, deficiency of iron and/or folate also occurs. Despite terminal ileal involvement in CD, megaloblastic anaemia due to vitamin B_{12} deficiency is unusual, although serum B_{12} levels can be below the normal range.
- ***Raised erythrocyte sedimentation rate (ESR) and C-reactive protein (CRP)*** are found, as are raised white cell and platelet counts.
- ***Hypoalbuminaemia*** is present in severe disease as part of an acute phase response to inflammation associated with a raised CRP.
- ***Liver biochemistry*** may be abnormal.
- ***Blood cultures*** are required if septicaemia is suspected.
- ***Serological tests*** may reveal negative perinuclear ANCA (pANCA) and positive ASCA (see 'Pathology [microscopic changes]' section, above).

Stool tests

Stool cultures, including *C. difficile* toxin assay, should always be performed if diarrhoea is present. Microscopy for parasites is essential in patients with a relevant travel history. Faecal calprotectin and lactoferrin, both neutrophil-derived proteins indicating intestinal inflammation, are raised in active intestinal disease. Faecal calprotectin is useful for disease monitoring in IBD.

Endoscopy and radiological imaging

- ***Colonoscopy*** is performed if colonic involvement is suspected, except in patients presenting with severe disease (in whom a limited unprepared sigmoidoscopy should be carried out). The findings vary from mild, patchy, superficial (aphthoid) ulceration to more widespread, larger, and deeper ulcers that produce a cobblestone appearance (see **Fig. 22.42**). Endoscopic assessment of the terminal ileum is essential in all patients with suspected CD. Two biopsies should be performed in five areas, including the rectum and terminal ileum.
- ***Upper gastrointestinal endoscopy*** is required to exclude oesophageal and gastroduodenal disease in patients with relevant symptoms and is increasingly being performed in all patients at diagnosis to define the extent of disease accurately as a guide to prognosis.
- ***Small bowel imaging*** is mandatory in patients with suspected CD. The technique used will depend on availability and local expertise. Techniques include barium follow-through, CT scan with oral contrast, small bowel ultrasound, or MRI enteroclysis. An asymmetrical alteration in the mucosal pattern with deep ulceration, and areas of narrowing or structuring may be found. Although disease is commonly confined to the terminal ileum (**Fig. 22.43**), other areas of the small bowel can be involved, and skip lesions with normal bowel are seen between affected sites. Axial imaging allows the diagnosis of extraintestinal sepsis in patients presenting acutely and is therefore preferred in this situation.
- ***Ultrasound scanning*** provides a convenient radiation-free method for assessing disease activity in the ileum and colon and can be performed by appropriately trained gastroenterologists at the bedside.
- ***Perianal MRI or endoanal ultrasound*** is used to evaluate perianal disease.
- ***Capsule endoscopy*** is used in CD patients when radiological examination is normal. A patency capsule assessment is often performed first to exclude strictures of the small bowel that would constitute a contraindication to subsequent capsule endoscopy.
- ***Radionuclide scans*** with indium- or technetium-labelled leucocytes are used in some centres to identify small intestinal and colonic disease inflammation and to localize extraintestinal abscesses.

Disease activity

This can be assessed using simple parameters such as haemoglobin, white cell count, inflammatory markers (raised ESR, CRP, and platelet count), and serum albumin. Formal clinical activity indices (e.g., CD Activity Index or Harvey Bradshaw Index) are used in research studies. Faecal calprotectin is a simple, cheap,

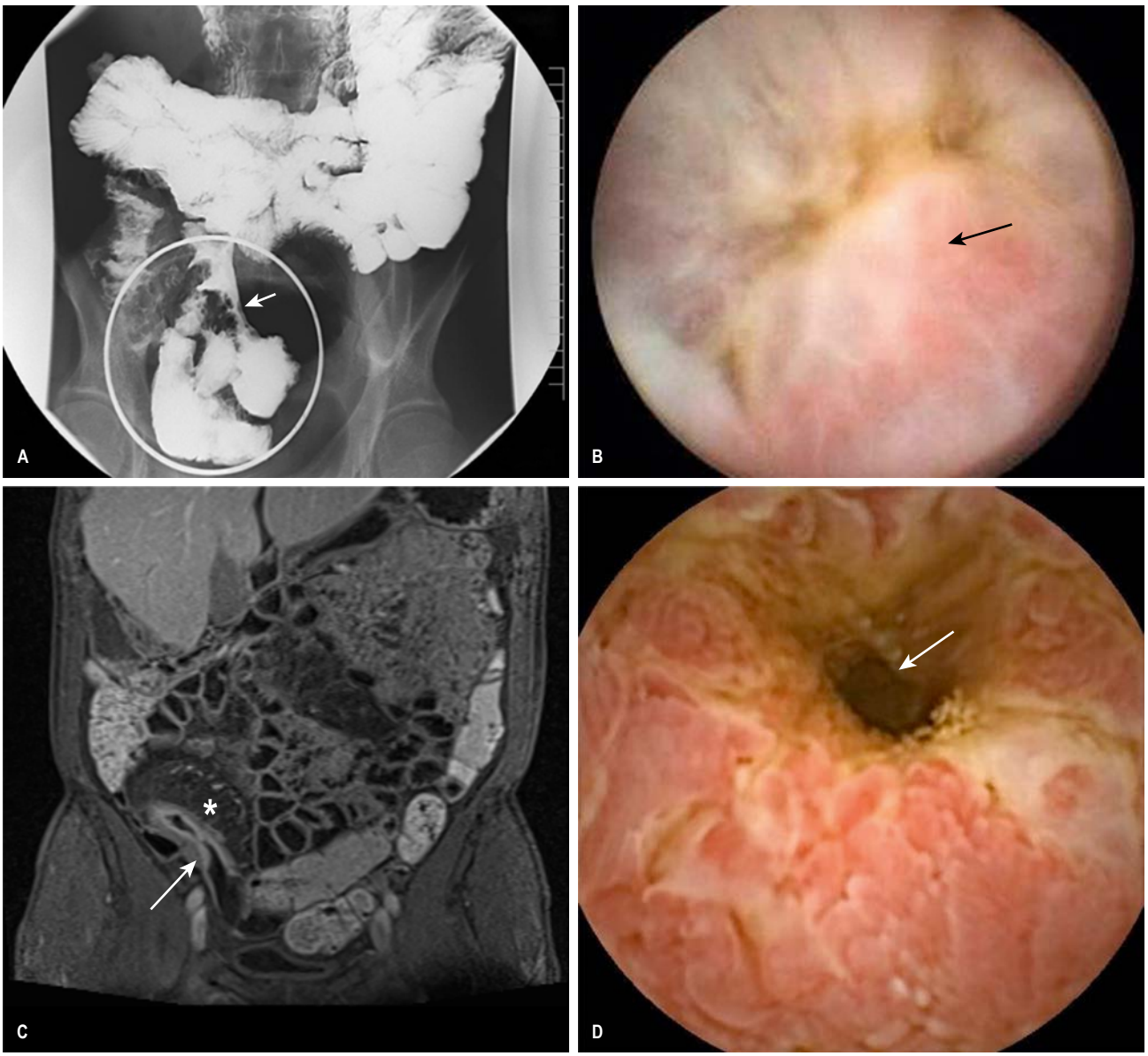


Fig. 22.43 Imaging of the small bowel in Crohn's disease. **(A)** Barium follow-through showing an ulcerated stricture *(arrowed)*, pre-stenotic dilation, and separation of the small bowel loops due to fibro-fatty proliferation. Small bowel follow-through has generally been superseded by small bowel MRI. **(B)** Capsule endoscopy picture of an ileal ulcer in a patient with Crohn's disease *(arrowed)*. **(C)** Post-contrast T1 fat-suppressed coronal image from an MRI enterography examination demonstrates bowel wall thickening of the terminal ileum and mucosal hyperenhancement consistent with active inflammation *(arrow)*. There is also fibrofatty proliferation (*) adjacent to the inflamed bowel consistent with chronic disease and, therefore, the overall appearances represent chronic active Crohn's disease. **(D)** Capsule endoscopy showing an ileal Crohn's disease stricture *(arrowed)*.

non-invasive marker of intestinal inflammation and is widely used, particularly in the outpatient setting, to monitor disease activity and response to treatment.

Medical management of Crohn's disease

See **Box 22.37**.

General considerations

The aims of management are to induce and then maintain clinical remission and to achieve mucosal healing in order to prevent disease progression and complications. Alternative causes for symptoms, such as gastroenteritis, extraintestinal sepsis, stricture formation, functional gastrointestinal disease, or bile salt malabsorption, must be excluded before commencing immunosuppressive therapy. Patients with mild symptoms and no evidence of extensive disease may require symptomatic treatment only. Cigarette smoking should be stopped. Anaemia, if due to vitamin B_{12}, folic acid, or iron deficiency, should be treated with the appropriate replacement. Patients who are intolerant of oral iron should receive an intravenous iron infusion. Most patients can be treated as outpatients, although severe attacks may require admission, and prophylaxis for thromboembolism (see Chapter 40) should be given to all inpatients.

Box 22.37 Options for the medical treatment of Crohn's disease

Induction of remission
- Oral or i.v. glucocorticosteroids
- Enteral nutrition
- Anti-tumour necrosis factor (TNF) antibodies

Maintenance of remission
- Azathioprine, mercaptopurine, methotrexate
- Anti-TNF antibodies

Perianal disease
- Surgical drainage of sepsis
- Ciprofloxacin and metronidazole
- Azathioprine
- Anti-TNF antibodies

Induction of remission

Glucocorticosteroids

These are commonly used to induce remission in moderate and severe attacks of CD (oral prednisolone 30–60 mg/day). Mild to moderate ileocaecal disease should be treated with controlled-release corticosteroids, such as budesonide, which has reduced systemic availability and is associated with a lower frequency and intensity of steroidal side-effects.

Overall remission/response rates vary from 60% to 90%, depending on type, site, and extent of disease. Steroids should be avoided in patients with penetrating intestinal disease or perianal sepsis.

Aminosalicylates

These have been used but there is little evidence to support their efficacy in CD.

Antibiotics

Antibiotics (ciprofloxacin and metronidazole) are used for treating secondary complications of CD (e.g., abscess and perianal disease).

Exclusive enteral nutrition

This consists of pre-mixed nutritional drinks which provide all of a patient's nutritional needs. It is the traditional treatment for moderate to severe attacks of CD in paediatric practice, but is under-utilized in adults due to issues with compliance to the diet. If administered as the sole source of nutrition for 28 days, rates of induction of remission are similar to those obtained with steroids. Relapse rates are high, however, particularly in those with colonic involvement.

Refractory or fulminant disease

Patients with symptoms that do not respond to conventional therapy should be re-assessed to exclude an alternative diagnosis such as a stricture or penetrating abscess. In patients with disease limited to the terminal ileum, surgical resection may be appropriate. In those with more extensive disease, remission should be induced with an anti-TNF agent, either as monotherapy or preferably in combination with an immunosuppressant such as azathioprine (see below).

Maintenance of remission

All patients require regular monitoring to exclude persistent intestinal inflammation. Patients with disease that has a good prognosis (older age at diagnosis, no perianal disease, limited ulceration at index investigations, non-smoker) may not require maintenance therapy. Patients with disease that has a poor prognosis (young age at diagnosis, extensive small bowel disease, deep colonic ulceration, perianal/rectal disease, smoker) or that flares up after induction therapy is withdrawn require long-term maintenance immunosuppression. The goal of maintenance therapy is to prevent disease progression, as well as to reduce the need for corticosteroids, which are associated with a high burden of side-effects. Therapies that induce mucosal healing result in better outcomes. All maintenance therapies require careful monitoring to ensure optimal disease control and prevent side-effects. If there is ongoing evidence of disease activity, adherence should be confirmed and dose optimization or therapy escalation undertaken.

Conventional maintenance therapies

These include ***azathioprine*** (AZA; 2–2.5 mg/kg per day), ***mercaptopurine*** (MP; 1–1.5 mg/kg per day), and ***methotrexate*** (25 mg **once a week** for 8–12 weeks [duration of induction dose dependent on clinical response] then reduce to 15 mg per week) (see **Box 22.37**). Long-term treatment with these drugs is necessary, as the rate of relapse on discontinuation is high. Patient education regarding side-effects and appropriate monitoring for complications is essential and may increase adherence. The key enzyme involved in AZA and MP metabolism is thiopurine methyl transferase (TPMT). This enzyme has a significant genetic variation and deficiencies can result in high circulating levels of thioguanine nucleotides with an increased risk of bone marrow depression. Assays of TPMT activity are available and should be performed before treatment. TPMT deficiency is not the only cause of bone marrow depression, so 3-monthly blood counts should be performed on all patients. Metabolite measurement can be undertaken to assess adherence and ensure optimal dosing.

Anti-TNF agents

These have clear evidence of benefit in the induction and maintenance of remission in patients with CD. They are indicated in patients with disease refractory to conventional immunosuppressive therapy. Early use of anti-TNF therapy is indicated in selected patients with disease that has a poor prognosis (see above). They are also used to treat complex perianal/rectal disease once sepsis has been drained. Available anti-TNF agents include ***infliximab*** (a chimeric anti-TNF-α IgG1 monoclonal antibody), ***adalimumab*** (a fully humanized anti-TNF IgG1 monoclonal antibody), and ***certolizumab pegol*** (a PEGylated Fab′ fragment of a humanized anti-TNF antibody). They neutralize soluble TNF-α, bind to membrane-bound TNF-α, and induce immune cell apoptosis, although the exact mechanism of action is not defined. In clinical trials, they have been shown to exert a steroid-sparing effect and bring about complete mucosal healing in up to one-third of patients in the long term. This results in a reduced need for hospital admission and surgery. These agents should always be used on a regular basis as maintenance therapy, as they are less effective and induce anti-drug antibodies if used episodically. In patients who are naïve to azathioprine, combination therapy increases efficacy and reduces immunogenicity. Their use should be limited to clinicians experienced in the management of CD, as they are associated with significant complications, including opportunistic infections (such as TB), demyelination, and malignancy.

Novel biological therapies

Advances in mucosal immunology have highlighted a variety of targets to resolve the inflammation and symptoms of IBD. Numerous investigational therapies concentrate on inhibiting, suppressing, or

Box 22.38 Biological therapies in IBD

Current therapies	Target molecule	Mechanism of action
Infliximab	Chimeric mAb against TNF-α	Neutralizes soluble TNF-α, binds to membrane-bound TNF-α and induces immune cell apoptosis
Adalimumab	Humanized mAb against TNF-α	Neutralizes soluble TNF-α, binds to membrane-bound TNF-α and induces immune cell apoptosis
Golimumab	Humanized IgG1 kappa mAb against TNF-α	Binds to both the soluble and transmembrane bioactive forms of human TNF-alpha, neutralizing activity
Ustekinumab	Human IgG1 kappa mAb binds to p40 subunit shared by IL-12 and IL-23	Interferes with binding of pro-inflammatory cytokines IL-12 and IL-23 to their cell surface receptors, interfering with NK cell activation and differentiation/activation of $CD4^+$ T cells
Risankizumab	Humanized IgG1 mAb blocks p19 subunit of IL-23	Neutralizes IL-23 preventing its interaction with the IL-23 receptor and activation of downstream signalling cascades
Vedolizumab	Anti-α4β7 integrin	Reduces lymphocyte recruitment to the inflamed intestine
Tofacitinib	Reversibly inhibits Janus kinase (JAK) 1 and JAK3 *in vitro* and to a lesser extent inhibits JAK2 and TYK2	Inhibition of JAK3, results in inhibition of cytokines IL-2, IL-4, IL-7, IL-9, IL-15, and IL-21 Inhibition of JAK1, results in inhibition of IL-6, type 1 IFNs, and IFN-γ Some effect on natural killer cell depletion
Upadacitinib	Selective and reversible Janus kinase (JAK) inhibitor inhibiting signalling by JAK1 or JAK 1/3 with functional selectivity over cytokine receptors that signal via pairs of JAK2	JAK inhibition results in inhibition of IL-6- and IFN-γ-related signalling; some effect on natural killer cell depletion
Ozanimod	Sphingosine-1 phosphate receptor modulator	Prevents lymphocyte regression from lymph nodes

altering T-cell differentiation. These biological therapies are summarized in **Box 22.38**. This section covers the novel biological therapies that are now widely used for both CD and UC.

- Vedolizumab is an anti-α4β7 integrin therapy which acts to reduce leucocyte recruitment to the inflamed intestine. This is used in patients with moderately to severely active Crohn's disease or UC who are intolerant to an anti-TNF or in whom it is contraindicated, or in those patients who have had an inadequate response or lost response to an anti-TNF-α agent.
- Ustekinumab targets the IL-12/IL-23 pathway, and is used to treat moderate to severely active Crohn's disease and UC, in those patients who cannot have, or have had an inadequate response to, an anti-TNF-α agent.
- Risankizumab is a humanized IgG1 monoclonal antibody blocking the p19 subunit of IL-23. It is delivered subcutaneously and is effective in achieving and maintaining remission in patients with moderately to severely active Crohn's disease. It is recommended for patients who have failed, not tolerated, or lost response to a previous biological agent, or where anti-TNF-alpha agents are not suitable.
- Ozanimod is an oral sphingosine-1 phosphate receptor modulator which prevents lymphocyte egression from lymph nodes and used to treat adults with moderately to severely active UC.
- Janus kinases (JAKs) are protein tyrosine kinases (TYKs) that cohere to the intracellular domains of transmembrane cytokine receptors of numerous growth factors and cytokines. Once bound, the JAKs become activated leading to phosphorylation of the receptor and the associated signal transducer and activator of transcription (STAT) proteins, which then react and form a dimer which enters the nucleus. They then bind to transcriptional regulatory sites and control cellular responses central in immune defence and in immune-mediated disease.
  - Upadacitinib is a selective JAK-1 inhibitor; it inhibits IL-6- and IFN-γ-related signalling.
  - It is an oral agent used to treat moderate to severely active Crohn's or UC that has not responded well to a previous biologic or if anti-TNF alphas are contraindicated.
  - Tofacitinib reversibly inhibits JAK1 and JAK3 *in vitro* and to a lesser extent inhibits JAK2 and TYK2. Tofacitinib, by inhibition of JAK3, inhibits cytokines IL-2, IL-4, IL-7, IL-9, IL-15, and IL-21; and by JAK1 inhibition, inhibits IL-6, type 1 IFNs, and IFN-γ. Tofacitinib is an oral agent used to treat moderately to severely active UC.
- Important adverse events with JAK inhibitors are venous thromboembolism, major cardiovascular events, herpes zoster infection, and malignancy.

Surgical management of Crohn's disease

Studies prior to the introduction of biological agents suggest that up to 80% of patients will require an operation at some time during the course of their disease. Nevertheless, surgery should be avoided if possible and only minimal resections undertaken, as recurrence (15% per year) is almost inevitable without prophylactic maintenance therapy. The *indications for surgery* are:

- failure of medical therapy, with acute or chronic symptoms producing ill-health
- complications (e.g., toxic dilation, obstruction, perforation, abscesses, enterocutaneous fistula)
- failure to grow in children despite medical treatment
- presence of perianal sepsis; an examination is performed under anaesthetic, the sepsis is drained, and a seton is inserted to ensure ongoing drainage.

In patients with small bowel disease some strictures can be widened (stricturoplasty), whereas others require resection and anastomosis. It is essential to consider postoperative maintenance therapy in patients undergoing ileocolonic resection and anastomosis. All patients should refrain from smoking. There is some evidence that a 3-month course of metronidazole reduces relapse rates. Patients with a high risk of relapse (previous surgery, smoker, penetrating disease at the time of index surgery) should be evaluated for maintenance immunosuppressive therapy. Patients should undergo an ileocolonoscopy to assess the anastomosis for disease

Box 22.39 Definition of a severe attack of ulcerative colitis

- Stool frequency: >6 stools/day with blood +++
- Fever: >37.5°C
- Tachycardia: >90 beats/min
- Erythrocyte sedimentation rate: >30 mm/h
- Anaemia: <100 g/L haemoglobin
- Albumin: <30 g/L

recurrence 6 months after surgery, when further therapy can be agreed with the patient.

When colonic CD involves the entire colon and the rectum is spared or minimally involved, a subtotal colectomy and ileorectal anastomosis may be performed. An eventual recurrence rate of 60% to 70% in the ileum, rectum, or both is to be expected; however, two-thirds of these patients retain a functional rectum for 10 years. If the whole colon and rectum are involved, a panproctocolectomy with an end ileostomy is the standard operation. CD patients are not suitable for a pouch operation (see 'Surgical management of ulcerative colitis' section), as recurrence in the pouch is high.

Problems associated with ileostomies include:

- mechanical problems
- dehydration, particularly if there is a short length of small bowel remaining
- psychosexual problems
- erectile dysfunction in males and reduced fecundity in females (due to prior pelvic surgery)
- recurrence of CD.

Prognosis

The majority of patients have inflammatory disease at diagnosis, although up to 20% will present with complicated disease, including strictures and penetrating disease. If disease is left untreated, its natural history is progression from inflammatory to stricturing and penetrating disease that may require surgery. Up to 50% of patients will require a surgical resection within the first 5 years of disease. The goal of therapy is to target early aggressive therapy at those patients with a poor prognosis. However, predicting disease prognosis at diagnosis is not easy. Features that imply a poor prognosis include young age at diagnosis (<20 years of age), extensive small bowel disease, complex perianal disease and deep ulceration at index colonoscopy.

Ulcerative colitis

Clinical features

The major symptom in UC is diarrhoea with blood and mucus, sometimes accompanied by lower abdominal discomfort. General features include malaise, lethargy and anorexia with weight loss, although these features are not as severe as with CD. Aphthous ulceration in the mouth may be seen. The disease can be mild, moderate or severe (**Box 22.39**), and in most patients runs a course of remissions and exacerbations. Disease extent is defined as limited to the rectum (proctitis), left-sided or extensive (see **Fig. 22.41**).

Proctitis is characterized by the frequent passage of blood and mucus, urgency and tenesmus. There are normally few constitutional symptoms and the stool, when passed, may be solid. Patients are nevertheless greatly inconvenienced by the frequency of defecation.

In an acute attack of left-sided or extensive UC, patients have bloody diarrhoea, passing up to 10 to 20 liquid stools per day. Diarrhoea also occurs at night, with urgency and incontinence that is severely disabling for the patient. Patients with an acute severe flare of colitis require urgent admission for intensive therapy (see Chapter 11).

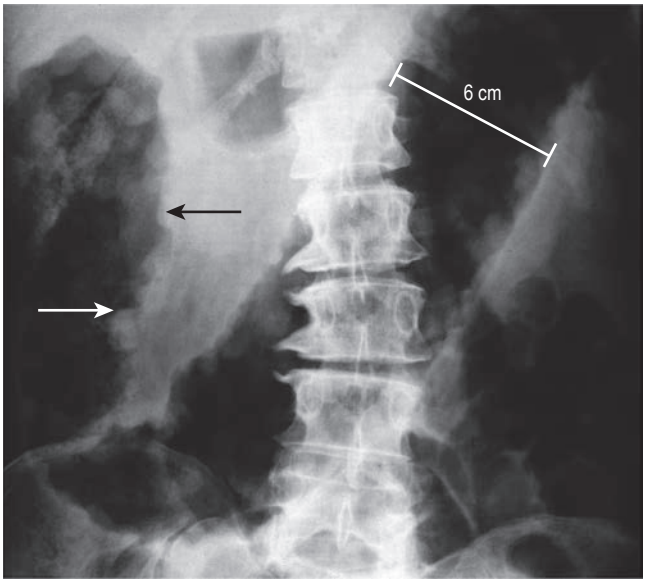


Fig. 22.44 Plain abdominal X-ray showing toxic dilation. The *arrows* indicate mucosal islands.

Toxic megacolon is a serious complication associated with acute severe colitis. The plain abdominal X-ray shows a dilated, thin-walled colon with a diameter of more than 6 cm; it is gas-filled and contains mucosal islands (**Fig. 22.44**). It is a particularly dangerous stage of advanced disease, with impending perforation and a high mortality (15%–25%). Urgent surgery is required in all patients in whom toxic dilation has not resolved within 48 hours, with intensive therapy as above. The differential diagnosis includes an infectious colitis, for example, with *C. difficile* and cytomegalovirus.

Examination

In general, there are no specific signs in UC. The abdomen may be slightly distended or tender to palpation. Tachycardia and pyrexia are signs of severe colitis and mandate admission. The anus is usually normal. Rectal examination will reveal the presence of blood. Sigmoidoscopy is usually abnormal, showing an inflamed, bleeding, friable mucosa. Very occasionally, rectal sparing occurs, with normal proctoscopy.

Investigations

Blood tests

- ***White cell and platelet counts*** are commonly raised in moderate to severe attacks, and iron deficiency anaemia is present.
- ***ESR and CRP*** are often raised; liver biochemistry may be abnormal, with hypoalbuminaemia occurring in severe disease.
- ***pANCA*** may be positive. This is contrary to CD, where pANCA is usually negative (see 'Pathology [microscopic changes]' section, above).

Stool tests and *C. difficile* toxin

These should always be performed to exclude infective causes of colitis. Stool microscopy to exclude amoebiasis is mandated in patients with a relevant travel history. Faecal calprotectin/lactoferrin will be elevated.

Colonoscopy

Endoscopy with mucosal biopsy is the 'gold standard' investigation for the diagnosis of UC. Colonoscopy also allows assessment of disease

activity and extent. In patients with long-term colitis, chromoendoscopy is used to diagnose dysplasia. Full colonoscopy should not be performed in severe attacks of disease for fear of perforation; instead, a limited unprepared FS should be used to confirm diagnosis.

Imaging

A plain abdominal X-ray is essential in patients suffering acute severe attacks to exclude colonic dilation. However, the extent of disease is not reliably assessed using this investigation. Other imaging modalities are rarely used in the assessment of patients with UC, as endoscopy is preferred. However, inflammation of the colonic wall is detected on ultrasound, as is the presence of free fluid within the abdominal cavity.

Medical management of ulcerative colitis

Wherever possible, patients with IBD should be managed in patient-focused IBD clinics with access to a full multidisciplinary team (MDT). The mainstay of treatment for mild and moderate disease of any extent is an aminosalicylate, which acts topically in the colonic lumen. The active moiety of these drugs is 5-aminosalicylic acid (5-ASA), which is absorbed in the small intestine. Therefore, the various aminosalicylate preparations are designed to deliver the active 5-ASA to the colon. This is achieved by binding of 5-ASA with an azo bond to sulfapyridine (sulfasalazine), 4-aminobenzoyl-β-alanine (balsalazide), or to 5-ASA itself (olsalazine), coating with a pH-sensitive polymer, packaging of 5-ASA in microspheres, or a combination of these. The azo bonds are broken down by colonic bacteria to release 5-ASA within the colon.

The mode of action of 5-ASA in IBD is unknown, although it may involve the intracellular peroxisome proliferator-activated receptor (PPAR)-γ signalling pathway. The aminosalicylates have been shown to be effective in inducing remission in mild to moderately active disease, and maintaining remission in all forms of disease. There is also evidence that they are chemopreventive for UC-associated colorectal cancer. 5-ASA can rarely cause renal disease.

Proctitis

Rectal 5-ASA suppositories are the first-line treatment. Topical steroids are less effective than 5-ASA preparations. Oral 5-ASA can be added to increase remission rates. Some cases of proctitis do not respond to 5-ASA treatment and require oral prednisolone.

Left-sided colitis

Topical 5-ASA enemas are the first-line treatment. The addition of an oral 5-ASA will increase remission rates. Patients who do not respond to this or have worsening symptoms require oral prednisolone.

Extensive colitis

Patients with mild to moderate symptoms can be treated with an oral 5-ASA at an adequate dose. The additional of a 5-ASA enema increases remission rates. Patients who do not respond to this or have worsening symptoms require oral prednisolone.

Refractory/severe colitis of any extent

Patients with colitis that is refractory to standard conventional therapies should be investigated to ensure adherence to therapy, confirm active colitis, and exclude infection, including biopsies to exclude reactivation of CMV colitis. They should then receive biological therapy with either an anti-TNF agent (infliximab, adalimumab, or golimumab) or an anti-integrin therapy (vedolizumab), or tofacitinib, an oral inhibitor of Janus kinase 1-3, which became available in 2018 for the treatment of adults with moderately to severely active UC. See 'Novel biological therapies' section for further discussion.

Patients with severe colitis (**Box 22.39**) and those who do not respond to oral prednisolone should be admitted to hospital for treatment (see Chapter 11).

Salvage therapy

Salvage therapy to avoid colectomy is required for patients with a CRP higher than 45 mg/L or more than eight bowel motions after 3 days of intravenous hydrocortisone. Continuing steroid therapy alone in this situation will delay the inevitable colectomy and increase mortality. Salvage medical therapies with clear evidence of benefit in controlled clinical trials are intravenous infliximab infusion or ciclosporin 2 mg/kg per day as a continuous infusion. The majority of patients are commenced on infliximab rather than ciclosporin. Patients with extensive ulceration may require higher induction doses of infliximab. Patients with a low albumin level have been shown to have a lower response rate to salvage therapy. These agents should only be used by experienced gastroenterologists who are part of an MDT with colorectal surgeons. Patients should be weaned off steroids rapidly, once salvage therapy has commenced, to reduce morbidity. Those who respond should be treated with oral ciclosporin or further infliximab infusions, as appropriate, while maintenance thiopurine therapy is commenced.

Potential future therapies

Alicaforsen is an inhibitor of intracellular adhesion molecule-1. Alicaforsen enemas have demonstrated efficacy in the treatment of distal ulcerative colitis as well as pouchitis in clinical trials.

An autologous stem cell transplantation trial in refractory Crohn's disease—low-intensity therapy evaluation (ASTIClite) is currently underway. It is looking to assess whether stem cell transplantation, with a low-intensity chemotherapy regimen, can reduce the symptoms and activity of Crohn's disease and enhance quality of life, compared to current standard care. The trial will also assess the safety of the procedure, explore the mechanisms involved in immune reconstitution, and determine whether patients who experience recurrent disease after the stem cell transplant will respond to treatments that had not worked previously. Finally, the trial aims to assess the long-term safety and efficacy in patients undergoing stem cell transplantation over a minimum of a further 4 years.

Surgical management of ulcerative colitis

While the treatment of UC remains primarily medical, surgery continues to have a central role because it may be life-saving, is curative, and eliminates the long-term risk of cancer. The main indications for surgery are severe colitis that fails to respond to medical therapy, and chronic active therapy-refractory disease. Other indications are listed in **Box 22.40**. In expert centres,

Box 22.40 Indications for surgery in ulcerative colitis

Fulminant acute attack
- Failure of medical treatment
- Toxic dilation
- Haemorrhage
- Imminent perforation

Chronic disease
- Incomplete response to medical treatment/steroid-dependent
- Dysplasia on surveillance colonoscopy

laparoscopic surgery is often used to improve postoperative pain, recovery time, and cosmesis.

In acute disease, subtotal colectomy with end ileostomy and preservation of the rectum is the operation of choice. At a later date, a number of surgical options are available and are best carried out in a specialist colorectal centre. These include proctectomy with a permanent ileostomy; to avoid a permanent ileostomy, an ileo-anal pouch procedure can be performed (Fig. 22.45). The ileo-anal pouch is anastomosed to the anus at the dentate line following excision of the remaining rectum. One-third of patients, however, will experience ***'pouchitis'***, inflammation of the pouch mucosa with clinical symptoms of diarrhoea, bleeding, fever and, at times, exacerbation of extracolonic manifestations (Fig. 22.46). The incidence of pouchitis is twice as high in patients with primary sclerosing cholangitis and is also raised in patients with a positive ANCA and backwash ileitis prior to colectomy. Two-thirds of pouchitis cases will recur as either acute relapsing or chronic unremitting forms. The mainstay of treatment is antibiotics (metronidazole with or without ciprofloxacin). Treatment is not always satisfactory and steroids may be required. The probiotic VSL#3 has been shown to be effective in preventing the onset of pouchitis and in maintaining remission in patients with antibiotic-treated pouchitis to induce mucosal healing.

Course and prognosis

One-third of patients with distal inflammatory proctitis due to UC will develop more proximal disease, with 5% to 10% developing total colitis. One-third of patients with UC will have a single attack and the others will follow a relapsing course. One-third of patients with UC will undergo colectomy within 20 years of diagnosis.

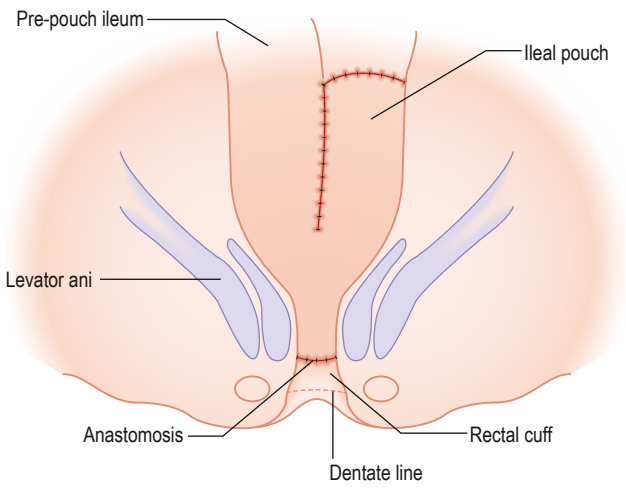


Fig. 22.45 An ileo-anal pouch for ulcerative colitis.

Cancer in inflammatory bowel disease

Patients with UC and extensive Crohn's colitis have an increased incidence of developing dysplasia and subsequent colon cancer. The risk of dysplasia is related to the extent and duration of disease, as well as the presence of untreated mucosal inflammation. A family history of colorectal cancer and the presence of primary sclerosing cholangitis also increase the risk. Appropriate colonoscopic screening strategies according to guidelines are used by many, although evidence for overall benefit is still uncertain. A 40-year analysis of colonoscopy surveillance for neoplasia in UC, conducted at St Mark's Hospital in the UK, suggested that surveillance may have a significant role in reducing the risk of advanced and interval colorectal cancer.

Mortality in inflammatory bowel disease

Population-based studies demonstrate that mortality in UC is similar to that in the general population. The two exceptions are patients with severe colitis, who have a slightly higher mortality in the first year after diagnosis, and patients aged over 60 at the time of diagnosis. Although it is currently unclear whether there is a slightly higher overall mortality in patients with CD, those with extensive jejunal and ileal disease and those with gastric and duodenal disease have been shown to have a relatively higher mortality.

Microscopic colitis

Patients with this group of disorders present with chronic or fluctuating watery diarrhoea. Although the macroscopic features on colonoscopy are normal, the histopathological findings on biopsy are abnormal. There are three distinct forms of microscopic inflammatory colitis:

- ***Microscopic UC*.** There is a chronic inflammatory cell infiltrate in the lamina propria, with deformed crypt architecture and goblet cell depletion, with or without crypt abscesses. Treatment is as for UC; many patients respond to treatment with aminosalicylates alone.
- ***Microscopic lymphocytic colitis*.** There is surface epithelial injury, prominent lymphocytic infiltration in the surface epithelium, and increased lamina propria mononuclear cells.
- ***Microscopic collagenous colitis*.** There is a thickened subepithelial collagen layer (>10 μm) adjacent to the basal membrane, and increased infiltration of the lamina propria with lymphocytes and plasma cells, and surface epithelial cell damage. It is predominantly a disorder of middle-aged or elderly females, and is associated with

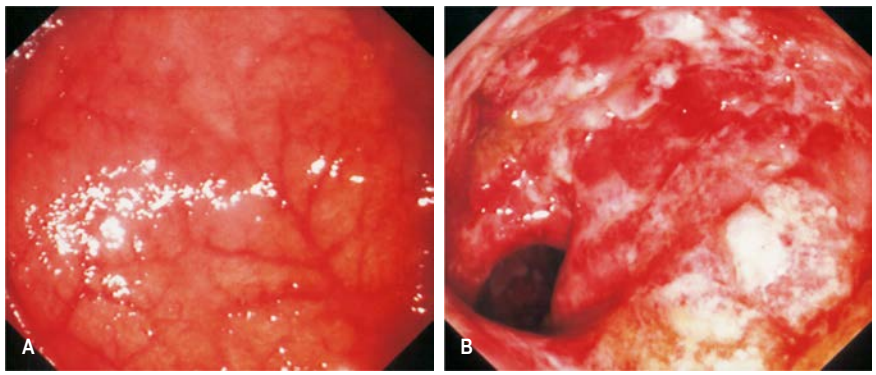


Fig. 22.46 Pouchitis: fibreoptic sigmoidoscopic appearances. **(A)** Mild changes. **(B)** Severe haemorrhagic ulceration.

a variety of autoimmune disorders (arthritis, thyroid disease, limited cutaneous scleroderma, and primary biliary cirrhosis).

The incidence of both microscopic lymphocytic and collagenous colitis is increased in patients with coeliac disease and this must be excluded. Treatment of microscopic and collagenous colitis is usually with budesonide. There is also evidence of benefit for aminosalicylates, bismuth-containing preparations and, if refractory, prednisolone, and azathioprine. A small number of patients with microscopic lymphocytic and collagenous colitis have coexisting bile acid malabsorption and can thus respond to colestyramine. Prognosis is good.

Further reading

Ananthakrishnan AN. Epidemiology and risk factors for IBD. *Nat Rev Gastroenterol Hepatol* 2015; 12:205–217.
Clarke WT, Feuerstein JD. Colorectal cancer surveillance in inflammatory bowel disease: practice guidelines and recent developments. *World J Gastroenterol* 2019; 25(30):4148–4157.
Graham DB, Xavier RJ. Pathway paradigms revealed from the genetics of inflammatory bowel disease. *Nature* 2020; 578:527–539.
Lamb CA, Kennedy NA, Raine T et al. British Society of Gastroenterology consensus guidelines on the management of inflammatory bowel disease in adults. *Gut* 2019; 0:1–106.
Loftus EV Jr, Panés J, Lacerda AP et al. Upadacitinib induction and maintenance therapy for Crohn's disease. *N Engl J Med* 2023; 388(21):1966.
Sands BE, Sandborn WJ, Panaccione R et al. Ustekinumab as induction and maintenance therapy for ulcerative colitis. *N Engl J Med* 2019; 381:1201–1214.

COLON AND RECTUM

Miscellaneous colonic conditions

Clostridium difficile-associated diarrhoea and pseudomembranous colitis

Clostridium difficile infection (CDI) is a common hospital-acquired infection and can be a significant cause of morbidity, especially among the elderly. It frequently occurs following antibiotic therapy due to disruption of the normal gut flora. The following antibiotics are particularly associated with CDI:

- fluoroquinolones
- clindamycin
- cephalosporins
- penicillins
- (macrolides and co-trimoxazole are less frequently associated).

Clinical features include profuse diarrhoea, which can lead to hypovolaemia and shock if severe. The white blood cell count is often significantly elevated and endoscopically there may be pseudomembranous colitis (severe inflammation of the bowel wall lining). Complications include ileus, toxic megacolon, and bowel perforation. If megacolon or perforation is suspected then patients should have urgent imaging (X-ray or CT).

Diagnosis is made with a positive test for *C. difficile* toxin A and B. A positive test for glutamate dehydrogenase (GDH) antigen is not necessarily indicative of active infection, simply that the patient is a carrier of the pathogen.

Treatment

Initial treatment regimens are detailed in **Box 22.9**. For recurrent or refractory CDI, antibiotics such as rifaximin or fidaxomicin can be useful. However, recently faecal microbiota transplantation (FMT) has become increasingly common. For this procedure, a solution of donor stool is given to the patient either through a nasogastric tube or via FS. The donor stool has been carefully selected from a centralized stool bank where all donors have had their blood and stool extensively screened for infection.

Megacolon

The term 'megacolon' is used to describe a number of congenital and acquired conditions in which the colon is dilated. In many instances, it is secondary to chronic constipation; in some parts of the world, Chagas' disease is a common cause.

In all young patients with megacolon, Hirschsprung's disease should be excluded. In this condition, which presents in the first years of life, an aganglionic segment of the rectum (megarectum) gives rise to constipation and subacute obstruction. Occasionally, Hirschsprung's disease affecting only a short segment of the rectum can be missed in childhood. A preliminary rectal biopsy is performed and stained for ganglion cells in the submucosal plexus. In doubtful cases, a full-thickness biopsy should be obtained. A frozen section is stained for acetylcholinesterase, which is elevated in Hirschsprung's disease. Manometric studies show failure of relaxation of the internal sphincter, which is diagnostic of Hirschsprung's disease. This condition can be successfully treated surgically.

Treatment of other causes of a megacolon is similar to that of slow-transit constipation, but saline washouts and manual removal of faeces are sometimes required.

Faecal incontinence

Of the healthy population over the age of 65, 7% experience a degree of incontinence. Incontinence is classified as minor (inability to control flatus or liquid stool, causing soiling) or major (frequent and inadvertent evacuation of stool of normal consistency); see **Box 22.41** for common causes. Obstetric injury is a common cause and sphincter defects have been found in up to 30% of primiparous women. Endoanal ultrasonography or pelvic MRI is the investigation of choice in the assessment of anal sphincter damage (**Fig. 22.47**). Neurophysiological investigation of pudendal nerve function, anal sensation, and anal sphincter function may be required to elicit the cause of the problem.

Initial ***management*** of minor incontinence is bowel habit regulation. Loperamide is the most potent antidiarrhoeal agent, which also increases internal sphincter tone.

? Box 22.41 Aetiology of faecal incontinence

- Congenital:
  - For example, surgery for imperforate anus
- Anal sphincter dysfunction:
  - Structural damage: surgery—anorectal, vaginal hysterectomy; obstetric injury during childbirth; trauma; radiation; perianal Crohn's disease
  - Pudendal nerve damage: childbirth
  - Perineal descent: prolonged straining at stool
- Rectal prolapse
- Faecal impaction with overflow diarrhoea
- Severe diarrhoea:
  - For example, ulcerative colitis, functional diarrhoea, irritable bowel syndrome
- Neurological and psychological disorders:
  - Spinal trauma (S2–S4)
  - Spina bifida
  - Stroke
  - Multiple sclerosis
  - Diabetes mellitus (with autonomic involvement)
  - Dementia
  - Psychological illness

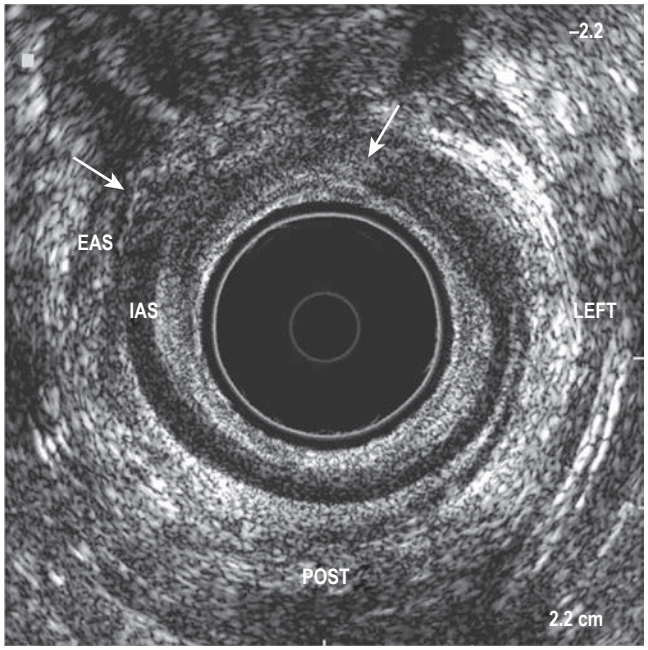


Fig. 22.47 Endoanal ultrasound scan. Axial mid-canal image, showing a large tear between 10 and 1 o'clock *(arrowed)* following vaginal delivery, involving the external anal sphincter (EAS) and internal anal sphincter (IAS), resulting in faecal incontinence. (*Courtesy of Professor Clive Bartram, Princess Grace Hospital, London.*)

Biofeedback is effective in some people with faecal incontinence associated with impaired function of the puborectalis muscle and the external anal sphincter. Sacral spinal nerve stimulation has been shown to be effective in the treatment of patients with a functionally deficient but morphologically intact external anal sphincter. Surgery may be required for anal sphincter trauma and should only be carried out in specialist centres.

Ischaemic disease of the colon (ischaemic colitis)

Occlusion of branches of the superior mesenteric artery (SMA) or inferior mesenteric artery (IMA), often in the older age group, commonly presents with sudden onset of abdominal pain and the passage of bright red blood *per rectum*, with or without diarrhoea. There may be signs of shock and evidence of underlying cardiovascular disease. The anatomy of the vascular supply to the colon results in a watershed area at the splenic flexure, which is therefore the most common site affected. This condition has also been described in females taking the contraceptive pill, patients on nicorandil, and those with thrombophilia and small- or medium-vessel vasculitis.

Examination

On examination, the abdomen may be distended and tender. A straight abdominal X-ray often shows thumb-printing (a characteristic sign of ischaemic disease) at the splenic flexure. Patients are likely to display signs of cardiovascular shock and may have a lactic acidosis.

Differential diagnosis and investigations

The differential diagnosis includes other causes of acute colitis. Patients often require an urgent CT scan to exclude perforation. An unprepared FS is the diagnostic investigation of choice; biopsies showing epithelial cell apoptosis and lamina propria fibrosis are characteristic. A colonoscopy should be performed when the patient has fully recovered to exclude the formation of a stricture at the site of disease and confirm mucosal healing. Patients without evidence of underlying cardiovascular disease should be screened for thrombophilia and vasculitis.

Management

Most patients settle on symptomatic treatment. A few patients show progressive signs of peritonism and imminent perforation, and require urgent surgery.

Diverticular disease

Diverticula are frequently found in the colon and occur in 50% of patients over the age of 50 years. They are most frequent in the sigmoid, but can be present throughout the whole colon.

Terminology:

- ***diverticulosis*** indicates the presence of diverticula
- ***diverticulitis*** implies that these diverticula are inflamed
- ***diverticular colitis*** refers to crescentic inflammation on the folds in areas of diverticulosis.

It is perhaps better to use the more general term ***diverticular disease***, as it is often difficult to be sure whether the diverticula are inflamed. The precise mechanism of diverticula formation is not known. There is thickening of the muscle layer and, because of high intraluminal pressures, pouches of mucosa extrude through the muscular wall through weakened areas near blood vessels to form diverticula. An alternative explanation is cholinergic denervation with increasing age, which leads to hypersensitivity and increased uncoordinated muscular contraction. Diverticular disease seems to be related to the low-fibre diet eaten in high income countries and is rare in rural Africa.

Diverticulitis occurs when faeces obstruct the neck of the diverticulum, causing stagnation and allowing bacteria to multiply and produce inflammation. This can then lead to bowel perforation (peridiverticulitis), abscess formation, fistulae into adjacent organs, haemorrhage and even generalized peritonitis.

Clinical features and investigations

Diverticular disease is asymptomatic in 95% of cases and is usually discovered incidentally on colonoscopy or barium enema examination. No treatment other than advice to increase dietary fibre is required in those patients. In symptomatic patients, intermittent left iliac fossa pain or discomfort and an erratic bowel habit commonly occur, which are difficult to differentiate from the irritable bowel syndrome. In severe disease, luminal narrowing results in severe pain and constipation. In the absence of clinical signs of acute diverticulitis, a colonoscopy or 'virtual colonoscopy' (see 'Endoscopy' section) is the investigation of choice (**Fig. 22.48**). Barium enema combined with FS is also used.

Management

Management of uncomplicated symptomatic disease is with a well-balanced (soluble and insoluble) fibre diet (20 g/day), with smooth muscle relaxants if required. Antibiotics and admission to hospital are not required for uncomplicated disease.

Acute diverticulitis

The pathophysiology of diverticulitis is associated with altered gut motility, increased luminal pressure, and a disordered colonic

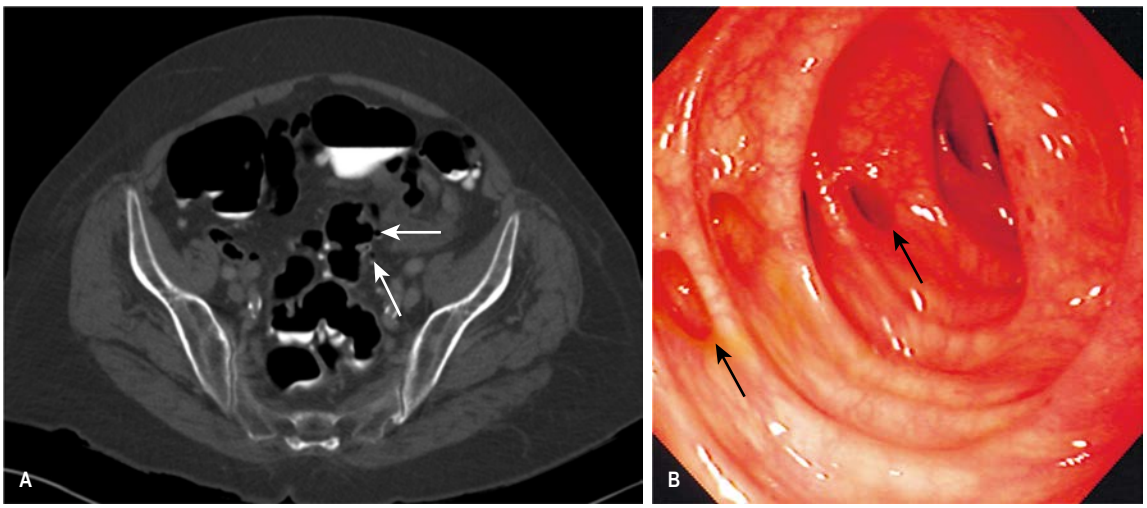


Fig. 22.48 Diverticular disease. **(A)** Axial CT colonography image demonstrating diverticular disease (diverticulosis) of the sigmoid colon *(arrows)*. **(B)** Diverticula *(arrowed)* seen on colonoscopy.

microenvironment. Acute diverticulitis most commonly affects diverticula in the sigmoid colon. It presents with severe pain in the left iliac fossa, often accompanied by fever and constipation. These symptoms and signs are similar to those of appendicitis but are located on the left side. On examination, the patient is often febrile with a tachycardia. ***Abdominal examination*** shows tenderness, guarding, and rigidity on the left side of the abdomen. A palpable tender mass is sometimes felt in the left iliac fossa.

Investigations

- ***Blood tests*** often reveal a polymorphonuclear leucocytosis. The ESR and CRP are raised.
- ***CT colonography*** (Fig. 22.49) will show colonic wall thickening, diverticula, and often pericolic collections and abscesses. There is usually a streaky increased density extending into the immediate pericolic fat with thickening of the pelvic fascial planes. These findings are diagnostic of acute diverticulitis (95% sensitivity and specificity) and differ from those of malignant disease. Sigmoidoscopy and colonoscopy are not performed during an acute attack.
- ***Ultrasound examination*** is often more readily available and is cheaper. It can demonstrate thickened bowel and large pericolic collections, but is less sensitive than CT.

Management

Mild attacks can be treated on an outpatient basis using oral antibiotics such as ciprofloxacin and metronidazole. Patients with signs of systemic upset (fevers, tachycardia), significant abdominal pain, or co-morbidity require bowel rest, intravenous fluids, and intravenous antibiotic therapy. Recent large trials of 5-ASA therapy have shown minor benefit in preventing recurrent diverticulitis. Repeat attacks often require surgery.

Complications of diverticular disease

- ***Perforation*** usually occurs in association with acute diverticulitis, and can lead to formation of a paracolic or pelvic abscess or generalized peritonitis. Surgery may be required.
- ***Fistula formation*** into the bladder, causing dysuria or pneumaturia, or into the vagina, causing discharge.
- ***Intestinal obstruction*** (see 'Intestinal obstruction' section), usually after repeated episodes of acute diverticulitis.
- ***Bleeding*** is sometimes massive. In most cases, the bleeding stops and the cause of the bleeding can be established by colonoscopy and sometimes angiography. In rare cases, emergency segmental colectomy is required.
- ***Mucosal inflammation*** occurs in areas of diverticula, giving the appearance of a segmental colitis at endoscopy that may resemble Crohn's disease.

Anorectal disorders

Pruritus ani

Pruritus ani, or an itchy bottom, is common. Perianal excoriation results from scratching. Usually, the condition results from haemorrhoids or overactivity of sweat glands. ***Management*** consists of enhancing toilet hygiene, keeping the area dry, and avoiding the use of perfumed moisturizing creams. Secondary causes include threadworm *(Enterobius vermicularis)* infestation, fungal infections (e.g., candidiasis), and perianal eczema, which should be treated appropriately.

Haemorrhoids

Haemorrhoids (primary—internal; second degree—prolapsing; third degree—prolapsed) usually cause rectal bleeding, discomfort, and pruritus ani. Patients may notice red blood on their toilet paper and blood on the outside of their stools. They are the most common cause of rectal bleeding (see Fig. 22.33). Diagnosis is made by inspection, rectal examination, and proctoscopy.

Management

If symptoms are minor, no treatment is required apart from advice about avoiding constipation. Suppositories containing a local anaesthetic and corticosteroids are helpful. If symptoms are more severe, rubber band ligation or injection of a sclerosant can be used. Haemorrhoidal artery ligation operation (HALO) uses Doppler ultrasound to identify and ligate feeding arteries to the haemorrhoids and is being used more frequently in place of surgery.

Anal fissures

An anal fissure is a tear in the sensitive skin-lined lower anal canal distal to the dentate line, which produces pain on defecation. It can

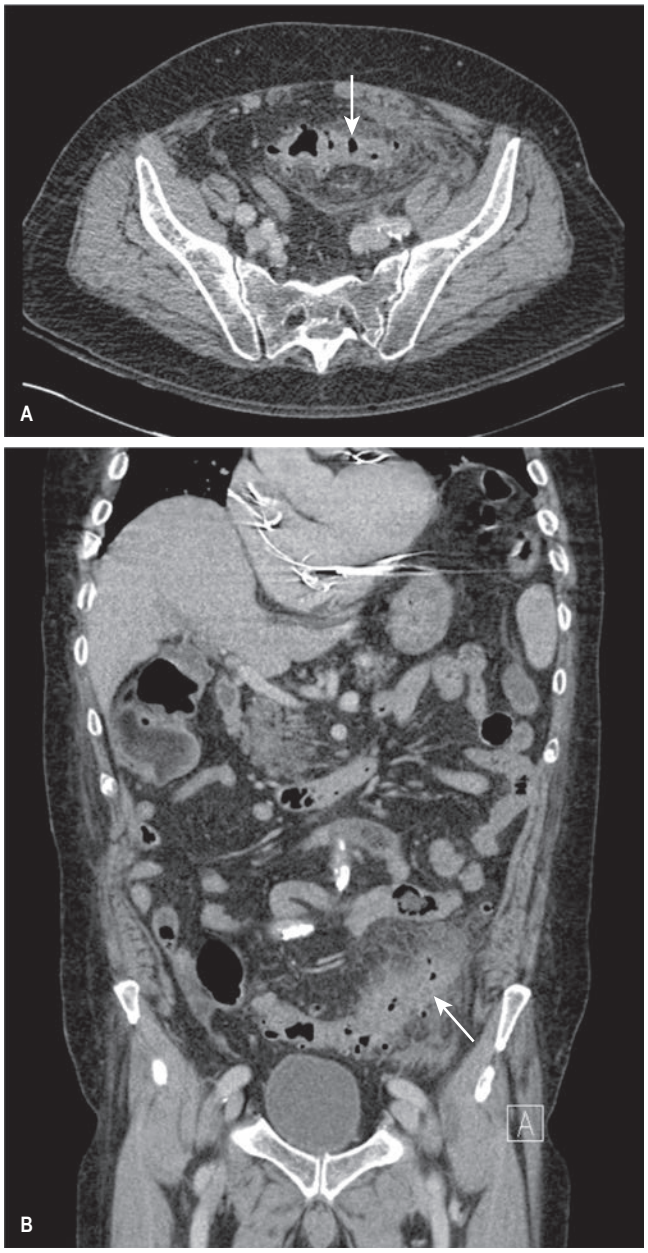


Fig. 22.49 CT colonography. Axial **(A)** and coronal **(B)** CT images demonstrate bowel wall thickening involving the sigmoid colon at the site of diverticular disease and inflammation of the adjacent mesenteric fat consistent with acute diverticulitis *(arrow)*.

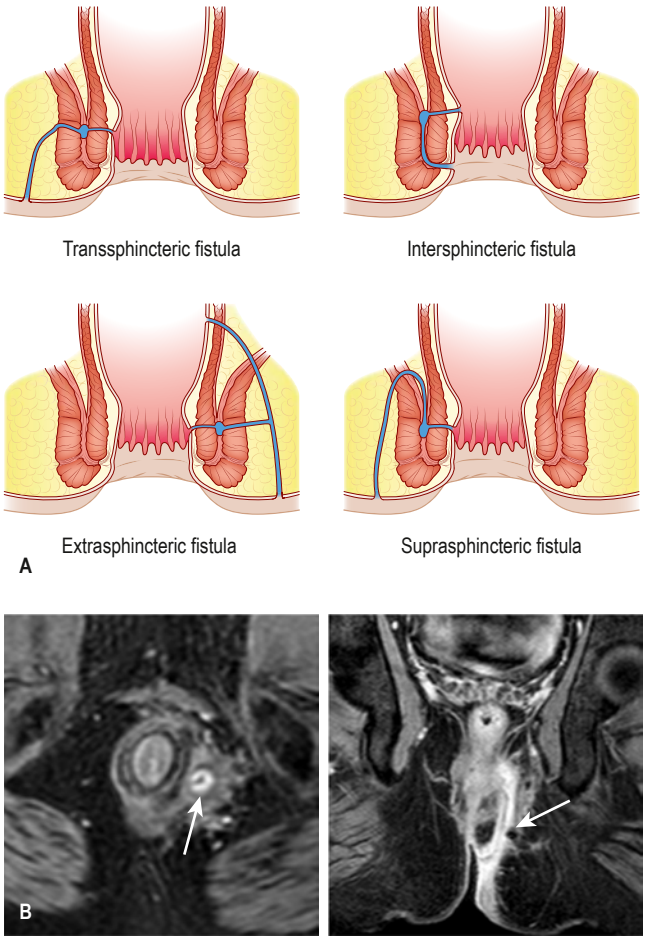


Fig. 22.50 Perianal fistulae. **(A)** Common sites. **(B)** Fistulae and sepsis in Crohn's disease *(arrowed)*. An MRI of the pelvis showing a complex extrasphincteric horseshoe tract with extension to the ischio-anal fossae (right; *arrowed*).

be an isolated primary problem in young to middle-aged adults or may occur in association with Crohn's disease or ulcerative colitis, in which case perianal abscesses and anal fistulae can complicate the fissure.

Diagnosis can usually be made on the history alone and confirmed on perianal inspection. Rectal examination is often not possible because of pain and sphincter spasm. The spasm not only causes pain but also impairs wound healing. In severe cases, proctoscopy and sigmoidoscopy should be performed under anaesthesia to exclude other anorectal diseases. Initial treatment is with local anaesthetic gel and stool softeners. Use of 0.4% glyceryl trinitrate and 2% diltiazem ointments is of benefit. Botulinum toxin is used in chronic fissures but lateral subcutaneous internal sphincterotomy is also employed for severe cases.

Fistula in ano and anorectal abscesses

The anatomy of perianal fistulae may be simple or complex (Fig. 22.50). The fistulae usually present as abscesses and heal after the abscess is incised. In other cases, a small, discharging pilonidal sinus may be noted by the patient.

Endoanal ultrasonography, MRI, and/or examination under anaesthetic are usually required to define the primary and any secondary tracks, exclude sepsis, and detect any associated disease, such as Crohn's disease and tuberculosis. Management is with surgical incision and drainage with antibiotics.

Rectal prolapse, intussusception, and solitary rectal ulcer syndrome

All these conditions are thought to be related, rectal prolapse being the unifying pathology. Some patients with solitary rectal ulcer syndrome (SRUS) do not have prolapse but strain excessively and ulcerate the anterior rectal wall, which is forced into the anus during attempts at defecation. Constipation and chronic straining may be precipitating causes. Patients commonly present with slight bleeding and mucus on defecation, tenesmus, and a sensation of anal obstruction.

SRUS is commonly located on the anterior wall of the rectum within 13 cm of the anal verge, and is sometimes difficult to distinguish from cancer and Crohn's disease during endoscopic

examination. SRUS has typical histological features of non-specific inflammatory changes with bands of smooth muscle extending into the lamina propria.

Asymptomatic SRUS should not be treated. Symptomatic patients should be advised to stop straining and measures should be taken to soften the stool. If rectal prolapse can be demonstrated during defecation, this should be repaired; in severe cases, surgical treatment by rectopexy may be indicated. Surgical treatment for complete rectal prolapse is also required.

Colonic tumours

Colon polyps and polyposis syndromes

A colonic polyp is an abnormal growth of tissue projecting from the colonic mucosa. Polyps range from a few millimetres to several centimetres in diameter and are single or multiple, pedunculated, sessile, or 'flat' (Fig. 22.51).

Many histological types of polyp are found in the colon (Box 22.42). However, adenomas are the precursor lesions in most cases of colon cancer.

Classification of colorectal polyps

Classification is summarized in Box 22.42.

Sporadic adenomas

An adenoma is a benign, dysplastic tumour of columnar cells or glandular tissue. Adenomas have tubular, tubulovillous, or villous morphology. The vast majority of adenomas are not inherited and are termed 'sporadic'. Although many sporadic adenomas do not become malignant in the patient's lifetime, they have a tendency to progress to cancer via increasing grades of dysplasia due to progressive accumulation of genetic changes (adenoma–carcinoma sequence). Factors favouring malignant transformation in colorectal polyps, and the relation between adenoma size and likelihood of cancer, are shown in Box 22.43.

The progression from benign polyp to cancer is shown in Fig. 22.52.

The likelihood of an adenoma being present increases with age; it is rare before the age of 30 years. By the age of 60 to 70,5% of asymptomatic subjects will have a polyp equal to or longer than 1 cm, or cancer with no symptoms, and up to 50% will have at least one small (less than 1 cm) adenoma. Removal of polyps at colonoscopy and subsequent surveillance reduce the risk of development of colon cancer by approximately 80%. Techniques such as chromoscopy, using dye spray or narrow band imaging, are being used to assist in their detection (flat adenomas account for approximately 12% of all adenomas).

Polyps in the rectum and sigmoid often present with rectal bleeding. More proximal lesions rarely produce symptoms and most are diagnosed on barium enema, CT colonography, or on colonoscopy performed for screening or for other reasons. Large villous adenomas can present with profuse diarrhoea with mucus and hypokalaemia.

Once a polyp has been found, it is almost always possible to remove it endoscopically. Surveillance guidelines dictate the frequency of repeat investigations:

- at 5 years, if 1 or 2 adenomas <1 cm are found
- at 3 years, if there are 3 to 4 small adenomas or at least 1 that is >1 cm
- at 1 year, if there are ≥5 small adenomas or there are ≥3, at least 1 of which is >1 cm.

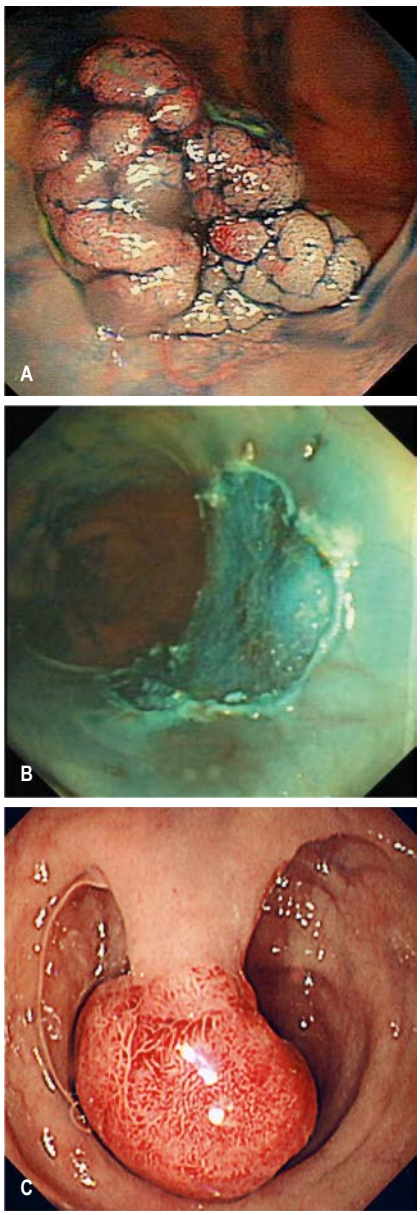


Fig. 22.51 Colonoscopic appearance of sessile and pedunculated polyps. **(A)** A sessile flat polyp with Indigo carmine. **(B)** Colon after resection of a sessile polyp. **(C)** A large pedunculated polyp.

If any doubt exists about the completeness of excision of any polyp, then an earlier repeat examination is suggested.

Sessile serrated adenomas

Serrated polyps form a heterogeneous group of colorectal lesions that includes the benign hyperplastic polyps (HPs), and the premalignant sessile serrated adenoma (SSA) and traditional serrated adenoma (TSA). They are characterized by the saw-tooth appearance of the crypt epithelium. It is now recognized that approximately 30% of colorectal cancers (particularly those in the right colon) originate from these lesions. Progression of serrated polyps is thought to occur via a distinct pathway from the adenoma–carcinoma sequence, with involvement of *BRAF* gene mutations and gene promoter hypermethylation (CpG island methylator phenotype, or

Box 22.42 Classification of colorectal polyps and polyposis syndromes

Histology	Polyposis syndrome	Defective gene	Inheritance	CRC risk
Hyperplastic	Hyperplastic polyposis	*BRAF*		Yes
Hamartoma	Juvenile polyposis	*MADH4* or *BMPR1A*	AD	10%–70%
	Peutz–Jeghers syndrome	*STK11*	AD	Yes
	Cowden's syndrome	*PTEN*	AD	10%?
	Lhermitte–Duclos disorder	*PTEN*	AD	
	Bannayan–Riley–Ruvalcaba syndrome	*PTEN*	AD	
Inflammatory	None	None		No
Lymphoid	Benign lymphoid polyposis	Unknown		No
Adenoma	FAP	*APC*	AD	100%
	AFAP			Yes
	Gardner's syndrome			Yes
	Turcot's syndrome			Yes
	MYH-AP	*MYH-AP*	AR	
Adenoma	Lynch syndrome (HNPCC)	Mismatch repair genes *(MSH-2, MLH-1)*	AD	70%–80%

AD, Autosomal dominant; *AFAP,* attenuated FAP; *APC,* adenomatous polyposis coli; *AR,* autosomal recessive; *CRC,* colorectal cancer; *FAP,* familial adenomatous polyposis; *HNPCC,* hereditary non-polyposis colorectal cancer; *MLH-1,* MutL homologue 1; *MSH-2,* MutS homologue 2; *MYH-AP,* MUT Y homologue-associated polyposis; *PTEN,* phosphatase and tensin homologue.

Box 22.43 Factors affecting risk of malignant change in an adenoma

Factor	Higher risk	Lower risk
Size	>1.5 cm	<1 cm
Type	Sessile or flat	Pedunculated
Histology	Severe dysplasia	Mild dysplasia
	Villous architecture	Tubular architecture
	Squamous metaplasia	
Number	Multiple polyps	Single polyp

CIMP; see 'Colorectal cancer [genetics]' section, later). Microsatellite instability can be detected in both the adenoma–carcinoma and the serrated pathways. Endoscopic resection of SSAs and TSAs with appropriate surveillance is recommended.

Inherited polyposis syndromes

About 5% of colorectal cancers have a well-defined single-gene basis.

Familial adenomatous polyposis

Familial adenomatous polyposis (FAP) is an autosomal dominant condition arising from germline mutations of the *APC* gene, located on chromosome 5q21–q22. More than 825 different mutations have been identified. Penetrance is virtually 100%. It is characterized by the presence of hundreds to thousands of colorectal and duodenal adenomas. The mean age of adenoma development is 16 years; the average age at which colorectal cancer develops is 39 years. Tracing and screening of relatives are essential, and affected individuals should be offered a prophylactic colectomy. Surgical options include colectomy and ileorectal anastomosis, which requires lifelong surveillance of the rectal stump, or a restorative proctocolectomy or pouch procedure with complete removal of rectal mucosa.

Cystic gland polyps, predominantly in the proximal stomach, and duodenal adenomas are frequently found in FAP, as well as other extraintestinal lesions such as osteomas, epidermoid cysts, and desmoid tumours. The duodenal adenomas may progress to cancer and are the most common cause of death in colectomized patients with FAP. Congenital hypertrophy of the retinal pigment epithelium (CHRPE) occurs in many families with FAP. Other cancers in FAP include thyroid, pancreatic, and hepatoblastomas.

APC gene mutations can be found in about 80% of families with FAP. Once the mutation has been identified in an index case, other family members can be tested for the mutation, and screening can then be directed at mutation carriers. If a mutation cannot be found in a known FAP case, all family members should undergo clinical screening with regular colonoscopy.

Attenuated FAP may be missed as it presents later (average age 44 years) and has fewer polyps (<100), which tend to occur on the right side of the colon rather than on the left. It may be indistinguishable from sporadic cases but the gene mutation is in the *APC* germline.

MYH-associated polyposis

MUT Y homologue-associated polyposis (MYH-AP) is an autosomal recessive inherited syndrome of multiple colorectal adenomas and cancer. *MYH* is a base-excision-repair gene that corrects oxidative DNA damage. MYH-AP may account for 7% to 8% of families with the FAP phenotype in whom *APC* mutations cannot be found. Subjects with multiple adenomas or an FAP phenotype without *APC* mutations and with a family history compatible with a recessive pattern of inheritance should be tested for MYH-AP.

Lynch syndrome

In Lynch syndrome (previously called hereditary non-polyposis colon cancer, HNPCC), polyps are formed in the colon and may progress rapidly to colon cancer. It affects 1:5000 people, causing 3% to 10% of colorectal cancer cases.

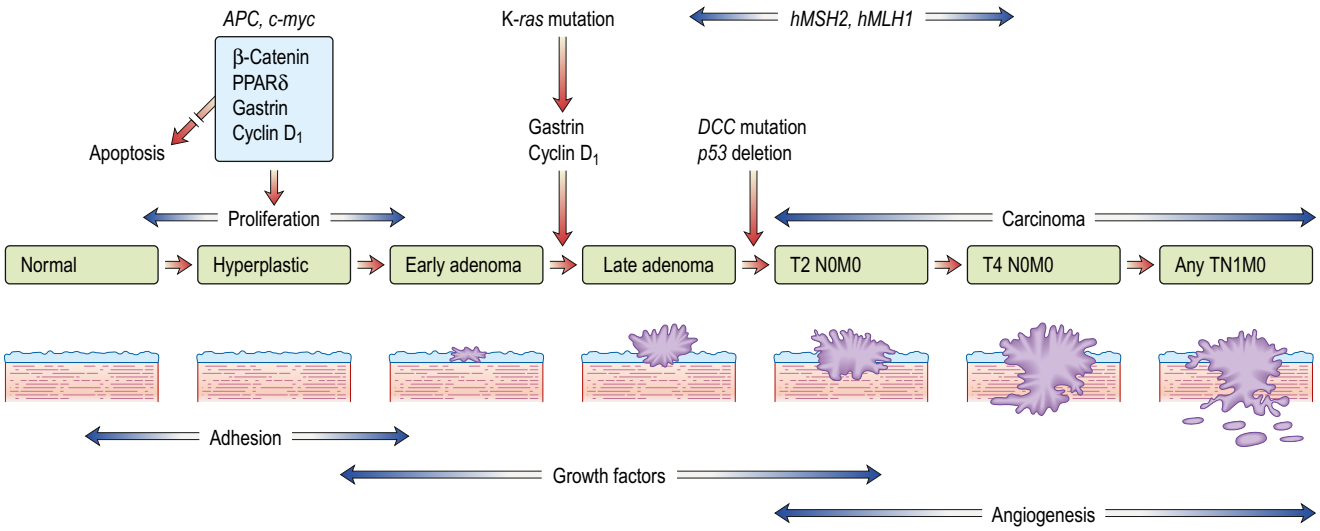


Fig. 22.52 The adenoma–carcinoma progression sequence, showing genetic mutations at different stages. *APC,* Adenomatous polyposis coli; *PPARδ,* peroxisome proliferator-activated receptor delta.

Box 22.44 Diagnostic criteria for Lynch syndrome

Modified Amsterdam criteria

- One individual diagnosed with CRC (or extracolonic Lynch tumours) before age 50 years
- Two affected generations
- Three affected relatives, one a first-degree relative of the other two
- FAP excluded
- Tumours verified by pathological examination

Bethesda guidelines

- CRC diagnosed in patient who are younger than 50 years
- Presence of synchronous, metachronous CRC, or other Lynch tumours, irrespective of age
- CRC with the MSI-H histology diagnosed in a patient who is younger than 60 years
- CRC diagnosed in one or more first-degree relative with a Lynch-related tumour, with one of the cancers being diagnosed under the age of 50 years
- CRC diagnosed in two or more first- or second-degree relatives with Lynch-related tumours, irrespective of age

The disease is caused by a mutation in one of the DNA mismatch repair genes, usually *hMSH2* or *hMLH1*, although others (*hMSH6, PMS1*, and *PMS*) have been reported. Mismatch repair genes are responsible for maintaining the stability of DNA during replication. Inheritance is autosomal dominant. The defect in function of the mismatch repair mechanism causes naturally occurring, highly repeated, short DNA sequences known as microsatellites to be shorter or longer than normal, a phenomenon called microsatellite instability (MSI).

Onset of cancer is earlier than in sporadic cases, at age 40 to 50 or younger. Tumours have a predilection for the right colon, in contrast to sporadic cases. In contrast to FAP, the lifetime risk of colon cancer (penetrance of the gene) in mutation carriers is 70% to 80%. Other cancers are also more common in Lynch syndrome: stomach, small intestine, bladder, skin, brain, and hepatobiliary system. Female patients are at risk for endometrial and ovarian cancer.

The diagnosis is made from the family history of colon cancer at a young age and the presence of associated cancers in the family. These are formalized in the various editions of the Amsterdam and the Bethesda criteria (**Box 22.44**).

Turcot's syndrome

This consists of FAP or Lynch syndrome (HNPCC) with brain tumours.

Gardner's syndrome

This involves, in addition to FAP, desmoid tumours, osteomas of the skull, and other lesions.

Hamartomatous polyps

These are commonly large and stalked. The inherited syndromes show autosomal dominant inheritance and include:

- ***Juvenile polyps***, which occur mainly in children and teenagers, and are found mainly in the colon; histologically, they show mucus retention cysts. Most are sporadic, but a syndrome of juvenile polyposis is defined as: more than three to five juvenile colonic polyps, juvenile polyps throughout the gastrointestinal tract, or any number of polyps with a family history. This is an autosomal dominant condition and the relevant gene has been identified (see **Box 22.42**). The polyps are a cause of bleeding and intussusception in the first decade of life. There is also an increased risk of colonic cancer (relative risk (RR) of 34), and surveillance and removal of polyps must be undertaken.
- ***Peutz–Jeghers syndrome*** (see 'Peutz–Jeghers syndrome' section).
- ***PTEN hamartoma–tumour syndrome*** (PHTS), which includes Cowden's syndrome, Bannayan–Riley–Ruvalcaba syndrome, and all syndromes caused by germline phosphatase and tensin homologue *(PTEN)* mutations. Cowden's (multiple hamartoma) syndrome is associated with characteristic skin stigmata, and by intestinal polyps regarded as hamartomas but with a mixture of cell types. These patients have an increased risk of various extraintestinal malignancies (thyroid, breast, uterine and ovarian). These syndromes are uncommon and together account for less than 1% of colon cancer cases.

Colorectal carcinoma

Colorectal cancer (CRC) is the third most common cancer worldwide and the second most common cause of cancer death in the UK.

Each year approximately, 40 000 new cases are diagnosed in England and Wales (68% colon, 32% rectal cancer), and CRC is

Box 22.45 Risk factors in colorectal cancer

Increased risk
- Increasing age
- Animal fat (saturated) and red meat consumption
- Sugar consumption
- Colorectal polyps
- Family history of colon cancer or colonic polyps
- Chronic inflammatory bowel disease
- Obesity (body and abdominal)
- Smoking
- Alcohol misuse
- Acromegaly
- Abdominal radiotherapy
- Ureterosigmoidostomy

Decreased risk
- Vegetable, garlic, milk, calcium consumption
- Exercise (colon only)
- Aspirin (including low-dose) and other NSAIDs

NSAIDs, Non-steroidal anti-inflammatory drugs.

registered as the cause of death in about half this number. The prevalence rate per 100 000 (at all ages) is 53.5 for males and 36.7 for females. The incidence increases with age; the average age at diagnosis is 60 to 65 years. Approximately 20% of patients in the UK have distant metastases at diagnosis. The disease is much more common in Westernized countries than in Asia or Africa.

Factors related to risk of colorectal cancer are shown in **Box 22.45**.

Genetics

Most colorectal cancers develop as a result of progression from normal mucosa to adenoma to invasive cancer. This progression is controlled by the accumulation of abnormalities in a number of critical growth-regulating genes and can be divided into three main pathways:

- ***Chromosomal instability (CIN).*** CIN is the most common cause of conventional adenomas throughout the colon. This pathway involves the sequential accumulation of genetic mutations in tumour suppressor genes, usually initiated by a mutation in the gene encoding adenomatous polyposis coli *(APC).*
- ***CpG island methylator phenotype.*** CpG island methylator phenotype (CIMP) tumours arise via the serrated neoplasia pathway and have a marked predilection for the proximal colon. Following an initiating genetic mutation in the genes encoding *BRAF* or *KRAS*, these lesions progress via *epigenetic* silencing of tumour suppressor and mismatch repair (MMR) genes by promoter methylation. This pathway is epitomized by the serrated polyposis syndrome (see 'Sessile serrated adenomas' section). The stepwise accumulation of genetic mutations in onco- and tumour suppressor genes that underpins the SCRC carcinogenesis pathway is well established and has significantly altered worldwide clinical practice. For example, mutations in KRAS appear to have importance as predictive factors for lack of response to certain oncological therapies. Activating KRAS mutations, which are identified in approximately 40% to 45% of CRCs, are associated with resistance to treatments that target the epidermal growth factor receptor (EGFR), such as cetuximab and panitumumab.
- ***Microsatellite instability.*** Microsatellite instability (MSI) tumours are also more commonly located in the proximal colon. They arise from defective DNA repair through inactivation of mismatch repair genes, epitomized by the germline mutation of MMR genes seen in Lynch syndrome (HNPCC).

This molecular classification can help to distinguish clinical characteristics, such as patient demographics, tumour distribution, response to therapy, and prognosis.

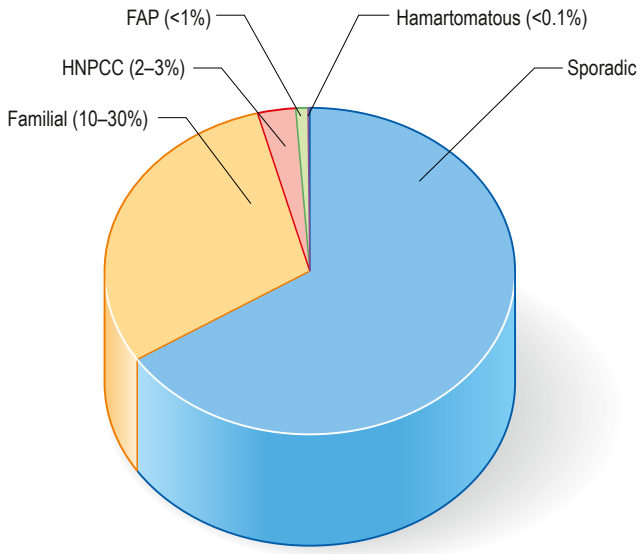


Fig. 22.53 Percentages of colon cancer according to family risk. *FAP,* Familial adenomatous polyposis; *HNPCC,* hereditary non-polyposis colorectal cancer (Lynch syndrome).

Box 22.46 Lifetime risk of colorectal cancer (CRC) in first-degree relatives of a CRC patient

Individuals affected	Risk
Population risk	1 in 50
One first-degree relative affected (any age)	1 in 17
One first-degree and one second-degree relative affected	1 in 12
One first-degree relative affected (age < 45)	1 in 10
Two first-degree relatives affected	1 in 6
Autosomal dominant pedigree	1 in 2

From Houlston RS, Murday V, Harocopos C et al. Screening and genetic counselling for relatives of patients with colorectal cancer in a family cancer clinic. Br Med J *1990; 301:366–368.*

Cancer families

A family history of CRC confers an increased risk to relatives. Family history is, next to age, the most common risk factor for CRC. FAP (**Fig. 22.53**) is the best-recognized syndrome predisposing to CRC but represents less than 1% of all colorectal cancers. Lynch syndrome (HNPCC) accounts for 3% to 10% of familial cancer (see 'Lynch syndrome' section).

Additionally, some colon cancers arise, at least in part, from an inherited predisposition, so-called familial risk (**Box 22.46**). Estimates of their frequency range from 10% to 30% of all CRC but the genes involved have yet to be identified. The risk of CRC can be estimated from a family history matched with empirical risk tables, so that appropriate advice regarding screening can be offered.

Most CRCs are, however, sporadic and occur in individuals without a strong family history (**Fig. 22.54**).

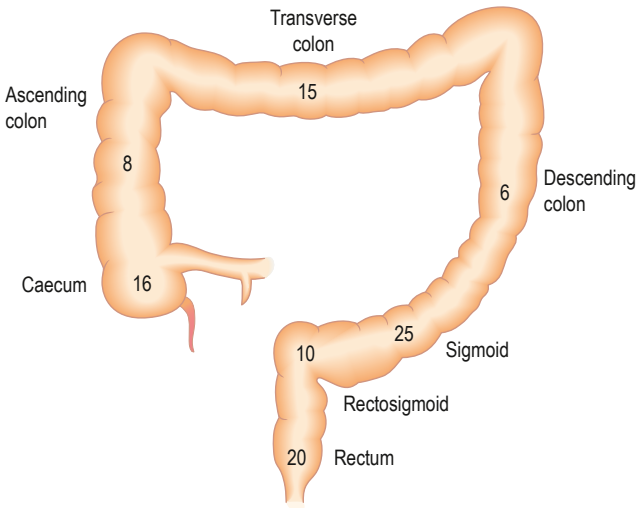


Fig. 22.54 Distribution of sporadic colorectal cancer. Only 60% are in the range of flexible sigmoidoscopy.

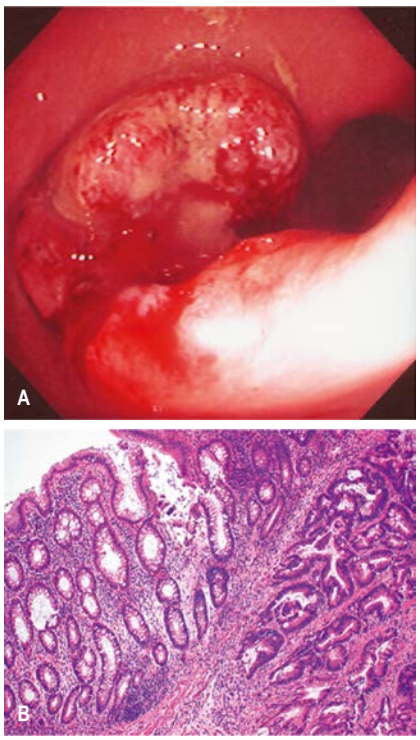


Fig. 22.55 Carcinoma in the ascending colon. (A) Colonoscopic appearance of a large irregular ulcer. **(B)** Histopathology showing an adenocarcinoma.

Pathology

CRC, which usually takes the form of a polypoid mass with ulceration, spreads by direct infiltration through the bowel wall. It involves lymphatics and blood vessels with subsequent spread, most commonly to the liver and lung. Synchronous cancers are present in 2% of cases. Histology is adenocarcinoma with variably differentiated glandular epithelium and mucin production. 'Signet ring' cells, in which mucin displaces the nucleus to the side of the cell, are relatively uncommon and generally have a poor prognosis.

Clinical features

Symptoms suggestive of colorectal cancer include change in bowel habit with looser and more frequent stools, rectal bleeding, tenesmus, and symptoms of anaemia. A rectal or abdominal mass may be palpable. Cancers arising in the caecum and right colon are often asymptomatic until they present as an iron deficiency anaemia. Cancer may present with intestinal obstruction.

Patients aged over 35 to 40 years presenting with new large bowel symptoms should be investigated. Digital examination of the rectum is essential and examination of the colon should be performed in all cases.

Screening for CRC

The UK has introduced a national screening programme using stool testing for blood in patients aged 60 to 64 years, in which positive tests have identified that 10% have cancer and 40% have adenomas. A randomized trial in Norway has found an increased number of early-stage cancers in the screened population but a high incidence of interval cancers, arising between biennial screens. Patients with an increased risk of developing bowel cancer (those with a family history of familial adenomatous polyposis or Lynch's syndrome [hereditary non-polyposis colorectal cancer]) will have a more intensive screening programme, starting at an earlier age.

- ***FOB tests*** have been studied as a screening test for colorectal cancer. Several large randomized studies have demonstrated a reduction in cancer-related mortality of 15% to 33%. Immunologically based FOB tests are superior to the conventional guaiac-based systems. The disadvantage of screening with FOB is its relatively low sensitivity, which means many negative colonoscopies. In FOB test screen-positive patients in the UK National Bowel Cancer Screening Programme (NHS BCSP), about 10% have cancer, 40% have adenomas and the colon is normal in 50%. Faecal immunochemical tests (FIT) for blood—FIT for haemoglobin are more specific than guaiac tests because FIT detects only human globin and therefore does not detect upper gastrointestinal bleeding (since the globin is digested in transit). Additionally, foods with peroxidase activity do not produce a positive reaction. FIT has replaced FOBt in the bowel cancer screening programme in the UK since April 2018.
- ***Flexible sigmoidoscopy*** screening has been shown to reduce the mortality from CRC, but not overall mortality.
- ***Colonoscopy*** is the 'gold standard' technique for examination of the colon and rectum and is the investigation of choice for high-risk patients. Universal screening strategies have been implemented in the UK for subjects between 60 and 69 who have a positive FOB. Cancer has been detected in 8% to 12% of patients, 75% of which was located in the left colon; 72% of detected cancers are at an 'early' stage (10% polyp cancer, 32% stage I, and 30% stage II).
- ***CT colonography*** ('virtual colonoscopy', see Fig. 22.10) is increasingly being used.

Investigations

- ***Colonoscopy*** is the 'gold standard' for investigation and allows biopsy for histology. Biopsy of the tumour is mandatory, usually at endoscopy (Fig. 22.55).
- ***Double-contrast barium enema*** can visualize the large bowel but has now been superseded by CT colonography.
- ***Endoanal ultrasound and pelvic MRI*** are used for staging rectal cancer.

- ***Chest, abdominal and pelvic CT*** scanning evaluate tumour size, local spread, and liver and lung metastases, contributing to tumour staging.
- ***PET scanning*** is useful for detecting occult metastases and for evaluation of suspicious lesions found on CT or MRI.
- ***MRI*** is also useful for evaluating suspicious lesions found on CT or ultrasound, especially in the liver.
- ***Serum carcinoembryonic antigen (CEA)*** is of little use for primary diagnosis and should not be performed as a screening test. It is useful for follow-up; rising levels suggest recurrence.
- ***Faecal occult blood (FOB) tests*** are used for mass population screening.

Management

The management of CRC is discussed in the Chapter 41.

Follow-up

All patients who have surgery should have a total colonoscopy performed before surgery to look for additional lesions. If total colonoscopy cannot be achieved before surgery, a second 'clearance' colonoscopy within 6 months of surgery is essential. Patients with stage II or III disease should be followed up with regular colonoscopy and CEA measurements; rising levels of CEA suggest recurrence. Annual CT scanning of the chest and abdomen to detect operable liver metastases should be performed for up to 3 years post surgery.

Further reading

Advani A, Kopetz S. Ongoing and future directions in the management of metastatic colorectal cancer. *J Surg Oncol* 2019; 642–652.
Assman SA, Keszthelyi, Kleijnen J et al. Guideline for the diagnosis and treatment of Faecal Incontinence—a joint European collaboration. *UEG J* 2022; 10(3):251–286.
Bevan R, Rutter MD. Colorectal cancer screening—who, how and when? *Clin Endosc* 2018; 51(1):37–49.
Bharucha AE, Knowles CH, Mack I et al. Faecal incontinence in adults. *Nat Rev Dis Primers* 2022; 8:53.
Bharucha AE, Lacy BE. Mechanisms, evaluation, and management of chronic constipation. *Gastroenterology* 2020; 158(5):1232–1249.
Ewing I, Hurley JJ, Josephides E et al. The molecular genetics of colorectal cancer. *Frontline Gastroenterology* 2014; 5:26–30.
Hanna MH, Kaiser AM. Update on the management of sigmoid diverticulitis. *World J Gastroenterol* 2021; 27(9):760–781.
Soler-González G, Sastre-Valera J, Viana-Alonso A et al. Update on the management of elderly patients with colorectal cancer. *Clin Transl Oncol* 2023; s12094.
Strate LL, Morris AM. Epidemiology, pathophysiology, and treatment of diverticulitis. *Gastroenterology* 2019; 156(5):1282–1289.

THE ACUTE ABDOMEN

This section deals with the acute abdominal conditions that cause the patient to be hospitalized within a few hours of the onset of pain (Box 22.47). If they are recognized quickly as an emergency, a reduction in morbidity and mortality can be achieved. Although a specific diagnosis should be attempted, the immediate problem in management is to decide whether an 'acute abdomen' exists and whether surgery is required.

Diagnosis

History

This should include previous operations, any gynaecological problems, and presence of any concurrent medical condition.

Box 22.47 Common causes of acute abdominal pain

Diagnosis	%[a]
Non-specific abdominal pain	35
Acute appendicitis	30
Gall bladder disease	10
Gynaecological disorders	5
Intestinal obstruction	5
Perforated ulcer/dyspepsia	5
Renal colic	2
Urinary tract infection	2
Diverticular disease	2
Other diagnoses	4

[a]Percentages are approximate and vary in different communities.

Pain

The onset, site, type, and subsequent course of the pain should be determined as accurately as possible. In general, the pain of an acute abdomen can be either constant (usually owing to inflammation) or colicky (because of a blocked 'tube'). The inflammatory nature of a constant pain will be supported by a raised temperature, tachycardia, and/or a raised white cell count. If these are normal, then other causes (e.g., musculoskeletal, aortic aneurysm), even rare ones (e.g., porphyria), should be considered. Colicky pain can be due to an obstruction of the gut, biliary system, urogenital system, or uterus. These cases will probably require conservative management initially, along with analgesics. If a colicky pain becomes a constant pain, then inflammation of the organ may have supervened (e.g., strangulated hernia, ascending cholangitis, or salpingitis).

A ***sudden onset*** of pain suggests:

- perforation (e.g., of a duodenal ulcer)
- rupture (e.g., of an ectopic pregnancy)
- torsion (e.g., of an ovarian cyst)
- acute pancreatitis
- infarction (e.g., mesenteric).

Back pain suggests:

- pancreatitis
- rupture of an aortic aneurysm
- renal tract disease.

Inflammatory conditions (e.g., appendicitis) produce a more gradual onset of pain. With peritonitis, the pain is continuous and may be made worse by movement. Many inflammatory conditions can progress to those listed as having a sudden onset due to complications.

Vomiting

Vomiting may accompany any acute abdominal pain but, if persistent, suggests an obstructive lesion of the gut. The character of the vomit should be asked about. Does it contain blood, bile, or small bowel contents?

Other symptoms

Any change in bowel habit or of urinary frequency should be documented and, in females, a gynaecological history, including last menstrual period, should be taken.

Physical examination

The general condition of the person should be noted. Do they look ill or shocked? Large volumes of fluid may be lost from the vascular

compartment into the peritoneal cavity or into the lumen of the bowel, giving rise to hypovolaemia: that is, a pale, cold skin, a weak, rapid pulse, and hypotension.

The abdomen

- ***Inspection.*** Look for the presence of scars, distension, or masses.
- ***Palpation.*** The abdomen should be examined gently for sites of tenderness and the presence or absence of guarding. Guarding is involuntary spasm of the abdominal wall and indicates peritonitis. This can be localized to one area or may be generalized, involving the whole abdomen.
- ***Bowel sounds.*** Increased high-pitched, tinkling bowel sounds indicate fluid obstruction; this occurs because of fluid movement within the dilated bowel lumen. Absent bowel sounds suggest peritonitis. In an obstructed patient, absent bowel sounds may be due to strangulation, ischaemia, or ileus. It is essential for the hernial orifices to be examined if intestinal obstruction is suspected.

Vaginal and rectal examination

Vaginal examination can be very helpful, particularly in diagnosing gynaecological causes of an acute abdomen (e.g., a ruptured ectopic pregnancy). Rectal examination is less helpful, as localized tenderness may be due to any cause; it may show blood on the glove.

Other observations

- ***Temperature.*** Fever is more common in acute inflammatory processes.
- ***Urine.*** Examine for:
  - blood—suggests urinary tract infection or renal colic
  - glucose and ketones—ketoacidosis can present with acute pain
  - protein and white cells—to exclude acute pyelonephritis.
- ***Medical causes.*** These should be borne in mind (Box 22.48).

Investigations

- ***Blood count.*** A raised white cell count occurs in inflammatory conditions.
- ***Serum lipase or amylase.*** High levels (more than five times normal) indicate acute pancreatitis. Raised levels below this can occur in any acute abdomen and should not be considered diagnostic of pancreatitis.
- ***Serum electrolytes.*** These are not particularly helpful for diagnosis but are useful for general evaluation of the patient.
- ***Pregnancy test.*** A urine dipstick is used for females of childbearing age.
- ***X-rays.*** An erect chest X-ray is useful to detect air under the diaphragm caused by a perforation (Fig. 22.56). Dilated loops of bowel or fluid levels are suggestive of obstruction on abdominal X-ray (Fig. 22.57).
- ***Ultrasound.*** This is useful in the diagnosis of acute cholangitis, cholecystitis, and aortic aneurysm, and in expert hands it is reliable in the diagnosis of acute appendicitis. Gynaecological and other pelvic causes of pain can be detected.
- ***CT scan.*** Spiral CT of the abdomen and pelvis is the most accurate investigation in most acute emergencies. It should be used more often to avoid unnecessary laparotomies.
- ***Laparoscopy.*** This is used increasingly as a diagnostic tool prior to proceeding with surgery, particularly in men and women over the age of 50 years. In addition, therapeutic procedures, such as appendicectomy, can be performed.

Box 22.48 Medical causes of acute abdomen

- Referred pain:
  - Pneumonia
  - Myocardial infarction
- Functional gastrointestinal disorders
- Renal causes:
  - Pelviureteric colic
  - Acute pyelonephritis
- Metabolic causes:
  - Diabetes mellitus
  - Acute intermittent porphyria
  - Lead poisoning
  - Familial Mediterranean fever
- Haematological causes:
  - Haemophilia and other bleeding disorders
  - Henoch–Schönlein purpura
  - Sickle cell crisis
  - Polycythaemia vera
  - Paroxysmal nocturnal haemoglobinaemia
- Vasculitis

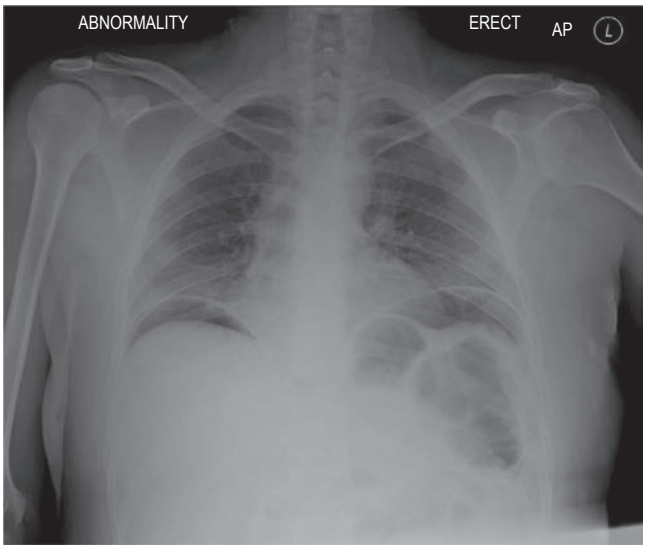


Fig. 22.56 Erect chest radiograph demonstrating free gas under the diaphragm (pneumoperitoneum).

Acute appendicitis

This is a common surgical emergency and affects all age groups. Appendicitis should always be considered in the differential diagnosis if the appendix has not been removed.

Acute appendicitis mostly occurs when the lumen of the appendix becomes obstructed with a faecolith; however, in some cases, there is only generalized acute inflammation. If the appendix is not removed at this stage, gangrene occurs with perforation, leading to a localized abscess or to generalized peritonitis.

Clinical features

Most patients present with abdominal pain; in many, it starts vaguely in the centre of the abdomen, becoming localized to the right iliac fossa in the first few hours. Nausea, vomiting, anorexia, and occasional diarrhoea can occur.

Examination of the abdomen usually reveals tenderness in the right iliac fossa, with guarding due to the localized peritonitis. There may be a tender mass in the right iliac fossa.

Fig. 22.57 Abdominal radiograph demonstrating multiple dilated loops of small bowel consistent with small bowel obstruction.

Although raised white cell counts, ESR and CRP are helpful, other laboratory tests can be less valuable. An ultrasound scan can detect an inflamed appendix and can also indicate an appendix mass or other localized lesion. CT is highly sensitive (98.5%) and specific (98% negative predictive value; 99.5% positive predictive value), and reduces the incidence of removal of a 'normal' appendix. With the use of these investigations, the incidence of 'normal' appendix histology has fallen to 15% to 20%.

Differential diagnosis

- Non-specific mesenteric lymphadenitis—may mimic appendicitis.
- Acute terminal ileitis due to Crohn's disease or *Yersinia* infection.
- Gynaecological causes:
  - inflamed Meckel's diverticulum
  - functional bowel disease.

Management

The appendix is removed by laparoscopic surgery. If an appendix mass is present, the patient is usually treated conservatively with intravenous fluids and antibiotics. The pain subsides over a few days and the mass usually disappears over a few weeks. Interval appendicectomy is recommended at a later date to prevent further acute episodes.

Gynaecological causes of an acute abdomen

Ruptured ectopic pregnancy

The fallopian tube is the most common extrauterine site of implantation. Delayed diagnosis is the major cause of morbidity. Most patients will present with recurrent low abdominal pain associated with vaginal bleeding. Diagnosis is usually made with abdominal and transvaginal ultrasound. Most patients can be managed by laparoscopic salpingostomy or salpingectomy.

Ovarian causes

- Rupture of 'functional' ovarian cysts in the middle of the cycle (Mittelschmerz).
- Torsion or rupture of ovarian cysts.

Acute salpingitis

Most cases are associated with sexually transmitted infection. Patients commonly present with bilateral low abdominal pain, a fever, and vaginal discharge. In the ***Fitz-Hugh–Curtis syndrome***, *Chlamydia* infection tracks up the right paracolic gutter to cause a perihepatitis. Patients can present with acute right hypochondrial pain, fever, and mildly abnormal liver biochemistry.

Acute peritonitis

Localized peritonitis

There is virtually always some degree of localized peritonitis with all acute inflammatory conditions of the gastrointestinal tract (e.g., acute appendicitis, acute cholecystitis). Pain and tenderness are largely features of this localized peritonitis. The treatment is for the underlying disease.

Generalized peritonitis

This is a serious condition, resulting from irritation of the peritoneum owing to infection (e.g., perforated appendix) or from chemical irritation due to leakage of intestinal contents (e.g., perforated ulcer). In the latter case, superadded infection gradually occurs; *E. coli* and *Bacteroides* are the most common organisms.

The peritoneal cavity becomes acutely inflamed, with production of an inflammatory exudate that spreads throughout the peritoneum, leading to intestinal dilation and paralytic ileus.

Clinical features

In perforation, the onset is sudden with acute, severe abdominal pain, followed by general collapse and shock. The patient may improve temporarily, only to become worse later as generalized toxaemia occurs.

When peritonitis is secondary to inflammatory disease, the onset is less rapid, the initial features being those of the underlying disease.

Investigations

Investigations should always include an erect chest X-ray. X-ray is used to detect free air under the diaphragm, and serum amylase is measured to diagnose acute pancreatitis, which is treated conservatively. Imaging with ultrasound and/or CT should always be performed for diagnosis.

Management

Peritonitis is treated surgically after adequate resuscitation and the re-establishment of a good urinary output. This includes insertion of a nasogastric tube, intravenous fluids, and antibiotics. Surgery has a two-fold objective:

- peritoneal lavage of the abdominal cavity
- specific treatment of the underlying condition.

Complications

Any delay in the treatment of peritonitis produces more profound toxaemia and septicaemia, which may lead to the development of

Box 22.49 Causes of intestinal obstruction

Small intestinal obstruction
- Adhesions (80% in adults)
- Hernias
- Crohn's disease
- Intussusception
- Obstruction due to extrinsic involvement by cancer

Colonic obstruction
- Carcinoma of the colon
- Sigmoid volvulus
- Diverticular disease

multiorgan failure (see Chapter 13). Local abscess formation can occur and should be suspected if a patient continues to remain unwell postoperatively, with a swinging fever, high white cell count and continuing pain. Abscesses are commonly pelvic or subphrenic, and can be localized and drained using ultrasound and CT scanning techniques.

Intestinal obstruction

Most intestinal obstruction is due to a mechanical block. Sometimes the bowel does not function, leading to a paralytic ileus. This occurs temporarily after most abdominal operations and with peritonitis. The most common cause in adults is adhesions (**Box 22.49**).

Obstruction of the bowel leads to bowel distension above the block, with increased secretion of fluid into the distended bowel. Bacterial contamination takes place in the distended stagnant bowel. In strangulation, the blood supply is impeded, leading to gangrene, perforation, and peritonitis unless urgent treatment of the condition is undertaken.

Clinical features

The patient complains of abdominal colic, vomiting, and constipation without passage of wind. In upper gut obstruction the vomiting is profuse, but in lower gut obstruction it may be absent.

Examination of the abdomen reveals distension with increased bowel sounds. Marked tenderness suggests strangulation, and urgent surgery is necessary. Examination of the hernial orifices and rectum must be performed. X-ray of the abdomen reveals distended loops of bowel proximal to the obstruction. Fluid levels are seen in small bowel obstruction on an erect film. In large bowel obstruction, the caecum and ascending colon are distended. An instant, water-soluble Gastrografin enema without air insufflation may help to demonstrate the site of the obstruction. CT can localize the lesion accurately and is the investigation of choice.

Management

Initial management is by resuscitation with intravenous fluids (mainly 0.9% saline with potassium) and decompression. Many cases will settle on conservative management, but an increasing temperature, raised pulse rate, increasing pain, and a rising white cell count require urgent scanning and possible exploratory laparotomy.

Laparotomy with removal of the obstruction will be necessary in some cases of small bowel obstruction. If the bowel is gangrenous owing to strangulation, gut resection will be required. A few patients (e.g., those with Crohn's disease) may have recurrent episodes of incomplete intestinal obstruction that can be managed conservatively. In large bowel obstruction due to malignancy, a self-expanding metal stent can be used, followed by elective surgery. In critically ill patients, a defunctioning colostomy may be needed. Volvulus of the sigmoid colon can be managed by the passage of a flexible sigmoidoscope or a rectal tube to un-kink and deflate the bowel, but recurrent volvulus may require sigmoid resection.

Acute colonic pseudo-obstruction

A clinical picture mimicking mechanical obstruction may develop in patients who do not have a mechanical cause. In more than 80% of cases, it complicates other clinical conditions, such as:
- intra-abdominal trauma, pelvic, spinal and femoral fractures
- postoperative states (abdominal, pelvic, cardiothoracic, orthopaedic, neurosurgical)
- intra-abdominal sepsis
- pneumonia
- metabolic disorders (e.g., electrolyte disturbances, malnutrition, diabetes mellitus, Parkinson's disease)
- drugs—opiates (particularly after orthopaedic surgery), antidepressants, antiparkinsonian drugs.

Patients present with rapid and progressive abdominal distension and pain. X-ray shows a gas-filled large bowel. Management is of the underlying problem (e.g., withdrawal of opiate analgesia), together with a trial of intravenous neostigmine therapy. Patients should be monitored and consideration should be given to surgery if the diameter of the caecum exceeds 14 cm.

Peritoneum

Anatomy and physiology

The peritoneal cavity is a closed sac lined by mesothelial cells; these produce surfactant, which acts as a lubricant within the peritoneal cavity. The cavity contains <100 mL of serous fluid containing <30 g/L of protein.

The mesothelial cells lining the diaphragm have gaps that allow communication between the peritoneum and the diaphragmatic lymphatics. Approximately one-third of fluid drains through these lymphatics, the remainder through the parietal peritoneum. These mechanisms allow particulate matter to be removed rapidly from the peritoneal cavity.

Complement activation is an early defence mechanism and is followed rapidly by upregulation of the peritoneal mesothelial cells and migration of polymorphonuclear neutrophils and macrophages into the peritoneum.

Mast cells release potent mediators of inflammation, including histamine and eicosanoids, and interact with T cells to generate an immune response.

The peritoneum-associated lymphoid tissue includes the omental milky spots, the lymphocytes within the peritoneal cavity, and the draining lymph nodes. B cells with a unique $CD5^+$ are common. This defence system plays a major role in localizing peritoneal infection.

Disorders affecting the peritoneum

Conditions that can affect the peritoneum are shown in **Box 22.50**.

Peritonitis can be acute or chronic, as seen in TB. Most cases of infective peritonitis are secondary to gastrointestinal disease, but it occurs occasionally without intra-abdominal sepsis in ascites due to liver disease. Very rarely, fungal and parasitic infections (e.g., amoebiasis, candidiasis) can also cause primary peritonitis. Peritonitis is discussed further in the 'Acute peritonitis' section.

Box 22.50 Diseases of the peritoneum

Infective (bacterial) peritonitis
- Secondary to gut disease, for example, appendicitis
- Perforation of any organ
- Chronic peritoneal dialysis
- Spontaneous, usually in ascites with liver disease
- Tuberculosis

Neoplasia
- Secondary deposits (e.g., from ovary, stomach)
- Primary mesothelioma

Vasculitis
- Rheumatic autoimmune disease
- Polyserositis (e.g., familial Mediterranean fever)

The peritoneum can be involved by secondary malignant deposits, and the most common cause of ascites in a young to middle-aged woman is an ovarian carcinoma.

A ***subphrenic abscess*** is usually secondary to infection in the abdomen and is characterized by fever, malaise, pain in the right or left hypochondrium, and shoulder-tip pain. An erect chest X-ray may show gas under the diaphragm, impaired movement of the diaphragm on screening, and/or a pleural effusion. Ultrasound is usually diagnostic. Percutaneous catheter drainage inserted under CT or ultrasound guidance and antibiotics constitute highly successful therapy.

Ascites is associated with all diseases of the peritoneum. The fluid that collects is an exudate with a high protein content. It is also seen in liver disease. The mechanism, causes, and investigation of ascites are discussed in Chapter 11.

Peritoneal adhesions

Adhesions form as a result of abdominal or pelvic surgery, or inflammation in the abdominoperitoneal cavity. They cause a variety of conditions, including adhesive small bowel obstruction (ASBO), chronic abdominal pain, complications during future surgery, and female infertility when they involve the fallopian tubes or ovaries. There is no satisfactory medical or surgical treatment and so surgical techniques have been developed to minimize peritoneal injury.

Retroperitoneal fibrosis (peri-aortitis)

This is a rare condition, in which there is a marked fibrosis over the posterior abdominal wall and retroperitoneum. It is associated with raised serum IgG4 levels and is described in Chapter 26.

Tuberculous peritonitis

This is the second most common form of abdominal TB. Three subgroups can be identified: wet, dry, and fibrous.
- In patients with the ***wet*** type, ascitic fluid should be examined for protein concentration (>20 g/L) and tubercle bacilli (rarely found).
- In the ***dry*** form, patients present with subacute intestinal obstruction, which is due to tuberculous small bowel adhesions.
- In the ***fibrous*** form, patients present with abdominal pain, distension, and ill-defined, irregular, tender abdominal masses.

The diagnosis of peritoneal TB can be supported by findings on ultrasound or CT screening (mesenteric thickening and lymph node enlargement). A histological diagnosis is not always required before instituting treatment. In some patients, careful laparoscopy (to avoid perforation) may have to be performed, and rarely laparotomy.

Management

Drug treatment is similar to that for pulmonary TB (see Chapter 20) and should be supervised by chest physicians who have experience in dealing with contacts.

Bibliography

Feldman M, Friedman LS, Brandt LL. *Sleisenger and Fordtran's Gastrointestinal and Liver Disease*, 10th edn. Philadelphia: WB Saunders; 2015.

Significant websites

http://www.bsg.org.uk *British Society of Gastroenterology.*
http://www.coeliac.co.uk *Coeliac UK.*
http://www.corecharity.org.uk *Gastric ulcer and GORD.*
https://www.crohnsandcolitis.org.uk *Crohn's and Colitis UK.*

Questions for this chapter can be found on Elsevier eBooks+.

23 Hepatology

Graham R. Foster and Alastair O'Brien

Clinical skills 680
History 680
Examination 680
Investigations 680
Background 682
Anatomy of the liver and biliary system 682
Liver 682
Biliary system 683
Functions of the liver 683
Protein metabolism 683
Carbohydrate metabolism 683
Lipid metabolism 683
Formation of bile 684
Hormone and drug inactivation 686
Immunological function 686
Investigations in the patient with liver disease 686
Jaundice 690
Hepatitis 692
Viral hepatitis 694
Acute hepatitis due to other infectious agents 703
Acute hepatic failure 704
Autoimmune hepatitis 705
Drug-induced chronic hepatitis 706
Chronic hepatitis of unknown cause 706
Metabolic dysfunction–associated steatotic liver disease 707
Cirrhosis 708
Liver transplantation 711
Complications and effects of cirrhosis 712
Types of cirrhosis 718
Alcohol and alcoholic liver disease 722
Budd–Chiari syndrome 726
Hepatic sinusoidal obstruction syndrome 726
Fibropolycystic diseases 726
Liver abscess 727
Other infections of the liver 728
Liver disease in pregnancy 728
Liver tumours 728
Secondary liver tumours 728
Primary malignant tumours 729
Benign tumours 730
Miscellaneous conditions of the liver 730
Drugs and the liver 732
Drug metabolism 732
Drug hepatotoxicity 732
Drug prescribing for patients with liver disease 733

CORE SKILLS AND KNOWLEDGE

Liver disease is common and its incidence is rising, predominantly because of viral hepatitis in low and middle income countries, and alcohol and obesity in developed ones, although obesity and alcohol are increasingly common causes of liver disease in low and middle income countries. It remains a major cause of mortality and morbidity, with death occurring secondary to cirrhosis and hepatocellular carcinoma.

Hepatologists divide their time between inpatient work (decompensated chronic liver disease and acute liver failure both require intensive inpatient management) and clinic work (managing patients with stable liver disease and diagnosing patients newly referred from primary care). Many also perform procedures such as endoscopic retrograde cholangiopancreatography (ERCP) or endoscopic ultrasound (EUS); and some subspecialize in the management of alcohol cessation, viral hepatitis, or liver transplantation.

Key skills in hepatology include:

- Managing patients with decompensated chronic liver disease, including variceal bleeding, ascites, and encephalopathy, and those presenting with acute liver failure
- Appreciating the role of history, examination, imaging, blood tests, and liver biopsy in the diagnosis of patients newly presenting with evidence of liver disease
- Understanding the natural history of hepatitis B and C infection and the effective available treatments
- Empathy and understanding to tackle alcoholism and obesity.

Opportunities to learn hepatology include attending ward rounds on inpatient liver units, participating in outpatient clinics, observing the work of specialist nurses who manage increasing numbers of patients with chronic liver disease, learning how to perform paracentesis safely on medical day units, and watching interventional endoscopy (variceal banding) and ERCP procedures being performed.

CLINICAL SKILLS

History

Both acute and chronic liver disease are often asymptomatic, especially in the early stages. Taking a history in hepatology involves not only asking about symptoms but also establishing the presence of common risk factors for liver disease to help establish a diagnosis (Box 23.1).

Asking about alcohol

This requires gentleness and tact (Box 23.2). Patients often underestimate their alcohol intake or degree of dependence. This history should be non-judgemental, and help should be offered through referral to alcohol support services if patients express a desire to become abstinent.

Examination

In acute liver disease there may be a few signs, apart from jaundice and an enlarged liver. In the cholestatic phase of the illness, pale stools and dark urine are present.

The figure on the next page illustrates the various signs associated with chronic liver disease, although physical examination is occasionally normal even in patients with advanced disease.

Investigations

A detailed overview of the diagnostic investigations used in liver disease begins at the 'Investigations in the patient with liver disease' section. Fig. 23.1 illustrates how these investigations may be employed in an outpatient setting for a patient referred from primary care with raised liver transaminases.

Box 23.1 Taking a history in liver disease

Symptoms of acute liver disease
- May be asymptomatic and anicteric
- Symptomatic disease, often viral, producing malaise, anorexia, and fever
- Jaundice (see later), often appearing as the illness progresses

Symptoms of chronic liver disease
- Non-specific symptoms, particularly weakness, anorexia, and fatigue
- Right hypochondrial pain due to liver distension
- Abdominal distension due to ascites
- Ankle swelling due to fluid retention
- Haematemesis and melaena from variceal haemorrhage
- Pruritus due to cholestasis—often an early symptom of primary biliary cholangitis (PBC)
- Endocrine dysfunction: gynaecomastia, loss of libido and amenorrhoea
- Confusion and drowsiness due to hepatic encephalopathy

Past medical history
- Any previous disease affecting the liver, biliary tract, pancreas, or gut
- Heart failure (risk of congestive hepatopathy)
- History of autoimmune disease (↑risk of autoimmune liver disorders)
- Features of metabolic syndrome: obesity, diabetes, gout, hypercholesterolaemia (risk of NAFLD)

Medications
- Recent antibiotic use (several cause cholestasis)
- Paracetamol overdose
- Full medication history, establishing whether any cause liver damage

Family history
- History of jaundice, liver disease, or specific diseases such as haemochromatosis

Alcohol use
- Take a detailed history (see later)

Risk factors for viral hepatitis
- Childhood spent in countries where hepatitis B or C is endemic
- Contaminated needles—unsafe tattoos, injecting drug use
- High-risk unprotected sexual activity

Box 23.2 Taking an alcohol history in patients with liver disease

Extent of use
- Preferred type of alcohol (beer, wine, spirits); strength where relevant ('super-strength' beers can be up to 10% ethanol)
- Daily consumption (in cans, bottles or glasses, subsequently converted to units)
- Number of abstinent days per week
- Frequency of consuming more than eight units on one occasion
- Duration of alcohol use, and changing use over this time

Degree of dependence
- Number of times alcohol has interfered with regular activities (e.g., going to work)
- Number of times the person has been unable to stop drinking once they have started
- Need for a morning drink to 'get going'
- Feelings of guilt about alcohol consumption
- Concern on the part of family and friends about alcohol use
- Injuries to self or others caused by alcohol

Views on abstinence
- Awareness of risks of current level of alcohol use (if this exceeds target levels)
- Willingness to cut down versus abstaining completely
- Previous attempts to give up
- Previous experiences of support groups, including Alcoholics Anonymous, and cognitive behavioural therapy
- Sources of support: family members, close friends, social or religious groups

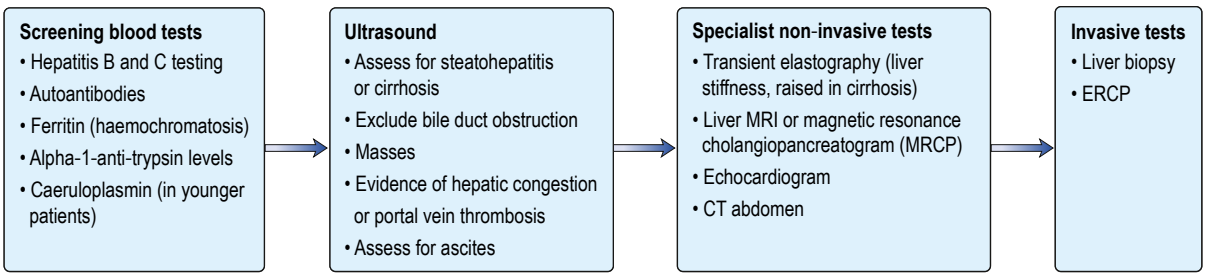
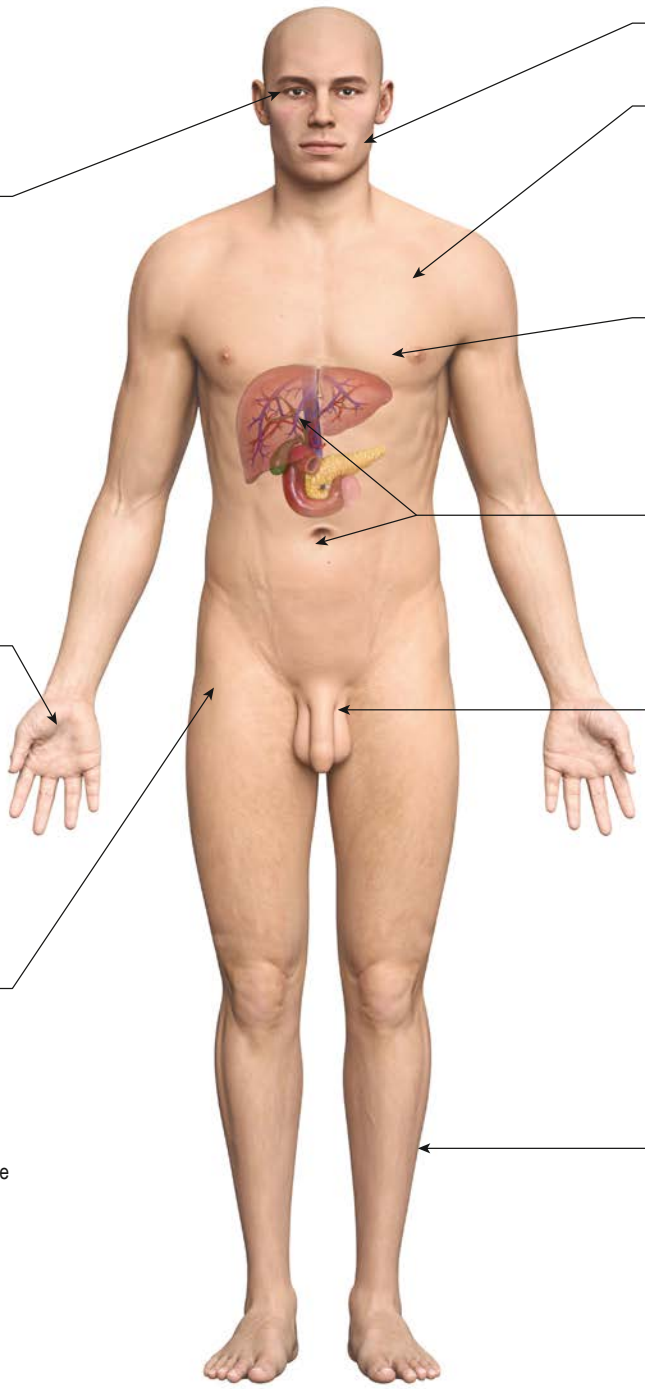


Fig. 23.1 Use of diagnostic investigations in an outpatient setting. This scheme would be appropriate for a patient referred from primary care with raised liver transferases.



General
Jaundice
Smell of alcohol/fetor hepaticus
Encephalopathy
Weight loss

Eyes
- Jaundice
- Kayser–Fleischer rings (Wilson's disease)
- Xanthelasma (on eyelids)

Hands
Flapping tremor
Palmar erythema
Dupuytren's contracture (alcohol)
Nails
- Clubbing
- Leuconychia

Skin
Scratch marks

NB: Often patients with liver disease have very few signs.

Parotid swelling

Spider naevi

Gynaecomastia

Abdomen
Scars
Distension (ascites – shifting dullness)
Dilated superficial veins
Hepatomegaly (small in cirrhosis)
Splenomegaly
Tumour
Palpable gall bladder

Testicular atrophy

Oedema (pitting)
Bruises

BACKGROUND

The aetiology of liver disease differs from region to region. In the developed world, liver inflammation is most often due to obesity, the metabolic syndrome (metabolic dysfunction–associated steatotic liver disease or MASLD, formerly referred to as non-alcoholic fatty liver disease or NAFLD), metabolic dysfunction–associated steatohepatitis (MASH, formerly non-alcoholic steatohepatitis) and alcohol excess. In the developing world, chronic viral infection with either hepatitis B or hepatitis C is the leading cause of liver mortality, although there are now excellent treatments for hepatitis C while obesity is increasing globally. Liver disease is the third leading cause of premature death in the UK, where there has been a 63.6% increase in liver disease–related premature deaths over the past two decades, with alcohol-related liver disease deaths increasing by 74.2%. Complications of liver disease account for 1 million deaths worldwide, with increasing prevalence in recent years and globally, about half a billion people suffer from chronic viral hepatitis. Effective health education and the improvement in public health, along with vaccination programmes, should help to stop the spread of viral infections and hopefully reduce risk factors for the metabolic syndrome (diabetes, hypertension, and obesity).

Cirrhosis represents the final common pathway for liver diseases and is characterized by progressive fibrosis of the liver parenchyma, which leads to portal hypertension (elevated blood pressure within the portal vein) and deterioration of liver function. In decompensated cirrhosis (ascites, hepatic encephalopathy, or variceal bleeding) the median overall survival is 2 years, which is a far worse prognosis than for many cancers.

Imaging techniques enable the liver, biliary tree, and pancreas to be visualized with precision, which can increase earlier diagnosis. Liver transplantation is an established therapy for both acute and chronic liver disease.

ANATOMY OF THE LIVER AND BILIARY SYSTEM

Liver

The liver is the body's largest internal organ (1.2–1.5 kg) and is situated in the right hypochondrium. A functional division into the larger right lobe (containing caudate and quadrate lobes) and the left lobe is made by the middle hepatic vein. The liver is further subdivided into eight segments (Fig. 23.2) by divisions of the right, middle, and left hepatic veins. Each segment has its own portal pedicle, permitting individual segment resection at surgery.

The hepatic blood supply constitutes 25% of the resting cardiac output and is delivered via two main vessels, entering via the liver hilum (porta hepatis):

- ***The hepatic artery***, a branch of the coeliac axis, supplies 25% of the hepatic blood flow. The hepatic artery autoregulates flow, ensuring a constant total blood flow.
- ***The portal vein*** drains most of the gastrointestinal tract and the spleen. It supplies 75% of hepatic blood flow. The normal portal pressure is 5 to 8 mmHg; flow increases after meals.

The blood from these vessels is distributed to the segments and flows into the sinusoids via the portal tracts.

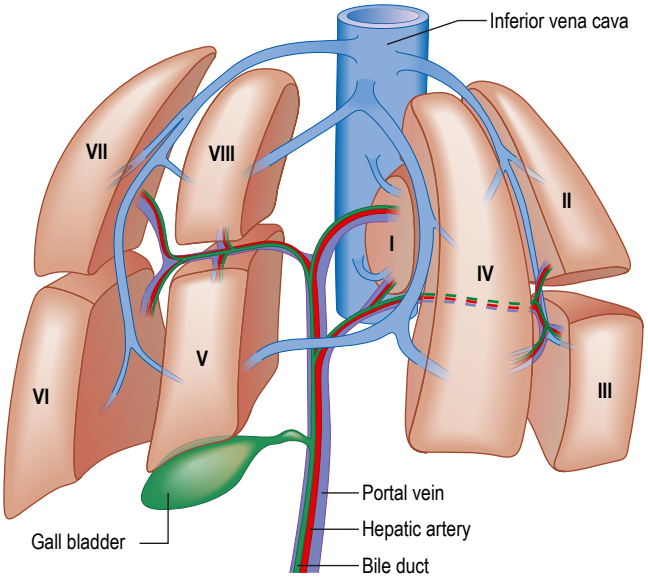


Fig. 23.2 Segmental anatomy of the liver. The eight hepatic segments are shown: I, caudate lobe; II–IV, left hemiliver; V–VIII, right hemiliver.

Blood leaves the sinusoids, entering branches of the hepatic vein, which join into three main branches before entering the inferior vena cava.

The ***caudate lobe*** is an autonomous segment, as it receives an independent blood supply from the portal vein and hepatic artery, and its hepatic vein drains directly into the inferior vena cava.

Lymph, formed mainly in the perisinusoidal space, is collected in lymphatics that are present in the portal tracts. These small lymphatics enter larger vessels, which eventually drain into the portal system.

The ***acinus*** is the functional hepatic unit. This consists of parenchyma supplied by the smallest portal tracts containing portal vein radicles, hepatic arterioles, and bile ductules (Fig. 23.3). The hepatocytes near this triad (zone 1) are well supplied with oxygenated blood and are more resistant to damage than the cells nearer the terminal hepatic (central) veins (zone 3).

The ***sinusoids*** lack a basement membrane and are loosely surrounded by specialist fenestrated endothelial cells and Kupffer cells (phagocytic cells). Sinusoids are separated by plates of liver cells (hepatocytes). The subendothelial space between the sinusoids and hepatocytes is the space of Disse, which contains a matrix of basement membrane constituents and stellate cells (see Fig. 23.21).

Stellate cells store retinoids in their resting state and contain the intermediate filament, desmin. When activated (to myofibroblasts), they are contractile and regulate sinusoidal blood flow. Endothelin and nitric oxide play a major role in modulating stellate cell contractility. Stellate cells are activated by a wide variety of inflammatory cytokines (such as tumour necrosis factor-alpha, TNF-α); once activated, they generate extracellular matrix proteins, including collagen, leading to fibrosis and eventually cirrhosis. Under appropriate conditions, stellate cells can also produce proteases that degrade the extracellular matrix, leading to reversal of fibrosis. The balance between collagen production and degradation is critical to the

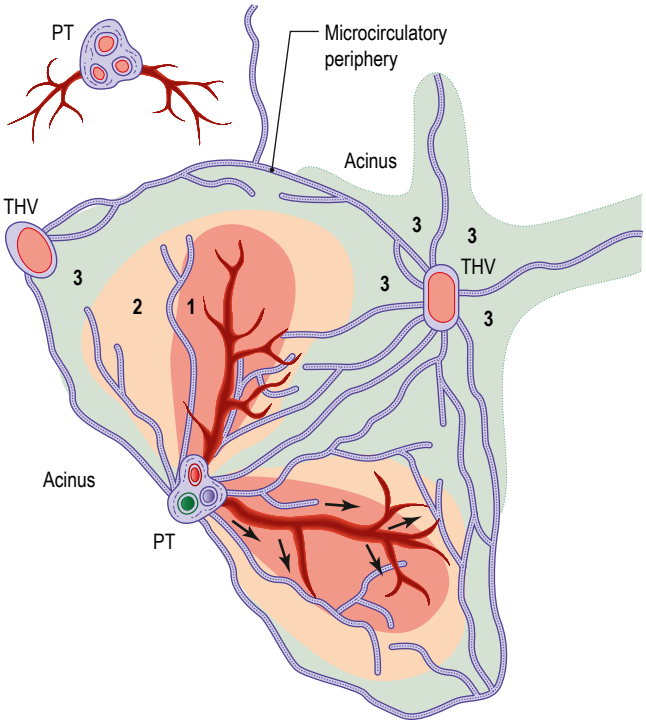


Fig. 23.3 The acinus. Zones 1, 2, and 3 represent areas supplied by blood, zone 1 being best oxygenated. Zone 3 is supplied by blood remote from afferent vessels and is in the microcirculatory periphery of the acinus. The perivascular area (the star-shaped green area around the terminal hepatic venule, THV) is formed by the most peripheral parts of zone 3 and is made up of several adjacent acini; it is the least well-oxygenated area. *PT,* Portal triad.

progression of liver scarring and cirrhosis development/resolution (see 'Cirrhosis [pathogenesis]' section).

Biliary system

Bile canaliculi form a network between the hepatocytes. These join to form thin bile ductules near the portal tract, which, in turn, enter the bile ducts in the portal tracts. These then combine to form the right and left hepatic ducts, which leave each liver lobe. The hepatic ducts join at the porta hepatis to form the common hepatic duct. The cystic duct connects the gall bladder to the lower end of the common hepatic duct. The gall bladder lies under the right lobe of the liver and stores and concentrates hepatic bile; its capacity is approximately 50 mL. The common bile duct is formed at the junction of the cystic and common hepatic ducts and is 8 mm in diameter or less, passing through the head of the pancreas, and narrowing at its lower end to pass into the duodenum. The common bile duct and pancreatic duct open into the second part of the duodenum, most often through a common channel at the ampulla of Vater, which contains the muscular sphincter of Oddi. This contracts rhythmically and prevents all of the bile from entering the duodenum, by maintaining a higher pressure than the gall bladder in the fasting state.

FUNCTIONS OF THE LIVER

Protein metabolism

See also Chapter 25.

Synthesis and storage

The liver is the principal site of synthesis of all circulating proteins, apart from γ-globulins (produced in lymphoid tissue). The liver receives amino acids from the intestine and muscles and, by controlling the rate of gluconeogenesis and transamination, regulates plasma levels. Plasma contains 60 to 80 g/L of protein, mainly albumin, globulin, and fibrinogen.

Albumin has a half-life of 16 to 24 days, and 10 to 12 g is synthesized daily. Its main functions are to maintain intravascular oncotic (colloid osmotic) pressure, and to transport water-insoluble substances such as bilirubin, hormones, fatty acids, and drugs. Reduced synthesis of albumin over prolonged periods produces hypoalbuminaemia and is seen in chronic liver disease and malnutrition. Hypoalbuminaemia is also found in hypercatabolic states (e.g., trauma, burns, and sepsis) and in diseases associated with an excessive loss (e.g., nephrotic syndrome or protein-losing enteropathy).

Transport or carrier proteins, such as transferrin and caeruloplasmin, acute phase, and other proteins (e.g., α_1-antitrypsin and α-fetoprotein) are also produced in the liver.

The liver also synthesizes all coagulation factors (except for one-third of factor VIII)—that is, fibrinogen, prothrombin, factors V, VII, IX, X, and XIII, proteins C and S, and antithrombin (see Chapter 38), as well as components of the complement system. The liver stores large amounts of certain vitamins, particularly A, D, and B_{12}, lesser amounts of others (vitamin K and folate), and minerals—iron in ferritin and haemosiderin, and copper.

Degradation (nitrogen excretion)

Amino acids are degraded by transamination and oxidative deamination to produce ammonia, which is then converted to urea and excreted by the kidneys. This is the major pathway for the elimination of nitrogenous waste. Failure of this process occurs in severe liver disease.

Carbohydrate metabolism

Glucose homeostasis and maintenance of blood sugar are major functions of the liver. It stores approximately 80 g of glycogen. In the immediate fasting state, blood glucose is maintained either by glucose release from glycogen breakdown (glycogenolysis) or by synthesis of new glucose (gluconeogenesis). Sources for gluconeogenesis are lactate, pyruvate, amino acids from muscles (mainly alanine and glutamine), and glycerol from lipolysis of fat stores. In prolonged starvation, ketone bodies and fatty acids are used as alternative sources of fuel as body tissues adapt to a lower glucose requirement (see Chapter 25).

Lipid metabolism

Fats are insoluble in water and are transported in plasma as protein–lipid complexes (lipoproteins). These are discussed in detail in Chapter 19.

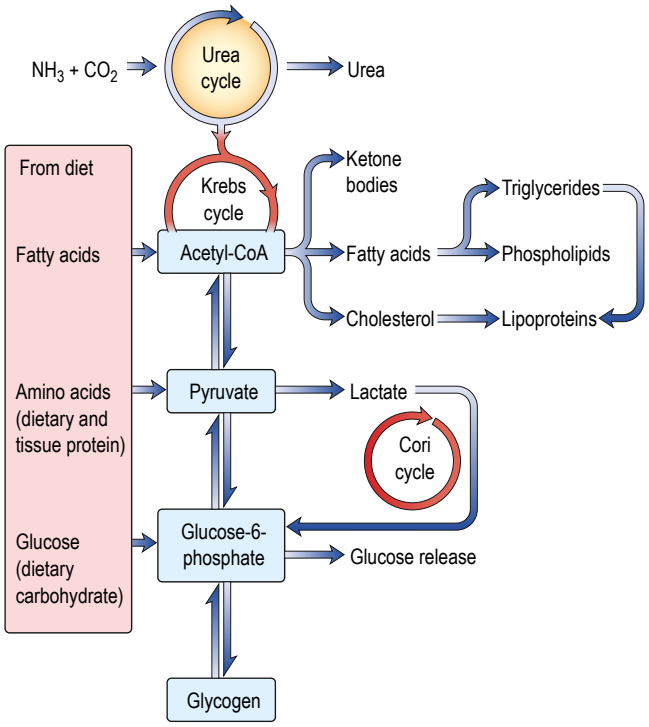


Fig. 23.4 Interrelationships of protein, carbohydrate, and lipid metabolism in the liver.

The liver has a major role in the metabolism of lipoproteins. It synthesizes very low-density lipoproteins (VLDLs) and high-density lipoproteins (HDLs). HDLs are the substrate for lecithin-cholesterol acyltransferase (LCAT), which catalyzes the conversion of free cholesterol to cholesterol ester (see later). Hepatic lipase removes triglyceride from intermediate-density lipoproteins (IDLs) to produce low-density lipoproteins (LDLs), which are degraded by the liver after uptake by specific cell-surface receptors (see Fig. 19.3).

Triglycerides are mainly of dietary origin but are also formed in the liver from circulating free fatty acids (FFAs) and glycerol, and incorporated into VLDLs. Oxidation or *de novo* synthesis of FFAs occurs in the liver, depending on availability of dietary fat.

Cholesterol may be of dietary origin but most is synthesized from acetyl-coenzyme A (acetyl-CoA) in the liver, intestine, adrenal cortex, and skin. It either occurs as free cholesterol or is esterified with fatty acids; this reaction is catalyzed by LCAT. This enzyme is reduced in severe liver disease, increasing the ratio of free cholesterol to ester, which alters membrane structures. One result of this is the red cell abnormalities (e.g., target cells) seen in chronic liver disease. Phospholipids (e.g., lecithin) are synthesized in the liver. The complex interrelationships between protein, carbohydrate, and fat metabolism are shown in Fig. 23.4.

Formation of bile

Bile secretion and bile acid metabolism

Bile consists of water, electrolytes, bile acids, cholesterol, phospholipids, and conjugated bilirubin. Two processes are involved in bile secretion across the canalicular membrane of the hepatocyte—***bile salt-dependent*** and ***bile salt-independent*** processes—each contributing about 230 mL/day. Another 150 mL/day is produced by bile ductule epithelial cells.

Bile formation requires uptake of bile acids and other organic and inorganic ions across the basolateral (sinusoidal) membranes by multiple transport proteins (sodium taurocholate co-transporting polypeptide [NTCP] and sodium-independent organic anion-transporting polypeptide 2 [OATP2]; Fig. 23.5). This process is driven by sodium/potassium adenosine triphosphatase (Na^+/K^+-ATPase) in the basolateral membranes. Intracellular transport across hepatocytes is partly through microtubules and partly by cytosol transport proteins.

Bile acids are also synthesized in hepatocytes from cholesterol, the rate-limiting step being those catalyzed mainly by cholesterol-7α-hydroxylase and the P450 enzymes (CYP7A1 and CYP8B1).

The farnesoid X bile acid receptor blocks bile acid formation from cholesterol and also regulates the transport proteins (NTCP, OATP2) that increase bile acid uptake by the liver. It is a target for a new class of therapeutic drugs, farnesoid X receptor (FXR) agonists (see section on 'Bile acid receptors in liver disease').

The canalicular membrane contains multispecific organic anion transporters, mainly ATPase-dependent (ATP binding cassette), the multidrug-resistant protein 2 (MRP2), multidrug-resistant protein 3 (MDR3), and the bile salt excretory pump (BSEP), which carry a broad range of compounds including bilirubin diglucuronide, glucuronidated and sulphated bile acids, and other organic anions against a concentration gradient into the biliary canaliculus. Na^+ and water follow the passage of bile salts by diffusion across the tight junction between hepatocytes (a bile salt-dependent process). In the bile salt-independent process, water flow is due to other osmotically active solutes such as glutathione and bicarbonate.

Secretion of a bicarbonate-rich solution is stimulated mainly by secretin and inhibited by somatostatin. This involves several membrane proteins, including the Cl^-/HCO_3^- exchanger and the cystic fibrosis transmembrane conductance regulator that controls Cl^- secretion, and water channels (aquaporins) in cholangiocyte membranes.

The bile acids are excreted into bile and pass via the common bile duct into the duodenum. The two primary bile acids—cholic acid and chenodeoxycholic acid (see Fig. 23.5)—are conjugated with glycine or taurine, which increases their solubility. Intestinal bacteria convert these acids into secondary bile acids, deoxycholic and lithocholic acid. Fig. 23.6 shows the enterohepatic circulation of bile acids.

The average total bile flow is 600 mL/day. During fasting, half flows into the duodenum and half is diverted into the gall bladder. The gall bladder mucosa absorbs 80%–90% of the water and electrolytes but is impermeable to bile acids and cholesterol. Following a meal, the I cells of the duodenal mucosa secrete cholecystokinin, which stimulates contraction of the gall bladder and relaxation of the sphincter of Oddi, allowing bile to enter the duodenum. An adequate bile flow is dependent on bile salts being returned to the liver by the enterohepatic circulation.

Bile acids act as detergents; their main function is lipid solubilization. Bile acid molecules have both a hydrophilic and a hydrophobic end. In aqueous solutions they form micelles, with

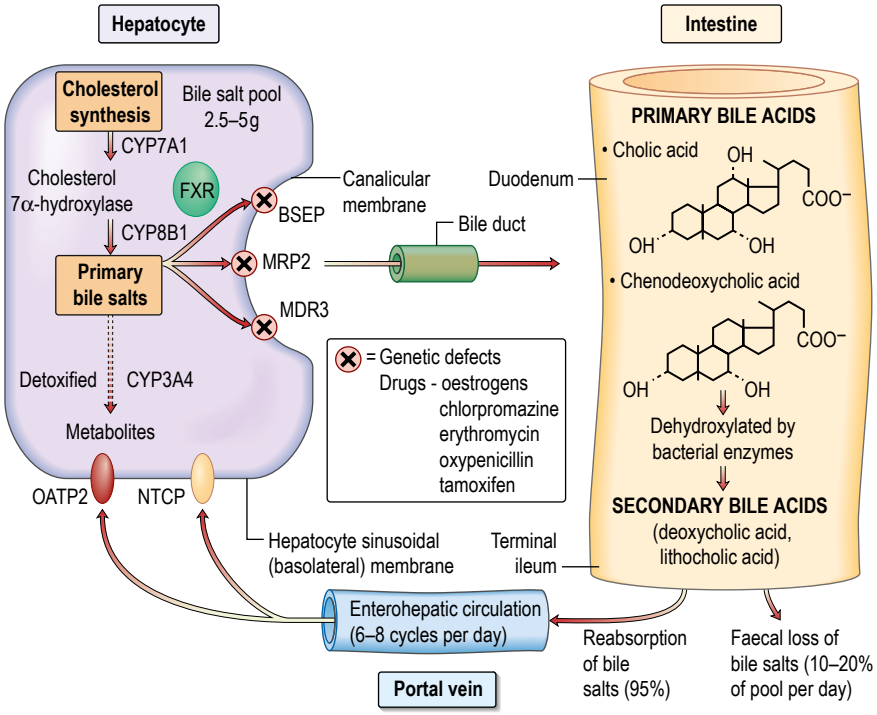


Fig. 23.5 Cholesterol synthesis and its conversion to primary and secondary bile acids. All bile acids are normally conjugated with glycine or taurine. *BSEP,* Bile salt export pump; *FXR,* nuclear farnesoid X receptor; *MDR3,* multidrug-resistant protein 3; *MRP2,* multidrug-resistant protein 2; *NTCP,* sodium taurocholate co-transporting polypeptide; *OATP2,* sodium-independent organic anion-transporting polypeptide 2.

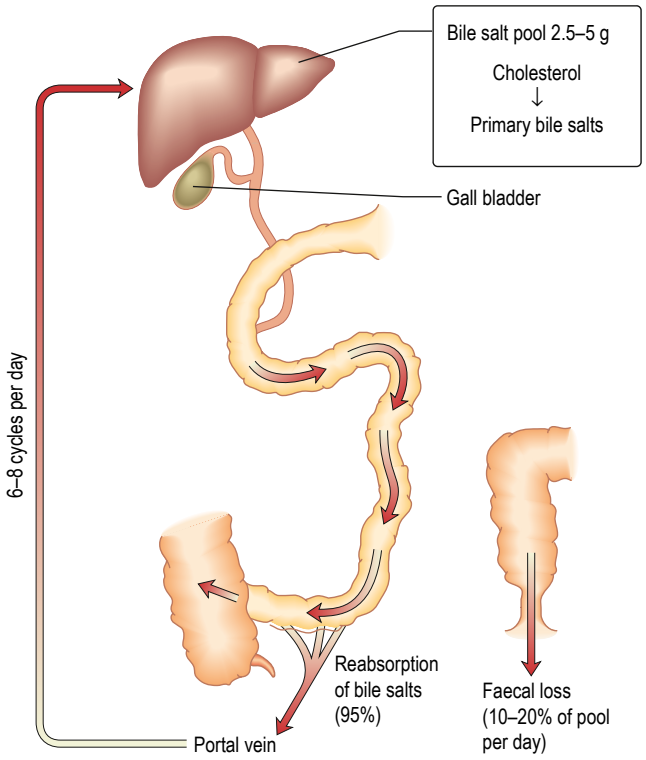


Fig. 23.6 Recirculation of bile acids. The bile salt pool is relatively small and the entire pool recycles six to eight times via the enterohepatic circulation. Synthesis of new bile acids compensates for faecal loss.

their hydrophobic (lipid-soluble) ends in the centre. Micelles are expanded by cholesterol and phospholipids (mainly lecithin), forming mixed micelles.

Bile acid receptors in liver disease

Bile acids have been identified as crucial cell signalling molecules that regulate multiple biological processes. Bile acids are endogenous ligands for FXR and TGR5, a G-protein coupled receptor. Gain- and loss-of-function studies have demonstrated that both are involved in the regulation of lipid and carbohydrate metabolism and inflammatory responses. These receptors may therefore be potential targets for treatment of MASLD (see 'Metabolic dysfunction–associated steatotic liver disease' section), and phase III trials of the FXR agonist obeticholic acid have completed in MASH with the drug outperforming placebo in helping more patients with pre-cirrhotic MASH achieve at least one stage of liver fibrosis improvement with no worsening of MASH. However, the drug failed to significantly resolve MASH and in addition, the drug had no benefit in more severe cirrhotic MASH patients. To date, it has not been approved for MASH therapy. Obeticholic acid has been approved for the second-line treatment of primary biliary cholangitis (PBC) following trial results demonstrating an improvement in liver blood tests. Finally, obeticholic acid showed potential beneficial effect for primary sclerosing cholangitis (PSC) in a phase II study. In cirrhotic animal models, this drug reduced both portal hypertension and gut bacterial translocation. However, hepatic decompensation, liver failure, and death have been reported when Child B or C cirrhosis patients are dosed more frequently than recommended, with the US Food and Drug Administration (FDA) releasing a black box warning for the use of obeticholic acid in

patients with decompensated cirrhosis and so it remains uncertain whether it will emerge as a treatment for portal hypertension.

Bilirubin metabolism

Bilirubin is produced mainly from the breakdown of mature red cells by Kupffer cells in the liver and reticuloendothelial system; 15% of bilirubin is formed from catabolism of other haem-containing proteins, such as myoglobin, cytochromes, and catalases.

Normally, 250 to 300 mg (425–510 mmol) of bilirubin are produced daily. The iron and globin are removed from haem and reused. Biliverdin is formed from haem and reduced to form bilirubin. The bilirubin produced is unconjugated and water insoluble, due to internal hydrogen bonding, and is transported to the liver attached to albumin. Bilirubin dissociates from albumin and is taken up by hepatic cell membranes and transported to the endoplasmic reticulum by cytoplasmic proteins, where it is conjugated with glucuronic acid and excreted into bile. The microsomal enzyme uridine diphosphoglucuronosyl transferase catalyzes the formation of bilirubin monoglucuronide and then diglucuronide. This conjugated bilirubin is water soluble; it is actively secreted into biliary canaliculi and excreted into the intestine within bile (see **Fig. 38.8**). It is not absorbed from the small intestine because of its large molecular size. In the terminal ileum, bacterial enzymes hydrolyze the molecule, releasing free bilirubin, which is then reduced to urobilinogen; some of this is excreted in the stools as stercobilinogen. The remainder is absorbed by the terminal ileum, passes to the liver via the enterohepatic circulation, and is re-excreted into bile. Urobilinogen bound to albumin enters the circulation and is excreted in urine via the kidneys. When hepatic excretion of conjugated bilirubin is impaired, a small amount is strongly bound to serum albumin and is not excreted by the kidneys; it accounts for persisting hyperbilirubinaemia after cholestasis has resolved.

Hormone and drug inactivation

The liver catabolizes hormones such as insulin, glucagon, oestrogens, growth hormone, glucocorticoids, and parathyroid hormone. It is also the prime target organ for many hormones (e.g., insulin). It is the major site for the metabolism of drugs (see 'Drug metabolism' section) and alcohol (see 'Alcohol and alcoholic liver disease' section). Fat-soluble drugs are converted to water-soluble substances that facilitate their excretion in the bile or urine. Cholecalciferol is converted to 25-hydroxycholecalciferol.

Immunological function

The liver plays a key role in the innate immune response and acts as a 'sieve' for bacterial and other antigens carried to it by the portal vein from the gastrointestinal tract (see 'Gut–liver axis' section).

Hepatocytes are responsible for biosynthesis of 80%–90% of innate immune proteins, including complement components and many secreted pattern-recognition receptors (PRRs) expressed by host cells. Kupffer cells, which account for 80%–90% of the total population of fixed tissue macrophages, are a critical component of the mononuclear phagocytic system and are central to both the hepatic and the systemic responses to pathogens. Following stimulation by endotoxin, the Kupffer cells release interleukin (IL)-6, IL-8, and TNF-α, and, in combination with liver sinusoidal cells, are responsible for the elimination of molecular wastes from the body.

In addition, liver lymphocytes are rich in innate immune cells, including natural killer and T-cell receptor $\gamma\delta$ T cells. Finally, nonparenchymal liver cells express high levels of membrane-bound PRRs, such as Toll-like receptor cells (TLRs). Not only does innate immunity in the liver play a key role in host defence against microbial infection and tumour formation but it also contributes to the pathogenesis of acute and chronic liver diseases, including alcoholic liver disease (see 'Alcohol and alcoholic liver disease' section).

Further reading

Dooley JS, Lok ASF, Garcia-Tsao G et al (eds). *Sherlock's Diseases of the Liver and Biliary System*, 13th edn. Hoboken: Wiley–Blackwell; 2018.
Li Y, Jadhav K, Zhang Y. Bile acid receptors in non-alcoholic fatty liver disease. *Biochem Pharmacol* 2013; 86:1517–1524.
Stofan M, and Grace L. Guo, GL. Bile Acids and FXR: Novel Targets for Liver Diseases. *Front Med (Lausanne)*. 2020; 7:544.

INVESTIGATIONS IN THE PATIENT WITH LIVER DISEASE

Investigative tests can be divided into:

- **Blood tests:**
  - Liver 'function' tests: serum albumin and bilirubin; prothrombin time (PT).
  - Liver biochemistry: aspartate aminotransferase (AST) and alanine aminotransferases (ALT)—an increase reflects hepatocellular damage; serum alkaline phosphatase (ALP) and gamma-glutamyl transpeptidase (γ-GT)—an increase reflects cholestasis; total protein.
  - Viral markers.
  - Additional blood investigations: haematological, biochemical, immunological, markers of liver fibrosis, and genetic analysis.
- ***Imaging techniques***—to define gross anatomy.
- ***Liver biopsy***—for histology.

Blood tests ordered for 'liver function' are usually processed by an automated multichannel analyzer to produce serum levels of bilirubin, aminotransferases, ALP, γ-GT, and total proteins. These routine tests are markers of liver damage but not actual tests of 'function' *per se*. Subsequent investigations are often based on these tests.

Blood tests

Useful blood tests for certain liver diseases are shown in **Box 23.3**.

Liver function tests

Serum albumin

This is a marker of synthetic function and is useful for gauging the severity of chronic liver disease, with a falling serum albumin, which is a poor prognostic sign. In acute liver disease, initial albumin levels may be normal. Interpretation of a low albumin can be difficult when other causes of hypoalbuminaemia (e.g., sepsis) are present.

Bilirubin

Serum bilirubin is normally almost all unconjugated. In liver disease, increased serum bilirubin (conjugated) is usually accompanied by other abnormalities in liver biochemistry. Differentiation between conjugated or unconjugated bilirubin is only necessary in congenital disorders of bilirubin metabolism (see later) or to exclude haemolysis.

Prothrombin time

PT is also a marker of synthetic function. Because of its short half-life, it is a sensitive indicator of both acute and chronic liver disease. Vitamin K deficiency should be excluded as the cause of a prolonged PT by giving an intravenous bolus (10 mg) of vitamin K.

Box 23.3 Useful blood and urine tests for certain liver diseases

Test	Disease
Anti-mitochondrial antibody	Primary biliary cholangitis
Anti-nuclear, smooth muscle (actin), liver/kidney microsomal antibody	Autoimmune hepatitis
Raised serum immunoglobulins:	
IgG	Autoimmune hepatitis
IgG4	Autoimmune hepatitis/cholangiopathy and pancreatitis
IgM	Primary biliary cholangitis
Viral markers	Hepatitis A, B, C, D, E, and others
α-Fetoprotein	Hepatocellular carcinoma
Serum iron, transferrin saturation, serum ferritin	Hereditary haemochromatosis
Serum and urinary copper, serum caeruloplasmin	Wilson's disease
α_1-Antitrypsin	α_1-Antitrypsin deficiency (cirrhosis ± emphysema)
Anti-nuclear cytoplasmic antibodies (ANCA)	Primary sclerosing cholangitis
Markers of liver fibrosis (see 'Markers of liver fibrosis' section)	Non-alcoholic fatty liver disease
	Hepatitis C
Genetic analyses	For example, *HFE* gene (hereditary haemochromatosis), α_1-antitrypsin

Vitamin K deficiency commonly occurs in biliary obstruction, as the low intestinal concentration of bile salts results in poor absorption of vitamin K.

Prothrombin times vary in different laboratories, depending on the thromboplastin used in the assay. The international normalized ratio (INR) was developed to standardize anticoagulation with coumarin derivatives but is very variable in liver disease and causes large differences when included in prognostic scores for cirrhosis across different centres. A rising INR in patients with liver disease that is not corrected by vitamin K is a poor prognostic sign.

Liver biochemistry

Aminotransferases

These enzymes (often referred to as transaminases) are contained in hepatocytes and leak into the blood with liver cell damage. Two enzymes are measured:

- ***Aspartate aminotransferase*** (AST) is primarily a mitochondrial enzyme (80%; 20% in cytoplasm) and is also present in heart, muscle, kidney, and brain. High levels are seen in hepatic necrosis, myocardial infarction, and muscle injury, for example, following very strenuous exercise and congestive cardiac failure.
- ***Alanine aminotransferase*** (ALT) is a cytosol enzyme, more specific to the liver, so that a rise occurs only with liver disease. The ALT:AST ratio is a useful clinical indicator:
- In viral hepatitis, ALT is greater than AST unless cirrhosis is present, in which case AST is greater than ALT.
- In alcoholic liver disease and steatohepatitis, the AST is often greater than the ALT.
- In patients with viral hepatitis, an AST:ALT ratio of more than 1 indicates cirrhosis.
- In patients with liver disease without cirrhosis, in whom AST is greater than ALT, alcohol or obesity is the most likely aetiological agent.

Alkaline phosphatase

Alkaline phosphatase (ALP) is present in hepatic canalicular and sinusoidal membranes, and also in bone, intestine, and placenta. If necessary, its origin can be determined by electrophoretic separation of isoenzymes or bone-specific monoclonal antibodies. In clinical practice, if the γ-GT is also abnormal, the ALP is presumed to come from the liver.

Serum ALP is raised in both intrahepatic and extrahepatic cholestatic disease of any cause, due to increased synthesis. In cholestatic jaundice, levels may be four to six times the normal limit. Raised levels also occur with hepatic infiltrations (e.g., metastases) and in cirrhosis, frequently in the absence of jaundice.

Gamma-glutamyl transpeptidase

This is a microsomal enzyme present in liver and also in many tissues. Its activity can be induced by numerous drugs such as phenytoin, warfarin, and rifampicin, and by alcohol. If the ALP is normal, a raised serum γ-GT can be a useful guide to alcohol intake. However, mild elevations of γ-GT are common, even with minimal alcohol consumption, and it is also raised in fatty liver disease. In the absence of other liver function test abnormalities, a slightly raised γ-GT can safely be ignored. In cholestasis, the γ-GT rises in parallel with the ALP, as it has a similar pathway of excretion.

Total proteins and globulin fraction

The globulin fraction is often raised in autoimmune hepatitis; if it falls, this indicates successful therapy.

Viral markers

Viruses are a major cause of liver disease. Virological studies have a key role in diagnosis; markers are available for most common viruses that cause hepatitis.

Haematological tests

A full blood count may show thrombocytopenia. Thrombocytopenia is a common finding in cirrhosis and is often aggravated by alcohol-induced bone marrow suppression. A low platelet count (below the lower limit of normal—150 × 10^9/L) should be regarded as indicative of cirrhosis, unless another cause can be found. In alcohol excess, red blood cells are often macrocytic.

Biochemical tests

- ***α_1-Antitrypsin*** enzyme deficiency can produce cirrhosis.
- ***α-Fetoprotein*** is normally produced by the fetal liver. Its reappearance in increasing and high concentrations in adults indicates hepatocellular carcinoma (HCC). Increased concentrations in pregnancy in blood and amniotic fluid suggest fetal neural tube defects. Blood levels are also slightly raised with regenerative liver tissue in patients with hepatitis or chronic liver disease and teratomas.
- ***Urinary copper*** is raised, and serum copper and caeruloplasmin are low in Wilson's disease (see 'Wilson's disease [progressive hepatolenticular degeneration]' section).

Immunological tests

Serum immunoglobulins

Increased γ-globulins are thought to result from reduced phagocytosis by sinusoidal and Kupffer cells of the gut-absorbed antigens.

These antigens then stimulate antibody production in the spleen, lymph nodes, and portal tract lymphoid and plasma cell infiltrates. In PBC the predominant raised serum immunoglobulin is IgM, while in autoimmune hepatitis it is IgG. IgG4 is raised in autoimmune pancreatitis/cholangitis (see Chapter 24).

Serum autoantibodies

- ***Anti-mitochondrial antibody (AMA)*** in serum is found in over 95% of patients with PBC (see 'Primary biliary cholangitis' section). Several different AMA subtypes are described, depending on their antigen specificity, and are also found in autoimmune hepatitis and other autoimmune diseases. AMA is demonstrated by an immunofluorescent technique and is neither organ- nor species-specific. The M2 subtype is specific for PBC.
- ***Nucleic, smooth muscle (actin), liver/kidney microsomal antibodies*** can be found in serum, often in high titre, in patients with autoimmune hepatitis. These serum antibodies are also present in other autoimmune conditions and other liver diseases.
- ***Anti-nuclear cytoplasmic antibodies (ANCA)*** can be found in the serum of patients with PSC (see 'Primary sclerosing cholangitis' section).

Markers of liver fibrosis

Fibrosis plays a key role in the outcome of many chronic liver diseases, and accurate assessment of fibrosis is critical for the appropriate management of many liver disorders. A variety of systems have been developed to assess the extent of liver fibrosis and these range from algorithms of varying degrees of complexity that use standard haematological and biochemical tests, to novel biomarkers. Simple algorithms include the ***APRI*** (*a*spartate aminotransferase to *p*latelet *r*atio *i*ndex) score or FIB-4 index (that uses age, ALT, AST, and platelet count), while more complex commercial tests include the ***fibrotest algorithm***. Novel biomarker–based algorithms include those based on measurements of hyaluronic acid, procollagen III amino terminal peptide, and tissue inhibitor of metalloproteinase I (TIMP-1), all combined in the ***enhanced liver fibrosis (ELF) test***. Many of these markers were developed for chronic hepatitis C initially but are often successfully applied to other liver disorders, especially non-alcoholic fatty liver disease. The current assays have a high sensitivity/specificity for the detection or absence of cirrhosis but are less effective at detecting intermediate levels of fibrosis. Combining mechanical non-invasive fibrosis tests, such as transient elastography (see below), with fibrosis markers allows many patients to avoid a liver biopsy to assess fibrosis.

Genetic analysis

These tests are performed routinely for haemochromatosis (*HFE* gene) and for α_1-antitrypsin deficiency. Markers are also available for the most frequent abnormal genes in Wilson's disease (see 'Wilson's disease [progressive hepatolenticular degeneration]' section).

Imaging techniques

Ultrasound examination

This non-invasive, safe, and relatively cheap technique involves analysis of the reflected ultrasound beam detected by a probe moved across the abdomen. The normal liver appears as a relatively homogeneous structure. The gall bladder, common bile duct, pancreas, portal vein, and other structures in the abdomen can be visualized. ***Abdominal ultrasound*** is useful in:

- Detection of extrahepatic obstruction (the bile duct is usually dilated, particularly in advanced disease). Note that opiates may cause biliary dilation without obstruction, and so scans in injecting drug users often show extrahepatic biliary dilation.
- Assessment of a jaundiced patient (to exclude obstruction) (see 'Differential diagnosis of jaundice' section).
- Assessment of hepatomegaly/splenomegaly.
- Detection of gallstones (see **Fig. 24.3**).
- Assessment of focal liver disease to detect lesions larger than 1 cm.
- Assessment of portal and hepatic vein patency.
- Assessment of the hepatic parenchyma—diffuse fatty infiltration often leads to a 'bright' appearance on ultrasound but experience is required to distinguish this from normal variation.
- Identification of cirrhosis—in advanced cirrhosis, the liver edge is irregular and the spleen is often enlarged, however, a normal ultrasound does not exclude cirrhosis.
- Assessment of lymph node enlargement.

Other abdominal masses can be delineated and biopsies obtained under ultrasonic guidance.

Colour Doppler ultrasound

This will demonstrate vascularity within a lesion, and the direction of portal and hepatic vein blood flow.

Hepatic stiffness (transient elastography)

Using an ultrasound transducer, a vibration of low frequency and amplitude is passed through the liver, the velocity of which correlates with hepatic stiffness. Stiffness (measured in kilopascal, kPa) increases with worsening liver fibrosis with sensitivity and specificity of 80% to 95%, compared to liver biopsy. Elastography can reliably exclude cirrhosis and, in patients with cirrhosis, increasing liver stiffness is associated with a higher risk of complications. Elastography is less effective for determining lesser degrees of fibrosis. It cannot be employed in the presence of ascites and morbid obesity, and is affected by inflammatory tissue and congestion, for example, in the presence of right-sided cardiac failure.

Acoustic radiation force impulse is also incorporated into standard B mode ultrasonography and has similar physical principles to transient elastography.

Computed tomography examination

Computed tomography (CT), during or immediately after intravenous contrast, shows both arterial and portal venous phases of enhancement, enabling more precise characterization of a lesion and its vascular supply (**Fig. 23.7**). Retrospective analysis of data allows multiple overlapping slices to be obtained with no increase in the radiation dose, providing excellent visualization of the size, shape, and density of the liver, pancreas, spleen, lymph nodes and lesions in the porta hepatis. Multiplanar and three-dimensional reconstruction in the arterial phase can create a CT angiogram, often making formal invasive angiography unnecessary. CT also provides guidance for biopsy. It has advantages over ultrasound in detecting calcification and is useful in obese subjects, although ultrasound is usually the first imaging modality to be used to investigate liver disease.

Magnetic resonance imaging

Magnetic resonance imaging (MRI) produces cross-sectional images in any plane within the body and does not involve radiation. It is the most sensitive investigation for focal liver disease but can also assess fibrosis. Diffuse liver disease alters the T1 and T2 characteristics. Other fat-suppression modes, such as ***short T1 inversion recovery (STIR)***, allow good differentiation between haemangiomas and other lesions. Contrast agents such as intravenous gadolinium, which allow further characterization of lesions, are suitable for those with iodine allergy, and provide angiography and

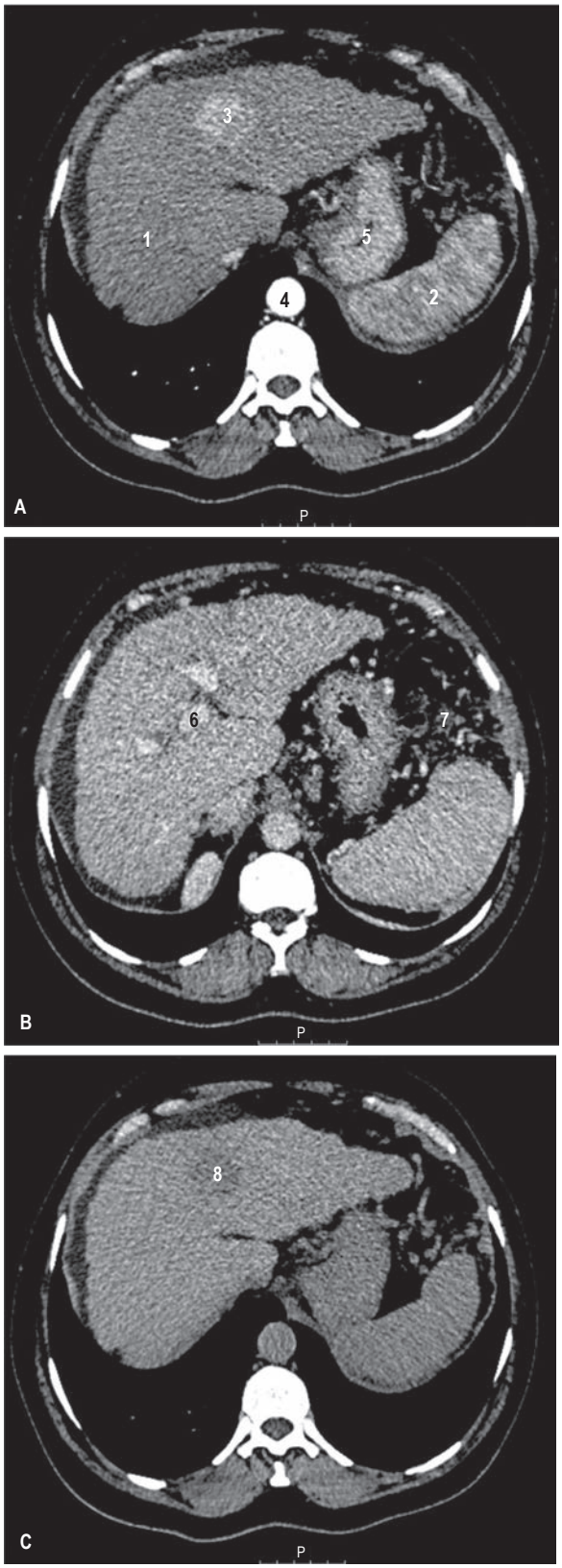


Fig. 23.7 Triple-phase liver CT showing a hepatocellular carcinoma. The liver has an irregular contour consistent with cirrhosis and there are splenic and gastric varices. The arterial phase **(A)** demonstrates an enhancing lesion, which is poorly seen on the portal venous phase **(B)** but demonstrates washout on the delayed phase imaging **(C)** consistent with hepatocellular carcinoma. (1) Liver; (2) spleen; (3) lesion; (4) aorta; (5) stomach; (6) portal vein; (7) varices; (8) washout of lesion.

venography of the splanchnic circulation. Use of these agents has superseded direct arteriography.

Magnetic resonance cholangiopancreatography

Magnetic resonance cholangiopancreatography (MRCP) involves the manipulation of data acquired by MRI. A heavily T2-weighted sequence enhances visualization of the 'water-filled' bile ducts and pancreatic ducts to produce high-quality images of ductal anatomy. This non-invasive technique is replacing diagnostic (but not therapeutic) endoscopic retrograde cholangiopancreatography (see Chapter 24) and is usually the next test to be applied if a biliary abnormality is present on ultrasound examination.

Endoscopy

Upper gastrointestinal endoscopy is used for diagnosis and treatment (band ligation) of varices, detection of portal hypertensive gastropathy, and detection of associated lesions such as peptic ulcers. Colonoscopy may show portal hypertensive colopathy. Capsule endoscopy can identify small intestinal varices.

Endoscopic retrograde cholangiopancreatography

Endoscopic retrograde cholangiopancreatography (ERCP) outlines the biliary and pancreatic ducts (see Chapter 24).

Angiography

This is performed by selective catheterization of the coeliac axis and hepatic artery. It outlines the hepatic vasculature and the abnormal vasculature of hepatic tumours, but CT and MRI have replaced diagnostic angiography. It may be used to identify lower gastrointestinal bleeding, including variceal, and treat this by embolizing the culprit vessel using small metal coils.

Radionuclide imaging and hepatic venous cannulation have been superseded by cross-sectional imaging and transjugular intrahepatic measurement of portal pressure.

Liver biopsy

Histological examination of the liver is valuable in the differential diagnosis of diffuse or localized parenchymal disease and its severity. Liver biopsy is routinely performed as a day-case procedure. The indications and contraindications are shown in **Box 23.4**. The mortality rate is less than 0.02% when the technique is performed by experienced operators.

Liver biopsy guided by ultrasound or CT is performed under a local anaesthetic via a percutaneous approach in the right intercostal space. A transjugular approach is used when liver histology is essential for management but coagulation abnormalities or ascites prevent the percutaneous approach.

Most ***complications*** of liver biopsy occur within 24 hours (usually in the first 2 hours). They are often minor and include abdominal or shoulder pain that settles with analgesics. Minor intraperitoneal bleeding can occur, but this settles spontaneously. Rare complications include major intraperitoneal bleeding, haemothorax and pleurisy, biliary peritonitis, haemobilia, and transient septicaemia. Haemobilia produces biliary colic, jaundice, and melaena within 3 days of the biopsy.

Further reading

Azizaddini S, Mani N. Liver Imaging. [Updated 2023 Feb 20]. In: *StatPearls* [Internet]. Treasure Island (FL): StatPearls Publishing; 2023 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK557460.

Tripodi A, Mannucci PM. The coagulopathy of chronic liver disease. *N Engl J Med* 2011; 365:147–156.

Box 23.4 Indications for and contraindications to liver biopsy

Diagnosis
- Multiple parenchymal liver diseases
- Abnormal liver tests of unknown aetiology
- Fever of unknown origin
- Focal or diffuse abnormalities on imaging studies (rarely for hepatocellular carcinoma)
- Unexplained hepatomegaly
- Drug-related liver disease
- Post liver transplant

Prognosis
- Staging of known parenchymal liver disease, e.g., hepatitis B/C, non-alcoholic fatty liver disease, primary biliary cholangitis, autoimmune hepatitis, primary sclerosing cholangitis
- Haemochromatosis and alcohol-induced liver disease

Management
- Development of treatment plans based on histological analysis

Usual contraindications to percutaneous needle biopsy
- Uncooperative patient
- Prolonged INR (>1.5)
- Platelets $<60 \times 10^9$/L
- Ascites
- Extrahepatic cholestasis

JAUNDICE

Jaundice (icterus) is detectable clinically when the serum bilirubin is over 50 μmol/L (3 mg/dL). It may be divided into:
- ***haemolytic jaundice***—increased bilirubin load for the liver cells
- ***congenital hyperbilirubinaemias***—defects in conjugation
- ***hepatic jaundice***—including hepatocellular (parenchymal) liver disease and biliary duct obstruction.

Haemolytic jaundice

The increased breakdown of red cells (see **Fig. 38.8**) leads to an increase in production of bilirubin. The resulting jaundice is usually mild (serum bilirubin of 68–102 μmol/L, or 4–6 mg/dL), as normal liver function can easily manage the increased bilirubin. Unconjugated bilirubin is not water-soluble and therefore does not pass into urine: hence 'acholuric jaundice'. Urinary urobilinogen is increased.

The causes are those of haemolytic anaemia (see Chapter 38), and clinical features of anaemia, jaundice, splenomegaly, gallstones, and leg ulcers may be seen.

Investigations show anaemia with markers of haemolysis (e.g., reticulocytosis, elevated lactate dehydrogenase) and elevated unconjugated bilirubin, but normal serum ALP, transferases, and albumin. Serum haptoglobulins are low if haemolysis is intravascular.

Congenital hyperbilirubinaemias (non-haemolytic)

Unconjugated types

Gilbert's syndrome

This is the most common familial unconjugated hyperbilirubinaemia and affects 2% to 7% of the population. It is asymptomatic and usually detected incidentally with a raised bilirubin (17–102 μmol/L, or 1–6 mg/dL). All other liver biochemistry is normal and there are no signs of liver disease. There is a family history of jaundice in 5% to 15% of patients. Most patients have reduced levels of UDP-glucuronosyl transferase (UGT-1) activity; this is the enzyme that conjugates bilirubin with glucuronic acid.

Mutations occur in the gene (*UGT1A1* promoter region) encoding this enzyme, with an expanded nucleotide repeat consisting of two extra bases in the upstream 5′ promoter element. This abnormality appears to be required for the syndrome to occur but is not in itself sufficient for clinical manifestation (phenotypic expression).

Establishing the diagnosis is necessary to provide reassurance and prevent unnecessary investigations. The raised unconjugated bilirubin is diagnostic and rises on fasting and during mild illness. The reticulocyte count is normal, excluding haemolysis, and no treatment is necessary.

Crigler–Najjar syndrome

This is very rare. Only patients with type II (autosomal dominant) disease, with a decrease in, rather than absence (type I—autosomal recessive) of, UGT survive into adult life. Liver histology is normal. Transplantation is the only effective treatment but advances in gene therapy may soon provide alternative options (see Chapter 5).

Conjugated types

Dubin–Johnson and Rotor's syndromes

Dubin–Johnson and Rotor's syndromes (autosomal recessive) are due to defects in hepatic bilirubin handling. The prognosis is good in both. In the Dubin–Johnson syndrome the liver is black owing to melanin deposition.

Benign recurrent intrahepatic cholestasis

This is rare and presents in early adulthood. Recurrent attacks of acute cholestasis occur without progression to chronic liver disease. Jaundice, severe pruritus, steatorrhoea, and weight loss may develop. Serum γ-GT is normal. Benign recurrent intrahepatic cholestasis may be associated with intrahepatic cholestasis of pregnancy (see Chapter 48). Genetic testing is available and there are two forms: BRIC1 is caused by mutations in the *ATP8B1* gene and BRIC2 by mutations in the *ABCB11* gene. Both are autosomal recessive. Treatment includes medications to manage symptoms and occasionally specialized therapies, such as nasobiliary drainage, to shorten episodes.

Progressive familial intrahepatic cholestasis syndromes

Progressive familial intrahepatic cholestasis (PFIC) syndromes are a heterogeneous group of autosomal recessive conditions defined by defective secretion of bile acids due to mutations in various transporter genes (see **Figs. 23.5** and **24.2**).
- In ***type 1*** (PFIC1), with cholestasis in the first weeks of life, the γ-GT is normal and the genetic defect lies in the *ATP8B1* gene.
- In ***type 2*** (PFIC2), there is frequently a non-specific giant cell hepatitis that progresses to cholestasis; again, γ-GT is normal and the gene involved is *ABCB11.*
- In ***type 3*** (PFIC3), deficient canalicular phosphatidylcholine transport and accumulation of toxic bile acids cause liver damage, which can lead to cirrhosis and is due to mutations in the *ABCB4* gene.

Liver transplantation is currently the only cure for these syndromes.

Hepatic jaundice (acquired)

This condition can be divided into jaundice due to parenchymal liver damage or cholestasis, which may be intra- or extrahepatic. The causes are shown in **Fig. 23.8**.

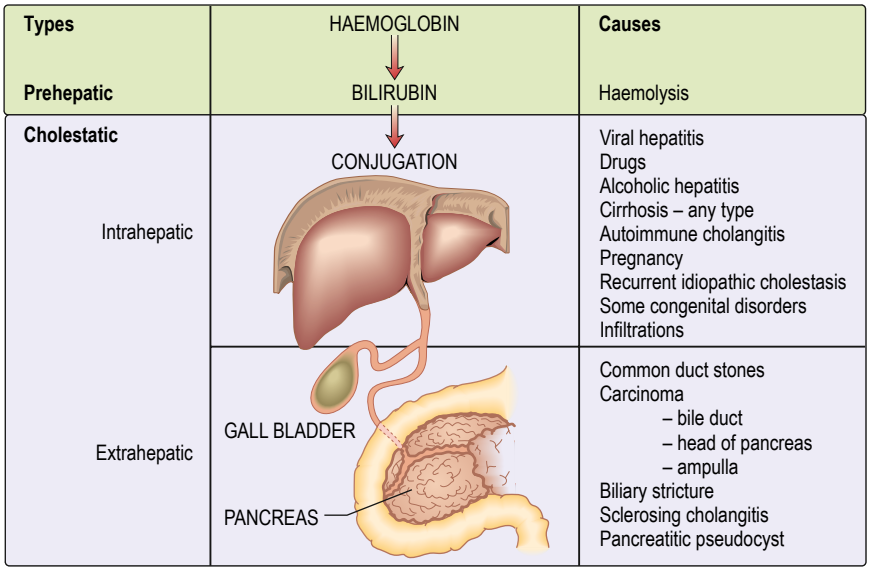


Fig. 23.8 **Causes of jaundice.**

- ***Hepatic jaundice*** is due to damage to hepatocytes leading to a failure in both bilirubin uptake and excretion.
- ***Extrahepatic cholestasis*** is due to large duct obstruction of bile flow at any point in the biliary tract distal to the bile canaliculi.
- ***Intrahepatic cholestasis*** occurs because of failure of bile secretion, which may be caused by intrinsic defects in bile secretion or inflammation in the intrahepatic ducts.

Clinically in cholestatic jaundice the stools are pale with dark urine, and the serum bilirubin is conjugated and these features are much less marked in hepatic jaundice. Intrahepatic and extrahepatic cholestatic jaundice must be differentiated, as their clinical management is entirely different.

Differential diagnosis of jaundice

Hepatic jaundice may occur in previously healthy people who have an acute hepatic illness or may develop in patients with cirrhosis who have 'decompensated'. The management of decompensated cirrhosis is described in the 'Cirrhosis' section but all patients with jaundice should be questioned closely about risk factors for chronic liver disease to determine whether this is a true acute presentation rather than an 'acute on chronic' illness. The history often gives a clue to the diagnosis. Certain causes of jaundice are more likely in particular categories of people. For example, a young person is more likely to have infectious hepatitis, so questions should be asked about drug and alcohol misuse, and sexual behaviour. An elderly person with gross weight loss is more likely to have a carcinoma. All patients may complain of malaise. Abdominal pain occurs in patients with biliary obstruction by gallstones, and sometimes with an enlarged liver there is pain resulting from distension of the capsule.

Questions should be appropriate to the particular situation, and the following aspects of the history should be covered:

- ***Country of origin.*** The incidence of hepatitis B virus (HBV) and hepatitis C virus (HCV) infection is increased in many parts of the world (see 'Hepatitis B' and 'Hepatitis C' sections).
- ***Duration of illness.*** A history of jaundice with prolonged weight loss in an older patient suggests malignancy. A short history, particularly with a prodromal illness of malaise, suggests an infectious hepatitis.
- ***Recent outbreak of jaundice.*** An outbreak in the community suggests hepatitis A virus (HAV).
- ***Intravenous drug use, or recent injections or tattoos.*** These all increase the chance of HBV and HCV infection.
- ***Men having sex with men.*** There is an increased chance of HBV and HCV infection.
- ***Female sex workers.*** There is an increased chance of HBV infection.
- ***Medical treatment in the developing world.*** There is an increased risk of HBV and HCV due to poorly sterilized equipment or administration of unscreened blood or blood products.
- ***Alcohol consumption.*** A history of drinking habits should be taken, although many patients often understate their consumption.
- ***Drugs (particularly those taken in the previous 2–3 months).*** Many drugs, including over-the-counter and herbal preparations, cause jaundice (see 'Drugs and the liver' section).
- ***Travel.*** Certain areas have a high risk of HAV infection, as well as hepatitis E virus (HEV), but HAV is common in the UK and HEV is common in travellers to the Indian subcontinent.
- ***Family history.*** Patients with, for example, Gilbert's disease may have family members who experience recurrent jaundice.
- ***Recent surgery.*** Surgery on the biliary tract or for carcinoma is relevant.
- ***Environment.*** People engaged in recreational activities in rural areas, as well as farm and sewage workers, are at risk for leptospirosis, hepatitis E, and exposure to chemicals.
- ***Fevers or rigors.*** These are suggestive of cholangitis or possibly a liver abscess.

Clinical features

The signs of acute and chronic liver disease should be looked for (see figure in 'Clinical skills' section). Certain additional signs may be helpful:

- ***Hepatomegaly.*** A smooth, tender liver is seen in hepatitis and in extrahepatic obstruction, but a knobbly, irregular liver suggests metastases or cirrhosis. Causes of hepatomegaly are shown in **Box 23.5**.
- ***Splenomegaly.*** This indicates portal hypertension when signs of chronic liver disease are present. The spleen can also be 'tipped' occasionally in viral hepatitis. In alcoholic cirrhosis, in particular, the spleen may not be grossly enlarged and may not be palpable.

Box 23.5 Causes of hepatomegaly

Apparent
- Low-lying diaphragm
- Riedel's lobe
- Cirrhosis (early)

Inflammation
- Hepatitis
- Schistosomiasis
- Abscesses (pyogenic or amoebic)

Cysts
- Hydatid
- Polycystic

Metabolic
- Fatty liver
- Amyloid deposition
- Glycogen storage disease

Haematological
- Leukaemias
- Lymphoma
- Myeloproliferative disorders
- Thalassaemia

Tumours
- Primary and secondary carcinoma

Venous congestion
- Heart failure
- Constrictive pericarditis
- Hepatic vein occlusion

Biliary obstruction
- Particularly extrahepatic

- ***Ascites.*** This is found in cirrhosis but can also be due to other causes (see Box 23.20).
- ***Palpable gall bladder.*** This occurs with a carcinoma of the pancreas obstructing the bile duct.
- ***Generalized lymphadenopathy.*** This suggests a lymphoma.
- ***Cold sores*** are often seen with a herpes simplex virus hepatitis.

Investigations

Jaundice is not itself a diagnosis and the cause should always be sought. The cause is either excessive haemolysis, parenchymal liver damage (hepatic jaundice), or biliary obstruction—either within or outside the liver. Liver biochemistry confirms the jaundice and may help in the diagnosis, but the most useful test is an ultrasound examination which identifies extrahepatic obstruction.

Ultrasound will demonstrate:
- the size of the bile ducts, which are dilated in extrahepatic biliary obstruction (Fig. 23.9)
- the level of the obstruction
- the cause of the obstruction in virtually all patients with tumours and in 75% of patients with gallstones.

The pathological diagnosis of any mass lesion can be made by fine-needle aspiration cytology (sensitivity approximately 60%) or by needle biopsy (sensitivity approximately 90%).

A flow diagram for the general investigation of the jaundiced patient is shown in Fig. 23.10.

Liver biochemistry

In hepatitis, serum AST and ALT tend to be high early in the disease, with only a small rise in serum ALP. Conversely, in extrahepatic obstruction, the ALP is high, with a smaller rise in aminotransferases. However, these findings alone cannot be relied on to make a diagnosis in an individual case and in obstructive jaundice the ALP changes may follow the ALT changes. The PT is often prolonged in longstanding liver disease and the serum albumin is low.

Haematological tests

In haemolytic jaundice the bilirubin is raised and the other liver biochemistry is normal. A raised white cell count may indicate infection (e.g., cholangitis). A leucopenia often occurs in viral hepatitis, while abnormal mononuclear cells suggest infectious mononucleosis (Epstein–Barr virus) and a Monospot test should be performed.

Other blood tests

These include tests to confirm the presence of specific causes of acute or chronic liver disease, such as AMA for PBC. HIV status should be established.

HEPATITIS

- ***Acute parenchymal liver damage*** can be caused by many agents (Fig. 23.11).
- ***Chronic hepatitis*** is defined as any hepatitis lasting for 6 months or longer and is classified according to the aetiology (Box 23.6). Chronic viral hepatitis is the principal cause of chronic liver disease, cirrhosis, and HCC worldwide.

Acute hepatitis

Pathology

Although some histological features are suggestive of the aetiological factor, most changes are non-specific. Hepatocytes show degenerative changes (swelling, cytoplasmic granularity, vacuolation) and undergo necrosis (becoming shrunken, containing eosinophilic Councilman bodies). The distribution of these changes may vary with aetiology but necrosis is usually maximal in zone 3. The extent of the damage is very variable between individuals, even when they are affected by the same agent; at one end of the spectrum, single and small groups of hepatocytes die (spotty or focal necrosis), while at the other end, there is multiacinar necrosis involving a substantial part of the liver (massive hepatic necrosis) and resulting in acute hepatic failure. Between these extremes, there is limited confluent necrosis with collapse of the reticulin framework, leading to linking (bridging) between the central veins, between the central veins and portal tracts, and between the portal tracts. The extent of the inflammatory infiltrate is also variable but portal tracts and lobules are infiltrated mainly by lymphocytes. Other variable features include cholestasis in zone 3 and fatty change, the latter being prominent in hepatitis due to alcohol or certain drugs.

Chronic hepatitis

Pathology

The pathological features are often diagnostic. Chronic inflammatory cell infiltrates, comprising lymphocytes, plasma cells, and sometimes lymphoid follicles, are usually present in the portal tracts. The amount of inflammation varies from mild to severe. In addition, there may be:
- loss of definition of the portal/periportal limiting plate—interface hepatitis (damage is due to apoptosis rather than necrosis)
- lobular change, focal lytic necrosis, apoptosis, and focal inflammation
- confluent necrosis
- fibrosis, which may be mild, bridging (across portal tracts) or severe cirrhosis.

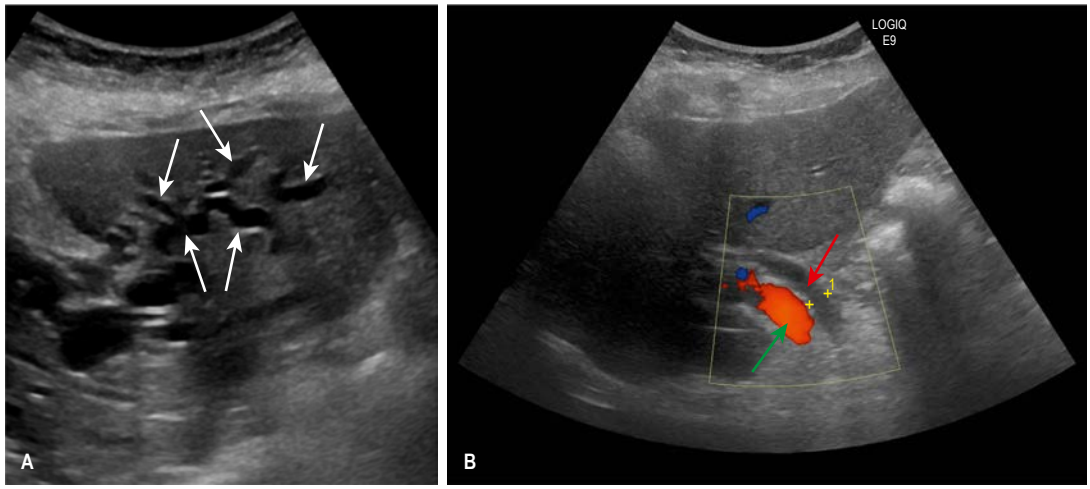


Fig. 23.9 Liver ultrasound. (A) Markedly dilated intrahepatic ducts *(arrowed)*. **(B)** A dilated common bile duct *(red arrow)*. A duct with a diameter of 6 mm or less is normal; however, it can be of larger calibre and still not represent pathology in some circumstances, such as in elderly patients and after cholecystectomy. Several factors must therefore be considered when deciding what is normal for a patient. Normal colour flow is demonstrated in the portal vein *(green arrow)*.

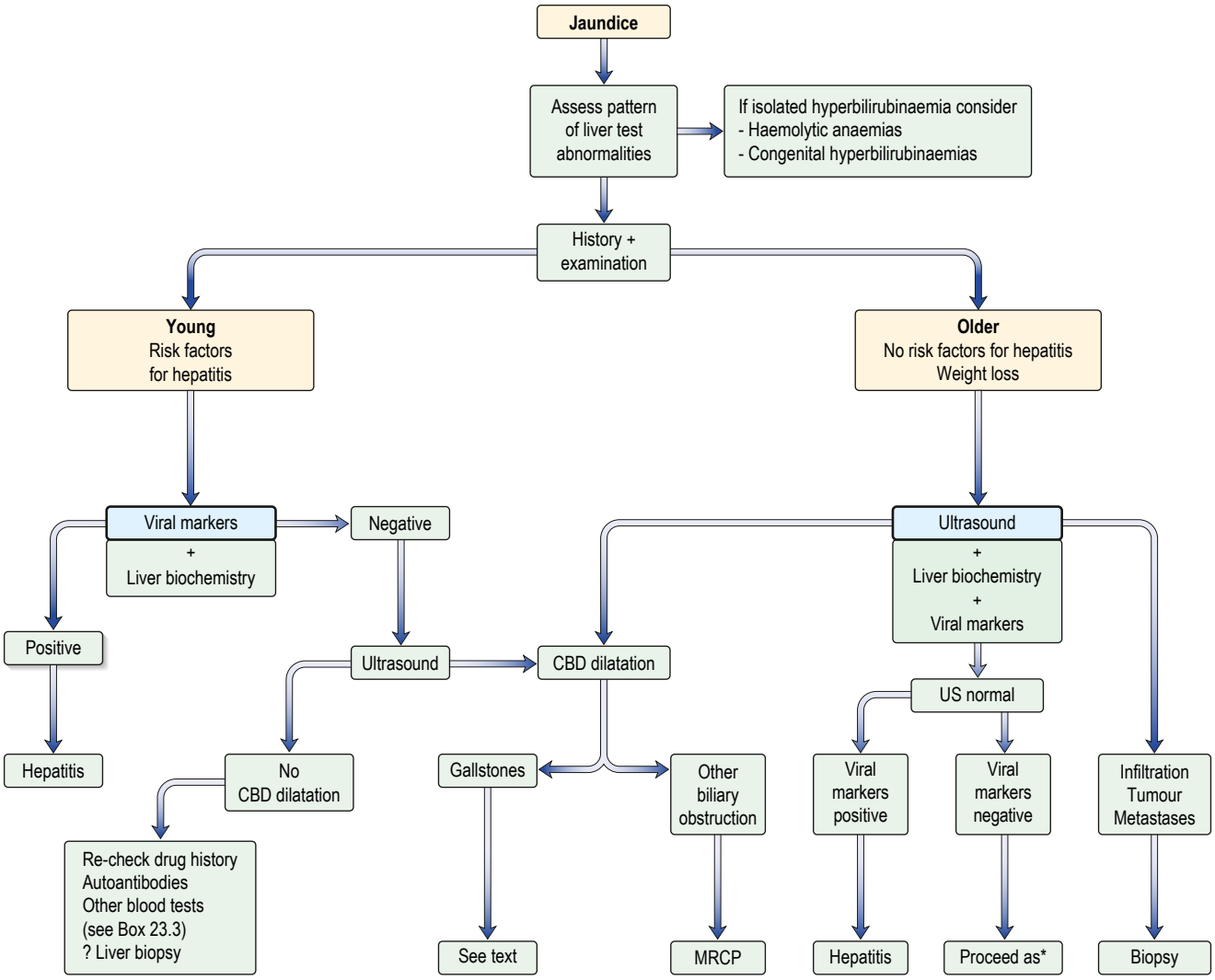


Fig. 23.10 Approach to the patient with jaundice. *CBD,* Common bile duct; *MRCP,* magnetic resonance cholangiopancreatography; *US,* ultrasound. *Proceed as in bottom left box (Re-check drug history…)

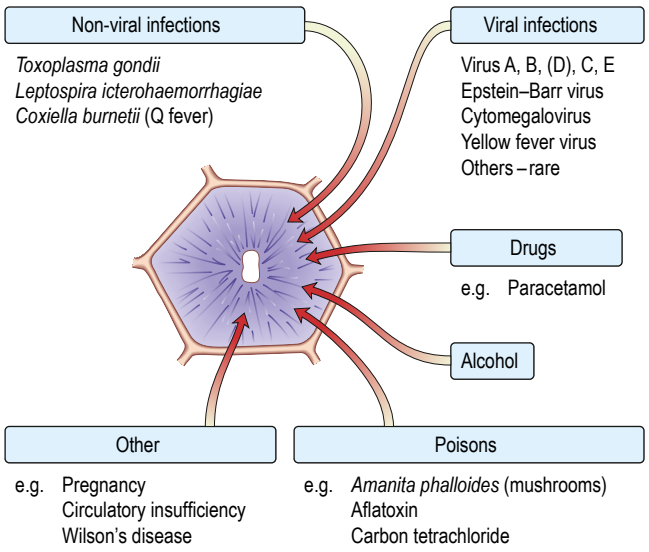


Fig. 23.11 Some causes of acute parenchymal damage.

Box 23.6 Causes of chronic hepatitis

- Metabolic:
  - Non-alcoholic fatty liver disease
- Alcohol-induced
- Viral:
  - Hepatitis B ± hepatitis D
  - Hepatitis C
  - Hepatitis E (immunosuppressed)
- Drugs:
  - (e.g., methyldopa, isoniazid, ketoconazole, nitrofurantoin)
- Autoimmune
- Hereditary:
  - Wilson's disease, haemochromatosis.

Unusual causes include infections (syphilis, tuberculosis, various tropical infections), infiltrative diseases, including amyloidosis and lymphoma, and ingestion of toxins.

The overall severity of the hepatitis is judged by the degree of hepatitis and inflammation (grading), and the severity of fibrosis or cirrhosis (staging). In chronic viral hepatitis there are various scoring systems. For example, the ***Knodell scoring system*** (histological activity index) uses the sum of four factors (periportal or bridging necrosis, intralobular degeneration and focal necrosis, portal inflammation, and fibrosis). The ***Ishak score*** stages fibrosis from 0 (none) to 6 (cirrhosis). The ***METAVIR system*** has four stages. Scoring systems are used for drug trials and for assessing progression of disease, but are not quantitative measures of fibrosis and different systems may be used for different diseases (for example, the ***Brunt scoring system*** is usually applied to NAFLD, and the METAVIR system is reserved for hepatitis C).

Viral hepatitis

The differing features of the common forms of viral hepatitis are summarized in **Box 23.7**. Two different patterns are recognized: acute hepatitis with rapid onset of infection and, usually, rapid resolution; and chronic viral hepatitis, which is asymptomatic and often detected on routine blood tests or during screening for infection. Hepatitis A always and hepatitis E usually cause acute infections, while hepatitis B, C, and D may cause acute or chronic disease.

Hepatitis A

Epidemiology

Hepatitis A is common worldwide, often occurring in epidemics. The disease is usually seen in the autumn and affects children and young adults. Spread of infection is mainly by the faeco-oral route and arises from ingestion of contaminated food or water (e.g., shellfish). Overcrowding and poor sanitation facilitate spread, and outbreaks of infection have occurred in men who have sex with men. There is no carrier state. In the UK it is a notifiable disease.

Hepatitis A virus

Hepatitis A virus (HAV) is a picornavirus (**Fig. 23.12**). It has a single serotype, as only one epitope is immunodominant. It replicates in the liver, is excreted in bile, and then excreted in the faeces for about 2 weeks before the onset of clinical illness and for up to 7 days after. The disease is maximally infectious just before the onset of jaundice. HAV particles can be demonstrated in the faeces by electron microscopy.

Clinical features

The viraemia causes the patient to feel unwell, with non-specific symptoms that include nausea and anorexia. Many recover at this stage and remain anicteric. An anicteric infection is common in children and confers lifetime immunity. In the developing world, improvements in hygiene have reduced early infection and, paradoxically, led to an increase in symptomatic infection in exposed adults.

After 1 or 2 weeks, some patients become jaundiced and symptoms often improve. Persistence of nausea, vomiting, or any mental confusion warrants assessment in hospital. As the jaundice deepens, the urine becomes dark and the stools are pale. The liver is moderately enlarged and the spleen may be palpable. Occasionally, tender lymphadenopathy is seen, with a transient rash in some cases. Thereafter, the jaundice lessens and, in the majority of cases, the illness is over within 3 to 6 weeks. Extrahepatic complications are rare but include arthritis, vasculitis, myocarditis, and acute kidney injury. A biphasic illness occasionally occurs, with the return of jaundice and, classically, a more severe 'second phase', which is cholestatic (cholestatic viral hepatitis) with an increase in the ALP rather than the aminotransferases; thus, it runs a prolonged course of 7 to 20 weeks. Rarely, the disease may be very severe with acute hepatitis, liver coma, and death; this is more common in the elderly. The typical sequence of events after HAV exposure is shown in **Fig. 23.13**.

Investigations

Liver biochemistry

- ***Prodromal stage:*** The serum bilirubin is usually normal. A raised serum AST or ALT, which can sometimes be over many thousands, precedes the jaundice.
- ***Icteric stage:*** The serum bilirubin reflects the level of jaundice. Serum AST reaches a maximum of 1 to 2 days after the appearance of jaundice, and may rise above 500 IU/L. Serum ALP is usually less than 300 IU/L.

After the jaundice has subsided, the aminotransferases may remain elevated for some weeks and occasionally for up to 6 months.

Haematological tests

There is leucopenia with a relative lymphocytosis. Very rarely, there is a Coombs'-positive haemolytic anaemia or an associated aplastic anaemia. The PT is prolonged in severe cases. The erythrocyte sedimentation rate (ESR) is raised.

Box 23.7 Some features of viral hepatitis

	A	B	C	D	E
Virus	RNA	DNA	RNA	RNA	RNA
	27 nm	42 nm	Approx. 50 nm	36 nm (with HBsAg coat)	27–35 nm
	Picornaviridae	*Hepadnaviridae*	*Flaviviridae*	*Deltaviridae*	*Hepeviridae*
Spread					
Faeco-oral	Yes	No	No	No	Yes
Blood/blood products	Rare	Yes	Yes	Yes	Rare
Vertical	No	Yes	Rare	Occasional	No
Saliva	Yes	Yes	Yes	? No	?
Sexual	Rare	Yes	Yes (rare)	Rare	No
Incubation	Short (2–3 weeks)	Long (1–5 months)	Long	Intermediate	Short
Age	Young	Any	Any	Any	Any
Carrier state	No	Yes	Yes	?	No
Chronic liver disease	No	Yes	Yes	Yes	No[a]
Liver cancer	No	Yes	Yes	Yes	No
Mortality (acute)	<0.5%	<1%	<1%	<1%	1%–2% (pregnant women 10%–20%)
Immunization					
Passive	Normal immunoglobulin serum i.m. (0.04–0.06 mL/kg)	Hepatitis B immunoglobulin (HBIg)	No	No	No
Active	Vaccine	Vaccine	No	HBV vaccine to prevent co-infection	Vaccine

[a]Chronic hepatitis in immunosuppressed patients.
HBsAg, Hepatitis B surface antigen.

Viral markers: antibodies to HAV

IgG antibodies are common in the general population but an anti-HAV IgM means an acute infection. In areas of high prevalence, most children have antibodies by the age of 3 years following asymptomatic infection.

Other tests

Further tests are not necessary in the presence of an IgM antibody but liver biochemistry must be followed to establish a return to normal levels. Severity should be assessed by changes in PT and not by the level of AST/ALT.

Differential diagnosis

Differentiation must be made from all other causes of jaundice, but in particular from other types of viral and drug-induced hepatitis.

Prognosis

The prognosis is excellent, with most patients making a complete recovery. The mortality in young adults is 0.1% but increases with age. Death is due to acute hepatic failure.

There is no reason to stop alcohol consumption, other than for the few weeks when the patient is ill. Patients may complain of debility for several months following resolution of the symptoms and biochemical parameters. This is known as post-hepatitis syndrome and treatment is by reassurance. HAV hepatitis never progresses to chronic liver disease.

Management

There is no specific treatment, and rest and dietary measures are unhelpful. Corticosteroids have no benefit. Admission to hospital is not usually necessary.

Prevention

Control of hepatitis depends on good hygiene. The virus is resistant to chlorination but is killed by boiling water for 10 minutes.

Active immunization

A formaldehyde-inactivated HAV vaccine is given to people travelling frequently to endemic areas, patients with chronic liver disease, those with haemophilia, and workers in frequent contact with hepatitis cases (e.g., in residential institutions for patients with learning disabilities). Community outbreaks can be interrupted by vaccination. A single dose produces antibodies that persist for at least 1 year, with immunity lasting beyond 10 years. This obviates the need for a booster injection in healthy individuals.

Passive immunization

Normal human immunoglobulin is used in line with the manufacturers' recommendations if exposure to HAV is for less than 2 weeks. HAV vaccine should also be given.

Hepatitis B

Epidemiology

The hepatitis B virus (HBV) is present worldwide and there are an estimated 220 million carriers. The UK and the USA have a low

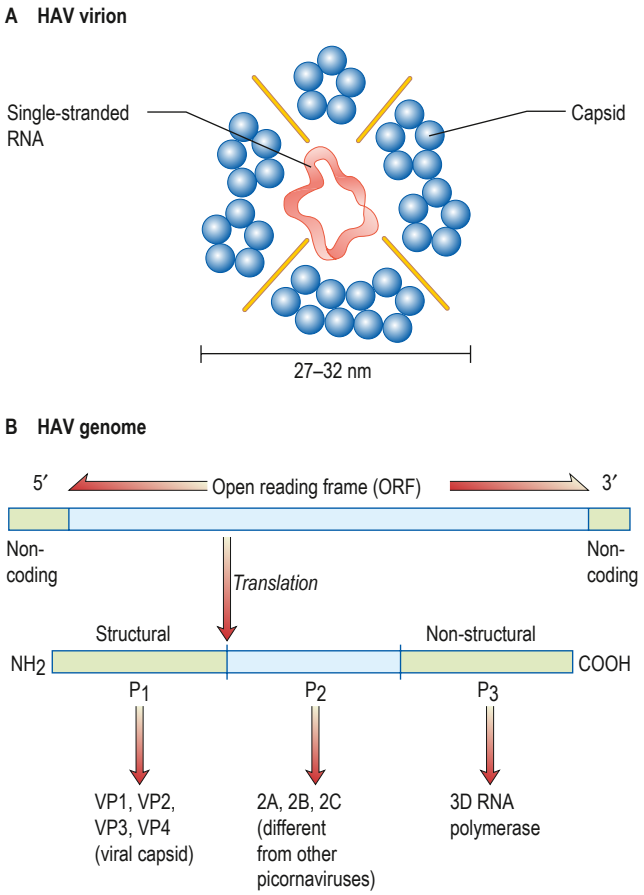


Fig. 23.12 Hepatitis A. (A) The hepatitis A (HAV) virion consists of four polypeptides (VP1–VP4), which form a tight protein shell, or capsid, containing the RNA. The major antigenic component is associated with VP1. **(B)** Arrangement of the HAV genome.

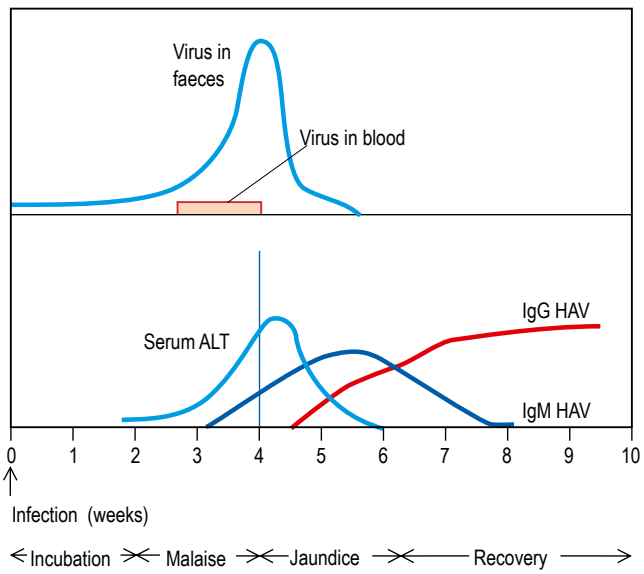


Fig. 23.13 Hepatitis A virus (HAV): sequence of events after exposure. *ALT,* Alanine aminotransferase; *Ig,* immunoglobulin.

carrier rate (0.5%–2%) but this rises to 10% to 20% in parts of Africa and the Middle and Far East and immigrants from these countries retain the high prevalence rates seen in their country of origin.

Vertical transmission from mother to child, usually during parturition or soon after birth, and rarely *in utero*, is the usual means of transmission worldwide, although in Africa transmission from other infected children is common during the early childhood years. HBV is not transmitted by breast-feeding.

Horizontal transmission occurs, particularly in children, through minor abrasions or close contact with other children, and HBV can survive on household articles, such as toys or toothbrushes, for prolonged periods. Childhood chronic HBV is associated with very high levels of viral replication (up to 10^{10} IU/mL), ensuring that high levels of virus are present in minute amounts of blood thus facilitating viral spread.

HBV spread also occurs by the intravenous route (e.g., by transfusion of infected blood or blood products, or by contaminated needles employed by drug users, tattooists, or acupuncturists) or by close personal contact, such as during sexual intercourse, particularly in men who have sex with men. The virus can be found in semen and saliva.

Hepatitis B virus

The complete infective virion or Dane particle is a 42 nm particle comprised of an inner core or nucleocapsid (27 nm), surrounded by an outer envelope of surface protein (hepatitis B surface antigen, HBsAg). This surface coat is produced in excess by the infected hepatocytes and can exist separately from the whole virion in serum and body fluid as 22 nm particles or tubules.

The HBV genome is variable, and genetic sequencing can be used to define the different HBV genotypes, A–H. There is a strong correlation between genotypes and geographical areas. ***Genotype A*** is found in Northwest Europe, North America, and Central Africa; ***B*** in South-east Asia (including China, Taiwan, and Japan); ***C*** in South-east Asia; ***D*** in southern Europe, India, and the Middle East; ***E*** in West Africa; ***F*** in South and Central America, in American Indians and in Polynesia; ***G*** in France and the USA; and ***H*** in Central and South America. These genotypes may influence the chance of responding to interferon treatment (A more than B; C more than D) but all genotypes respond equally well to nucleoside analogues.

The core or nucleocapsid is formed of core protein (HBcAg), containing incompletely double-stranded circular DNA and DNA polymerase/reverse transcriptase. One strand is almost a complete circle and contains overlapping genes that encode both structural proteins (pre-S, surface [S], core [C]) and replicative proteins (polymerase [P] and X). The other strand is variable in length. DR1 and DR2 are direct repeats necessary for HBV synthesis during viral replication (Fig. 23.14).

HBeAg is a protein formed via specific self-cleavage of the pre-core/core gene product, which is secreted separately by the cell.

Hepatitis B mutants

Mutations occur in the various reading frames of the HBV genome (Fig. 23.14). These mutants can emerge in patients with chronic HBV infection (escape mutants) or can be acquired by infection.

HBsAg mutants are produced by alterations in the 'a' antigenic determinants of the HBsAg proteins, usually with a substitution of glycine for arginine at position 145. This results in changes in the antibody-binding domain and may render the vaccine ineffective, although spread of vaccine-resistant virus has not proved problematic. In patients with some HBV genotypes (particularly genotype D), a mutation in the ***pre-core region***, when a guanosine (G) to adenosine (A) change creates a stop codon (TAG), prevents the production of HBeAg (the secreted form of HBcAg) but the synthesis of HBcAg is unaffected. This mutation may be associated with

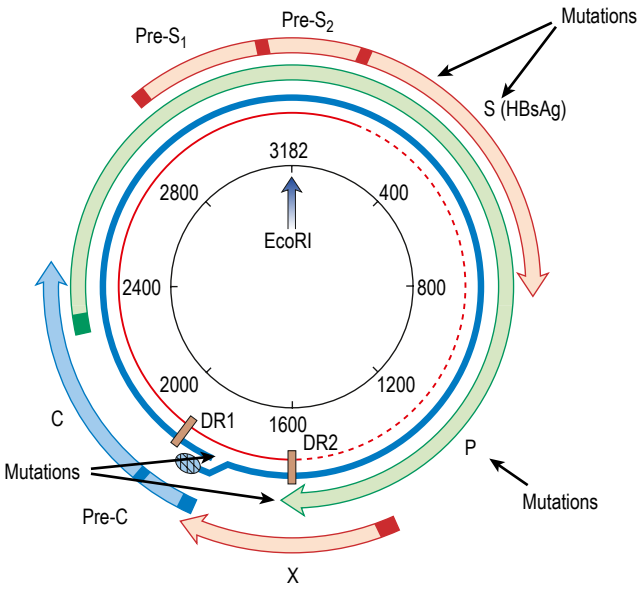


Fig. 23.14 Hepatitis B virus (HBV) genome. The viral DNA is partially double-stranded (*red incomplete circle* and *blue circle*). The long strand *(blue)* encodes seven proteins from four overlapping reading frames (S, surface [Pre-S_1, Pre-S_2, S]; C, core [Pre-C, C]; P, polymerase [P]; and X gene [X]). EcoRI restriction enzyme-binding site is included as a reference point. *DR,* Direct repeat; *HBsAg,* hepatitis B surface antigen.

HBeAg-negative disease but other mutations in the core promoter region of the virus also give rise to HBeAg-negative disease.

DNA polymerase mutants occur, particularly following treatment with the first generation of directly acting antiviral drugs, such as lamivudine.

Pathogenesis

Pre-S_1 and pre-S_2 regions are involved in attachment to the hepatocyte receptor—the sodium taurocholate co-transporting polypeptide (NTCP). After penetration into the cell the virus loses its coat and the virus core is transported to the nucleus without processing. The transcription of HBV into messenger RNA takes place when the HBV DNA is converted into a closed covalent circular form (cccDNA), which acts as a template for RNA transcription. The transcribed RNA is then reverse-transcribed by the viral polymerase protein to HBV DNA, which is circularized and packaged. Hence HBV, like HIV, replicates via reverse transcription that introduces errors into the replication cycle (reverse transcriptase enzymes are error-prone) but allows production of viral DNA that can integrate into the host genome. Unsurprisingly, some drugs (e.g., lamivudine and tenofovir) that are active against the HIV polymerase also have effects on the HBV polymerase.

Translation into HBV proteins (Box 23.8), as well as replication of the genome, takes place in the endoplasmic reticulum; the proteins are then packaged together and exported from the cell. There is excess production of non-infective HBsAg particles, which are extruded into the circulation.

HBV is not usually directly cytopathic (although high replication levels in immunosuppressed individuals can lead to direct toxicity) and liver damage is produced by the host immune response.

HBV-specific cytotoxic CD8 T cells recognize the viral antigen via human leucocyte antigen (HLA) class I molecules on the infected hepatocytes. However, suppressor or regulatory T cells inhibit these cytotoxic cells, leading to viral persistence and chronic

Box 23.8 Hepatitis B virus (HBV) proteins

HBV protein	Significance
Core	Protein of core particle; forms nucleocapsids
Pre-core (HBeAg)	Pre-core/core cleaves to HBeAg; good marker of active replication and role in inducing immunotolerance
Surface (HBsAg)	Envelope protein of HBV; basis of current vaccine
Pre-S_2Pre-S_1	HBV binding and entry into hepatocytes
Polymerase	Viral replication
X protein	Transcriptional and transactivator activity

HBV infection. Th1 responses (IL-2, interferon-gamma) are thought to be associated with viral clearance, and Th2 (IL-4, 5, 6, 10, 13) responses with the development of chronic infection and disease severity. Viral persistence in patients with a poor cell-mediated response leads to an asymptomatic, inactive, chronic HBV infective state. However, a 'better' response results in continuing hepatocellular damage, with the development of chronic hepatitis.

Chronic HBV infection progresses through a series of four distinct phases (Fig. 23.15).

- ***Phase 1 HBeAg-positive chronic HBV infection.*** The natural history of childhood-acquired HBV is shown in the figure. In this first early phase there is high-level viral replication, with HBeAg, which is not associated with an immune-mediated liver injury response and does not cause liver damage. This was previously called the 'immunotolerant phase' of HBV but it is now clear that an immune response does develop, although why this does not lead to an obvious hepatitis has not yet been established. This phase matures into the next adolescent phase.
- ***Phase 2 HBeAg-positive chronic HBV infection.*** During adolescence, liver damage often develops, with fluctuating raised transferases (ALT) in the presence of high levels of HBeAg-positive infection. This was previously termed the 'immunoactive phase' but it is now recognized that the change from 'inactive' to 'active' disease is complex and not simply related to changes in immune tolerance. In some patients this disease phase will progress to cirrhosis and treatment in this phase may lead to seroconversion from HBeAg-positive to HBeAg-negative.
- ***Phase 3 HBeAg-negative chronic HBV infection.*** In many patients the disease evolves to a quiescent phase (previously termed the 'inactive carrier' phase), when host immune responses suppress viral replication, leading to low-level HBV DNA (<2000 IU/mL), absence of HBeAg and normal ALTs. In some patients the virus is eventually cleared with the loss of HBsAg.
- ***Phase 4 HBeAg-negative chronic HBV infection.*** In a proportion of patients with inactive HBV, disease will reactivate as they age, and this final phase is characterized by high-level viral replication (HBV DNA $10^5/10^6$) but negative HBeAg and raised ALT. In this phase, fluctuating ALT levels may lead to misdiagnosis and so liver function and HBV DNA should be tested four times a year to establish the diagnosis in patients who are HBeAg-negative. This disease phase is often associated with viral mutations (see earlier).

HCC can develop at all stages of disease but is more common in those with high levels of HBV DNA.

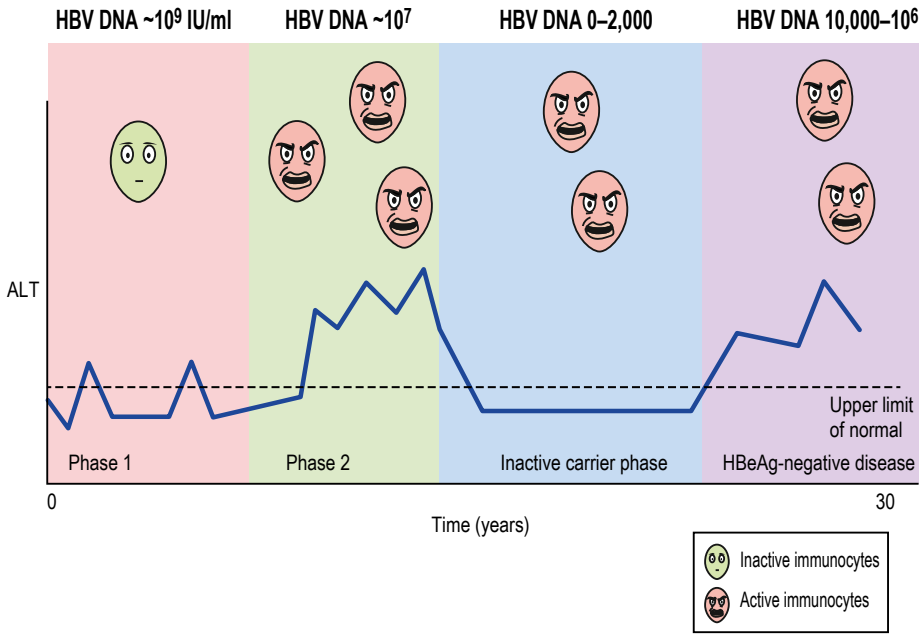


Fig. 23.15 Natural history of childhood-acquired hepatitis B virus (HBV). **Phase 1:** HBeAg-positive chronic HBV infection: early disease is characterized by high-level viraemia with HBeAg (the secreted form of HBcAg) and a minimal immune response, not leading to any significant liver damage (i.e., there is no acute hepatitis). **Phase 2:** HBeAg-positive chronic HBV infection: fluctuating liver function tests (tests of alanine aminotransferase, ALTs) and ongoing liver damage that may lead to cirrhosis. Treatment in this phase may lead to seroconversion. In many the immunoactive phase is followed by a period of immune control **(the 'inactive carrier' phase)**, when host immune responses suppress viral replication, leading to low-level HBV DNA (<2000 IU/mL), absence of HBeAg and normal ALTs. In some patients the virus is eventually cleared, with the loss of HBsAg; in many, however, viral reactivation occurs with viral mutations, leading to **HBeAg-negative disease** with high levels of viral replication (HBV DNA $10^5/10^6$) and ALT in the absence of HBeAg. Note that, in this phase, fluctuating ALTs may lead to misdiagnosis, and so the ALTs and HBV DNA should be tested four times a year to establish the diagnosis in patients who are HBeAg-negative. (N.B. Unlike children, adults commonly develop an acute hepatitis leading to viral clearance in the overwhelming majority [see Fig. 23.16]).

Immunosuppression, such as occurs during chemotherapy, aggravates all phases of HBV and a particular problem is seen in patients with HBeAg-negative, inactive disease, where the presence of normal liver biochemistry provides a false sense of security. HBV reactivation is common in such patients and has a high mortality. It is essential for all those due to receive chemotherapy to be screened for HBsAg, and those who have chronic HBV should receive prophylactic antiviral therapy, such as tenofovir or entecavir. Profound immunosuppression, such as occurs during bone marrow transplantation, or rituximab therapy, can lead to disease reactivation in patients who have lost all traces of HBV, and in those with evidence of HBV exposure (i.e., those with antibodies against hepatitis B—anti-hepatitis B core or anti-hepatitis B surface antibodies); consideration of prophylactic therapy or close monitoring is required.

Clinical features of acute hepatitis B infection

The sequence of events following acute HBV infection is shown in Fig. 23.16. However, in many, the infection is subclinical. When HBV infection is acquired perinatally, an acute hepatitis does not usually occur and chronic infection develops. In adults, an acute infection is common and the virus is cleared in approximately 99% of patients. The clinical picture is the same as that found in HAV infection, although the illness may be more severe. In addition, a serum sickness-like immunological syndrome may be seen. This consists of rashes (e.g., urticaria or a maculopapular rash) and polyarthritis affecting small joints in up to 25% of cases in the prodromal period. Fever is usual. Extrahepatic immune complex-mediated conditions, such as an arteritis or glomerulonephritis, are occasionally seen.

Investigations

These are generally the same as for hepatitis A.

Specific tests

The markers for HBV are shown in Box 23.9. HBsAg is looked for initially; if it is found, a full viral profile is performed. In acute infection, as HBsAg may be cleared rapidly and anti-HBc IgM is diagnostic, patients must be tested for both HBsAg and anti-core antibodies if HBV is suspected. HBV DNA is the most sensitive index of viral replication. In all patients with HBsAg testing for hepatitis delta (HDV) should be performed.

Prognosis

The majority of patients recover completely, acute hepatic failure occurring in up to 1%. Some patients go on to develop chronic hepatitis and the outcome depends on several factors, chiefly the age of the patient.

Management of acute hepatitis

This is mainly symptomatic. However, patients should have their HBV markers monitored. Several experts suggest that entecavir

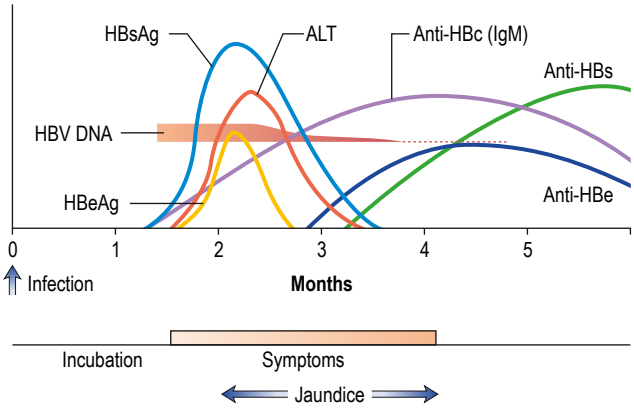


Fig. 23.16 Time course of the events and serological changes following acute infection with hepatitis B virus (HBV)—such as typically occurs in infected adults. ***Antigens:*** HBsAg appears in the blood from about 6 weeks to 3 months after an acute infection and then disappears. HBeAg rises early and usually declines rapidly. ***Antibodies:*** Anti-HBs appear late and indicate immunity. Anti-HBc is the first antibody to appear and high titres of IgM anti-HBc suggest an acute and continuing viral replication. It persists for many months. IgM anti-HBc may be the only serological indicator of recent HBV infection in a period when HBsAg has disappeared and anti-HBs are not detectable in the serum. Anti-HBe appears after the anti-HBc and its appearance relates to a decreased infectivity: that is, a low risk. *ALT,* Alanine aminotransferase.

Box 23.9 Significance of viral markers in hepatitis B

Marker	Significance
Antigens	
HBsAg	Acute or chronic infection
HBeAg	Acute hepatitis B Persistence implies: Continued infectious state Development of chronicity
HBV DNA	Implies viral replication Found in serum and liver Levels indicate response to antiviral treatment
Antibodies	
Anti-HBs	Immunity to HBV; previous exposure; vaccination
Anti-HBe	Seroconversion
Anti-HBc	
IgM	Acute hepatitis B (high titre) Chronic hepatitis B (low titre)
IgG	Past exposure to hepatitis B (HBsAg-negative)

or tenofovir should be given for the persistent presence of HBeAg beyond 12 weeks, and in some patients who are very ill.

Prevention

A highly effective vaccine prevents HBV infection and this is now widely deployed, although some countries still rely on the avoidance of risk factors (see earlier) and selective vaccination of high-risk groups. The important role of vaccination in the prevention of mother-to-child transmission is discussed later. Infectivity is highest in those with the e antigen and/or HBV DNA in their blood. Recent changes in the use of antiviral agents for chronic HBV infection recommend therapy for most people with high levels of viraemia, which is likely to reduce viral transmission still further.

Immunization

Vaccination is obligatory in most high income countries (the UK recently added HBV to the childhood vaccination schedule), as well as countries with high endemicity, and has been shown to reduce mortality and morbidity.

Active and passive (combined) prophylaxis with vaccination and immunoglobulin should be given to non-vaccinated healthcare staff with accidental needlestick injury; all newborn babies of HBsAg-positive mothers; and unvaccinated regular sexual partners of HBsAg-positive patients who have been found to be HBV-negative.

For adults, a dose of 500 IU of specific hepatitis B immunoglobulin (HBIg) (200 IU for newborns) is given; the vaccine (i.m.) is given at another site.

Prevention of mother-to-child transmission of HBV offers the potential to arrest transmission of this infection and, eventually, lead to the elimination of HBV as a global public health issue. The WHO has issued recommendations on the elimination of HBV and a key component of this is the prevention of mother-to-child transmission by identification of infected mothers and vaccination of all children. In women with high levels of virus additional steps to prevent transmission to their children include treatment during pregnancy with oral antiviral agents to lower the viral load and passive vaccination of children with antibodies against HBsAg.

Active immunization

This is with a recombinant yeast vaccine, produced by insertion of a plasmid containing the gene of HBsAg into a yeast.

Dosage regimen

Three injections (at 0, 1, and 6 months) are given into the deltoid muscle; this gives long-term protection in over 90% of patients. People who are over 50 years of age or clinically ill and/or immunocompromised (including those with human immunodeficiency virus [HIV] infection or acquired immunodeficiency syndrome [AIDS]) have a poor antibody response; more frequent and larger doses may be required and a small proportion of people do not respond to current vaccines. It is not cost-effective to check antibody levels prior to active immunization. There are few side-effects from the vaccine.

Chronic hepatitis B virus infection

Following an acute HBV infection, which may be subclinical, approximately 1% to 10% of patients will not clear the virus and will develop a chronic HBV infection. Infected children invariably develop chronic infection and mother-to-child transmission explains the very high prevalence of chronic HBV worldwide. Many patients with chronic infection are asymptomatic, but may develop symptoms and signs of chronic liver disease or extrahepatic manifestations of chronic HBV infection as described in 'Clinical features of acute hepatitis B infection'.

Investigations

These may show a moderate rise in aminotransferases but infection with normal aminotransferases is common. The serum bilirubin is often normal. HBsAg and HBV DNA are found in the serum, sometimes with HBeAg. Increasingly quantitation of the amount of HBsAg is performed since those with low levels are more likely to lose virus with antiviral treatment.

Histologically, there is a full spectrum of changes, from near normal with only a few lymphocytes and interface hepatitis to full-blown cirrhosis. HBsAg may lend a 'ground-glass' appearance to

the cytoplasm on haematoxylin and eosin staining, and this can be confirmed on orcein staining or, more specifically, immunohistochemical staining. HBcAg can also be demonstrated in hepatocytes by appropriate immunohistochemical staining.

Management of chronic hepatitis B

The proven safety of the oral antiviral medications used to treat chronic HBV infection has led to recommendations that the majority of patients with HBV DNA levels greater than 2000 IU/mL should be offered therapy. In the past therapy was restricted to those with high-level viral replication (HBV DNA levels >10 000 IU/mL) who also had evidence of ongoing liver damage (raised ALT levels or evidence of fibrosis on liver biopsy or non-invasive assessment, such as transient elastography) but there is an emerging consensus that this approach, which requires repeated, at least annual, assessment of patients is too restrictive and that most viraemic patients should be treated. Patients who have normal liver function and very low levels of viraemia (HBV DNA levels less than 2000) do not require treatment but must be reviewed regularly (every 6 or 12 months) with an assessment of HBV DNA and liver function tests to determine whether or not they have developed activation of their disease.

The lifetime risk of malignancy is increased in all patients who are HBsAg-positive, and males over 40 and females over 50 should be offered 6-monthly screening by abdominal ultrasound and α-fetoprotein testing for early detection of HCC.

Aim of therapy

The goals are to prevent disease progression and, ideally, eliminate HBsAg.

- ***Oral nucleotides*** suppress viral replication and are used for prolonged periods of time.
- ***Interferon*** is an immunostimulator that induces an immune response, leading to prolonged remission after discontinuation of therapy. It is now rarely used but may be re-introduced as a component of curative treatment regimens with emerging antiviral drugs.

In patients receiving long-term oral antiviral agents, liver fibrosis regresses. Even those with cirrhosis may recover, and the liver may remodel and lose all traces of fibrosis. The risk of liver cancer appears to be reduced in those taking long-term oral nucleotides but a residual risk remains and screening for liver cancer should continue in those receiving treatment.

Antiviral agents

Oral antiviral therapy for HBV (entecavir and tenofovir-containing agents) is very effective and almost all compliant patients respond with a decrease in HBV DNA to undetectable levels and a reduction in liver inflammation (**Box 23.10**). Both HBeAg-positive and HBeAg-negative patients respond equally well. Long-term viral suppression has been shown to reverse fibrosis, and even patients with cirrhosis respond with reversion of fibrosis. Resistance is not seen with tenofovir and is extremely rare with entecavir in patients not previously treated with older agents. Resistance-prone drugs, like ***lamivudine***, are no longer recommended. A small proportion of patients develop an immune response leading to loss of HBeAg and, very rarely, loss of HBsAg. However, the majority of patients who commence oral antiviral agents will require very prolonged treatment, perhaps life-long. Some patients with low levels of HBsAg who discontinue therapy after several years go on to eliminate HBsAg, often with a transient increase in liver inflammation, but a tiny proportion develop liver failure when treatment is stopped. Studies are in progress to determine the factors that predict beneficial outcomes following treatment discontinuation. ***Entecavir*** and ***tenofovir*** are the drugs of choice for HBV, and both agents are associated with very few side-effects and an excellent response. The affordable, off-patent, tenofovir disoproxil fumarate (TDF) may occasionally provoke renal Fanconi syndrome (a syndrome of renal tubular dysfunction accompanied by urinary phosphate wasting and reduction of bone mineral density) but this is most often seen when using TDF in patients with HIV infection and is extremely rare in people with HBV disease alone. This rare side-effect may be avoided by the use of the more expensive tenofovir alafenamide. Combination therapy has little benefit and a single drug should be used. New treatments that target alternative steps in the viral life cycle (e.g., polymerization of the core protein) or augment the immune response (e.g., 'checkpoint inhibitors', see Chapter 41) are under evaluation, as they carry the possibility of finite therapy with loss of HBsAg).

Pegylated interferon-alfa-2a (180 μg once a week s.c.) is now rarely used in people with chronic HBV infection and is usually restricted to those who are HBeAg-positive with active disease. Some patients (25%–45%, depending on genotype—A and B respond best) lose HBeAg and move to the 'inactive' HBeAg-negative phase of disease. A proportion then goes on to lose HBsAg some years after treatment discontinuation. Patients with higher serum aminotransferase values (three times the upper limit of normal), who are younger and have viral loads below 10^7 IU/mL, respond best to treatment. Patients with concomitant HIV respond poorly and those with cirrhosis should not receive interferon. Response can be assessed during therapy by measuring the serum levels of HBsAg: if these fall after 3 months, a favourable outcome is likely, and most doctors stop therapy after 3 months if the HBsAg level remains unchanged. In patients with HBeAg-negative HBV, pegylated interferon-alfa-2a is occasionally used, as it offers a finite duration of therapy. A proportion of patients with active disease (high ALT, increased HBV DNA) convert to inactive disease, and response can be assessed by an early decline in HBsAg.

Side-effects of treatment include an acute influenza-like illness occurring 6 to 8 hours after the first injection. This usually disappears after subsequent injections but malaise, headaches, and myalgia are common; depression, reversible hair loss, and bone marrow depression and infection may also occur.

Box 23.10 Factors predictive of a sustained response to treatment in chronic hepatitis B

Duration of disease
- Short

Liver biochemistry
- High serum aminotransferases

Histology
- Active liver disease (mild to moderate)

Viral levels
- Low HBV DNA levels

Other
- Absence of immunosuppression
- Female gender
- Acquisition in adulthood
- Delta virus negative
- Rapidity of response to oral therapy

Prognosis

The clinical course of hepatitis B is very variable; treatments have improved survival, stopped progression of fibrosis, and enabled regression of fibrosis. Established cirrhosis is associated with a poor prognosis. ***HCC*** is a frequent association and is one of the

most common carcinomas in HBV-endemic areas such as the Far East. Surveillance for HCC must continue, even when HBV DNA is negative in patients who have HBsAg and are not treated, and also in those rendered negative by therapy. The incidence of HCC is being reduced by routine HBV vaccination of all children.

Hepatitis D

This is caused by the hepatitis D virus (HDV or delta virus), an incomplete RNA particle enclosed in a shell of HBsAg; it belongs to the *Deltaviridae* family. The virus is unable to replicate on its own but is activated by the presence of HBV. It is common in some parts of the world, including Eastern Europe (Romania, Bulgaria), North Africa, and the Brazilian rainforest. Hepatitis D viral infection can occur either as a co-infection or as a superinfection.

- ***Co-infection*** of HDV and HBV is clinically indistinguishable from an acute icteric HBV infection but a biphasic rise of serum aminotransferases may be seen. ***Diagnosis*** is confirmed by finding serum IgM anti-HDV in the presence of IgM anti-HBc. IgM anti-delta appears at 1 week and disappears by 5 to 6 weeks (occasionally 12 weeks), when serum IgG anti-delta is seen. The HDV RNA is an early marker of infection. The infection may be transient but the clinical course is variable.
- ***Superinfection*** results in an acute flare-up of previously quiescent chronic HBV infection. A rise in serum AST or ALT may be the only indication of infection. ***Diagnosis*** is made by finding HDV RNA or serum IgM anti-HDV at the same time as IgG anti-HBc. Active HBV DNA synthesis is reduced by delta superinfection and patients are usually negative for HBeAg with low HBV DNA.

Acute hepatic failure can follow both types of infection but is more common after co-infection. HDV RNA in the serum and liver can be measured and is found in acute and chronic HDV infections.

Chronic hepatitis D

This is a relatively infrequent chronic hepatitis but spontaneous resolution is rare. Between 60% and 70% of patients will develop cirrhosis, and more rapidly than with HBV infection alone. In 15% the disease is rapidly progressive, with development of cirrhosis in only a few years. The diagnosis is made by finding anti-delta antibody in a patient with chronic liver disease who is HBsAg-positive. It can be confirmed by finding HDV in the liver or HDV RNA in the serum by reverse transcription polymerase chain reaction (PCR).

Management

Treatment of patients with active liver disease (raised ALT levels and/or inflammation on biopsy) is with pegylated interferon-alfa-2a for 12 months, although response rates are very low. The HBsAg entry inhibitor bulevirtide reduces HDV RNA levels when administered long term and is increasingly used, either in isolation or in combination with oral nucleotides or interferon, to reduce the rate of progression of hepatitis delta–associated liver disease. Therapy must be continued for more than 1 year and it is not yet clear whether prolonged remission after cessation of treatment can be achieved with very long-term treatment. The drug is self-administered as a subcutaneous daily injection and the chief side-effect is local injection site inflammation.

Hepatitis C

Epidemiology

An estimated 70 million people are infected with this virus worldwide. The prevalence rate of infection ranges from 0.4% in Europe and 1% to 3% in Southern Europe (possibly linked to intramuscular injections of vaccines or other medicines) to 6% in Africa; in Egypt, the rates were as high as 19%, owing to parenteral antimony treatment for schistosomiasis. The virus is transmitted by blood and blood products, and was common in people with haemophilia treated before screening of blood products was introduced. The incidence in intravenous drug users is high (40%–60%). The low rate of HCV infection in high-risk groups—such as men who have sex with men, sex workers, and attendees at sexually transmitted infection clinics—suggests a limited role for sexual transmission. Vertical transmission from a healthy mother to child can occur but is rare (approximately 5%). Other routes of community-acquired infection (e.g., close contact) are extremely rare. In 20% of cases the exact mode of transmission is unknown.

Hepatitis C virus

Hepatitis C virus (HCV) is a single-stranded RNA virus of the *Flaviviridae* family. The RNA genome is approximately 10 kb in length, encoding a polyprotein product consisting of structural (capsid and envelope) and non-structural viral proteins (**Fig. 23.17**). Comparisons of subgenomic regions, such as E1, NS4, or NS5, have allowed variants to be classified into six genotypes, which have differing geographical distributions. Variability is distributed throughout the genome, with the non-structural gene of different genotypes showing 30% to 50% nucleotide sequence disparity. Genotypes 1a and 1b account for 70% of cases in the USA and 50% in Europe, while genotype 3 is common in the Indian subcontinent and genotype 4 is prevalent in Egypt. There is a rapid change in envelope proteins, making it difficult to develop a vaccine. Antigens from the nucleocapsid regions have been used to develop enzyme-linked immunosorbent assays (ELISAs). The current assay, ELISA-3, incorporates antigens NS3, NS4, and NS5 regions.

Clinical features

Most acute infections are asymptomatic, about 10% of patients having a mild influenza-like illness with jaundice and a rise in serum aminotransferases. Most patients will not be diagnosed until they present, years later, with evidence of abnormal transferase values at health checks or with chronic liver disease.

Investigations

Investigation is by evaluation of HCV RNA and HCV antibodies; HCV RNA can be detected from 1 to 8 weeks after infection. Anti-HCV tests are usually positive 8 weeks from infection. Patients with acute HCV infection should be tested on several occasions, as many have fluctuating viraemia during the first few months of infection, with periods of undetectable HCV RNA followed by virological relapse. Viral clearance is confirmed by multiple tests for HCV RNA over a period of many months.

Management

In acute infection spontaneous viral clearance is relatively common but therapy is effective and most experts recommend early treatment to reduce the risk of chronic infection and prevent transmission. Therapy almost always leads to viral clearance (see ‘Chronic hepatitis C infection [management]’, later). Needle-stick injuries sustained from patients with HCV must be followed and treated early, although the vast majority (>97%) do not go on to develop viraemia. For the treatment of patients with co-infection with HIV, see Chapter 44.

Prognosis

Some 85% to 90% of asymptomatic patients develop chronic liver disease. A higher percentage of symptomatic patients ‘clear’ the virus, with only 48% to 75% going on to chronic liver disease.

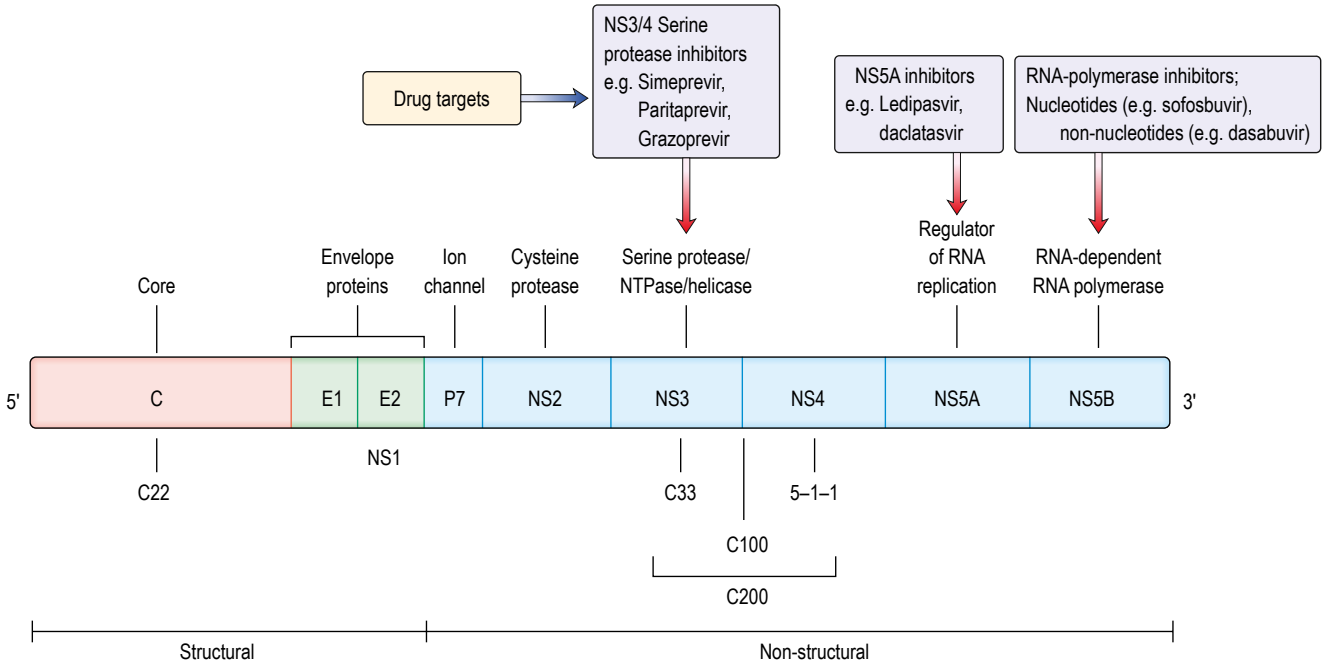


Fig. 23.17 Hepatitis C virus. This is a single-stranded RNA virus with viral proteins. Targets for recent drugs are shown. *C,* Core; *E,* envelope; *NS,* non-structural.

Chronic hepatitis C infection

Pathogenesis

As with hepatitis B infection, cytokines in the Th2 phenotypes are both profibrotic and cause the development of chronic infection. A dominant CD4 Th2 response, with a weak CD8 interferon-gamma response, may lead to rapid fibrosis. Th1 cytokines are antifibrotic and thus a dominant CD4 Th1 and CD8 cytolytic response may cause less fibrosis.

Clinical features

Patients with chronic HCV infection are usually asymptomatic, the disease only being discovered following a routine biochemical test when mild elevations in the aminotransferases (usually ALT) are noticed (50%). The elevation in ALT may be minimal and fluctuating (Fig. 23.18), and some patients have a persistently normal ALT (25%), the disease being detected by checking HCV antibodies (e.g., in blood donors). Non-specific malaise and fatigue are common in chronic infection and often reverse following viral clearance. Extrahepatic manifestations are seen, including arthritis, cryoglobulinaemia with or without glomerulonephritis, and porphyria cutanea tarda. There is a higher incidence of diabetes, and associations with lichen planus, sicca syndrome, and non-Hodgkin lymphoma.

Chronic HCV infection causes slowly progressive fibrosis that leads, over decades, to cirrhosis. After 20 years of infection, 16% of patients have developed cirrhosis; the proportion that will ultimately develop cirrhosis after a lifetime of infection remains unknown but is likely to be higher than the proportions reported after short-term follow-up. Factors associated with rapid progression of HCV fibrosis include excess alcohol consumption, co-infection with HIV, obesity, diabetes, and infection with genotype 3. Once cirrhosis has developed, some 3% to 5% per year will develop decompensated cirrhosis and approximately 3% will develop liver cancer. Unlike HBV infection, HCV rarely causes liver cancer in the absence of cirrhosis. In most parts of the world, HCV was spread (either by injection drug use or by poorly sterilized medical devices) in the 1970s. Given the slow development of cirrhosis, current mathematical models of disease burden predict a massive increase in HCV-related end-stage liver disease over the next decade which may be avoided by widespread treatment with current antiviral agents.

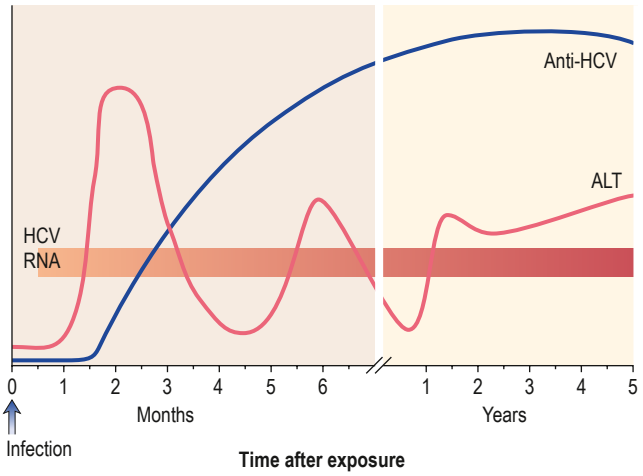


Fig. 23.18 Time course of the events and serological changes seen following infection with hepatitis C virus. *ALT,* Alanine aminotransferase.

Investigations

Diagnosis is made by finding HCV antibody in the serum using third-generation ELISA-3 tests. A proportion of patients with spontaneous clearance will have undetectable HCV RNA in the serum (measured by PCR) but most individuals who are antibody-positive will be viraemic. The level of viraemia varies from a few thousand to many millions of viral copies per millilitre, although current practice is to present viral load measurements in international units (IU). Disease progression and treatment outcome are not influenced by the viral load. The availability of pan-genotypic treatments removes the

need to assess viral genotype pre-treatment but this is still necessary when genotype specific treatments are in use.

Management

The aim of treatment is to eliminate the HCV RNA from the serum in order to:

- stop the progression of active liver disease
- prevent transmission of the virus
- prevent the development of HCC.

A clinical cure is determined by a sustained virological response (SVR), defined by a negative HCV RNA by PCR 3 months after the end of therapy.

Antiviral agents

Treatment for HCV infection is now based on direct-acting antiviral agents that are administered orally. These drugs target enzymes in the virus—usually the polymerase, NS5a protein or NS3/4 protease. Therapy may involve a nucleotide inhibitor of the polymerase (sofosbuvir) combined with an NS5a inhibitor. This combination may be pan-genotypic (sofosbuvir/velpatasvir or sofosbuvir/daclatasvir) or specific for genotypes 1 and 4 (sofosbuvir/ledipasvir), and involves a single tablet given once a day for 8 to 12 weeks. An alternative strategy is to use a protease inhibitor–based regimen (either glecaprevir or grazoprevir) combined with an NS5a inhibitor. Glecaprevir is combined with pibrentasvir and involves three tablets a day for 8 weeks in patients without cirrhosis or for 12 weeks in those with cirrhosis. The regimen is effective in all viral genotypes. Grazoprevir is combined with elbasvir and is used as a single tablet that is taken for 12 weeks and is effective in patients with genotypes 1 and 4. All of the treatment regimens eliminate virus in more than 95% of patients; in the very few who do not respond to therapy a 'rescue' treatment involving sofosbuvir/velpatasvir and the pan-genotypic protease inhibitor voxilaprevir is licensed and eliminates virus in over 90% of patients. Following viral clearance, liver fibrosis progression is halted and may even reverse, but in patients with cirrhosis the risk of liver cancer persists (albeit at a reduced level). Patients who do not have cirrhosis can be discharged from follow-up after successful therapy, defined as undetectable virus 12 weeks after discontinuation of treatment, but those with cirrhosis should continue on a program of surveillance for malignancy.

The extraordinary effectiveness of the oral antiviral regimens against HCV, combined with their excellent safety profile, has led to campaigns to identify and treat all patients with HCV, and the World Health Organization has a goal of eliminating HCV by 2030. If this can be achieved, it will represent the first time that a viral infection has been controlled without an effective vaccine. Elimination of hepatitis C requires widespread screening for infection, rapid introduction of pan-genotypic treatments, and minimal monitoring of patients while on therapy, although patients with decompensated cirrhosis may require additional support. This approach has been adopted in Egypt and has led to virtual elimination of the infection with a Presidential Program involving mobile treatment units and access to sofosbuvir/daclatasvir for those found to be infected. Other countries are adopting similar approaches and success will depend upon healthcare workers ensuring that all those at risk (i.e., born in a high-risk country or having past exposure to injecting drugs) are tested and treated.

Hepatitis E

Hepatitis E virus (HEV) is an RNA virus (*Hepeviridae* family; **Fig. 23.19**); it causes disease that is clinically very similar to hepatitis A. It is enterally transmitted and two patterns of disease are recognized.

- In countries with poor sanitation, outbreaks with HEV genotypes 1 and 2 occur, usually due to contaminated water.

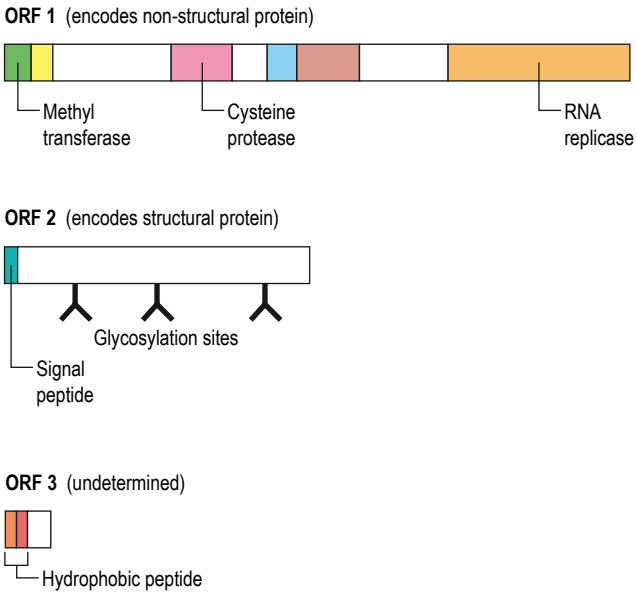


Fig. 23.19 Hepatitis E genome. A single-stranded RNA genome with three open reading frames (ORFs).

- Low-level ongoing infection via contaminated meat products (often pork or boar-containing sausages) is now frequently seen throughout the world due to genotype 3 infection and the majority of elderly people in the UK have evidence of exposure.

Epidemic HEV with genotypes 1 and 2 has a mortality from fulminant hepatic failure of 1% to 2%, which rises to 20% in pregnant women but hepatic failure is much less common in sporadic genotype 3 infection. There is no carrier state and infection does not progress to chronic liver disease, except in some immunosuppressed patients. Non-hepatic manifestations including neurological (Guillain-Barré and encephalitis), haematological (thrombocytopaenia), and renal (glomerulonephritis) complications are increasingly identified. An ELISA for IgG and IgM anti-HEV is available for diagnosis. HEV RNA can be detected in the serum or stools by PCR. Prevention and control depend on good sanitation and hygiene; a vaccine has been developed and used successfully in China. In patients who are immunocompromised, HEV infection may lead to chronic infection, which is treated by reduction of the immunosuppression and/or introduction of the broad-spectrum antiviral agent ribavirin.

Hepatitis non-A–E

Approximately 10% to 15% of acute viral hepatitides cannot be typed and are described as hepatitis non-A–E. GB agent (hepatitis G virus, HGV) and transfusion-transmitted virus (TTV) agents have not been documented as causing disease in humans.

Acute hepatitis due to other infectious agents

Abnormal liver biochemistry is frequently found in a number of acute infections. The abnormalities are usually mild and have no clinical significance. Further information about these diseases can be found in Chapter 42.

Infectious mononucleosis

Infectious mononucleosis is due to the Epstein–Barr (EB) virus. Mild jaundice, associated with minor abnormalities of liver biochemistry, is extremely common but 'clinical' hepatitis is rare. Hepatic

histological changes occur within 5 days of onset; the sinusoids and portal tracts are infiltrated with large mononuclear cells but the liver architecture is preserved. A Paul–Bunnell or Monospot test is usually positive, and atypical lymphocytes are present in the peripheral blood. Treatment is symptomatic.

Cytomegalovirus

Cytomegalovirus (CMV) can cause acute hepatitis, usually a 'glandular fever-type syndrome', in healthy individuals but is more severe in those with an impaired immune response. Only the latter need treatment with valganciclovir or ganciclovir.

CMV DNA is positive in blood; CMV IgM is also positive but there are false-positive reactions. Liver biopsy shows intranuclear inclusions and giant cells.

Herpes simplex

Very occasionally, the herpes simplex virus causes a generalized acute infection, particularly in the immunosuppressed patient and occasionally in pregnancy. Aminotransferases are usually massively elevated. Liver biopsy shows extensive necrosis. Aciclovir is used for treatment.

Toxoplasmosis

The clinical picture in toxoplasmosis is similar to that of infectious mononucleosis, with abnormal liver biochemistry, but the Paul–Bunnell test is negative.

Yellow fever

Yellow fever is a viral infection carried by the mosquito *Aedes aegypti*; it can cause acute hepatic necrosis. There is no specific treatment.

Adenovirus

There has been a recent surge in cases of acute severe hepatitis in previously healthy children under the age of 10, many reported from the UK. Growing evidence links these to adenovirus (subtype 41) and adeno-associated virus type 2 in genetically susceptible individuals. Adenovirus is also an occasional cause of hepatitis in immunosuppressed adult patients.

Further reading

European Association for the Study of the Liver. EASL Clinical Practice Guidelines: management of chronic hepatitis B infection. *J Hepatol* 2017; 67:370–398.
European Association for the Study of the Liver. EASL recommendations on treatment of hepatitis C 2018.*J Hepatol* 2020; 73:1170–1128.
National Institute for Health and Care Excellence. *NICE Clinical Guideline 165: Hepatitis B (Chronic): Diagnosis and Management of Chronic Hepatitis B in Children, Young People and Adults*. NICE 2013, updated 2017; https://www.nice.org.uk/guidance/cg165.
Razavi H, Waked I, Sarrazin C et al. The present and future disease burden of hepatitis C virus (HCV) infection with today's treatment paradigm. *J Viral Hepatitis* 2014; 21(Supplement S1):34–59.
Visvanathan K, Dusheiko G, Giles M et al. Managing HBV in pregnancy. *Gut* 2016; 65:340–350.
Wedemeyer H, Yurdaydin C, Dalekos GN et al, for the HIDIT Study Group. Peginterferon plus adefovir versus either drug alone for hepatitis delta. *N Engl J Med* 2011; 364:322–331.
World Health Organization. *Combating Hepatitis B and C to Reach Elimination by 2030.* Advocacy brief 2016, available from https://www.who.int/hepatitis/publications/hep-elimination-by-2030-brief/en/.

ACUTE HEPATIC FAILURE

Acute liver failure (ALF) is defined as acute liver injury with encephalopathy and deranged coagulation (INR > 1.5) in a patient with a previously normal liver. The time of development of jaundice to encephalopathy varies from 7 days (hyper-acute) to 8 to 28 days (acute) and 21 to 26 weeks (subacute). Occasionally, patients may have previous liver damage: for example, D virus superinfection in a previous carrier of HBsAg.

Box 23.11 Causes of acute hepatic failure

Viruses
- HAV, HBV, (HDV), HEV; rarely, HCV
- Cytomegalovirus
- Haemorrhagic fever viruses
- Herpes simplex virus
- Paramyxovirus
- Epstein–Barr virus

Drugs (examples)
- Paracetamol (acetaminophen)
- Antibiotics (ampicillin–clavulanate, ciprofloxacin, doxycycline, erythromycin, isoniazid, nitrofurantoin, tetracycline)
- Antidepressants (amitriptyline, nortriptyline)
- Antiepileptics (phenytoin, valproate)
- Anaesthetic agents (halothane)
- Lipid-lowering medications (atorvastatin, lovastatin, simvastatin)
- Immunosuppressive agents (cyclophosphamide, methotrexate)
- NSAIDs
- Salicylates (as a result of Reye's syndrome, see 'Hepatic mitochondrial injury syndromes')
- Disulfiram, flutamide, gold, propylthiouracil
- Illicit drugs (e.g., 'ecstasy' or cocaine)
- Herbal/alternative medicines (ginseng, pennyroyal oil, *Teucrium polium* chaparral or germander tea, kawa kawa)

Toxins
- Amanita phalloides mushroom toxin
- Bacillus cereus toxin
- Cyanobacteria toxin
- Organic solvents (e.g., carbon tetrachloride)
- Yellow phosphorus

Hepatic failure in pregnancy
- Acute fatty liver of pregnancy (AFLP)
- HELLP (*h*aemolysis, *e*levated *l*iver enzymes, *l*ow *p*latelets)

Vascular causes
- Ischaemic hepatitis
- Budd–Chiari syndrome
- Hepatic sinusoidal obstruction syndrome
- Portal vein thrombosis
- Hepatic arterial thrombosis (consider post transplant)

Metabolic causes
- α_1-antitrypsin deficiency
- Fructose intolerance
- Galactosaemia
- Lecithin-cholesterol acyltransferase deficiency
- Reye's syndrome
- Tyrosinaemia
- Wilson's disease

Malignancies
- Primary (usually HCC, rarely cholangiocarcinoma)
- Secondary (extensive hepatic metastases or infiltration)

Miscellaneous
- Adult-onset Still's disease
- Heatstroke
- Primary graft non-function in liver transplant

ALF is a rare but often life-threatening syndrome that is due to acute hepatitis from many causes (**Box 23.11**). These causes vary throughout the world; drug-induced liver injury (DILI) is the most common cause in 'Western' countries (paracetamol overdose is common in the UK—50% of cases). HCV only rarely causes acute hepatic failure.

Histologically, there is multiacinar necrosis involving a substantial part of the liver. Severe fatty change is seen in pregnancy (see Chapter 45) and Reye's syndrome (see 'Hepatic mitochondrial injury syndromes') or following intravenous tetracycline administration.

Clinical features

Examination shows a jaundiced patient with a small liver and signs of hepatic encephalopathy. The mental state varies from slight drowsiness, confusion, and disorientation (grades I and II) to unresponsive coma (grade IV) with convulsions. Fetor hepaticus is common but ascites and splenomegaly are rare. Fever, vomiting, hypotension,

and hypoglycaemia occur. Neurological examination shows spasticity and hyper-reflexia; plantar responses remain flexor until late. Cerebral oedema develops in up to 80% of patients with ALF but is far less common with subacute failure; its severe consequences, intracranial hypertension and brain herniation, have fallen substantially over the last 30 years and now only account for a minority of patient deaths. Other complications include bacterial and fungal infections, gastrointestinal bleeding, respiratory arrest, kidney injury (acute tubular necrosis), and pancreatitis.

Investigations

- Routine tests (see 'Investigations in the patient with liver disease' section).
- For suspected paracetamol overdoses, levels must be measured. Importantly, a low paracetamol level on admission is not uncommon, particularly in 'staggered' overdoses and may be misinterpreted as being below the treatment threshold for *N*-acetyl cysteine (NAC) infusions. If there is any doubt, NAC should be given (see also Chapter 15).
- There is hyperbilirubinaemia, high serum aminotransferases, and low levels of coagulation factors, including prothrombin and factor V. Aminotransferases are not useful indicators of the course of the disease, as they tend to fall along with the albumin with progressive liver damage.
- An electroencephalogram (EEG) is sometimes helpful in grading the encephalopathy.
- Ultrasound will define liver size and may indicate underlying liver pathology.

Management

There is no specific treatment, but patients should be managed in a specialized unit and early referral is encouraged. Transfer criteria to these units are shown in **Box 23.12**. The use of NAC has been extended to empirical treatment of non-paracetamol ALF, however, there is inconsistent evidence for a role. Given its excellent safety profile and that paracetamol may frequently be a co-factor for ALF, many will prescribe it. Supportive therapy, as for hepatic encephalopathy, is necessary (see 'Portosystemic encephalopathy [management]'). Continuous renal replacement therapy (RRT) lowers circulating ammonia concentration and is associated with improved survival in ALF and should be considered for those with hepatic encephalopathy. When signs of raised intracranial pressure (which may be measured directly) are present, 20% mannitol (1 g/kg body weight) should be infused intravenously; this dose may need to be repeated. Dexamethasone is not of value. Hypoglycaemia, hypokalaemia, hypomagnesaemia, hypophosphataemia, and hypocalcaemia should be anticipated and corrected with a 10% glucose infusion (checked by 2-hourly capillary blood glucose testing), potassium, magnesium, phosphate, and calcium supplements, respectively. Hyponatraemia should be corrected with hypertonic saline. Patients are often catabolic and enteral nutrition should be started if patients cannot maintain oral intake; protein feeding can be transiently held in hyperammonemic patients. Coagulopathy is managed with intravenous vitamin K, platelets, blood, or fresh frozen plasma and patients are given a proton pump inhibitor (PPI) to prevent gastrointestinal bleeding, although substantial bleeding complications are uncommon. Prophylaxis against bacterial and fungal infection is routine, as infection is a frequent cause of death and may preclude liver transplantation, although there is no strong evidence supporting this approach. Suspected infection should be treated immediately with appropriate antibiotics. Renal and respiratory failure should be treated as necessary. Liver transplantation has been a major advance for patients with ALF. It is difficult to judge the timing or the necessity for transplantation, but there are guidelines based on validated prognostic indices of survival (see later).

Box 23.12 Transfer criteria to specialized units for patients with acute liver injury

- INR > 3.0
- Presence of hepatic encephalopathy
- Hypotension after resuscitation with fluid
- Metabolic acidosis
- Prothrombin time (seconds) > interval (hours) from overdose (paracetamol cases)

Prognosis

Outcomes with medical management alone, at least for some causes of ALF, are now excellent, with a study supporting better long-term outcomes than emergency liver transplant. Poor prognostic variables indicating a need to discuss with a liver transplant unit, as incorporated in British and European acute liver failure guidelines, are shown in **Box 23.13**.

Further reading

Bernal W, Wendon J. Acute liver failure. *N Engl J Med* 2013; 369:2525–2534.
EASL Clinical Practical Guidelines on the management of acute (fulminant) liver failure. *J Hepatol.* 2017;66:1047–1081.
Vasques F, Cavazza A, Bernal W. Acute liver failure. *Curr Opin Crit Care* 2022; 28(2):198–207.

AUTOIMMUNE HEPATITIS

Autoimmune hepatitis (AIH) is a progressive inflammatory liver condition with a 75% female preponderance. Approximately 40% of AIH patients have a family history of autoimmune disease (e.g., pernicious anaemia, thyroiditis, or coeliac disease), and at least 20% have concomitant autoimmune diseases or will develop them during follow-up.

Pathogenesis

The pathogenesis of AIH is incompletely understood, although evidence increasingly demonstrates that genetic susceptibility, molecular mimicry, and impaired immunoregulatory networks contribute to the initiation and perpetuation of the autoimmune attack. Liver damage is thought to be mediated primarily by T-cell–mediated events ($CD4^+$ T cells) against liver antigens, producing a progressive necro-inflammatory process that leads to fibrosis and cirrhosis. However, no clear trigger mechanism has been found.

Box 23.13 Poor prognostic variables in acute hepatic failure indicating a need for discussion with a liver transplant unit

Non-paracetamol (acetaminophen) causes

- pH < 7.30 or HCO_3 < 18 mmol/L
- INR > 1.8
- Oliguria/renal failure or Na < 130 mmol/L
- Encephalopathy, hypoglycaemia, or metabolic acidosis
- Bilirubin > 300 μmol/L

Paracetamol overdose

- Arterial pH < 7.30 or HCO_3 < 18 mmol/L
- INR > 3.0 on day two or >4.0 thereafter
- Oliguria and/or acute kidney injury
- Altered level of consciousness
- Hypoglycaemia
- Elevated arterial lactate (>4 mmol/L) unresponsive to fluid resuscitation

INR, International normalized ratio.

Clinical features

There are two peaks in presentation. In the peri- and postmenopausal group, patients may be asymptomatic, or present with fatigue and abnormalities in liver biochemistry or the presence of chronic liver disease on examination. In the teens and early twenties, the disease (often type II) presents as an acute hepatitis with jaundice and very high aminotransferases, which do not improve with time. This age group often has clinical features of cirrhosis and patients who are ill may also have features of an autoimmune disease, such as fever, migratory polyarthritis, glomerulonephritis, pleurisy, pulmonary infiltration, or lung fibrosis.

There are rare overlap syndromes with PBC and PSC, existing concomitantly or developing consecutively.

Investigations

Liver biochemistry

The serum aminotransferases are high, with lesser elevations in the ALP and bilirubin. The serum γ-globulins are high: frequently twice normal, particularly the IgG. The biochemical pattern is similar in both types.

Haematology

A mild normochromic normocytic anaemia with thrombocytopenia and leucopenia is present, even before portal hypertension and splenomegaly. The PT is often high.

Autoantibodies

Two types of autoimmune hepatitis have been recognized:

- ***Type I with antibodies*** (titres >1:80):
  - anti-nuclear (ANA)
  - anti-smooth muscle (anti-actin).
- ***Type II with antibodies:*** anti-liver/kidney microsomal (anti-LKM1). The main target is cytochrome P4502D6 (CYP2D6) on liver cell plasma membranes.
- A third type of AIH (AIH-3) was proposed, positive for anti-soluble liver antigen (SLA) and negative for conventional autoantibodies. However, subsequent studies showed that anti-SLA is also present in typical cases of AIH-1 and AIH-2. Adult patients positive for anti-SLA only, and therefore having *bona fide* AIH-3, are rare. Approximately 13% of patients lack the autoantibodies listed.

Liver biopsy

Since transferases and IgG levels do not reflect disease activity, liver biopsy is mandatory to confirm the diagnosis and evaluate disease severity. The biopsy (**Fig. 23.20**) commonly shows chronic active hepatitis with inflammation and interface hepatitis. This reflects the changes of chronic hepatitis described in the 'Chronic hepatitis' section. The amount of interface hepatitis is variable but tends to be high in untreated patients. Lymphoid follicles are less often seen than in hepatitis C, and plasma cell infiltration is frequent. Fibrosis is present in all but the mildest forms of the disease and approximately one-third of patients have cirrhosis at presentation.

Management

Prednisolone 30 mg is given daily for at least 2 weeks, followed by slow reduction to a maintenance dose of 5 to 15 mg daily. Azathioprine should be added, 1 to 2 mg/kg daily, as a steroid-sparing agent and, in some, as sole long-term maintenance therapy. Levels of thiopurine methyltransferase (TPMT) should be obtained before treatment is started (see Chapter 22). Other agents that have been used in resistant cases include budesonide (in non-cirrhotic patients), mycophenolate, ciclosporin, and tacrolimus.

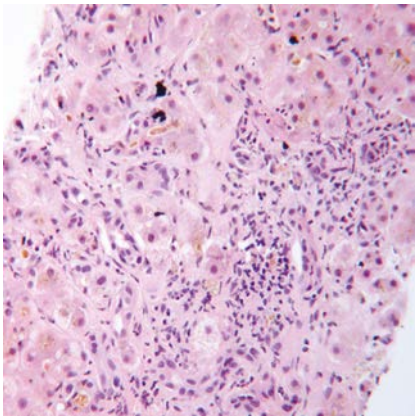


Fig. 23.20 Chronic active hepatitis. Note the inflammation with interface hepatitis (haematoxylin and eosin, ×10).

Prognosis

Steroid and azathioprine therapy induce remission in over 80%; indeed, this response forms part of the diagnostic criteria. Treatment is life-long in most, although withdrawal may be considered after 2 to 3 years of biochemical remission. Those with initial cirrhosis are more likely to relapse following withdrawal and require indefinite therapy. Liver transplantation is performed if treatment fails, although the disease may recur. HCC occurs less frequently than with viral-induced cirrhosis. The risk of malignancy (particularly of the skin) associated with chronic low-dose azathioprine therapy has been reported to be 1.4 times normal.

Further reading

Mieli-Vergani G, Vergani D, Czaja AJ et al. Autoimmune hepatitis. *Nat Rev Dis Primers* 2018; 4:18017.
Terziroli Beretta-Piccoli B, Mieli-Vergani G, Vergani D. Autoimmune hepatitis: standard treatment and systematic review of alternative treatments. *World J Gastroenterol* 2017; 23:6030–6048.

DRUG-INDUCED CHRONIC HEPATITIS

Several drugs can cause a chronic hepatitis similar to autoimmune hepatitis (see **Box 23.27**). Patients are often female, present with jaundice and hepatomegaly, and have raised serum aminotransferases and globulin levels. Improvement follows drug withdrawal but exacerbations can occur with re-introduction. Isoniazid, amiodarone and methotrexate can cause chronic histological changes. With rare exceptions, patients with pre-existing chronic liver disease are not more susceptible to drug injury.

CHRONIC HEPATITIS OF UNKNOWN CAUSE

As increasing numbers undergo routine blood tests, mild elevations in serum aminotransferases and γ-GT are commonly found. Many will have no symptoms or signs of liver disease. All known aetiological agents should be excluded, risk factors for MASLD evaluated, and tests carried out to exclude liver diseases. Liver biopsy should be performed if the elevation in aminotransferases (>100 IU/L) persists for over a year but often only non-specific changes are found.

METABOLIC DYSFUNCTION–ASSOCIATED STEATOTIC LIVER DISEASE

Metabolic dysfunction–associated steatotic liver disease or MASLD, formerly referred to as non-alcoholic fatty liver disease or NAFLD, is now the most common cause of chronic liver disease in many high income countries. MASLD is a spectrum of liver diseases comprised of metabolic dysfunction–associated steatotic liver (MASL) and metabolic dysfunction–associated steatohepatitis (MASH, formerly non-alcoholic steatohepatitis). Whereas MASL has negligible risk of progression, 10% to 30% of MASH patients may develop cirrhosis or HCC. MASL and MASH have traditionally been considered as two separate clinical entities, rather than two points on a disease continuum. However, studies that have evaluated sequential liver biopsies are challenging this notion and have found that both patients with MASL and those with MASH have the potential to develop progressive liver disease, suggesting that MASL, MASH and fibrosis progression are a continuum rather than separate diagnoses. Indeed, one large study showed no difference in liver-related adverse events between definite MASH and severe steatosis. However, patients with advanced fibrosis at presentation were much more likely to progress than those without, and it is these patients therefore that require follow-up. In this cohort, complications of cirrhosis were the third most common cause of death, following cardiovascular events and non-hepatic malignancies.

Risk factors for MASLD are obesity, hypertension, type 2 diabetes, and hyperlipidaemia, such that MASLD is considered to be the liver component of the metabolic syndrome (see Chapter 37) and reflected in the recent change in nomenclature. Most patients are asymptomatic, although hepatomegaly may be present. Mild increases in serum aminotransferases and/or γ-GT (with ALT greater than AST) are frequently the sole liver biochemical abnormality or it is often detected on routine abdominal ultrasound examination. Despite the high prevalence of MASLD, well-defined screening recommendations are currently lacking.

Pathogenesis

Histological changes follow a spectrum similar to that of alcohol-induced hepatic injury and range from simple fatty change to fat and inflammation (MASH), fibrosis, and cirrhosis.

Oxidative stress injury and other factors lead to lipid peroxidation in the presence of fatty infiltration and inflammation. Fibrosis may then occur, enhanced by insulin resistance, which induces connective tissue growth factor.

Investigations

- ***Ultrasound*** demonstration of steatosis in the absence of other injurious causes, such as alcohol, usually provides the diagnosis.
- ***Elastography*** (see 'Hepatic stiffness [transient elastography]' section) is frequently used to evaluate the degree of fibrosis but may not be technically possible in the morbidly obese.
- ***Liver biopsy*** allows staging of the disease. Although no definitive guidelines exist, clinicians may perform a biopsy if a diagnosis of MASH or advanced fibrosis is considered likely.

Management

- All MASLD patients require ***lifestyle advice*** aimed at weight loss, increased physical activity, and attention to cardiovascular risk factors. Although evidence is lacking for statins as treatment for MASLD, they are frequently indicated for co-morbid conditions (e.g., diabetes) and can be prescribed to patients with mild-to-moderately abnormal liver enzymes with a low risk of causing adverse liver events. Calorie restriction is recommended, aimed at losing 0.5 to 1 kg per week until target weight is achieved. A Mediterranean diet, high in vegetables, fruits, nuts, beans, cereals, grains, fish, and unsaturated fats such as olive oil, together with a lower intake of meat and dairy foods, may be the healthiest diet for MASLD. A reduction of more than 7% to 9% in body weight has been associated with reduced steatosis, hepatocellular injury, and hepatic inflammation. However, less then 35% of overweight or obese adults with MASLD receiving usual care in a large study in the USA achieved ≥5% weight reduction during a median follow-up of 39 months.
- ***Pioglitazone*** or vitamin E may be used for those with biopsy-proven MASH, when lifestyle intervention has failed. Meta-analysis data demonstrated that pioglitazone significantly improved liver steatosis, inflammation and, to a lesser degree, fibrosis. However, it is associated with weight gain and reports of congestive cardiac failure, bladder cancer, and reduced bone density. Conversely, pioglitazone reduces death, myocardial infarction, and stroke in diabetes patients. The risks and benefits to each patient should be evaluated accordingly and pioglitazone is infrequently used in clinical practice.
- ***Vitamin E*** (800 IU/day) is an antioxidant that improves steatohepatitis. It should be noted that various meta-analyses have raised concerns against the long-term use of vitamin E including increased risk of hemorrhagic stroke and prostate cancer. Again, this is not frequently used.
- Weight loss following ***bariatric surgery*** leads to reduced steatosis, steatohepatitis, and fibrosis. The optimum technique is unknown and long-term data are lacking, although initial concerns about worsening fibrosis have not been borne out. Bariatric surgery should be avoided in those with advanced cirrhosis and portal hypertension, but gastric bypass and sleeve gastrectomy (see Chapter 37) have been shown to achieve weight loss and improve obesity-related co-morbidities in Child–Pugh A cirrhotic patients.
- ***Semaglutide:*** use of this glucagon-like peptide 1 (GLP-1) agonist leads, on average, to more than 10% weight loss. In the UK NICE has recommended the use of semaglutide for obesity, alongside a reduced-calorie diet and increased physical activity to adults who have at least one weight-related co-morbidity and a body mass index (BMI) of at least 35 kg/m^2. A phase II trial of MASH F2/3 patients (i.e., with significant or advanced fibrosis) showed significantly improved MASH resolution but no effect on fibrosis. A further trial in MASH cirrhosis showed no effect on resolution or fibrosis stage.

Hepatocellular carcinoma

The yearly cumulative incidence of HCC is 2.6% in patients with MASH cirrhosis and ultrasound surveillance should therefore be performed 6-monthly. Hyperinsulinaemia and obesity are risk factors for many malignancies and HCC may occur in MAFL/MASH patients in the absence of cirrhosis, although this is rare.

Liver transplantation

MASH cirrhosis is the most rapidly growing indication for transplantation in the USA, the leading cause in females and the second leading cause overall. Survival is comparable to that in other indications, and although recurrence occurs post transplant (4%–25%), it does not appear to have an impact on graft survival. Patients frequently have multiple cardiovascular risk factors that must be managed aggressively. UK guidelines suggest that bariatric surgery could be considered at transplantation for the morbidly obese.

New treatments for MASH

There has been substantial industry investment in pharmacotherapies for MASH, with the pipeline currently crowded at mid-stage. Two companies are in the process of submitting approvals in the USA for use of their drugs in non-cirrhotic MASH, obeticholic acid (FXR agonist, see also 'Bile acid receptors in liver disease' section) and resmetirom (selective thyroid hormone receptor-β agonist; THR-beta agonist). Notably obeticholic acid did not show benefit for MASH cirrhosis. Other agents in trials include elafibranor, a dual peroxisome proliferator-activated receptor (PPAR)α/δ agonist; cenicriviroc, a dual antagonist of C-C chemokine receptor types 2 and 5; selonsertib, an apoptosis signal-regulating kinase 1 inhibitor; and liraglutide, a glucagon-like peptide. It is highly likely that the landscape of treatments for MASH will change substantially over the next few years.

However, it is imperative to tackle obesity at source: for example, through public health campaigns, taxation of high-sugar food, reduction in processed food within the diet, education, increased activity, and vegetable consumption. Prevention of obesity, especially in children and adolescents, will most likely be far superior to the effects of these drugs.

Further reading

Dufour J, Anstee QM, Bugianesi E, et al. Current therapies and new developments in NASH. *Gut* 2022; 71:2123–2134.

CIRRHOSIS

In cirrhosis, the liver architecture is diffusely abnormal and interferes with liver blood flow and function; this leads to the clinical manifestations of portal hypertension and liver failure.

Aetiology

Causes of cirrhosis are shown in **Box 23.14**. Alcohol is currently the most common cause in the West but is likely to be superseded by MASLD; viral infection is the most common aetiological factor worldwide. All patients with cirrhosis must be investigated to exclude treatable causes (e.g., viral or autoimmune hepatitis).

Pathogenesis

Although the liver has a remarkable capacity to adapt to injury through tissue repair, chronic injury results in inflammation, matrix deposition, necrosis, and angiogenesis, all of which lead to fibrosis (**Fig. 23.21**). Liver injury causes necrosis and apoptosis, releasing cell contents and reactive oxygen species (ROS). This activates hepatic stellate cells and tissue macrophages through the CC-chemokine ligand 2–CC-chemokine receptor 2 (CCL2–CCR2) axis. These cells phagocytose necrotic and apoptotic cells and secrete pro-inflammatory mediators, including transforming growth factor-beta (TGF-β); this leads to transdifferentiation of stellate cells to myofibroblasts and platelet-derived growth factor (PDGF), which stimulates myofibroblast proliferation. Macrophages degrade scar matrix by secretion of matrix metalloproteinases (MMPs), but this is inhibited by concurrent myofibroblast and macrophage production of tissue inhibitors of metalloproteinases (TIMPs). This results in progressive matrix deposition and scar accumulation. Increased gut permeability and hepatic lipopolysaccharide–Toll-like receptor 4 (LPS–TLR4) signalling also promotes fibrogenesis. Repetitive or chronic injury and inflammation perpetuate this process.

If the cause of fibrosis is eliminated (e.g., through treatment of viral hepatitis), resolution (complete reversal to near-normal liver architecture) of early fibrosis can occur. In cirrhosis, regression (improvement, not reversal) occurs, which improves clinical outcomes. Antifibrotic therapies are an active research area but currently liver transplantation is the only available treatment for liver failure.

? Box 23.14 Causes of cirrhosis

Common

- Alcohol
- Hepatitis B ± D
- Hepatitis C
- Non-alcoholic fatty liver disease (NAFLD)

Others

- Primary biliary cholangitis
- Secondary biliary cirrhosis
- Autoimmune hepatitis
- Hereditary haemochromatosis
- Hepatic venous congestion
- Budd–Chiari syndrome
- Wilson's disease
- Drugs (e.g., methotrexate)
- α_1-antitrypsin deficiency
- Cystic fibrosis
- Galactosaemia
- Glycogen storage disease
- Veno-occlusive disease
- Idiopathic (cryptogenic)
- ? Other viruses

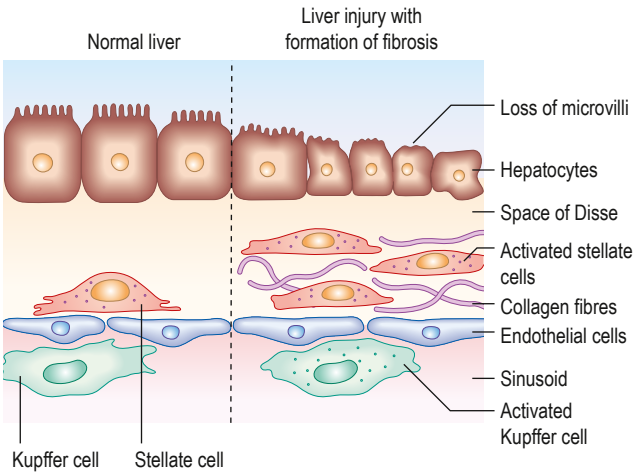


Fig. 23.21 Pathogenesis of fibrosis. The normal liver is shown on the left. Activation of the stellate cell is followed by proliferation of fibroblasts and deposition of collagen.

Pathology

The characteristic features of cirrhosis are regenerating nodules separated by fibrous septa, and loss of lobular architecture within the nodules (**Fig. 23.22A–C**). Two types are described:

- ***Micronodular cirrhosis.*** Regenerating nodules are usually less than 3 mm in size with uniform involvement of the liver; often caused by alcohol or biliary tract disease.
- ***Macronodular cirrhosis.*** The nodules are of variable size and normal acini may be seen within larger nodules; often caused by chronic viral hepatitis.

A mixed picture with small and large nodules occurs occasionally.

Symptoms and signs are described in the figure in 'Clinical skills' section.

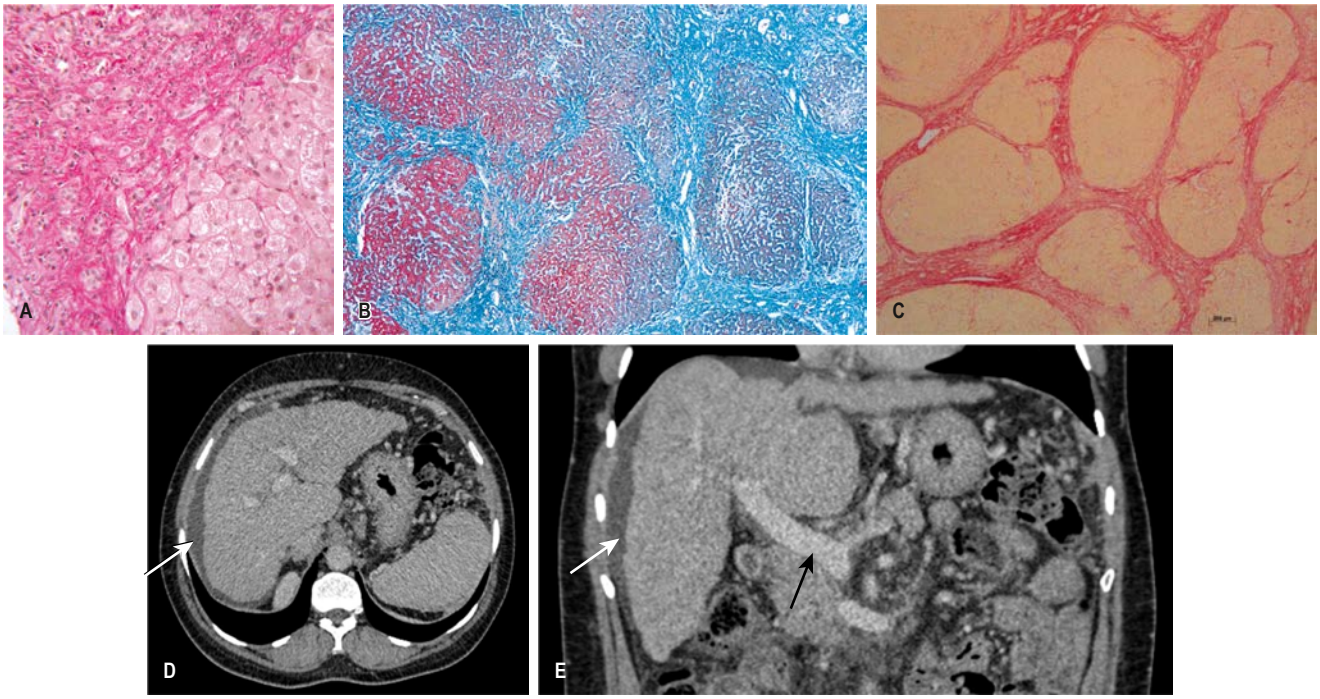


Fig. 23.22 Pathology of cirrhosis. **(A)** Haematoxylin and eosin histological view (×10). **(B)** Histological appearance (Giemsa stain) showing nodules of liver tissue of varying size, surrounded by fibrosis. **(C)** Picrosirius red stain of collagen used for morphometric evaluation of fibrosis. **(D)** Axial CT image showing an irregular contour of the liver consistent with cirrhosis. There are splenic, gastric and oesophageal varices and a small volume of ascites *(arrowed)*. **(E)** Coronal CT image of the liver demonstrates an irregular contour to the liver consistent with cirrhosis. The portal vein is patent *(black arrow)*. There is small-volume ascites *(white arrow)*.

Investigations

Investigations aim to assess the severity and type of liver disease.

Severity

- ***Liver function.*** Serum albumin and PT are the best indicators of liver function; the outlook is poor if the albumin level is less than 25 g/L and the degree to which the PT is prolonged is commensurate with disease severity.
- ***Liver biochemistry.*** This may be normal, depending on the severity of disease. In most cases, there is a slight elevation in the serum ALP and serum aminotransferases. In decompensated cirrhosis, all biochemistry is deranged.
- ***Serum electrolytes.*** Low sodium occurs in severe liver disease due to a defect in free water clearance or excess diuretic therapy.
- ***Serum creatinine.*** An elevated concentration of more than 130 μmol/L is a marker of poor prognosis.
- ***Biomarkers.*** In the enhanced liver fibrosis (ELF) test, which assesses fibrosis, a value of less than 7.7 indicates none to mild, 7.7 to 9.8 moderate, and 9.8 or more indicates severe disease.

In addition, a serum α-fetoprotein of more than >200 ng/mL is strongly suggestive of HCC.

Type

This can be determined by:

- viral markers
- serum autoantibodies
- serum immunoglobulins
- iron indices and ferritin
- copper and caeruloplasmin (see 'Wilson's disease [progressive hepatolenticular degeneration]/Investigations' section
- α_1-antitrypsin (see 'Alpha$_1$-antitrypsin deficiency' section)
- genetic markers.

Serum copper and serum α_1-antitrypsin should always be measured in young people with cirrhosis. Total iron-binding capacity (TIBC) and ferritin should be measured to exclude hereditary haemochromatosis; genetic markers are also available (see 'Hereditary haemochromatosis' section).

Imaging

- ***Ultrasound examination*** can demonstrate changes in the size and shape of the liver. Fatty change and fibrosis produce a diffusely increased echogenicity. In established cirrhosis, there may be marginal nodularity of the liver surface and distortion of the arterial vascular architecture. The patency of the portal and hepatic veins can be evaluated. Ultrasound is useful for detecting HCC.
- ***Transient elastography*** is increasingly used to avoid liver biopsy (see 'Hepatic stiffness [transient elastography]' section). Technical limitations preclude its use in patients with ascites or morbid obesity but it is suitable for most. If the reading is high (>25), portal hypertension is likely, and some experts recommend that endoscopy to identify varices should be restricted to patients with high fibroscan scores.
- ***CT scanning*** (see 'Computed tomography examination' section) is also helpful. **Fig. 23.22D and E** shows hepatosplenomegaly and the dilated collaterals seen in chronic liver disease. Arterial phase contrast-enhanced scans are useful for detecting HCC.
- ***Endoscopy*** is performed for the detection and treatment of varices and portal hypertensive gastropathy. If identified, varices

Box 23.15 Poor prognostic indicators in cirrhosis

Blood tests
- Low albumin (<28 g/L)
- Low serum sodium (<125 mmol/L)
- Prolonged prothrombin time >6 sec above normal value
- Raised creatinine >130 μmol/L

Clinical
- Persistent jaundice
- Failure of response to therapy
- Ascites
- Haemorrhage from varices, particularly with poor liver function
- Neuropsychiatric complications developing with progressive liver failure
- Small liver
- Persistent hypotension
- Aetiology (e.g., alcoholic cirrhosis, if the patient continues drinking)

can be treated by banding or injection sclerotherapy or medically with non-selective beta blockers. Colonoscopy is occasionally carried out for portal hypertensive colopathy.

- ***MRI scanning*** is useful in the diagnosis of both malignant and benign tumours such as haemangiomas, and MR cholangiography can demonstrate the biliary tree anatomy and MR angiography vascular anatomy.

Liver biopsy

This remains the 'gold standard' for confirming the type and severity of liver disease. Adequate samples, in terms of length and number of complete portal tracts, are necessary for diagnosis and staging/grading of chronic viral hepatitis; in macronodular cirrhosis, the core of liver may fragment, causing sampling errors. Immunocytochemical stains can identify viruses, bile ducts, angiogenic structures, and oncogenic markers. Chemical measurement of iron and copper is necessary to confirm a diagnosis of iron overload or Wilson's disease. Digital image analysis of picrosirius red staining can be used to quantitate collagen in biopsy specimens (see **Fig. 23.22C**).

Management

Patients should undergo 6-monthly ultrasound to screen for the early development of HCC (see 'Hepatocellular carcinoma' section), as all therapeutic strategies work best with small, single tumours.

Treatment of the underlying cause may arrest or improve cirrhosis for viral or autoimmune hepatitis. Patients with compensated cirrhosis should lead a normal life. Those at risk should receive hepatitis A and B vaccinations. The only dietary restriction is reduction of salt intake (≤2 g sodium per day, equivalent to ≤5 g dietary salt). Alcohol should be avoided for any cause of cirrhosis, as should non-steroidal anti-inflammatory drugs (NSAIDs), which may precipitate gastrointestinal bleeding or renal impairment.

For patients with decompensated cirrhosis, management focuses on prevention and treatment of its complications, such as hepatic encephalopathy, ascites, and variceal bleeding, described elsewhere in this chapter.

Prognosis

Prognosis is extremely variable. Ascites is the most common complication of cirrhosis, and this clinical presentation represents a watershed moment with patients' median survival falling from greater than 12 years for compensated cirrhosis to approximately 2 years. Poor prognostic indicators are shown in **Box 23.15**.

There are a number of prognostic classifications with the most commonly used, the Child–Pugh score (A, B, and C; **Box 23.16**) and the Model for End-stage Liver Disease (MELD; **Box 23.17**). MELD focuses on serum bilirubin, creatinine, and the INR and is widely used as a predictor of mortality in patients awaiting liver transplantation. Modifications of the MELD score (e.g., UKELD) are used in some countries.

Box 23.16 Scoring systems in cirrhosis: modified Child–Pugh classification

Score	1	2	3
Ascites	None	Mild	Moderate/severe
Encephalopathy	None	Mild	Marked
Bilirubin (μmol/L)	<34	34–50	>50
Albumin (g/L)	>35	28–35	<28
Prothrombin time (sec over normal)	<4	4–6	>6

Add above scores for your patient for survival figures below:
Child's A (<7): 82% survival at 1 year, 45% at 5 years, 25% at 10 years
Child's B (7–9): 62% survival at 1 year, 20% at 5 years, 7% at 10 years
Child's C (10+): 42% survival at 1 year, 20% at 5 years, 0% at 10 years

Box 23.17 Scoring systems in cirrhosis: Model of End-stage Liver Disease (MELD)

$$3.8 \times \mathrm{LN}(\text{bilirubin in mg/dL}) + 9.6 \times \mathrm{LN}(\text{creatinine in mg/dL}) + 11.2 \times \mathrm{LN}(\mathrm{INR}) + 6.4^{a}$$

To convert:
- bilirubin from μmol/L to mg/dL, divide by 17
- creatinine from μmol/L to mg/dL, divide by 88.4

MELD scores (with no complications):
- Score <10: 97% survival at 1 year
- Score 30–40: 70% at 1 year

[a]Check online.
INR, International normalized ratio; *LN*, natural logarithm.

Acute-on-chronic liver failure

Acute-on-chronic liver failure (ACLF) refers to the condition of patients hospitalized for an acute complication of cirrhosis accompanied by organ failure(s); it has a high short-term mortality. ACLF is considered distinct from traditional decompensated cirrhosis, based not only on the presence of organ failure(s) and high mortality rate, but also on younger age, alcohol aetiology, higher prevalence of certain precipitating events (e.g., bacterial infections or active alcohol excess), and a higher level of systemic inflammation. It is estimated to occur in a third of patients hospitalized with decompensated cirrhosis.

Liver assist devices

As the liver has great potential to regenerate, extracorporeal liver support devices may allow patients with liver failure to be bridged to recovery or transplantation. Current approaches include the use of biological devices that contain hepatocytes and those that function as detoxification devices, and artificial liver support systems. However, no device has yet shown benefit in large multicentre randomized trials. A recent phase II study using the DIALIVE device showed promise.

Gut–liver axis

The liver is exposed to gut-derived bacterial components that have little consequence in health, as an effective gut barrier limits the transport of bacterial components to the liver. In advanced

liver disease the intestinal barrier function is compromised due to changes in gut motility, an increase in intestinal permeability, and suppression of gut immunological functions. This leads to bacteria and bacterial components (e.g., endotoxin, lipopolysaccharide) entering the portal circulation and being transported to the liver, activating Toll-like receptors and producing an inflammatory response. This cross-talk between the intestinal microbiota and the liver is referred to as the gut–liver axis and is seen as a key pathophysiological mechanism in the progression of liver disease and development of the complications of cirrhosis. Antibiotics and non-selective beta-blockers intercept the gut–liver axis by ***blocking bacterial translocation***, which is likely to account for their beneficial effects in reducing portal pressure, variceal haemorrhage, and spontaneous bacterial peritonitis (SBP). However, absorbable antibiotics will lead to the selection of resistant bacteria. Rifaximin, a poorly absorbed antibiotic used to treat hepatic encephalopathy, specifically affects the gut flora and has a low risk for inducing resistance; it may therefore have a role for this indication.

Further reading

Agarwal B, Cañizares RB, Faouzi Saliba F. Randomized-controlled trial of the DIALIVE liver dialysis device vs. standard of care in patients with acute-on-chronic liver failure. *J Hepatol* 2023; S0168-8278(23)00188-5.
Campana L, John P Iredale JP. Regression of liver fibrosis. *Semin Liver Dis* 2017; 37(1):1–10.
Moreau R, Jalan R, Gines P et al. and the CANONIC Study Investigators of the EASL–CLIF Consortium. Acute-on-chronic liver failure is a distinct syndrome that develops in patients with acute decompensation of cirrhosis. *Gastroenterology* 2013; 144:1426–1437.
Ponziani FR, Zocco MA, Cerrito L, Gasbarrini A, Pompili M. Bacterial translocation in patients with liver cirrhosis: physiology, clinical consequences, and practical implications. *Expert Rev Gastroenterol Hepatol* 2018; 12(7):641–656.

Liver transplantation

This is the only established treatment for advanced liver disease. Shortage of donors is a major problem in all high income countries and in some, such as Japan or South Korea, living-related donors form the majority.

Indications

These include:

- ***acute liver disease:*** acute hepatic failure of any cause (see 'Acute hepatic failure' section)
- ***chronic liver disease:*** usually for complications of cirrhosis that are no longer responsive to therapy.

The timing of transplantation depends on donor availability. All patients with end-stage cirrhosis (Child's grade C; MELD score ≥20; UKELD score ≥49) and those with debilitating symptoms should be referred to a transplant centre. In addition, specific extrahepatic complications of cirrhosis, even with preserved liver function, such as hepatopulmonary syndrome and porto-pulmonary hypertension, can be reversed by transplantation.

- ***Primary biliary cholangitis.*** Patients should be transplanted when their serum bilirubin persistently exceeds 100 μmol/L or when they have intractable pruritus.
- ***Chronic hepatitis B if HBV DNA-negative*** or levels are falling with effective therapy. Following transplantation, recurrence of hepatitis is prevented by hepatitis B immunoglobulin and nucleoside analogues in combination.
- ***Chronic hepatitis C.*** Following the introduction of the directly acting antiviral drugs, rates of transplantation for HCV have fallen substantially; for example, a reduction from 10.5% in 2013 to 4.7% in 2016 was seen in the UK.
- ***Autoimmune hepatitis.*** In patients who have failed to respond to medical treatment; the disease can recur.
- ***Alcoholic liver disease.*** Well-motivated patients who have stopped drinking without improvement of liver disease are offered a transplant, with frequent counselling. It has been shown that transplantation may represent life-saving treatment in patients with severe alcoholic hepatitis who are not responding to medical therapy, however, further studies are awaited. Early liver transplantation is an emerging therapy for acute severe alcoholic hepatitis, and multiple small studies have demonstrated that 'highly selected' patients have comparable 1- and 3-year survival to those with chronic liver disease following transplantation. However, not all caregivers agree that the usual expectation of demonstrated alcohol abstinence prior to transplantation should be waived and further data are needed.
- ***Primary metabolic disorders.*** Examples are Wilson's disease, hereditary haemochromatosis, and α_1-antitrypsin deficiency.
- ***MASH cirrhosis.*** In view of the increasing rates of obesity, this is highly likely to become the most frequent indication for transplantation.
- ***Other conditions*** include PSC, polycystic liver disease, and metabolic diseases such as primary oxaluria.

Contraindications

Absolute contraindications include active sepsis outside the hepatobiliary tree, malignancy outside the liver, liver metastases (except neuroendocrine), and a lack of psychological commitment on the part of the patient.

Relative contraindications are mainly anatomical considerations that make surgery more difficult, such as extensive splanchnic venous thrombosis. With exceptions, patients aged 70 years or over are not usually given a transplant. In HCC the recurrence rate is high unless there are fewer than three small (<3 cm) lesions or a solitary nodule of less than 5 cm (Milan criteria).

Preparation for surgery

Pre-transplant work-up includes confirmation of the diagnosis, ultrasound and cross-sectional imaging, and radiological demonstration of the hepatic arterial and biliary trees, as well as assessment of cardiorespiratory and renal status. In view of the ethical and financial implications of transplantation, regular psychosocial and possibly psychiatric support is vital.

The donor should be ABO-compatible (HLA matching is not necessary) and have no evidence of active sepsis, malignancy, or HIV, HBV, or HCV infection. Younger donors (<50 years) lead to better graft function. The donor liver is cooled and stored on ice; its preservation time may be up to 20 hours. The recipient operation takes approximately 8 hours and rarely requires a large blood transfusion. Deceased donor livers (from donors after brain or circulatory death) may consist of whole graft, split grafts (for two recipients), or reduced grafts. Live donors may be healthy individuals or patients with, for example, familial amyloid polyneuropathy, whose livers can be transplanted into others when they themselves receive a donor liver (so-called domino transplant). The mortality of right lobe donors is between 1 in 200 and 1 in 400. Recent advances in devices for storing donated liver prior to transplantation may allow livers to be used that were previously considered unsuitable.

Box 23.18 Complications and effects of cirrhosis

- Portal hypertension and gastrointestinal haemorrhage
- Ascites
- Portosystemic encephalopathy
- Hepatocellular carcinoma
- Bacteraemias, infections
- Renal failure
- Hepatopulmonary syndrome

The operative mortality of liver transplant recipients is very low and most postoperative deaths occur in the first 3 months. Sepsis and haemorrhage can be serious complications. Opportunistic infections occur owing to immunosuppression. Various immunosuppressive agents have been used: tacrolimus—alone or in combination with azathioprine or mycophenolate mofetil—steroids, sirolimus, and micro-emulsified ciclosporin are the most common.

Rejection

- ***Acute or cellular rejection*** usually occurs 5 to 10 days post transplant; it can be asymptomatic or there may be a fever. On biopsy, a pleomorphic portal infiltrate is seen with prominent eosinophils, bile duct damage, and endothelialitis of the blood vessels. This responds well to immunosuppressive therapy.
- ***Chronic ductopenic rejection*** is seen between 6 weeks and 9 months post transplant, with disappearing bile ducts (vanishing bile duct syndrome, VBDS), and an arteriopathy with narrowing and occlusion of the arteries. Early ductopenic rejection is rarely reversed by immunosuppression and often requires retransplantation.
- ***Graft-versus-host disease*** is extremely rare.

Prognosis

Elective liver transplantation in low-risk patients has a 90% 1-year survival and a 70% to 85% 5-year survival. Patients require lifelong immunosuppression, although doses can be reduced over time without significant problems. Future strategies to reduce immunosuppression requirements after transplant may include infusion of autologous regulatory T cells. HCV cirrhosis, PSC, and HCC are conditions in which long-term survival after transplantation is compromised by disease recurrence.

Further reading

Burke NT, Maurice JB, Nasralla D, Potts J, Westbrook R. Recent advances in liver transplantation. *Frontline Gastroenterol.* 2021; 13(1):57–63.

Complications and effects of cirrhosis

These are shown in **Box 23.18**.

Portal hypertension

The portal vein is formed by the union of the superior mesenteric and splenic veins. Normal pressure is 5 to 8 mmHg with only a small gradient across the liver to the hepatic vein, in which blood is returned to the heart via the inferior vena cava. Portal hypertension is classified according to the site of obstruction:

- ***pre-hepatic***—blockage of the portal vein before the liver
- ***intrahepatic***—distortion of the liver architecture, either pre-sinusoidal (e.g., schistosomiasis) or post-sinusoidal (e.g., cirrhosis)

Box 23.19 Causes of portal hypertension

Prehepatic
- Portal vein thrombosis

Intrahepatic

Pre-sinusoidal
- Schistosomiasis; sarcoidosis
- Primary biliary cholangitis

Sinusoidal
- Cirrhosis (e.g., alcoholic)
- Partial nodular transformation
- Congenital hepatic fibrosis

Post-sinusoidal
- Veno-occlusive disease
- Budd–Chiari syndrome

Post-hepatic
- Right heart failure (rare)
- Constrictive pericarditis
- Inferior vena cava obstruction

- ***post-hepatic***—venous blockage outside the liver (rare).

As portal pressure rises above 10 to 12 mmHg, the compliant venous system dilates and collaterals form within the systemic venous system. The main sites of collaterals are the gastro-oesophageal junction, rectum, left renal vein, diaphragm, retroperitoneum, and the anterior abdominal wall via the umbilical vein.

The collaterals at the gastro-oesophageal junction (varices) are superficial and are prone to rupture whereas portosystemic anastomoses at other sites rarely cause symptoms. Rectal varices are frequently found (30%) if looked for and can be differentiated from haemorrhoids, which are lower in the anal canal. The microvasculature of the gut becomes congested, giving rise to portal hypertensive gastropathy and colopathy, in which there is punctate erythema and erosions, which can bleed.

Pathophysiology

Following liver injury and fibrogenesis, the contraction of activated myofibroblasts (mediated by endothelin, nitric oxide, and prostaglandins) contributes to increased resistance to blood flow. This increased resistance leads to portal hypertension and opening of portosystemic anastomoses in both pre-cirrhotic and cirrhotic livers. Neo-angiogenesis also occurs. The hyperdynamic circulation of cirrhosis (caused by nitric oxide, cannabinoids, and glucagon) leads to peripheral and splanchnic vasodilation. This, combined with plasma volume expansion due to sodium retention (see 'Ascites' section), leads to portal hypertension.

Aetiology

The most common cause is cirrhosis (**Box 23.19**). Others are described here.

Pre-hepatic causes

Extrahepatic blockage due to portal vein thrombosis can be caused by congenital portal venous abnormalities, neonatal sepsis of the umbilical vein, or inherited prothrombotic conditions, such as factor V Leiden or primary myeloproliferative disorders with or without *JAK2* mutations (see Chapter 38).

Patients present with gastrointestinal bleeding, often at a young age. They have normal liver function and prognosis following bleeding is therefore excellent.

The portal vein blockage can be identified by ultrasound with Doppler imaging; CT is also used.

Treatment for variceal bleeding is usually repeated endoscopic therapy or prophylaxis with non-selective beta-blockade. Splenectomy is performed only if there is isolated splenic vein thrombosis. Anticoagulation prevents further thrombosis and intestinal infarction and does not increase the risk of bleeding.

Intrahepatic causes

Cirrhosis is by far the most common cause but others include:

- ***Non-cirrhotic portal hypertension*** is characterized by mild portal tract fibrosis on liver histology. The aetiology is unknown but arsenic, vinyl chloride, anti-retroviral therapy, and other toxins have been implicated. A similar disease is frequently found in India. The liver lesion does not progress and prognosis is good.
- ***Schistosomiasis*** with extensive pipe-stem fibrosis occurs in endemic areas such as Egypt and Brazil. Often there is concomitant liver disease, such as HCV infection, which was transmitted by non-sterile equipment.
- ***Other causes*** include congenital hepatic fibrosis, and nodular regenerative hyperplasia and partial nodular transformation.

Post-hepatic causes

Prolonged severe heart failure with tricuspid incompetence or constrictive pericarditis can cause portal hypertension. The Budd–Chiari syndrome is described in the 'Budd–Chiari syndrome' section.

Clinical features

Patients are often asymptomatic, the only clinical evidence being splenomegaly, although features of chronic liver disease may exist (see the figure in 'Clinical skills' section).

Variceal haemorrhage

Approximately 90% of patients with cirrhosis will develop gastro-oesophageal varices over 10 years but only one-third of these will bleed. Bleeding is likely to occur with large varices, or in those with red signs at endoscopy, and in severe liver disease.

Management

Management can be divided into:

- the active bleeding episode
- prevention of rebleeding
- prophylactic measures to prevent the first haemorrhage.

Despite the therapeutic techniques available, prognosis ultimately depends on the severity of the underlying liver disease with overall 6-week mortality from variceal haemorrhage 15% to 25%, reaching 50% in Child's grade C.

Initial management of acute variceal bleeding

See **Fig. 23.23**, and also the general management of gastrointestinal haemorrhage in Chapter 11.

Resuscitation

- ***Assess*** the patient: pulse, blood pressure, and conscious state.
- Insert a ***large-bore intravenous line*** and obtain ***blood*** for group and crossmatching, haemoglobin, PT/INR, serum creatinine and electrolytes, liver biochemistry, and blood cultures.
- ***Restore blood volume*** with plasma expanders or, if possible, blood transfusion. See the treatment of shock for more detail (see Chapter 13). Prompt correction, but not over-correction, of hypovolaemia is necessary in cirrhosis patients, as their baroreceptor reflexes are diminished. A target haemoglobin of 80 g/L is sufficient, and this lessens the likelihood of early rebleeding.
- Perform an ***ascitic tap***.
- Monitor for alcohol withdrawal and treat with lorazepam or oxazepam (the half-life of chlordiazepoxide will increase in advanced liver disease, amplifying the risk of adverse events). Give intravenous thiamine.
- Start prophylactic antibiotics. These treat and prevent infection and reduce early rebleeding and mortality, although most of these studies were performed many years ago. Given the improvement of endoscopic techniques and the increasing burden of anti-microbial resistance, universal prophylaxis may no longer be required and patients that recover quickly should have antibiotics stopped at 48 or 72 hours unless there is evidence of ongoing infection.

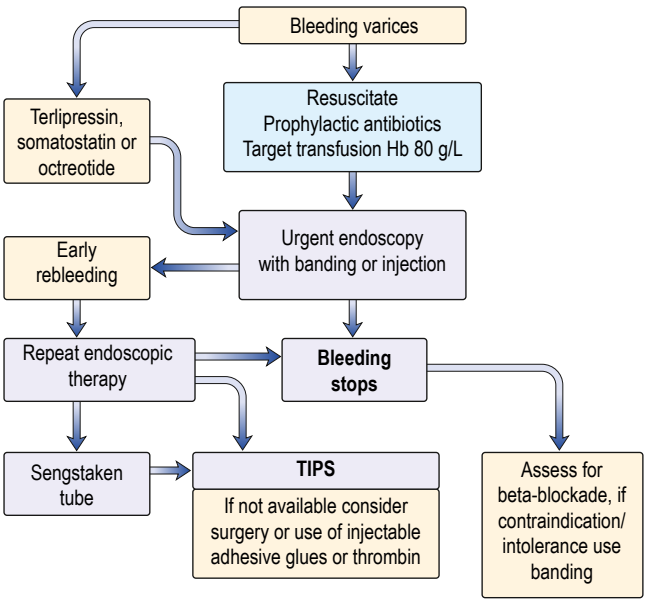


Fig. 23.23 Management of gastrointestinal haemorrhage due to oesophageal varices. *Hb,* Haemoglobin; *TIPS,* transjugular intrahepatic portosystemic shunt.

Urgent endoscopy

Endoscopy (**Fig. 23.24**) should be performed to confirm the diagnosis and to exclude bleeding from other sites (e.g., gastric ulceration) and portal hypertensive gastropathy/*g*astric *a*ntral *v*ascular *e*ctasia (GAVE). The latter describes chronic gastric congestion, punctate erythema, and gastric erosions, which may contribute to chronic anaemia. Portal hypertensive gastropathy and GAVE are distinct entities; management of portal hypertensive gastropathy is centred on reduction in portal pressures with beta-blockers, whereas treatment of GAVE is endoscopic and uses various ablative techniques.

Variceal banding

Banding of oesophageal varices is performed by mounting a band on the tip of the endoscope, sucking the varix into the end of the scope, and dislodging the band over the varix using a trip-wire mechanism.

Between 15% and 20% of bleeding comes from gastric varices, which are associated with a greater mortality (up to 40%); endoscopic injection of cyanoacrylate is the best treatment.

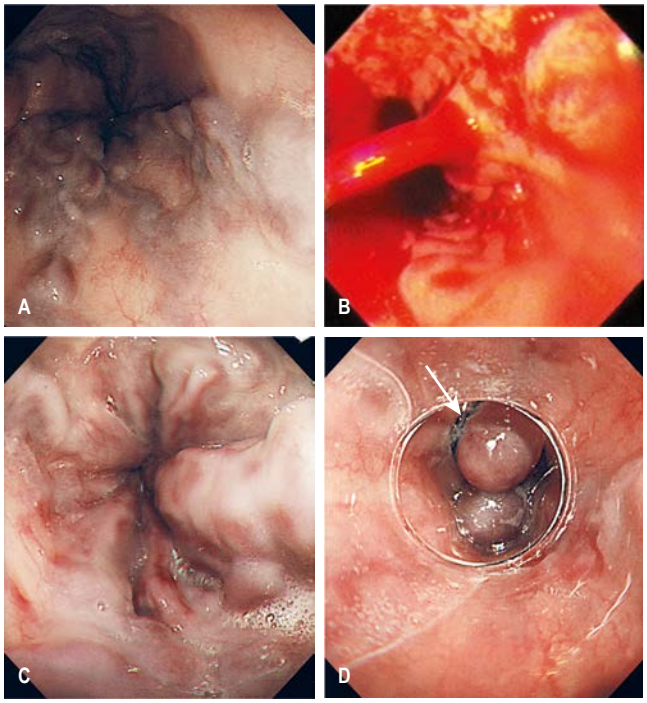


Fig. 23.24 Endoscopic views of oesophageal varices. **(A)** Gross varices. **(B)** Blood spurting from a varix. **(C)** Following a recent bleed. **(D)** Varices with a band in place *(arrowed)*. *(A, C, and D, Courtesy of Dr Peter Fairclough.)*

Overall, haemostasis is achieved in 80% to 90% of patients. Best practice is to perform the endoscopy with the patient under general anaesthetic to provide appropriate airway support, and with critical care support.

Other measures

Vasoconstrictor therapy

This is used to restrict portal inflow by splanchnic arterial constriction and has shown benefit when used in combination with endoscopic techniques.

- ***Terlipressin.*** This is the only vasoconstrictor shown to reduce mortality. The dose is 2 mg every 4 to 6 hours, with many units favouring the longer dosing interval to reduce adverse effects including painful vasoconstriction in the hands and feet. Treatment is typically stopped once satisfactory haemostasis has been achieved and longer courses have not been associated with a survival benefit. Terlipressin should not be given to patients with ischaemic heart disease. The patient may complain of abdominal colic, and may defecate and have facial pallor owing to generalized vasoconstriction.
- ***Somatostatin.*** This reduces bleeding with few side-effects and is given as an initial 250 µg bolus followed by a continuous infusion of 250 µg/hour. Somatostatin is reserved for patients with contraindications to terlipressin.

Balloon tamponade

Balloon tamponade is used if endoscopic therapy has failed or if there is exsanguinating haemorrhage. The usual balloon tube is a four-lumen Sengstaken–Blakemore, which should be left in place for no more than 12 hours and removed in the endoscopy room prior to follow-up endoscopy. The tube is passed into the stomach and the gastric balloon inflated with air and pulled back. It should be positioned in close apposition to the gastro-oesophageal junction to prevent the cephalad variceal blood flow to the bleeding point. The oesophageal balloon should be inflated only if bleeding is not controlled by the gastric balloon alone.

Haemostasis is achieved in up to 90%. However, the balloon may cause serious complications, such as aspiration pneumonia, oesophageal rupture, and mucosal ulceration. The procedure is also very unpleasant for the patient.

A self-expanding covered metal stent (Danis), which has a wire loop to enable removal and is introduced orally or endoscopically, can be placed over the varices. This is effective and has the advantages that it does not impair swallowing, cannot be removed by uncooperative patients, and allows post-endoscopic investigation. The stent is removed 7 days after insertion. Evidence supports that the Danis stent improves the short-term control of bleeding compared with a balloon tamponade when there is endoscopic failure and it can be left in place for longer, allowing time to stabilize the patient.

Additional management of the acute episode

- ***Measures to prevent encephalopathy.*** Portosystemic encephalopathy (PSE) can be precipitated by a large bleed (blood contains protein). Management is as outlined on in the 'Portosystemic encephalopathy/Management' section.
- ***Nursing.*** Patients require high-dependency/intensive care nursing. They should remain nil by mouth until bleeding has stopped.
- ***Reduction in acid secretion.*** PPIs are widely used but may cause increased incidence of *Clostridium difficile* and other infections. Sucralfate 1 g four times daily can reduce oesophageal ulceration following endoscopic therapy.

Management of an acute rebleed

Rebleeding occurs in approximately 15% to 20% within 5 days. The source should be established by endoscopy and this is sometimes related to ulceration or slippage of a ligation band. Endoscopy should be performed once only to control rebleeding. If haemostasis cannot be achieved, then transjugular intrahepatic portosystemic shunting (TIPS) will be necessary.

Transjugular portosystemic shunting

TIPS is performed when bleeding cannot be controlled either acutely or following a rebleed. Under X-ray guidance, a guidewire is passed from the jugular vein into the liver and the portal vein. After balloon expansion of the tract between the hepatic and portal veins, an expandable, covered metal shunt is placed over the wire to form a channel between the systemic and portal venous systems. It reduces the hepatic sinusoidal and portal vein pressure by creating a total shunt. There is an increased risk of portal-systemic encephalopathy. Stent stenosis or thrombosis is far less frequent now that 'covered' stents are used, as opposed to the 'bare metal' stents employed previously. Collaterals arising from the splenic or portal veins can be selectively embolized.

A recent observational study in 671 patients validated the survival benefit of 'pre-emptive' TIPS (within 72 hours of admission) in those with cirrhosis and acute variceal haemorrhage, and at high risk for treatment failure. Most of the benefit was seen in patients with Child–Pugh class C disease. However, widespread implementation of this approach represents a significant challenge for many healthcare systems, and despite strong supporting evidence, most centres have not adopted it at present. A randomized controlled trial will shortly begin in the UK to test this approach.

Emergency surgery

This is performed very rarely when other measures fail. Oesophageal transection and ligation of the feeding vessels to the bleeding varices or acute portosystemic shunt surgery is performed.

Prevention of recurrent variceal bleeding (secondary prophylaxis)

The risk of bleeding recurring without prophylaxis is 60% to 80% over a 2-year period, with an approximate mortality of 20% per episode.

Prophylactic long-term measures

Non-selective beta-blockade

Oral propranolol or carvedilol to reduce the resting pulse rate by 25% decreases portal pressure. Portal inflow is reduced by a decrease in cardiac output (β_1) and by blockade of β_2 vasodilator receptors on the splanchnic arteries, leaving an unopposed vasoconstrictor effect. Significant reduction of hepatic venous pressure gradient (HVPG, measured by hepatic vein catheterization) is associated with very low rates of rebleeding, particularly if below 12 mmHg. Data have emerged demonstrating that additional α_1-adrenergic blockade, causing vasodilation with carvedilol, may increase the number of patients with a reduction in HVPG compared to propranolol.

Endoscopic treatment

Repeated courses of banding at 2 to 4 weekly intervals lead to obliteration of varices and markedly reduce rebleeding. Between 30% and 40% of varices return per year, so follow-up endoscopy should be performed at 1 to 3 months after obliteration and then every 6 to 12 months. Complications of banding include oesophageal ulceration, mediastinitis, and rarely, strictures.

Combination therapy reduces overall bleeding compared to endoscopic therapy alone but with no overall improvement in mortality. A pragmatic approach is therefore to give combination therapy to those who can tolerate beta-blockade, and band ligation alone to those who cannot.

Surgery

- ***Surgical portosystemic shunting*** is associated with an extremely low risk of rebleeding and is used if TIPS is not available. Hepatic encephalopathy is a significant complication. The 'shunts' are usually an end-to-side portocaval anastomosis or a selective distal splenorenal shunt (Warren shunt).
- ***Devascularization procedures***, including oesophageal transection, do not produce encephalopathy and can be used when there is splanchnic venous thrombosis.
- ***Liver transplantation*** (see 'Liver transplantation' section) is the best option when there is poor liver function.

Prophylactic measures (primary prophylaxis)

Patients with cirrhosis and significant varices that have not bled should be prescribed non-selective beta-blockers. This reduces the chances of upper gastrointestinal bleeding by approximately 50%, may increase survival and is cost-effective. If there are contraindications or the patient cannot tolerate beta-blockers, variceal banding is an option. A large-scale trial comparing the efficacy of banding versus carvedilol for primary prophylaxis was unable to complete due to disruption caused by the COVID-19 pandemic.

Box 23.20 Serum–ascites albumin gradient

High serum–ascites albumin gradient (>11 g/L)

- Portal hypertension of any cause
  - Pre-hepatic e.g., portal vein thrombosis
  - Intrahepatic e.g., cirrhosis
  - Post-hepatic e.g., constrictive pericarditis
  - See **Box 23.19** for complete list

Low serum–ascites albumin gradient (<11 g/L)

- Ascites without portal hypertension
  - Peritoneal carcinomatosis
  - Peritoneal tuberculosis
  - Pancreatitis
  - Nephrotic syndrome

Ascites

Ascites, fluid within the peritoneal cavity, is a common complication of cirrhosis. Several factors underlie its pathogenesis:

- ***Sodium and water retention*** results from peripheral arterial vasodilation (secondary to increased nitric oxide, atrial natriuretic peptide, and prostaglandins), which causes a reduction in the effective blood volume. This reduction activates the sympathetic nervous system and the renin–angiotensin system, promoting salt and water retention (see **Fig. 27.5**).
- ***Portal hypertension*** exerts a local hydrostatic pressure, leading to increased hepatic and splanchnic production of lymph, and transudation of fluid into the peritoneal cavity.
- ***Low serum albumin*** (due to poor liver function) may further contribute by reducing plasma oncotic pressure.

In patients with ascites, urine sodium excretion rarely exceeds 5 mmol in 24 hours. Loss of sodium from extrarenal sites accounts for approximately 30 mmol in 24 hours. Under these circumstances, a normal daily sodium intake of 120 to 200 mmol results in a positive sodium balance of approximately 90 to 170 mmol in 24 hours (equivalent to 600–1300 mL of fluid retained).

Clinical features

Abdominal swelling may develop over days or several weeks. Precipitating factors include continued excessive alcohol consumption, infection/sepsis, development of an HCC, or splanchnic vein thrombosis. Mild abdominal pain and discomfort are common but, if more severe, should raise the suspicion of SBP (see later). Respiratory distress and difficulty eating accompany tense ascites.

The presence of fluid is confirmed clinically by demonstrating shifting dullness. Many patients also have peripheral oedema. A pleural effusion (usually right-sided) may be found and arises from the passage of ascites through congenital diaphragmatic defects.

Investigations

A diagnostic aspiration of 10 to 20 mL of fluid should be obtained for:

- ***Cell count.*** A neutrophil count of more than 250 cells/mm^3 is indicative of an underlying (usually spontaneous) bacterial peritonitis.
- ***Gram stain and culture.***
- ***Protein measurement.*** Serum–ascites albumin gradient (SAAG)—(serum albumin)—(ascitic fluid albumin) can be calculated to help distinguish between ascites due to portal hypertension (SAAG > 11 g/L) or ascites associated with non-liver disease–related abnormalities of the peritoneum, such as neoplasia (SAAG < 11 g/L) (**Box 23.20**). A high SAAG reflects relative

concentration of albumin in the circulation as fluid crosses the peritoneal membrane due to increased hydrostatic pressure resulting from portal hypertension; ascites due to inflammatory or malignant diseases of the peritoneum is not associated with this phenomenon, and SAAG is thus low.

- ***Cytology.*** A search should be made for malignant cells.
- ***Amylase.*** Pancreatic ascites should be excluded.

Management

The aim is both to reduce sodium intake and to increase renal sodium excretion, producing a net reabsorption of fluid from the ascites into the circulating volume. The maximum rate at which ascites can be mobilized is 500 to 700 mL in 24 hours (see later).

- ***Serum electrolytes, creatinine, and estimated glomerular filtration rate (eGFR).*** For hospitalized patients, check on alternate days; weigh the patient and measure urine output daily.
- ***Bed rest.*** This will cause a diuresis by improving renal perfusion but is rarely helpful.
- ***Dietary sodium restriction.*** It is possible to reduce sodium intake to 40 mmol in 24 hours and still maintain an adequate protein and calorie intake with a palatable diet.
- ***Drugs.*** Many contain significant amounts of sodium (up to 50 mmol daily). Examples include antacids and antibiotics (particularly penicillins and cephalosporins). Sodium-retaining drugs (NSAIDs, corticosteroids) should be avoided.
- ***Fluid restriction.*** This is unnecessary unless the serum sodium is lower than 128 mmol/L (see later).
- ***Diuretics.*** The diuretic of choice is the aldosterone antagonist spironolactone, starting at 100 mg daily. Chronic administration produces gynaecomastia. Eplerenone 25 mg once daily does not cause gynaecomastia.

The aim of diuretic therapy should be to produce a net loss of fluid approaching 700 mL in 24 hours (0.7 kg weight loss, or 1.0 kg if peripheral oedema is present). Although 60% of patients respond on this regimen, the spironolactone can be increased gradually to 400 mg daily if necessary, providing there is no hyperkalaemia. Urinary sodium may be used to titrate the dose. A loop diuretic, such as furosemide 20 to 40 mg or bumetanide 0.5 mg or 1 mg daily, is added if response is poor. These loop diuretics have several potential disadvantages, including hyponatraemia, hypokalaemia, and volume depletion.

Diuretics should be temporarily discontinued if a rise in serum creatinine occurs, representing over-diuresis and hypovolaemia, or if there is hyperkalaemia or worsening encephalopathy. Hyponatraemia almost always represents haemodilution secondary to a failure to clear free water (usually a marker of reduced renal perfusion), and diuretics should be stopped if the sodium falls below 128 mmol/L. Vaptans (see Chapter 27), vasopressin V_2-receptor antagonists that increase free water clearance, have a small beneficial effect on hyponatraemia and ascites but do not affect mortality, complications of cirrhosis or renal failure; routine use in cirrhosis cannot be recommended.

Shunts

A TIPS may be inserted to treat resistant ascites, providing there is no spontaneous PSE and there is minimal disturbance of renal function. Frequency of paracentesis and diuretic use is reduced, and nutrition is enhanced. Survival may also improve.

A peritoneo-bladder conduit, by means of an implantable, rechargeable, battery-powered pump, has been developed for use in patients with advanced cirrhosis and resistant ascites (Alfapump). This removes ascites from the peritoneal cavity into the urinary bladder, to be eliminated through urination. Studies have shown a reduction in the need for large-volume paracentesis but also several complications, including pain and infection, occur and this is not widely available.

In patients requiring end-of-life care, insertion of a long-term abdominal drain (LTAD) may allow ascites management outside of a hospital setting. These have been used successfully in patients whose ascites is due to malignancy and are currently being assessed in a large-scale randomized trial in the UK.

Spontaneous bacterial peritonitis

Spontaneous bacterial peritonitis (SBP) represents a serious complication that is an indication for referral for transplant assessment; it occurs in up to 18% of patients with ascites who have undergone decompensation. The infecting organisms gain access to the peritoneum by haematogenous spread; most are *Escherichia coli*, *Klebsiella*, or enterococci, although gram-positive organisms such as *Streptococcus pneumoniae* and staphylococcal species have increased significantly in recent years. The condition should be suspected in any patient with ascites who deteriorates, as pain and pyrexia are frequently absent. Diagnostic aspiration should always be performed. A raised neutrophil count in ascites is sufficient evidence alone to start immediate treatment. A broad-spectrum antibiotic is used, with subsequent alteration according to culture results *in combination with infusions of human albumin solution*. Evidence has emerged that use of non-selective beta-blockers should be carefully considered/stopped following diagnosis of SBP, as prescription may be associated with increased risk of renal impairment. Mortality is 10% to 15%. Recurrence is common (70% within a year) and secondary prevention—for example, with norfloxacin 400 mg daily—reduced incidence of SBP in a study of 80 patients. Primary prophylaxis of SBP in patients with low ascites protein levels below 10 to 15 g/L or severe liver disease may prevent hepatorenal syndrome and improve survival. However, the most studied agent, norfloxacin, is not available in the UK and fluoroquinolones have been given a boxed warning by the FDA, and restricted use has been advised by the Medicines and Healthcare products Regulatory Agency (MHRA) and (European Medicines Agency (EMA) because of increased incidence of musculoskeletal, vascular, and nervous system side-effects. There has also been an increase in anti-microbial resistance in liver cirrhosis patients over the last decade. A large-scale UK randomized trial of co-trimoxazole for primary prophylaxis is due to report in 2025. The benefits and harms of antibiotics must be carefully considered.

Portosystemic encephalopathy

Portosystemic encephalopathy (PSE) is a chronic neuropsychiatric syndrome that is secondary to cirrhosis. Acute encephalopathy can occur in acute hepatic failure (see 'Acute hepatic failure' section). PSE can arise in portal hypertensive patients due to spontaneous 'shunting', or in those with surgical or TIPS shunts.

Pathogenesis

In cirrhosis, the portal blood bypasses the liver via collaterals, and 'toxic' metabolites pass directly to the brain to produce encephalopathy. Ammonia-induced alteration of brain neurotransmitter balance, especially at the astrocyte–neuron interface, is considered to be the leading pathophysiological mechanism. Ammonia is produced by the breakdown of protein by intestinal bacteria. Other implicated substances are FFAs and mercaptans; accumulation of false neurotransmitters (octopamine) or activation of the γ-aminobutyric acid (GABA) inhibitory neurotransmitter system may

Box 23.21 Factors precipitating portosystemic encephalopathy

- High dietary protein
- Gastrointestinal haemorrhage
- Constipation
- Infection, including spontaneous bacterial peritonitis
- Fluid and electrolyte disturbance due to diuretic therapy or paracentesis
- Drugs (e.g., any central nervous system depressant)
- Portosystemic shunt operations, TIPS
- Any surgical procedure
- Progressive liver damage
- Development of hepatocellular carcinoma

also be responsible. Increased blood levels of aromatic amino acids (tyrosine and phenylalanine) and reduced branched-chain amino acids (valine, leucine, and isoleucine) also occur.

Clinical features

An acute onset often has a precipitating factor (**Box 23.21**). The patient becomes increasingly drowsy and comatose.

Chronically, there is a disorder of personality, mood, and intellect, with a reversal of normal sleep rhythm. These changes may fluctuate and a collateral history should be obtained. The patient is irritable, confused, and disorientated, and has slow, slurred speech. General features include nausea, vomiting, and weakness. There is hyper-reflexia and increased tone. Coma occurs as the encephalopathy becomes more marked. Convulsions are very rare, and if they do occur, other causes must be considered.

Signs include:

- fetor hepaticus (a sweet smell to the breath)
- a coarse flapping tremor seen when the hands are outstretched and wrists hyperextended (asterixis)
- constructional apraxia, with the patient being unable to write or draw a five-pointed star
- decreased mental function, which can be assessed by using the serial sevens test or a trail-making (or connection) test.

Investigations

Diagnosis is clinical.

Additional investigations

- ***EEG*** shows a decrease in the frequency of the normal α-waves (8–13 Hz) to 1.5 to 3 Hz. These changes occur before coma supervenes.
- ***Visual evoked responses*** (see Chapter 28) also detect subclinical encephalopathy.
- ***Arterial blood ammonia*** can be useful for the differential diagnosis of coma and for following a patient with PSE but is not always available.

Management

- ***Identify and remove the possible precipitating cause***, such as cerebral depressant drugs, constipation or electrolyte imbalance due to over-diuresis.
- ***Give purgation and enemas*** to empty the bowels of nitrogenous substances. Lactulose (10–30 mL three times daily) is an osmotic purgative that reduces the colonic pH and limits ammonia absorption. Lactilol (β-galactoside sorbitol 30 g daily) is metabolized by colonic bacteria and is comparable in efficacy.
- ***Maintain nutrition***, if necessary, via a fine-bore nasogastric tube, and do not restrict protein for more than 48 hours.
- **Rifaximin** is a poorly absorbed semisynthetic antibiotic based on rifamycin that is licenced for secondary prevention of PSE.
- ***Consider stopping or reduce diuretic therapy.***
- ***Give intravenous fluids*** if evidence of dehydration (beware of too much sodium).
- ***Treat infection.***
- ***Increase protein*** in the diet to the limit of tolerance as encephalopathy improves.
- ***Embolize collaterals.*** In selected patients, embolization of collaterals that bypass the liver may improve liver blood flow and reduce encephalopathy.

Prognosis

Acute encephalopathy in acute hepatic failure has a very poor prognosis associated with that of the disease itself. In cirrhosis, chronic PSE adversely affects prognosis, but the course is very variable. Very rarely with chronic portosystemic shunting, an organic syndrome with cerebellar signs or choreoathetosis develops, as well as myelopathy leading to a spastic paraparesis due to demyelination. These patients require referral to a liver transplant centre.

Renal failure (hepatorenal syndrome)

The hepatorenal syndrome typically occurs in patients with advanced cirrhosis, portal hypertension, jaundice, and ascites. The urine output is low with a low urinary sodium concentration, a maintained capacity to concentrate urine (i.e., intact tubular function), and almost normal renal histology. The renal failure is therefore described as 'functional'. It is often precipitated by over-vigorous diuretic therapy, NSAIDs, diarrhoea, paracentesis, and infection, particularly SBP.

The mechanism is similar to that of ascites, with extreme peripheral vasodilation leading to decreased effective blood volume and consequent hypotension (see 'Ascites' section). This causes increased plasma renin, aldosterone, noradrenaline (norepinephrine), and vasopressin, leading to renal vasoconstriction. There is an increased pre-glomerular vascular resistance that causes blood to be directed away from the renal cortex. This leads to a reduced GFR and plasma renin remains high. Salt and water retention occurs, with reabsorption of sodium from the renal tubules.

Eicosanoids have been incriminated in the pathogenesis, supported by precipitation of the syndrome by inhibitors of prostaglandin synthase, such as NSAIDs.

Diuretic therapy should be stopped and intravascular hypovolaemia corrected, for which albumin is recommended. Terlipressin with intravenous albumin improved renal function in 32% of patients compared to 17% with placebo, but not 3-month survival in a US large-scale trial with increased adverse events, especially pulmonary oedema seen in the terlipressin arm. Noradrenaline can be used as an alternative vasoconstrictor. Liver transplantation is the best option. In patients who are candidates for transplantation, haemodialysis can be used as a bridging option but is frequently difficult to perform, and survival is generally limited by the severity of the hepatic failure.

Hepatopulmonary syndrome

This is hypoxaemia in patients with advanced liver disease due to intrapulmonary vascular dilation with no evidence of primary pulmonary disease. The patients have features of cirrhosis with spider naevi and clubbing, as well as cyanosis. Most are asymptomatic, but with more severe disease become breathless on standing. Transthoracic echocardiography shows intrapulmonary shunting and arterial blood gases confirm hypoxaemia. These changes are improved with liver transplantation.

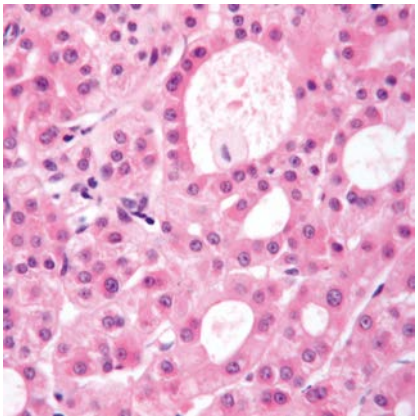


Fig. 23.25 Primary biliary cholangitis. Portal tract inflammation with lymphocytes and plasma cells (haematoxylin and eosin, ×10).

Porto-pulmonary hypertension

This occurs in 1% to 2% of patients with cirrhosis and portal hypertension. It may respond to medical therapy (e.g., intravenous epoprostenol, or oral bosentan and sildenafil). Severe pulmonary hypertension is a contraindication to liver transplantation.

Primary hepatocellular carcinoma

This is discussed in the 'Primary malignant tumours' section.

Further reading

Bass NM, Mullen KD, Sanyal A et al. Rifaximin Treatment in Hepatic Encephalopathy. *N Engl J Med* 2010: 362(12):1071–1081.
Rodriguez-Roisin R, Krowka MJ. Hepatopulmonary syndrome. *N Engl J Med* 2008; 358:2378–2387.
Tripathi D, Stanley AJ, Hayes PC et al. UK guidelines on the management of variceal haemorrhage in cirrhotic patients *Gut* 2015; 64:1680–1704.
Wijdicks EFM. Hepatic encephalopathy. *N Engl J Med* 2016; 375:1660–1670.

Types of cirrhosis

Alcohol related liver cirrhosis

This is discussed in the 'Alcohol and alcoholic liver disease' section.

Primary biliary cholangitis

Primary biliary cholangitis (PBC; Fig. 23.25) is a chronic disorder with progressive destruction of small interlobular bile ducts, leading to cirrhosis. Females aged 40 to 50 years constitute 90% of patients. PBC is diagnosed increasingly frequently in its milder forms. The prevalence is approximately 7.5 per 100 000, with a 1% to 6% increase in first-degree relatives. PBC has been called 'chronic non-suppurative destructive cholangitis', a term more descriptive of the early lesion, which emphasizes the fact that true cirrhosis occurs only in the later stages.

Aetiology

The aetiology is unknown, but an immunological basis is well described. Serum anti-mitochondrial antibodies (AMA) are found in almost all patients. The mitochondrial antigen M2 is specific to PBC and five M2-specific antigens have been identified. The presence of AMA in high titre is unrelated to the clinical or histological picture and its role in pathogenesis is unclear. Antibodies against nuclear antigens, such as anti-gp210, are present in 50% of patients and correlate with progression towards liver failure.

It seems likely that an environmental factor acts on a genetically predisposed host via molecular mimicry, initiating autoimmunity. *E. coli* and *Novosphingobium aromaticivorans* antibodies are present in high titre.

The synthesis of IgM is increased, which is thought to be due to the failure of the switch from IgM to IgG antibody synthesis. No specific associated class II major histocompatibility complex (MHC) loci have been found.

Clinical features

Asymptomatic patients are discovered on routine examination or screening, and may have hepatomegaly, a raised serum ALP or autoantibodies.

Pruritus is often the earliest symptom. Fatigue, which is frequently disabling, may accompany the pruritus, particularly in progressive cases. When jaundice appears, hepatomegaly is usually present. Pigmented xanthelasma on eyelids or deposits of cholesterol in the creases of the hands may be seen.

Associated disorders

Autoimmune disorders (e.g., Sjögren's syndrome, scleroderma, thyroid disease) occur with increased frequency. Keratoconjunctivitis sicca (dry eyes and mouth) is seen in 70% of cases. Renal tubular acidosis, membranous glomerulonephritis, coeliac disease and interstitial pneumonitis are also associations.

Investigations

- ***Mitochondrial antibodies***—measured routinely by ELISA (in titres >1:160)—are present in over 95% of patients; M2 antibody is 98% specific. Other non-specific antibodies (e.g., anti-nuclear factor and smooth muscle) may also be present.
- ***High serum alkaline phosphatase*** is often the only liver biochemistry abnormality.
- ***Serum cholesterol*** is raised.
- ***Serum IgM*** may be very high.
- ***Ultrasound*** can show a diffuse alteration in liver architecture.
- ***Liver biopsy*** shows characteristic histological features of a portal tract infiltrate, mainly of lymphocytes and plasma cells; approximately 40% have granulomas. Most of the early changes are in zone 1. Later, there is damage to and loss of small bile ducts with ductular proliferation. Portal tract fibrosis and, eventually, cirrhosis are seen.

Hepatic granulomas are also seen in sarcoidosis, tuberculosis, schistosomiasis, drug reactions, brucellosis, and parasitic infestation (e.g., strongyloidiasis).

Differential diagnosis

The classical picture presents little difficulty with diagnosis (and can be confirmed by biopsy, although this is only necessary if another cause is being considered). A group of patients with the histological changes of PBC but the serology of autoimmune hepatitis are termed as having autoimmune cholangitis and respond to steroids and azathioprine.

In the jaundiced patient, extrahepatic biliary obstruction should be excluded by ultrasound or MRCP.

Management

- ***Ursodeoxycholic acid*** (10–15 mg/kg) improves bilirubin and aminotransferase levels. It should be given early in the asymptomatic phase, as these patients benefit, whereas no benefit is achieved in advanced disease. In patients who do not respond

to ursodeoxycholic acid, obeticholic acid is approved for use and effectively reduces the serum alkaline phosphatase, a marker for long-term survival. The most clinically relevant adverse effect is dose-related pruritus.

- ***Steroids*** may improve biochemical and histological disease but cause osteoporosis and other side-effects, and should not be used.
- Malabsorption of ***fat-soluble vitamins (A, D, and K)*** occurs, and supplementation is required.
- ***Bisphosphonates*** are needed for osteoporosis. Despite raised serum lipid concentrations, PBC is not associated with an increased cardiovascular disease risk, and strategies for prevention of vascular events should be tailored to the individual.
- Pruritus is difficult to control; ***colestyramine*** is helpful, although unpalatable. ***Rifampicin***, and ***naloxone*** and ***naltrexone*** (opioid antagonists) and sertraline have been shown to be of benefit. Intractable pruritus may be relieved by plasmapheresis.
- The lack of effective medical therapy has made PBC a major indication for ***liver transplantation*** (see 'Liver transplantation' section).
- Fatigue is common and can be severely debilitating; there is no proven therapy. Liver transplant is associated with improvement but a substantial number of patients still suffer fatigue following transplant.

Complications

Complications are those of cirrhosis. In addition, osteoporosis and, rarely, osteomalacia and a polyneuropathy occur.

Prognosis

Prognosis is very variable. Asymptomatic patients and those presenting with pruritus only will survive for more than 20 years. Symptomatic patients with jaundice have a more rapidly progressive course and may die of liver failure or bleeding varices within 5 years. Liver transplantation should therefore be offered when the serum bilirubin is persistently above 100 μmol/L.

Primary sclerosing cholangitis

Primary sclerosing cholangitis (PSC) is a chronic cholestatic liver disease characterized by fibrosing inflammatory destruction of both the intra- and extrahepatic bile ducts. In 75% of patients, PSC is associated with inflammatory bowel disease (usually ulcerative colitis); it is not unusual for PSC to predate the onset of inflammatory bowel disease. The causes are unknown but genetic susceptibility to PSC is associated with the HLA-A1-B8-DR3 haplotype. The autoantibody pANCA (anti-neutrophil cytoplasmic antibody) is found in the serum of 60% of cases. Seventy per cent of patients are male and the average age of onset is approximately 40 years. Secondary PSC is seen in patients with HIV and cryptosporidium (see Chapter 44) and may follow ketamine misuse.

Clinical features

With increasing screening of patients who have inflammatory bowel disease, PSC is detected at an asymptomatic phase with abnormal liver biochemistry, usually a raised serum ALP. Symptomatic presentation is usually with fluctuating pruritus, jaundice, and cholangitis.

Diagnosis

The typical biliary changes associated with PSC may be identified by MRCP. The cholangiogram characteristically shows irregularity of calibre of both intra- and extrahepatic ducts, although either may be involved alone (**Fig. 23.26**).

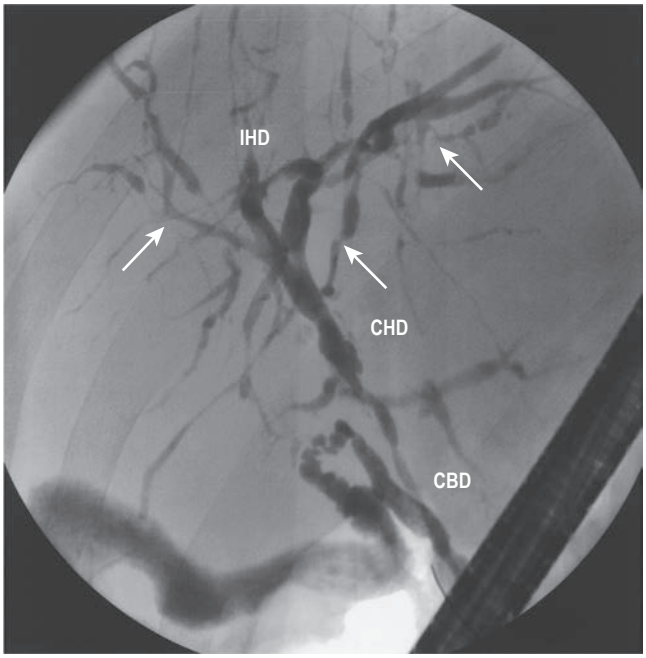


Fig. 23.26 Primary sclerosing cholangitis. An endoscopic cholangiogram showing the typical features of primary sclerosing cholangitis. There are calibre irregularities of the intrahepatic ducts (IHD). There is also minor stricturing of the extrahepatic ducts at the confluence between the common bile duct (CBD) and the common hepatic duct (CHD).

Pathology

Histology can be contributory; it shows inflammation of the intrahepatic biliary radicles with associated scar tissue, classically described as having the appearance of 'onion skin'. These changes range from minor inflammatory infiltrates to the level of established cirrhosis.

Management

PSC is a slowly progressive lesion (symptoms and biochemical tests may fluctuate), which ultimately leads to liver cirrhosis and associated decompensation. Recurrent cholangitis may be a feature before the onset of cirrhosis. Cholangiocarcinoma occurs in up to 15% of patients (see 'Cholangiocarcinoma' section and Chapter 24).

The only proven treatment is liver transplantation. Ursodeoxycholic acid has been evaluated extensively but there is no evidence of benefit and high-dose therapy (30 mg/kg) may be deleterious. In a small minority the dominant lesion is sited in the extrahepatic ducts (see **Fig. 23.26**). Such lesions may be amenable to endoscopic biliary intervention with balloon dilation and temporary stent placement (see Chapter 24).

Secondary biliary cirrhosis

Cirrhosis can result from prolonged (for months) large-duct biliary obstruction. Causes include bile duct strictures, gallstones, and sclerosing cholangitis. An ultrasound examination and MRCP, sometimes followed by ERCP or percutaneous transhepatic cholangiography (where the ducts are cannulated under ultrasound guidance through the skin) if cannulation is difficult, are performed to outline the ducts, and any remedial cause is treated.

Hereditary haemochromatosis

Hereditary haemochromatosis (HH) is an inherited disease characterized by excess iron deposition in various organs, leading to eventual fibrosis and functional organ failure. There are four main types of inherited disorders:

- ***type 1*** HFE: the ***HFE*** gene (mutation C282Y) is the most common and is on chromosome 6
- ***type 2A:*** the juvenile ***HJV*** gene (mutation G320V); ***type 2B:*** the juvenile ***HAMP*** gene (mutation 93delG)
- ***type 3*** TfR2: the ***TfR2*** gene (mutation Y250X)
- ***type 4*** ferroportin: the ***SLC40A1*** gene (mutation V162del).

All are transmitted by an autosomal recessive gene, apart from the ferroportin iron overload, which is dominantly transmitted.

HH has a prevalence in White people (homozygotes) of 1 in 400, but very variable phenotypic expression and a heterozygote (carrier) frequency of 1 in 10. It is the most common single-gene disorder in White people. Although HH occurs worldwide, it is less common in people without European ancestry.

Aetiology

Between 85% and 90% of patients with overt HH are homozygous for the Cys 282 Tyr (C282Y mutation): that is, type 1 *HFE*. A second mutation (His 63 Asp, H63D) occurs in about 25% of the population and is in complete linkage disequilibrium with Cys 282 Tyr.

HH Type 3 occurs in southern Europe and is associated with TfR2, a transferrin receptor isoform. The other types, ferroportin-related (type 4) and juvenile forms (2A and 2B), are much rarer.

Dietary intakes of iron and chelating agents (ascorbic acid) may be relevant. Iron overload may be present in alcoholics; alcohol excess *per se* does not cause HH, although there is a history of excess alcohol intake in 25% of patients.

Mechanism of damage

The *HFE* gene protein interacts with the transferrin receptor 1, which is a mediator in intestinal iron absorption (see Fig. 38.11). Iron is taken up by the mucosal cells inappropriately, exceeding the binding capacity of transferrin.

Hepcidin, a protein synthesized in the liver (Fig. 23.27), is central to the control of iron absorption; it is increased in iron deficiency states and decreased with iron overload. The mutations described disrupt hepcidin expression, leading to unregulated absorption of dietary iron from the gut via ferroportin channels on the basolateral surface of enterocytes and uninhibited iron overload.

Hepatic expression of the hepcidin gene is decreased in HFE haemochromatosis, facilitating liver iron overload. Excess iron is then gradually taken up by the liver and other tissues over a long period. It seems likely that it is the iron itself that precipitates fibrosis.

Pathology

In symptomatic patients the total body iron content is 20 to 40 g, compared with 3 to 4 g in a normal person. The iron content is particularly increased in the liver (Fig. 23.28) and pancreas (50–100 times normal) but is also higher in other organs (e.g., endocrine glands, heart, and skin).

In established cases the liver shows extensive iron deposition and fibrosis. Early in the disease, iron is deposited in the periportal hepatocytes (in pericanalicular lysosomes). Later, it is distributed widely throughout all acinar zones, biliary duct epithelium, Kupffer cells, and connective tissue. Cirrhosis is a late feature.

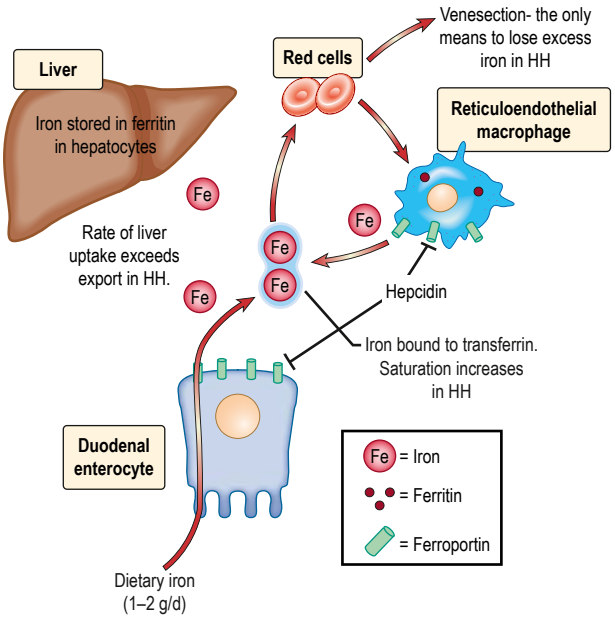


Fig. 23.27 The circulation of iron from the duodenal enterocyte to and from the liver, red cells, and reticuloendothelial macrophages is dysregulated in hereditary haemochromocytosis. Iron is absorbed from the intestine (see Fig. 38.11) via ferroportin channels on the basolateral surface of enterocytes and circulates bound to transferrin. The reticuloendothelial cells clear old erythrocytes and release iron to circulate and be stored as ferritin in the liver. The liver is the major site of production of the peptide hormone hepcidin which blocks ferroportin, preventing transit of iron across the gut and release of iron from reticuloendothelial cells. Reduced levels of hepcidin in HH result in unregulated absorption and release of iron to the circulation and progressive accumulation in tissue stores, particularly the liver. *Fe,* Iron.

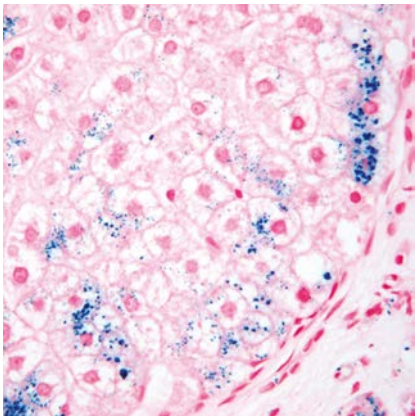


Fig. 23.28 Cirrhotic nodule in haemochromatosis. Note the iron overload *(blue)* (Perls, ×25).

Clinical features

The course of the disease depends on a number of factors, including sex, dietary iron intake, presence of associated hepatotoxins (especially alcohol), and genotypes. Overt clinical manifestations occur more frequently in males; the reduced incidence in females is probably explained by physiological blood loss and a smaller dietary intake of iron. Most affected individuals present in the fifth decade. The classic triad of bronze skin pigmentation (due to

melanin deposition), hepatomegaly, and diabetes mellitus is present only in cases of gross iron overload.

Hypogonadism secondary to pituitary dysfunction is the most common endocrine feature. Deficiency of other pituitary hormones is also found but symptomatic endocrine deficiencies, such as loss of libido, are very rare. Cardiac manifestations, particularly heart failure and arrhythmias, are common, especially in younger patients. Calcium pyrophosphate is deposited asymmetrically in both large and small joints (chondrocalcinosis), leading to an arthropathy. The exact relationship of chondrocalcinosis to iron deposition is uncertain.

Complications

A recent large study showed that among males, HFE p.C282Y homozygosity was significantly associated with increased risk for incident primary hepatic malignancy and excess mortality, but these increased risks were not seen for females. For male C282Y homozygotes up to age 75 years, the lifetime risk of primary hepatic malignancy was 7.2% (95% CI, 3.9–13.1) compared with 0.6% (95% CI, 0.4–0.7) for males with neither variant. HCC has been described only very rarely in non-cirrhotic patients in whom the excess iron stores have been removed.

Investigations

Homozygotes

- ***Serum iron*** is elevated (>30 μmol/L) in 90% with a reduction in the TIBC and a transferrin saturation of more than 45%.
- ***Serum ferritin*** is elevated (usually >500 μg/L or 240 nmol/L).
- ***Liver biochemistry*** is often normal, even with established cirrhosis.

Heterozygotes

Heterozygotes may have normal biochemical tests or modest increases in serum iron transferrin saturation (>45%) or serum ferritin (usually >400 μg/L).

Genetic testing

If iron studies are abnormal, genetic testing is performed.

Liver biopsy

This is not required for diagnosis but is useful to establish the extent of tissue damage, assess tissue iron and measure the hepatic iron concentration (>180 μmol/g dry weight of liver indicates haemochromatosis).

Mild degrees of parenchymal iron deposition in patients with other forms of cirrhosis, particularly if due to alcohol, can often cause confusion with true homozygous HH.

Magnetic resonance imaging

MRI shows a dramatic reduction in the signal intensity of the liver and pancreas owing to the paramagnetic effect of ferritin and haemosiderin. A highly T2-weighted, gradient recalled echo (GRE) technique detects all clinically relevant liver iron overload (>60 μmol/g of liver). In secondary iron overload (haemosiderosis), which involves the reticuloendothelial cells, the pancreas is spared, enabling distinction between these two conditions.

Management

Venesection

This prolongs life and may reverse tissue damage; the risk of malignancy still remains if cirrhosis is present. All patients should have excess iron removed as rapidly as possible. This is achieved using venesection of 500 mL performed twice weekly for up to 2 years: that is, 160 units with 250 mg of iron per unit, which equals 40 g removed. During venesection, serum iron and ferritin and the mean corpuscular volume (MCV) should be monitored. These fall only when available iron is depleted. Three or four venesections per year are required to prevent re-accumulation of iron. Serum ferritin should remain within the normal range.

Manifestations of the disease usually improve or disappear, except for diabetes, testicular atrophy, and chondrocalcinosis. The requirements for insulin often diminish in diabetic patients. Testosterone replacement is frequently helpful.

In the rare patient who cannot tolerate venesection (because of severe cardiac disease or anaemia), chelation therapy with desferrioxamine, either intermittently or continuously by infusion, has been successful in removing iron.

Screening

In all cases of HH, all first-degree family members must be screened to detect early and asymptomatic disease. *HFE* mutation analysis is performed with measurement of transferrin saturation and serum ferritin.

In the general population, serum iron and transferrin saturation are the best and cheapest tests available.

Wilson's disease (progressive hepatolenticular degeneration)

Dietary copper is normally absorbed from the stomach and upper small intestine. Copper is transported to the liver loosely bound to albumin; in the liver, it is incorporated into apocaeruloplasmin, forming caeruloplasmin, a glycoprotein synthesized in the liver, and secreted into the blood. The remaining copper is normally excreted in the bile and excreted in faeces.

Wilson's disease is a very rare inborn error of copper metabolism that results in copper deposition in various organs, including the liver, the basal ganglia of the brain, and the cornea. It is potentially treatable and all young patients with liver disease must be screened for this condition.

Aetiology

Wilson's disease is an autosomal recessive disorder with a molecular defect within a copper-transporting ATPase encoded by a gene (designated *ATP7B*) located on chromosome 13. It affects between 1 in 30 000 and 1 in 100 000 individuals. Over 300 mutations have been identified, the most frequent being His 1069 Gly (H1069Q), found in approximately 50% of White patients; compound heterozygotes are common. This mutation is rare in India and Asia. Wilson's disease occurs worldwide, particularly in countries where consanguinity is common. There is a failure of both incorporation of copper into procaeruloplasmin, which leads to low serum caeruloplasmin, and biliary excretion of copper. There is a low serum caeruloplasmin in over 80% of patients but this is not the cause of the copper deposition. The precise mechanism for the failure of copper excretion is not known.

Pathology

The liver histology is not diagnostic and varies from that of chronic hepatitis to macronodular cirrhosis. Stains for copper show a periportal distribution but this can be unreliable (see later). The basal ganglia are damaged and show cavitation, the kidneys demonstrate tubular degeneration, and erosions are seen in bones.

Clinical features

Children usually present with hepatic problems, whereas young adults have more neurological problems, such as tremor, dysarthria, involuntary movements, and eventually dementia. The liver disease varies from episodes of acute hepatitis, especially in children (which can go on to acute hepatic failure), to chronic hepatitis or cirrhosis.

Typical signs are those of chronic liver disease with neurological signs of basal ganglia involvement (see Chapter 28). A specific sign is the Kayser–Fleischer ring, caused by copper deposition in Descemet's membrane in the cornea. It appears as a greenish-brown pigment at the corneoscleral junction and frequently requires slit-lamp examination for identification. It may be absent in young children.

Investigations

- ***Serum copper and caeruloplasmin*** are usually reduced but can be normal.
- ***Urinary copper*** is usually increased to 100 to 1000 μg in 24 hours (1.6–16 μmol); normal levels are less than 40 μg (0.6 μmol).
- ***Liver biopsy*** aids diagnosis, which depends on measurement of the amount of copper in the liver (>250 μg/g dry weight), although high levels of copper are also found in the liver in chronic cholestasis.
- ***Haemolysis and anaemia*** may be present.
- ***Genetic analysis*** is limited but selected exons are screened according to population groups.

Management

- Lifetime treatment with ***penicillamine***, 1.5 to 2 g daily in divided doses, is effective in chelating copper. If treatment is started early, clinical and biochemical improvement can occur. Urinary copper levels should be monitored and the drug dose adjusted downwards after 2 to 3 years. Serious side-effects of the drug occur in 10% and include skin rashes, leucopenia, skin changes, and renal damage.
- ***Trientine*** (750–1250 mg per day in divided doses) and ***zinc acetate*** (150 mg/day) are used as maintenance therapy and for asymptomatic cases. All siblings and children of patients should be screened (*ATP7B* mutation analysis is useful) and treatment with zinc is given, even in the asymptomatic, if there is evidence of copper accumulation. A diet low in copper (i.e., excluding chocolate and peanuts) is advised.

Prognosis

Early diagnosis and effective treatment have improved the outlook. Neurological damage is, however, permanent. Acute hepatic failure or decompensated cirrhosis should be treated by liver transplantation.

Alpha$_1$-antitrypsin deficiency

A deficiency of α_1-antitrypsin (α_1-AT, see also Chapter 20) is sometimes associated with liver disease and pulmonary emphysema (particularly in smokers). Part of a family of serine protease inhibitors, or serpin superfamily, α_1-AT is a glycoprotein. Deficiency of α_1-AT is a genetic disorder and 1 in 10 northern Europeans carries an abnormal gene.

The protein is a 394-amino acid 52 kDa acute phase protein that is synthesized in the liver and constitutes 90% of the serum α_1-globulin seen on electrophoresis. Its main role is to inhibit the proteolytic enzyme, neutrophil elastase.

The gene is located on chromosome 14. The genetic variants of α_1-AT are characterized by their electrophoretic mobilities as medium (M), slow (S), or very slow (Z). The normal genotype is protease inhibitor MM (PiMM), the homozygote for Z is PiZZ, and the heterozygotes are PiMZ and PiSZ. S and Z variants are due to a single amino acid replacement of glutamic acid at positions 264 and 342 of the polypeptide, respectively. This results in decreased synthesis and secretion of the protein by the liver as protein–protein interactions occur between the reactive centre loop of one molecule and the β-pleated sheet of a second (loop sheet polymerization).

How this causes liver disease is uncertain. It is postulated that failure of secretion of the abnormal protein leads to an accumulation in the liver, leading to liver damage.

Clinical features

The majority of patients with clinical disease are homozygotes with a PiZZ phenotype. Some may present in childhood and a few require transplantation. Approximately 10% to 15% of adult patients will develop cirrhosis, usually over the age of 50 years, and 75% will have respiratory problems. Approximately 5% of patients die of their liver disease. Heterozygotes (e.g., PiSZ or PiMZ) may develop liver disease but the risk is small.

Investigations

- ***Serum α_1-antitrypsin*** is low, at 10% of the normal level in the PiZZ phenotypes, and 60% of normal in the S variant.

Histologically, periodic acid–Schiff (PAS)-positive, diastase-resistant globules that contain α_1-AT are seen in periportal hepatocytes. Fibrosis and cirrhosis can be present.

Management

There is no treatment, apart from managing the complications of liver disease. Patients with hepatic decompensation should be assessed for liver transplantation and should stop smoking.

Further reading

Andrews NC. Closing the iron gate. *N Engl J Med* 2012; 366:376–377.
Atkins JL, Pilling LC, Masoli JAH, et al. Association of hemochromatosis *HFE* p.C282Y homozygosity with hepatic malignancy. *JAMA* 2020;324:2048–2057.
European Association for the Study of the Liver. EASL Clinical Practice Guidelines: Wilson disease. *J Hepatol* 2012; 56:671–685.
Fiorella Murillo Perez C et al. Greater transplant-free survival in patients receiving obeticholic acid for primary biliary cholangitis in a clinical trial setting compared to real-world external controls. *Gastroenterology.* 2022;163(6):1630–1642.e3.
Hirschfield GM, Dyson JK, Alexander GJM, et al. The British Society of Gastroenterology/UK-PBC primary biliary cholangitis treatment and management guidelines. *Gut* 2018;67:1568–1594.
Karlsen TH, Folseraas T, Thorburn D, Vesterhus M. Primary sclerosing cholangitis—a comprehensive review. *J Hepatol* 2017 vol. 67 j 1298–1323.
Kowdley KV, Brown KE, Ahn J, Sundaram V. ACG Clinical Guideline: Hereditary Hemochromatosis. *Am J Gastroenterol.* 2019 Aug; 114(8):1202–1218.

ALCOHOL AND ALCOHOLIC LIVER DISEASE

Alcohol is consumed in large quantities all over the world and in many countries, alcohol consumption is a major medical and social problem (see Chapter 30). It increases morbidity and mortality in a variety of ways, including effects on heart disease, stroke, cancers, and neurological/psychiatric problems, as well as liver disease, and is associated with nutritional deficiencies and abnormal metabolism of drugs.

Box 23.22 The main pathways of ethanol oxidization

Alcohol dehydrogenase

$$\underset{\text{(ethanol)}}{CH_3CH_2OH} + NAD^+ \xrightarrow{ADH} \underset{\text{(acetaldehyde)}}{CH_3CHO} + NADH + H^+ \quad [1]$$

Liver microsomal ethanol oxidizing system (MEOS)

- Includes the specific P450 enzyme, CYP2EI, which is induced by ethanol

$$CH_3CH_2OH + NADPH + H^+ + O_2 \xrightarrow{MEOS} CH_3CHO + NADP + 2H_2O \quad [2]$$

Ethanol (ethyl alcohol) is oxidized, in the steps shown in **Box 23.22**, to acetaldehyde. Acetaldehyde is then converted to acetate, 90% in the liver mitochondria. Acetate is released into the blood and oxidized by peripheral tissues to carbon dioxide and water.

Alcohol dehydrogenases are found in many tissues and it has been suggested that enzymes present in the gastric mucosa may contribute substantially to ethanol metabolism.

Ethanol itself produces 29.3 kJ/g (7 kcal/g), but many alcoholic drinks also contain sugar, which increases their calorific value. For example, 1 pint of beer provides about 1045 kJ (250 kcal), so heavy drinkers will be unable to lose weight if they continue to drink.

Effects of excess alcohol consumption

Excess consumption of alcohol leads to two major problems, both of which can be present in the same patient:

- alcohol dependence syndrome (see Chapter 30)
- physical damage to various tissues.

Each unit of alcohol (defined as 10 mL [7.9 g] of pure ethanol) corresponds to about half a pint of normal beer, one single measure of spirit, or half a glass of wine (**Fig. 23.29**). An intake of less than 14 units per week is generally considered to be safe. All the long-term effects of excess alcohol consumption are due to excess ethanol, irrespective of the type of alcoholic beverage; that is, beer and spirits are no different in their long-term effects. Short-term effects, such as hangovers, depend on additional substances, particularly other alcohols such as isoamyl alcohol, which are known as congeners. Brandy and bourbon contain the highest percentage of congeners.

The amount of alcohol that produces damage varies and not everyone who drinks heavily will suffer physical damage. For example, only 20% of people who drink heavily develop cirrhosis of the liver. The effect of alcohol on different organs of the body is not the same; in some patients, the liver is affected; in others, the brain or muscle. The differences may be genetically determined. Thiamine deficiency contributes to neurological (confusion, Wernicke–Korsakoff's syndrome; see Chapter 28) and some of the non-neurological manifestations (cardiomyopathy).

The susceptibility of different organs to damage is variable and the figures given in **Box 23.23** are provided only as a guide to sensible drinking. Heavy drinkers who persist for many years are at greater risk than heavy sporadic drinkers. Women are advised not to drink alcohol at all during pregnancy because consumption of even small amounts of alcohol can lead to fetal growth restriction and may also increase the risk of miscarriage. The fetal alcohol syndrome is characterized by intellectual disabilities, dysmorphic features, and growth impairment; it occurs in fetuses of alcohol-dependent women.

A summary of the physical effects of alcohol is given in **Box 23.24**. Details of these diseases are discussed in the relevant chapters. The effects of alcohol withdrawal are discussed in Chapter 11.

Specific public health and clinical guidance on prevention and management of alcohol-related disorders has been produced by NICE. The clinical guidelines specifically address alcohol withdrawal, Wernicke's encephalopathy, and both alcohol-related liver and pancreatic diseases.

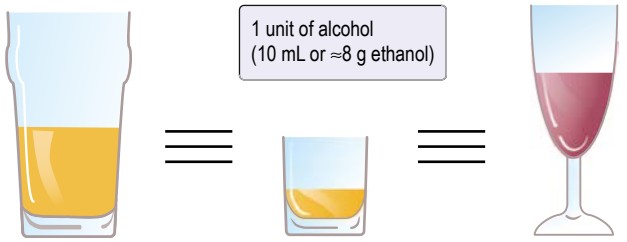


Fig. 23.29 Measures containing one unit of alcohol. The percentage refers to the proportion of alcohol by volume (vol/vol). The alcohol content of beers, spirits, and wines can vary substantially.

Box 23.23 Guide to sensible alcohol drinking

Weekly maximum

- 14 units for both men and women

To help achieve this

- Use a standard measure
- Do not drink during the daytime
- Have alcohol-free days each week

Remember

- Health can be damaged without being 'drunk'. Regular heavy intake is more harmful than occasional binges
- Do not drink to 'drown sorrows'
- In the UK, the drink-before-driving limit of alcohol in the blood is 800 mg/L
- 1 unit of alcohol is eliminated per hour; therefore spread drinking time
- Food decreases absorption and therefore results in a lower blood alcohol level
- Four to five units are sufficient to put the blood alcohol level over the legal driving limit in a 70 kg man (less in a lighter person)

Alcoholic liver disease

In general, the effects of a given intake of alcohol on the liver seem to be worse in females. The following figures are for males and should be reduced by about 30% for females:

- ***High risk:*** 160 g ethanol per day (20 single drinks)
- ***Medium risk:*** 80 g ethanol per day (10 single drinks)
- ***Little risk:*** 40 g ethanol per day (5 single drinks).

Ethanol is metabolized in the liver by two pathways, resulting in an increase in the NADH/NAD ratio. The altered redox potential causes increased hepatic fatty acid synthesis with decreased fatty acid oxidation; both events lead to hepatic accumulation of fatty acid, which is then esterified to glycerides.

The changes in oxidation–reduction also impair carbohydrate and protein metabolism and are the cause of the centrilobular necrosis of the hepatic acinus that is typical of alcohol damage. TNF-α release from Kupffer cells causes the release of reactive oxygen species, leading in turn to tissue injury and fibrosis.

Acetaldehyde is formed by the oxidation of ethanol, and its effect on hepatic proteins may well be a factor in producing liver cell damage. The exact mechanism of alcoholic hepatitis and cirrhosis is unknown, but since only 10% to 20% of people who drink heavily will develop cirrhosis, a genetic predisposition is recognized. Genome-wide association studies (GWAS) have identified

Box 23.24 Physical effects of excess alcohol consumption

Central nervous system
- Epilepsy
- Wernicke–Korsakoff's syndrome
- Polyneuropathy

Muscles
- Acute or chronic myopathy

Cardiovascular system
- Cardiomyopathy
- Beriberi heart disease
- Cardiac arrhythmias
- Hypertension

Metabolism
- Hyperuricaemia (gout)
- Hyperlipidaemia
- Hypoglycaemia
- Obesity

Endocrine system
- Pseudo–Cushing's syndrome

Respiratory system
- Chest infections

Gastrointestinal system
- Acute gastritis (including bleeding)
- Carcinoma of the oesophagus or large bowel
- Pancreatic disease
- Liver disease (fatty liver, hepatitis, cirrhosis)

Haemopoiesis
- Macrocytosis (due to direct toxic effect on bone marrow or folate deficiency)
- Thrombocytopenia
- Leucopenia

Bone
- Osteoporosis
- Osteomalacia

the *PNPLA3, TM6SF2,* and *HSD17B13* genes as risk factors for chronic liver disease progression. Immunological mechanisms have also been proposed, with the release of cytokines, particularly IL-8, which is a neutrophil chemoattractant and infiltration with neutrophils is a feature of alcoholic hepatitis.

Alcohol can enhance the effects of the toxic metabolites of drugs (e.g., paracetamol) on the liver, as it induces microsomal metabolism via the microsomal ethanol oxidizing system (MEOS; see **Box 23.23**).

Pathology

Alcohol can produce a wide spectrum of liver disease, from fatty change to hepatitis and cirrhosis.

Fatty liver

The metabolism of alcohol invariably produces fat in the liver (**Fig. 23.30**), mainly in zone 3. This is minimal with small amounts of alcohol, but with larger amounts the cells become swollen with fat (steatosis). There is no liver cell damage. The fat disappears on stopping alcohol. Steatosis is also seen in MASLD (see 'Metabolic dysfunction–associated steatotic liver disease' section).

In some cases, collagen is laid down around the central hepatic veins (perivenular fibrosis) and this can progress to cirrhosis without a preceding hepatitis. Alcohol directly affects stellate cells, transforming them into collagen-producing myofibroblast cells).

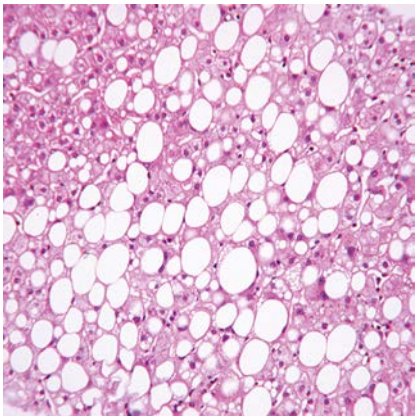


Fig. 23.30 Liver steatosis. (Haematoxylin and eosin, ×10.)

Alcoholic hepatitis

In addition to fatty change, there is infiltration by polymorphonuclear leucocytes and hepatocyte necrosis, mainly in zone 3. Dense cytoplasmic inclusions called Mallory bodies are sometimes seen in hepatocytes, and giant mitochondria are also a feature. Mallory bodies are suggestive of, but not specific for, alcoholic damage, as they can be found in other liver diseases, such as Wilson's disease and PBC. If alcohol consumption continues, alcoholic hepatitis may progress to cirrhosis.

Alcoholic cirrhosis

This is classically of the micronodular type, but a mixed pattern is also seen accompanying fatty change; evidence of pre-existing alcoholic hepatitis may be present.

Clinical features

Fatty liver

There are often no symptoms or signs. Vague abdominal symptoms of nausea, vomiting, and diarrhoea are due to the more general effects of alcohol on the gastrointestinal tract. Hepatomegaly, sometimes huge, can occur, together with other features of chronic liver disease.

Alcoholic hepatitis

The clinical features vary in degree:
- The patient may be well, with few symptoms, the hepatitis being apparent only on the liver biopsy in addition to fatty change.
- Mild to moderate symptoms of ill-health, occasionally with mild jaundice, may occur. Signs include all the features of chronic liver disease. Liver biochemistry is deranged and the diagnosis is made on liver histology.
- In the severe case, often superimposed on alcoholic cirrhosis, the patient is ill, with jaundice and ascites. Abdominal pain is frequently present and a high fever is associated with the liver necrosis. On examination there is deep jaundice, hepatomegaly, sometimes splenomegaly, and ascites with ankle oedema. The signs of chronic liver disease are also present.

Alcoholic cirrhosis

This represents the final stage of liver disease from alcohol use. Nevertheless, patients can be very well with few symptoms. On examination, there are usually signs of chronic liver disease (see the figure in 'Clinical skills' section). The diagnosis is confirmed by liver biopsy.

The patient usually presents with one of the complications of cirrhosis. In many cases there are features of alcohol dependency (see Chapter 30), as well as evidence of involvement of other systems, such as polyneuropathy.

Investigations

Fatty liver

An elevated MCV often indicates heavy drinking. Liver biochemistry shows mild abnormalities with elevation of both serum aminotransferase enzymes. The γ-GT level is a useful test for determining whether the patient is taking alcohol. With severe fatty infiltration, marked changes in all liver biochemical parameters can occur. Ultrasound or CT will demonstrate fatty infiltration, as will liver histology. Elastography (see 'Hepatic stiffness [transient elastography]' section) can be used to estimate the degree of fibrosis.

Alcoholic hepatitis

Investigations show a leucocytosis with markedly deranged liver biochemistry and elevated:

- serum bilirubin
- serum AST and ALT (AST often greater than ALT)
- serum ALP
- PT.

A ***low serum albumin*** may also be found. Rarely, hyperlipidaemia with haemolysis ***(Zieve syndrome)*** may occur.

Liver biopsy, if required, is performed by the transjugular route because of the prolonged PT.

Alcoholic cirrhosis

Investigations are as for cirrhosis in general.

Management and prognosis

General management

All patients should stop drinking alcohol. Delirium tremens (a withdrawal symptom) is treated with diazepam. Intravenous thiamine should be given empirically to prevent Wernicke–Korsakoff encephalopathy. Bed rest is necessary, along with a diet high in protein and vitamin supplements. Parenteral nutrition is sometimes required. Patients must be advised to participate in alcohol cessation programmes. The likelihood of abstention is dependent on many factors, particularly social and family ones.

Fatty liver

The patient is advised to stop drinking alcohol; the fat will disappear and the liver biochemistry usually returns to normal. Small amounts of alcohol can be drunk subsequently, as long as patients are aware of the problems and can control their consumption.

Alcoholic hepatitis

In severe cases the alcoholic hepatitis leads to acute decompensation and the patient requires admission to hospital. Nutrition must be maintained with enteral feeding, if necessary, and vitamin supplementation is given. Steroid therapy has been widely used in patients with a discriminant function (DF) score of more than 32 but a multicentre UK study from 2015, which included over 1000 patients, suggested that there was no survival benefit and there was an increased infection risk in those taking steroids. If steroids are prescribed, current guidelines emphasize early identification of non-responders to limit unnecessary exposure and minimize risk using tools such as the Lille score (>0.45 indicates poor response to steroids, which can therefore be stopped; **Box 23.25**).

Box 23.25 Lille score for alcoholic hepatitis

(*Calculator at* http://www.lillemodel.com)

R = 3.19 − (0.101 × age in years) + (0.147 × albumin on admission in g/L) + (0.0165 × change in bilirubin level from day 0 to day 7 in μmol/L) − (0.206 × renal insufficiency [0 if absent, 1 if present][a]) − (0.0065 × bilirubin day 0 in μmol/L) − (0.0096 × INR)

Score = EXP(−R)/[1 + EXP(−R)]

A score of <0.16 indicates a 96% chance of survival at 28 days; ≥0.56 indicates a 55% chance of survival at 28 days.

[a]Creatinine >115 μmol/L.
INR, International normalized ratio.

Box 23.26 Glasgow alcoholic hepatitis score

Score	1	2	3
Age	<50	>50	
White cell count (×10⁹/L)	<15	>15	
Urea (nmol/L)	<5	>5	
Bilirubin (μmol/L)	<125	125–250	>250
INR	<1.5	1.5–2.0	>2.0

Poor prognosis = total score >9.

INR, International normalized ratio.

Discriminant function

$$DF = [4.6 \times \text{prothrombin time above control in seconds}] + \text{bilirubin (mg/dL)}$$

Bilirubin mmol/L ÷ 17 to convert to mg/dL.

Severe = >32.

Other prognostic tools include the Glasgow score (**Box 23.26**), with a score of more than 9 indicating the 28-day mortality is 75% and the MELD score (see **Box 23.17**).

Infection must be excluded or concomitantly treated, although a large RCT did not support prophylactic antibiotics to improve survival. Treatment for encephalopathy and ascites may be required. Granulocyte colony–stimulating factor therapy may improve outcomes in severe AH. However, there was substantial heterogeneity observed in a meta-analysis with conflicting results between Asian and European studies, and G-CSF cannot currently be recommended for AH, especially in Europe.

Patients are advised to stop drinking for life, as this is undoubtedly a pre-cirrhotic or cirrhotic condition. The prognosis is variable: despite abstinence, liver disease is progressive in many patients but abstinence significantly improves survival. Unfortunately, many return to drinking. In severe cases the mortality is at least 50%; with a PT twice normal, progressive encephalopathy, and acute kidney injury, the mortality approaches 90%. Early liver transplantation for severe alcoholic hepatitis has been shown to have very acceptable short to intermediate-term patient survival and rates of alcohol use post-transplant, however, long-term outcomes are uncertain, as are how best to select and manage these patients before this becomes an accepted indication for liver transplant.

Alcoholic cirrhosis

The management of cirrhosis is described in the 'Cirrhosis/Management' section. Again, all patients are advised to stop drinking

for life. Abstinence from alcohol results in an improvement in prognosis, with a 5-year survival of 90%, but with continued drinking this falls to 60%. In advanced disease (i.e., jaundice, ascites, and haematemesis) the 5-year survival rate falls to 35% with most of the deaths occurring in the first year. Liver transplantation results in good survival. A trial of abstention to establish whether liver disease can improve is mandatory; prior to consideration to be added to the transplant list and patients often sign a contract with their clinicians regarding their abstinence, both before and after transplantation, and specific follow-up regarding alcohol use is recommended.

HCC is a complication, particularly in men.

Further reading

National Institute for Health and Care Excellence. *NICE Clinical Guideline 115: Alcohol-use Disorders: Diagnosis, Assessment and Management of Harmful Drinking and Alcohol Dependence*. NICE 2011; http://www.nice.org.uk/guidance/cg115.

Thursz MR, Richardson P, Allison M et al. Prednisolone or pentoxifylline for alcoholic hepatitis. *N Engl J Med* 2015; 372:1619–1628.

BUDD–CHIARI SYNDROME

In this condition there is obstruction to the venous outflow of the liver owing to occlusion of the hepatic vein. In one-third of patients the cause is unknown but specific causes include hypercoagulability states (e.g., paroxysmal nocturnal haemoglobinuria, polycythaemia vera) or thrombophilia (see Chapter 40), the contraceptive pill, and leukaemia. Other causes include occlusion of the hepatic vein owing to posterior abdominal wall sarcomas, renal or adrenal tumours, HCC, hepatic infections (e.g., hydatid cyst), congenital venous webs, radiotherapy, or trauma to the liver.

Clinical features

The acute form presents with abdominal pain, nausea, vomiting, tender hepatomegaly, and ascites (a fulminant form that occurs particularly in pregnant females). In the chronic form there is enlargement of the liver (particularly the caudate lobe), mild jaundice, ascites, a negative hepatojugular reflux, and splenomegaly with portal hypertension.

Investigations

Investigations show a high protein content in the ascitic fluid and characteristic liver histology with centrizonal congestion, haemorrhage, fibrosis, and cirrhosis. Ultrasound, CT, or MRI will demonstrate hepatic vein occlusion with diffuse abnormal parenchyma on contrast enhancement. The caudate lobe is spared because of its independent blood supply and venous drainage. There may be compression of the inferior vena cava. Pulsed Doppler sonography or colour Doppler is useful, as it shows abnormalities of flow in the hepatic vein. Thrombophilia screening is mandatory. Multiple defects of coagulation occur. Thrombosis of the portal vein is present in up to 20% of patients.

Differential diagnosis

A similar clinical picture can be produced by inferior vena caval obstruction, right-sided cardiac failure, or constrictive pericarditis, and appropriate investigations should be performed.

Management

Thrombolytic therapy can be given acutely. Ascites should be treated, as should any underlying cause (e.g., polycythaemia). Congenital webs should be treated radiologically or resected surgically. TIPS is the treatment of choice and surgical portocaval shunts are reserved for those who fail this treatment, providing there is no caval obstruction or severe caval compression when a caval stent can be inserted. Liver transplantation is the first-choice treatment for ***chronic*** Budd–Chiari syndrome and the fulminant form. Life-long anticoagulation is mandatory following TIPS and transplantation.

Prognosis

The prognosis depends on the aetiology however, a step-wise approach to therapy, as described earlier, can lead to a 5-year survival rate of nearly 85%.

Further reading

Mancuso A. An update on management of Budd–Chiari syndrome. *Ann Hepatol* 2014; 13:323–326.

Valla DC. Budd–Chiari syndrome/hepatic venous outflow tract obstruction. *Hepatol Int* 2018; 12(Suppl 1):168–180.

HEPATIC SINUSOIDAL OBSTRUCTION SYNDROME

Hepatic sinusoidal obstruction syndrome (SOS, previously known as veno-occlusive disease) is due to injury of the hepatic veins and presents similarly to the Budd–Chiari syndrome. It was originally described in Jamaica, where ingestion of toxic pyrrolizidine alkaloids in bush tea (made from plants of the genera *Senecio*, *Heliotropium*, and *Crotalaria*) caused damage to the hepatic veins, but it can be seen in other parts of the world. It also occurs as a complication of chemotherapy and total body irradiation, used before allogeneic bone marrow transplantation. The development of SOS after bone marrow transplantation carries a high mortality. Treatment is supportive, with control of ascites and hepatocellular failure. TIPS has been used in a few cases. Defibrotide is recommended for prophylactic use before bone marrow transplantation in children and adults with a high risk of SOS (e.g., pre-existing hepatic disease or prior treatment with gemtuzumab ozogamicin).

FIBROPOLYCYSTIC DISEASES

These diseases are usually inherited and lead to the presence of cysts or fibrosis in the liver, kidney, and occasionally the pancreas, and other organs.

Polycystic disease of the liver

Multiple cysts can occur in the liver as part of autosomal dominant polycystic disease of the kidney (see Chapter 26). These cysts are usually asymptomatic but occasionally cause abdominal pain and distension. Liver function is normal and complications such as oesophageal varices are very rare. The prognosis is excellent and depends on the kidney disease.

Solitary cysts

These are usually found by chance during imaging and are mainly asymptomatic.

Congenital hepatic fibrosis

In this rare condition the liver architecture is normal but there are broad collagenous fibrous bands extending from the portal tracts. Congenital hepatic fibrosis is often inherited as an autosomal recessive condition but can also occur sporadically. It usually presents in childhood with hepatosplenomegaly, and portal hypertension is common. It may present later in life and can be misdiagnosed as cirrhosis.

A wedge biopsy of the liver may be required to confirm the diagnosis. The outlook is good and the condition should be distinguished from cirrhosis. Patients who bleed do well after endoscopic therapy of varices (or a portocaval anastomosis) because of their good liver function.

Congenital intrahepatic biliary dilation (Caroli's disease)

In this rare, non-familial disease there are saccular dilations of the intrahepatic or extrahepatic ducts. It can present at any age (although it usually does so in childhood) with fever, abdominal pain, and recurrent attacks of cholangitis with gram-negative septicaemia. Jaundice and portal hypertension are absent. Diagnosis is by ultrasound, percutaneous transhepatic cholangiography, and MRCP. There is an increased risk of biliary malignancy.

LIVER ABSCESS

Pyogenic abscess

Pyogenic abscesses are uncommon; they may be single or multiple. The most common used to be a portal pyaemia from intra-abdominal sepsis (e.g., appendicitis or perforations) but nowadays the aetiology is not known in many cases. In the elderly, biliary sepsis is a common cause. Other causes include trauma, bacteraemia, and direct extension from, for example, a perinephric abscess.

The most common organism is *E. coli. Streptococcus milleri* and anaerobic organisms such as *Bacteroides* are often seen. Others include *Enterococcus faecalis*, *Proteus vulgaris*, and *Staphylococcus aureus*. Often the infection is mixed.

Clinical features

Some patients present with malaise lasting several days or even months. Others can present with fever, rigors, anorexia, vomiting, weight loss, and abdominal pain. In these patients a gram-negative septicaemia with shock can occur. On examination there may be little of note; alternatively, the patient may be toxic, febrile, and jaundiced. In such patients the liver is tender and enlarged, and there may be a pleural effusion or a pleural rub in the right lower chest.

Investigations

Patients are often investigated for 'pyrexia of unknown origin' and most investigations will be normal.

- ***Serum bilirubin*** is raised in 25% of cases.
- ***Normochromic normocytic anaemia*** may occur, usually accompanied by a polymorphonuclear leucocytosis.

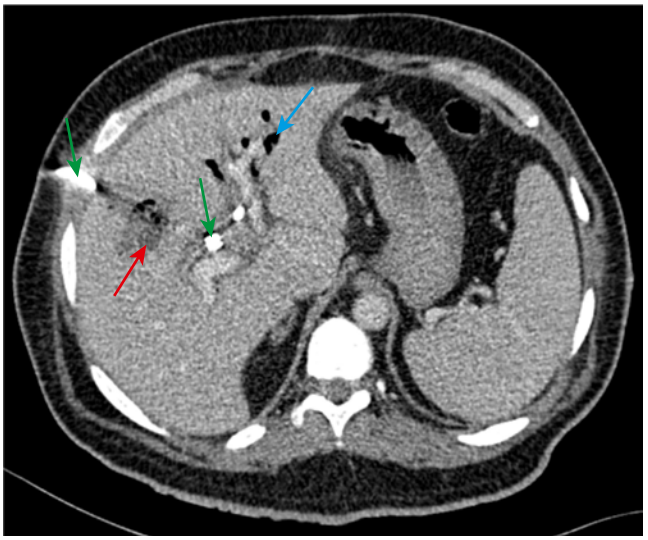


Fig. 23.31 Axial CT image demonstrating a fluid- and gas-containing lesion consistent with a liver abscess *(red arrow)*. The patient has an external biliary drain *in situ (green arrows)* and there is gas in the biliary tree (aerobilia) *(blue arrow)*.

- ***Serum ALP, ESR, and CRP*** are often raised.
- ***Serum vitamin B_{12}*** is very high, as it is stored in and subsequently released from the liver.
- ***Blood cultures*** are positive in only 30% of cases.

Imaging

Ultrasound is useful for detecting abscesses. A CT scan may be of value in complex and multiple lesions (Fig. 23.31). A chest X-ray will show elevation of the right hemidiaphragm, with a pleural effusion in the severe case. Depending on age, imaging of the colon and biliary tree may be necessary to find the source of infection.

Management

Aspiration of the abscess should be attempted under ultrasound control. Antibiotics should initially cover gram-positive, gram-negative, and anaerobic organisms until the causative organism is identified.

Further drainage via a large-bore needle under ultrasound control or surgically may be necessary if resolution is difficult or slow. Any underlying cause must be treated.

Prognosis

The overall mortality depends on the nature of the underlying pathology but has been reduced to approximately 16% with needle aspiration and antibiotics.

Amoebic abscess

This occurs worldwide and must be considered in patients travelling from endemic areas. *Entamoeba histolytica* (see Chapter 42) can be carried from the bowel to the liver in the portal venous system, leading to portal inflammation with the development of multiple microabscesses and, eventually, single or multiple large abscesses.

Clinically, the onset is usually gradual but may be sudden. There is fever, anorexia, weight loss, and malaise. There is often no history of dysentery. On examination the patient looks ill and has tender hepatomegaly and signs of an effusion or consolidation in the base of the right side of the chest. Jaundice is unusual.

Investigations

These are as for pyogenic abscess, plus:

- ***Serological tests for amoeba*** (e.g., haemagglutination, amoebic complement fixation test, ELISA). These are always positive, particularly if there are bowel symptoms; they remain positive after a clinical cure and therefore do not indicate current disease. A repeat negative test, however, is good evidence against an amoebic abscess.
- ***Diagnostic aspiration of fluid*** looking like anchovy sauce.

Management

Metronidazole 800 mg three times daily is given for 10 days. Aspiration is used in patients failing to respond, in those with multiple and sometimes large abscesses, and in those with abscesses in the left lobe of the liver or impending rupture. To clear infection from the intestine paromomycin or diloxanide must also be taken.

Complications

Complications include rupture, secondary infection, and septicaemia.

OTHER INFECTIONS OF THE LIVER

Schistosomiasis

Schistosoma mansoni and *S. japonicum* affect the liver but *S. haematobium* rarely does so (see also Chapter 42). During their life cycle the ova reach the liver via the venous system and obstruct the portal branches, producing granulomas, fibrosis, and inflammation, but not cirrhosis.

Clinical features and investigations

Clinically, there is hepatosplenomegaly and portal hypertension, which is particularly severe with *S. mansoni*.

Investigations show a raised serum ALP, and ova can be found in the stools (centrifuged deposits) and in rectal and liver biopsies. Skin tests and other immunological tests often give false results and may also be positive because of past infection.

Management

Treatment is with praziquantel, but fibrosis still remains with a potential risk of portal hypertension and associated bleeding, characteristically pre-sinusoidal due to intense portal fibrosis.

Hydatid disease

Cysts caused by *Echinococcus granulosus* are single or multiple. They usually occur in the lower part of the right lobe. The cyst has three layers: an outside layer derived from the host, an intermediate laminated layer, and an inner germinal layer that buds off brood capsules to form daughter cysts (see also Chapter 42).

Clinical features and investigations

Clinically, there may be no symptoms or there may be a dull ache and swelling in the right hypochondrium.

Investigations show a peripheral eosinophilia in 30% of cases and usually a positive hydatid complement fixation test or haemagglutination (85%). Plain abdominal X-ray may show calcification of the outer coat of the cyst. Ultrasound and CT scan demonstrate cysts and may show diagnostic daughter cysts within the parent cyst (**Fig. 23.32**).

Management

Medical treatment (e.g., with albendazole 10 mg/kg, which penetrates into large cysts) results in cysts becoming smaller. Puncture, aspiration, injection, re-aspiration (PAIR) has been used since the 1980s. Fine-needle aspiration is undertaken under ultrasound control with chemotherapeutic cover. Surgery can be performed with removal of the cyst intact, if possible, after first sterilizing the cyst with alcohol, saline, or cetrimide. Chronic calcified cysts can be left; there have been no well-designed clinical trials for any therapy.

Complications and prognosis

These include rupture into the biliary tree or other organs, or intraperitoneally, with spread of infection. The prognosis without any complications is good, although there is always a risk of rupture. ***Preventative measures*** include deworming pet dogs and preventing pets from eating infected carcasses, as well as veterinary control programmes.

Echinococcus multilocularis causes alveolar echinococcosis and is almost exclusively a hepatic disease, with a high mortality if not treated. Early diagnosis enables radical surgery and then continued chemosuppression.

Acquired immunodeficiency syndrome

The liver is often involved in AIDS and is a significant cause of morbidity or mortality. HIV itself is not the cause of the liver abnormalities. Clinical hepatomegaly is common (60% of patients). The following are seen, although less frequently in areas where anti-retroviral therapy (ART) is available:

- ***Pre-existing/coincidental viral hepatitis.*** The hepatitis (HBV, HCV, HDV) progresses more rapidly and is a leading cause of death.
- ***Neoplasia.*** Kaposi's sarcoma and non-Hodgkin lymphoma may be seen, and there is an increased risk of HCC.
- ***Opportunistic infection*** (e.g., *Mycobacterium tuberculosis*, *M. avium-intracellulare*, *Cryptococcus*, *Candida albicans*, toxoplasmosis).
- ***Drug hepatotoxicity.***
- ***Secondary sclerosing cholangitis***.
- ***Non-cirrhotic portal hypertension.*** This is associated with ART and is more common in people who have had profound immunosuppression and received early nucleotides, such as didanosine (DDI). The incidence is believed to be declining with the introduction of early ART and less toxic drugs.

Further reading

Joshi D, O'Grady J, Dieterich D et al. Increasing burden of liver disease in patients with HIV infection. *Lancet* 2011; 377:1198–1209.

LIVER DISEASE IN PREGNANCY

See Chapter 48.

LIVER TUMOURS

Secondary liver tumours

The most common liver tumour is a secondary (metastatic) tumour, particularly from the gastrointestinal tract (from the distribution

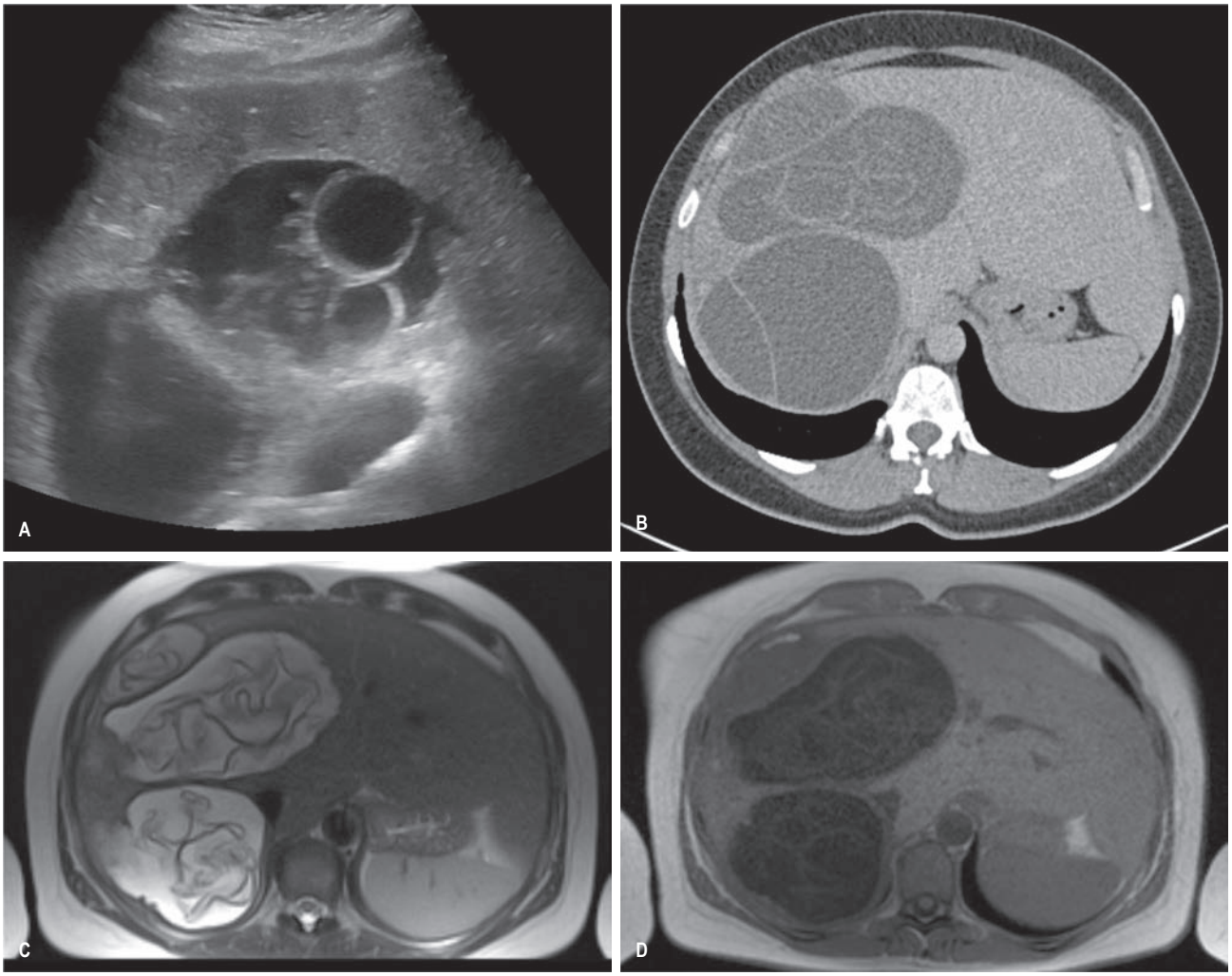


Fig. 23.32 Multimodality imaging from the same patient demonstrating hydatid cysts containing multiple daughter cysts. **(A)** Ultrasound. **(B)** CT. **(C)** T2-weighted MRI. **(D)** T1-weighted MRI.

of the portal blood supply), breast, or bronchus. Secondary liver tumours are usually multiple.

Clinical features

These are variable but usually include weight loss, malaise, upper abdominal pain, and hepatomegaly, with or without jaundice.

Investigations

Ultrasound is the primary investigation, with CT or MRI to define metastases and look for a primary. The serum ALP is almost invariably raised.

Management

This will depend on the site of the primary and the burden of liver metastases. The best results are obtained in colorectal cancer with few hepatic metastases. If the primary tumour is removed and hepatic resection is performed, reasonable survival rates are possible. Chemotherapy is used, particularly with breast cancer. Radiofrequency ablation of the metastases is an alternative to surgery. Thermal therapy and cryotherapy are also used.

Primary malignant tumours

Primary liver tumours may be benign or malignant.

Hepatocellular carcinoma

Hepatocellular carcinoma (HCC) is the sixth most common cancer worldwide and the fourth leading cause of cancer-related death globally.

Aetiology

HBV infection is the most prominent risk factor for HCC development, accounting for approximately 50% of cases. The risk attributed to HCV infection has substantially decreased due to the very high cure rates with antiviral drugs, although those with cirrhosis are still at high risk for HCC incidence even after HCV clearance. In areas where HBV is prevalent, 90% of patients with this cancer are positive for HBV with 80% of HCC cases having cirrhosis, whereas cirrhosis is always present in HCV patients that develop HCC. The development of HCC is related to the integration of viral HBV DNA into the genome of the host hepatocyte (see 'Hepatitis B virus [pathogenesis]'), and to the degree of viral replication (>10 000 copies/mL). Primary liver cancer is also associated with other forms of cirrhosis, such as alcoholic cirrhosis, NAFLD-associated cirrhosis, and haemochromatosis. It occurs two to three times more often in males than females, and males have poorer survival rates and higher recurrence rates after treatment. Other aetiological factors include aflatoxin (a metabolite of a fungus found in groundnuts) and androgenic anabolic steroids.

Pathology

The tumour is either single or occurs as multiple nodules throughout the liver. Histologically, it consists of cells resembling hepatocytes. It can metastasize via the hepatic or portal veins to the lymph nodes, bones, and lungs.

Clinical features

The clinical features include weight loss, anorexia, fever, an ache in the right hypochondrium, and ascites. The rapid development of these features in a cirrhotic patient is suggestive of HCC. On examination, an enlarged, irregular, tender liver may be felt. Increasingly, due to surveillance, HCC is found without symptoms in patients with cirrhosis.

Investigations

- Routine liver biochemistry, full blood count, serum creatinine, and electrolytes.
- ***Serum α-fetoprotein*** may be raised but is normal in approximately 40% of patients.
- ***Ultrasound*** scans show filling defects in 90% of cases.
- ***MRI with contrast or triple phase (see Fig. 23.33D–F).*** CT scans (**Fig. 23.33A–C**) identify HCC but it can be difficult to confirm the diagnosis in lesions smaller than 1 cm, where the diagnostic accuracy of MRI is higher.
- ***Tumour biopsy*** (**Fig. 23.34**), particularly under ultrasonic guidance, is rarely used for diagnosis, as imaging techniques readily show characteristic appearances (hypervascularity of the nodule and lack of portal vein washout) and because seeding along the biopsy tract can occur.

Management and prognosis

See Chapter 41.

Prevention

Persistent HBV infection, usually acquired after perinatal infection, is a strong risk factor for HCC in many parts of the world, such as South-east Asia. Widespread vaccination against HBV has reduced the annual incidence of HCC in Taiwan.

Cholangiocarcinoma

Cholangiocarcinomas are increasing in incidence and can be extrahepatic or intrahepatic (see also Chapters 24 and 41). Intrahepatic adenocarcinomas arising from the bile ducts account for approximately 10% of primary tumours of the liver and biliary tract. They are not associated with cirrhosis or hepatitis B. In the Far East, they may be associated with infestation with *Clonorchis sinensis* or *Opisthorchis viverrini*. The clinical features are similar to those of primary HCC, except that jaundice is frequent with hilar tumours, and cholangitis is more common. There is an increased association with inflammatory bowel disease and PSC (see Chapter 24).

Surgical resection is rarely possible and patients usually die within 6 months. Transplantation is contraindicated, outside of specialized protocols.

Benign tumours

The most common benign tumour is a ***haemangioma*** and occurs in approximately 10% of people. They are usually small and single but can be multiple and large. Haemangiomas are usually found incidentally on ultrasound, CT, or MRI, and have characteristic appearances. They require no treatment.

Hepatic adenomas are rare benign tumours, more common in females and associated with oral contraceptives. Obesity, non-alcoholic steatohepatitis, and metabolic syndromes are associated with hepatic adenomatosis. They are associated with risks of haemorrhage and malignant transformation and can present with abdominal pain or intraperitoneal bleeding. Elective resection is recommended in all males with adenomas (as malignant potential is greater) and in females with adenomas greater than 5 cm. Resection may also be required for patients with abdominal pain.

Further reading

Villanueva A. Hepatocellular carcinoma. *N Engl J Med* 2019;380:1450–1462.

MISCELLANEOUS CONDITIONS OF THE LIVER

Hepatic mitochondrial injury syndromes

These syndromes, in which there is mitochondrial damage with inhibition of β-oxidation of fatty acids, can be categorized as having the following causes:

- ***Genetic***, with abnormalities that include medium-chain acyl-coenzyme A dehydrogenase deficiency, leading to microsteatosis.
- ***Toxins*** leading to liver failure, which include aflatoxin and cerulide (produced by *Bacillus cereus*); the latter causes food poisoning (see Chapter 42).
- ***Drugs*** (e.g., i.v. tetracycline, valproic acid, and nucleoside reverse-transcriptase inhibitors), which can produce a fatal microsteatosis.
- ***Idiopathic***, the best known being fatty liver of pregnancy (see Chapter 48) and Reye's syndrome. The latter, caused by inhibition of β-oxidation and uncoupling of oxidative phosphorylation in mitochondria, leads in children to an acute encephalopathy and diffuse microvesicular fatty infiltration of the liver. Aspirin ingestion and viral infections have been implicated as precipitating agents. Mortality is about 50%, usually due to cerebral oedema.

Idiopathic adult ductopenia

This unexplained condition is characterized by pruritus and cholestatic jaundice. Histology of the liver shows a decrease in intrahepatic bile ducts in at least 50% of the portal tracts, together with the features of cholestasis and marked fibrosis or cirrhosis. In most, the disease is progressive and the only treatment is liver transplantation.

Indian childhood cirrhosis

This condition of children is seen in the Indian subcontinent. The cause is unknown. Eventually, there is development of a micronodular cirrhosis with excess copper in the liver.

Hepatic porphyrias

These are dealt with in Chapter 36.

Cystic fibrosis

Cystic fibrosis (see also Chapter 20) mainly affects the lung and pancreas, but patients can develop fatty liver, cholestasis, and cirrhosis. The aetiology of the liver involvement is unclear.

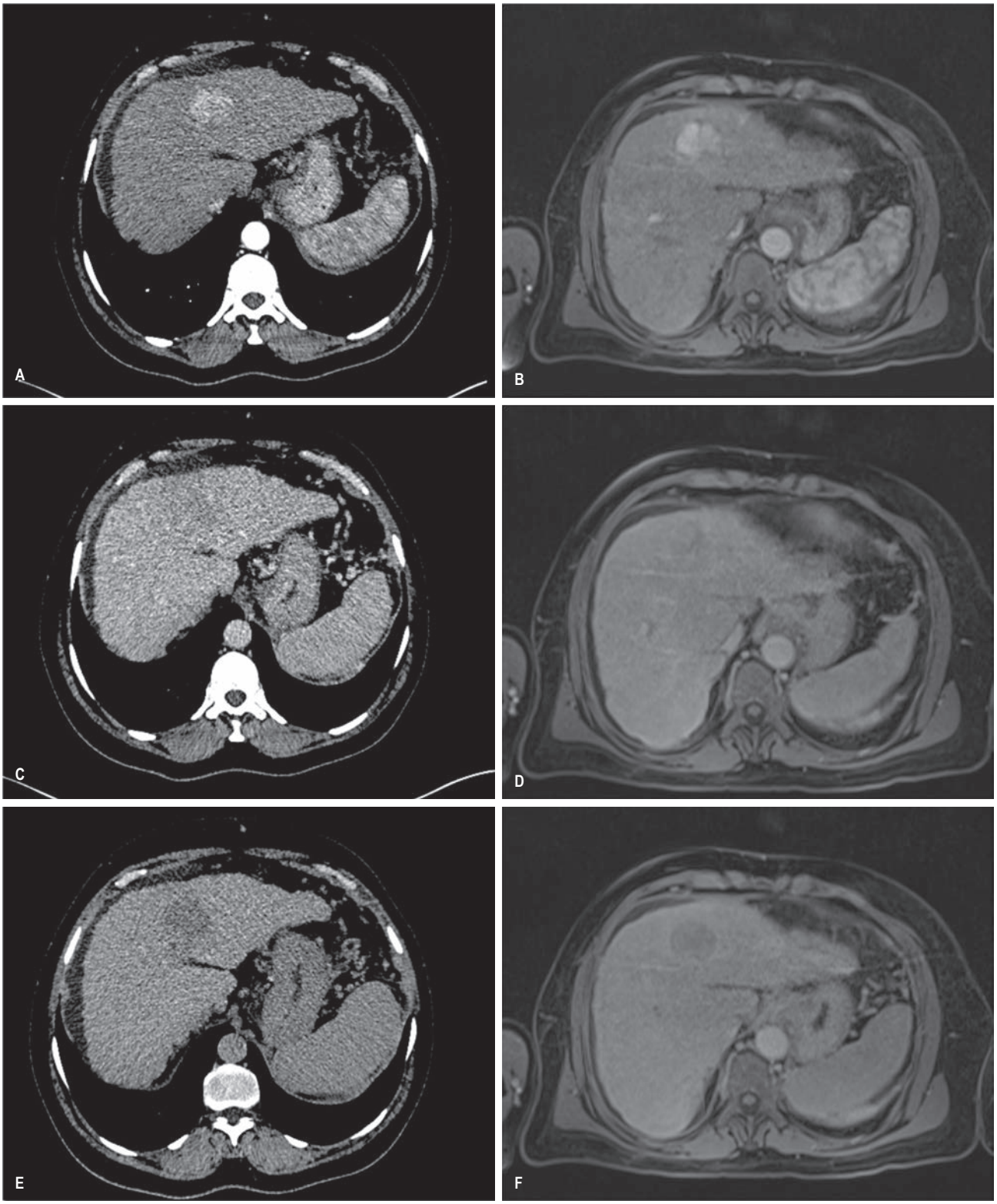


Fig. 23.33 Hepatocellular carcinoma (HCC). Two cases of HCC on **(A–C)** multiphasic contrast-enhanced CT and **(D–F)** MRI. HCC classically demonstrates enhancement on the arterial phase imaging (A, D) and washout on more delayed phases of imaging. In both cases there is subtle washout on the portal venous phase imaged (B, E) with more definite washout on the delayed phase imaging (C, F).

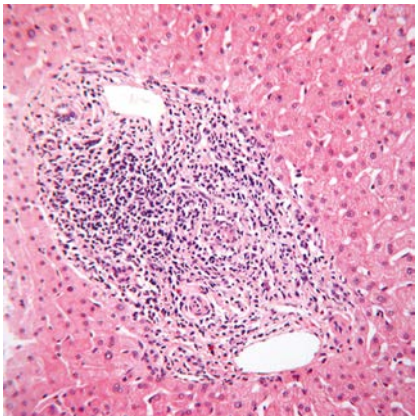


Fig. 23.34 Hepatocellular carcinoma. (Haematoxylin and eosin, ×25.)

Coeliac disease

Abnormal liver biochemical tests are common in coeliac disease (see also Chapter 22) and return to normal with a gluten-free diet. A tissue transglutaminase test should be performed if hepatic causes are not found when investigating abnormal liver biochemistry.

DRUGS AND THE LIVER

Drug metabolism

The liver is the major site of drug metabolism. Drugs are converted from fat-soluble to water-soluble substances that can be excreted in the urine or bile. This metabolism of drugs is mediated by a group of mixed-function enzymes (see Chapter 7).

Drug hepatotoxicity

Many drugs impair liver function. When abnormal liver biochemical tests are found, drugs should always be considered as a cause, particularly when other causes have been excluded. Damage to the liver by drugs (DILI) is usually classified as being either predictable (or dose-related) or non-predictable (not dose-related) (see **Box 7.9**). However, there is considerable overlap and at least ***six mechanisms*** may be involved in the production of damage:

1. disruption of intracellular calcium homeostasis
2. disruption of bile canalicular transport mechanisms
3. formation of non-functioning adducts (enzyme–drug), which may then lead to
4. presentation on the surface of the hepatocyte as new immunogens (attacked by T cells)
5. induction of apoptosis
6. inhibition of mitochondrial function, which prevents fatty acid metabolism and accumulation of both lactate and reactive oxygen species.

The predominant mechanism or combination of mechanisms determines the type of liver injury: that is, hepatitic, cholestatic, or immunological (skin rashes, fever, and arthralgia, i.e., serum sickness syndrome). Eosinophilia and circulating immune complexes and antibodies are occasionally detected.

When a small amount of hepatotoxic drug whose effect is dose-dependent (e.g., paracetamol) is ingested, a large proportion of it undergoes conjugation with glucuronide and sulphate, while the remainder is metabolized by microsomal enzymes to produce toxic derivatives that are immediately detoxified by conjugation with glutathione. If larger doses are ingested, the former pathway becomes saturated and the toxic derivative is produced at a faster rate. Once the hepatic glutathione is depleted, large amounts of the toxic metabolite accumulate and produce damage (see Chapter 15).

The 'predictability' of drugs to produce damage can, however, be affected by metabolic events preceding their ingestion. For example, chronic alcohol users may become more susceptible to liver damage because of the enzyme-inducing effects of alcohol, or ill or starving patients may become susceptible because of the depletion of hepatic glutathione produced by starvation. Many other factors, such as environmental or genetic effects, may be involved in determining the 'susceptibility' of certain patients to certain drugs.

The incidence of drug hepatotoxicity is 14 per 100 000 population with a 6% mortality. It is the most common cause of acute liver failure in the USA. Liver transplantation is used.

Hepatitic damage

The type of damage produced by various drugs is shown in **Box 23.27**. Most reactions occur within 3 months of starting the drug. Monitoring liver biochemistry in patients on long-term treatment, such as anti-tuberculosis therapy, is mandatory. If a drug is suspected of causing hepatic damage, it should be stopped immediately. Liver biopsy is of limited help in confirming the diagnosis, but occasionally, hepatic eosinophilia or granulomas may be seen. Diagnostic challenge with subtherapeutic doses of the drug is sometimes required after the liver biochemistry has returned to normal, to confirm the diagnosis.

Individual drugs

Paracetamol

In high doses, paracetamol produces liver cell necrosis (see earlier). The toxic metabolite binds irreversibly to liver cell membranes. Overdosage is discussed in Chapter 15.

Steroid compounds

Cholestasis is caused by natural and synthetic oestrogens, as well as methyltestosterone. These agents interfere with canalicular biliary flow by blocking MRP2 and MDR3 (see **Fig. 23.5**) and cause a pure cholestasis.

Cholestasis is rare with the contraceptive pill because of the low dosage used. However, the contraceptive pill is associated with an increased incidence of gallstones, hepatic adenomas (rarely, HCCs), the Budd–Chiari syndrome, and peliosis hepatis. The latter condition, which also occurs with anabolic steroids, consists of dilation of the hepatic sinusoids to form blood-filled lakes.

Phenothiazines

Phenothiazines (e.g., chlorpromazine) can produce a cholestatic picture owing to a hypersensitivity reaction. This occurs in 1% of patients, usually within 4 weeks of starting the drug. Typically, it is associated with a fever and eosinophilia. Recovery occurs on stopping the drug.

Anti-tuberculous chemotherapy

Isoniazid produces elevated aminotransferases in 10% to 20% of patients. Hepatic necrosis with jaundice occurs in a smaller percentage. The hepatotoxicity of isoniazid is related to its metabolites and is dependent on acetylator status. Rifampicin produces hepatitis, usually within 3 weeks of starting the drug, particularly

Box 23.27 Liver damage produced by some drugs

Types of liver damage	Drugs	
Zone 3 necrosis	Carbon tetrachloride *Amanita* mushrooms Paracetamol	Salicylates Piroxicam Cocaine
Zone 1 necrosis	Ferrous sulphate	
Microvesicular fat	Sodium valproate	Tetracyclines
Steatohepatitis	Amiodarone Synthetic oestrogens	Nifedipine
Fibrosis	Methotrexate Other cytotoxic agents Arsenic	Vitamin A Retinoids
Vascular		
Sinusoidal dilation	Contraceptive drugs	Anabolic steroids
Peliosis hepatis	Azathioprine Oral contraceptives	Anabolic steroids, e.g., danazol
Veno-occlusive	Pyrrolizidine alkaloids (*Senecio* in bush tea)	Cytotoxics—cyclophosphamide
Acute hepatitis	Isoniazid Rifampicin Methyldopa Atenolol Enalapril Verapamil Ketoconazole	Cytotoxic drugs Clonazepam Disulfiram Niacin Volatile liquid anaesthetics, e.g., halothane Infliximab
Chronic hepatitis	Methyldopa Nitrofurantoin	Fenofibrate Isoniazid
General hypersensitivity	Sulphonamides, e.g., Sulphasalazine Co-trimoxazole Fansidar Penicillins, e.g., Flucloxacillin Ampicillin Amoxicillin Co-amoxiclav NSAIDs, e.g., Salicylates Diclofenac	Allopurinol Anti-thyroid, e.g., Propylthiouracil Carbimazole Quinine, e.g., quinidine Diltiazem Anticonvulsants, e.g., phenytoin
Canalicular cholestasis	Sex hormones Ciclosporin Chlorpromazine Haloperidol Erythromycin Flucloxacillin	Fucidin Cimetidine/ranitidine Nitrofurantoin Imipramine Azathioprine Oral hypoglycaemics
Biliary sludge	Ceftriaxone	
Sclerosing cholangitis	Hepatic arterial infusion of 5-fluorouracil	
Hepatic tumours	Pills with high hormone content (adenomas)	
Hepatocellular carcinoma	Contraceptive pill	Danazol
Nodular regenerative hyperplasia	Azathioprine	Some anti-retroviral therapies

Note: Anti-HIV drugs, e.g., maraviroc, cause hepatic dysfunction.
NSAID, Non-steroidal anti-inflammatory drug.

in patients on high doses. Pyrazinamide produces abnormal liver biochemical tests and, rarely, liver cell necrosis.

Amiodarone

This leads to a steatohepatitis histologically, and liver failure if the drug is not stopped in time.

Sodium valproate

This causes mitochondrial injury with microvesicular steatosis. Intravenous carnitine should be used as an antidote.

Drug prescribing for patients with liver disease

The metabolism of drugs is impaired in severe liver disease (with jaundice and ascites), as the removal of many drugs depends on liver blood flow and the integrity of the hepatocyte. In general, therefore, the effect of drugs is prolonged by liver disease and also by cholestasis. This is further accentuated by portosystemic shunting, which diminishes the first-pass extraction of drugs. With hypoproteinaemia, there is decreased protein binding of some drugs,

and bilirubin competes with many drugs for the binding sites on serum albumin. In patients with PSE, care must be taken in prescribing drugs with a central depressant action, such as narcotics like codeine and anxiolytics. Other common drugs to be avoided in cirrhosis include angiotensin-converting enzyme (ACE) inhibitors, angiotensin receptor blockers (which may provoke hepatorenal syndrome), and NSAIDs (which cause bleeding).

Further reading

Björnsson E. Review article: drug-induced liver injury in clinical practice. *Aliment Pharmacol Ther* 2010; 32:3–13.
Hoofnagel JH, Bjornesson E. Drug induced liver injury. *N Engl J Med* 2019; 381;264–273

Bibliography

Ryder SD, Chapman RW (eds). Liver disorders: Parts 1 and 2. *Medicine* 2011; 39:507–634.
Williams R, Aspinall R, Bellis M et al. Addressing liver disease in the UK. *Lancet* 2014; 384:1953–1997.

Questions for this chapter can be found on Elsevier eBooks+.

24 Biliary tract and pancreatic disease

Christopher Antony Wadsworth

Background 735
Gall bladder and biliary system 736
Biliary system 736
Gallstones 736
Miscellaneous conditions of the biliary tract 743
Gall bladder 743
Extrahepatic biliary tract 744
Tumours of the biliary tract 744
Pancreas 746
Pancreatitis 749
Pancreatic cancer 756

CORE SKILLS AND KNOWLEDGE

In high-income countries, up to 15% of the adult population have gallstones, which are the leading cause of hepatobiliary disease; although in many they do not cause symptoms, acute pancreatitis and ascending cholangitis are serious conditions with a significant mortality. Cancers of the pancreas and gall bladder are not uncommon, and often present late.

Patients with diseases of the biliary tract and pancreas may be cared for as inpatients and outpatients by general surgeons, gastroenterologists, or hepatologists, depending on local services, with complex disease being looked after in specialist centres where there is hepatobiliary surgical, radiological, and endoscopic support. Endoscopic retrograde cholangiopancreatography (ERCP) and endoscopic ultrasound (EUS) are core procedures, allowing a range of interventions including stone extraction, stent placement, and direct biopsy of lesions.

Key skills in this field include:

- Managing emergency conditions, such as ascending cholangitis and acute pancreatitis
- Being aware of the need for vigilance in the early diagnosis of pancreatic cancer and cholangiocarcinoma.

Opportunities for learning about biliary tract and pancreatic disease include searching out, interviewing, and examining inpatients with these conditions (who may be cared for on general surgical or gastrointestinal wards), observing ERCP and EUS procedures, and seeking to understand the treatment strategies employed. Attending an upper gastrointestinal surgical clinic and acute surgical on call, a surgical list to observe a laparoscopic cholecystectomy, and an upper gastrointestinal multidisciplinary team meeting may all prove useful.

BACKGROUND

Biliary and pancreatic disease places a major burden on healthcare systems worldwide. In the West, acute pancreatitis and gallstone disease are the two most common gastrointestinal causes of urgent hospital admission. Pancreatic and biliary tree neoplasms carry a poor prognosis, pose significant diagnostic challenges, and are increasing in incidence.

Patients with upper abdominal pain (acute or chronic), cholestasis (painless or painful), or insidious onset of weight loss commonly present in primary and secondary care. Such presentations can represent major underlying pathology and should warrant prompt and appropriate investigation.

Key skills for managing pancreatobiliary disease include the ability to:

- recognize patterns of liver enzyme derangement, including cholestasis
- manage severe sepsis complicating biliary disease
- select a rational imaging modality for upper abdominal pain
- recognize and manage nutritional deficiency in patients with acute and chronic pancreatitis
- recognize and institute management of diabetes that is due to pancreatitis (see Chapter 35).

Management of pancreatobiliary disease may require the appropriate use of

- antibiotics in the management of biliary sepsis
- pancreatic enzyme supplements
- fat-soluble vitamin supplements (vitamins A, D, and K)
- proton pump inhibitors to improve pancreatic enzyme function
- chemotherapy and immunotherapy drugs in the management of pancreatobiliary adenocarcinoma.

Ultrasound, computed tomography (CT), and magnetic resonance imaging (MRI) technologies enable the biliary tree and pancreas to be visualized with increasing resolution. Widespread use of such imaging has led to improved diagnosis but also to a burgeoning need to characterize and manage incidental, small, often asymptomatic lesions, particularly cysts in the pancreas.

Endoscopic diagnosis of pancreatobiliary disease has been enhanced by improvements in EUS technology, and the

reemergence of direct cholangioscopy for visualization of the bile duct. Therapeutic endoscopy and laparoscopic surgery are now widely available and permit a minimally invasive approach to the management of almost all gallstone disease, and to the resection or palliation of pancreatobiliary cancer.

GALL BLADDER AND BILIARY SYSTEM

The structure and function of the biliary system and the formation of bile are discussed in Chapter 23.

BILIARY SYSTEM (FIG. 24.1)

Bile canaliculi form a network between the hepatocytes. These join to form thin bile ductules, which, in turn, enter the bile ducts in the portal tracts. These then combine to form the right and left hepatic ducts, which leave each liver lobe. The hepatic ducts join at the porta hepatis to form the common hepatic duct. The cystic duct connects the gall bladder to the lower end of the common hepatic duct. The gall bladder lies under the right lobe of the liver, and stores and concentrates hepatic bile; its capacity is approximately 50 mL. The common bile duct (CBD) is formed at the junction of the cystic and common hepatic ducts, and is 8 mm or less in diameter, passing through the head of the pancreas and narrowing at its lower end to pass into the duodenum. The CBD and pancreatic duct open into the second part of the duodenum, most often through a common channel at the ampulla of Vater, which contains the smooth muscle sphincter of Oddi. This contracts rhythmically and prevents all of the bile from entering the duodenum, by maintaining a higher pressure than the gall bladder in the fasting state.

Gallstones

Gallstone disease represents a major cause of patient morbidity, particularly in the Western world, and has a major impact on healthcare economics. Gallstones may be present at any age but are unusual before the third decade. Their prevalence is strongly influenced by both age and sex. There is a progressive increase in the presence of gallstones with age and the prevalence is two to three times higher in females than in males, although this difference is less marked in the sixth and seventh decades. At this age the prevalence ranges

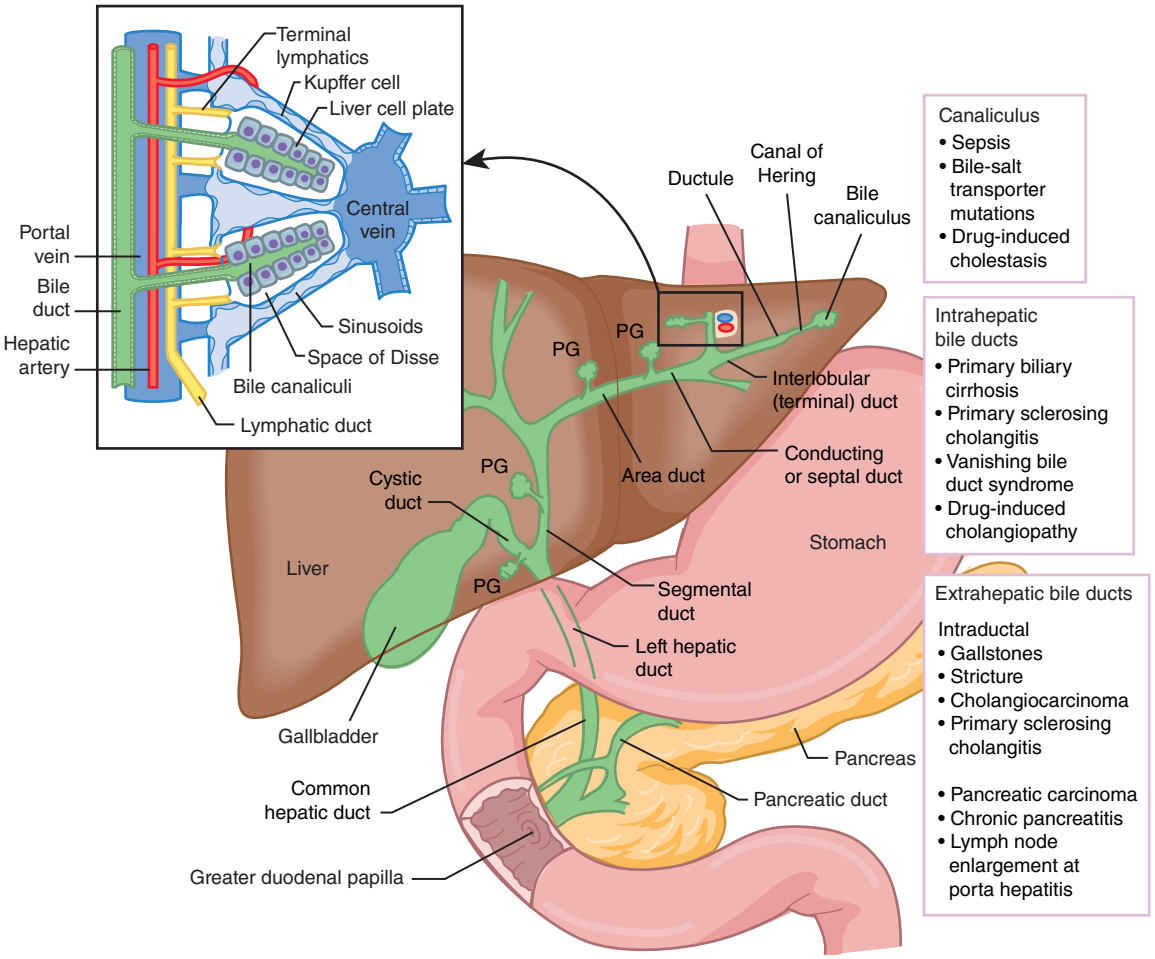


Fig. 24.1 Structure of the biliary system. The various structures of the biliary system are shown schematically. The large segmental ducts have small adjacent peribiliary glands (PG). The finer branches are highlighted in the box at the upper left of the image. The boxes at the right show various biliary tract diseases at the specific levels affected. *PBC,* Primary biliary cholangitis; *PSC,* primary sclerosing cholangitis. *(From Lefkowitch, J. H. [2021].* Scheuer's liver biopsy interpretation *[10th ed.]. Elsevier, p. 59.)*

between 20% and 30%. The increase in life expectancy is reflected in a greater burden of symptomatic gallstone disease. There are considerable racial differences, gallstones being more common in Scandinavians, South Americans, and Native North Americans but less so in Asian and African groups. These racial differences may reflect both genetic and dietary factors. Some of these differences are being eroded by the adoption of Western diets containing high cholesterol in countries with emerging economies.

Pathogenesis

Types of gallstone

The large majority of gallstones are of two types: cholesterol stones (containing >50% cholesterol) and, less frequently, pigment stones, predominantly composed of calcium bilirubinate or polymer-like complexes with calcium, copper, and some cholesterol.

Cholesterol gallstones

This type of stone accounts for 85% of gallstones in the Western world. The formation of cholesterol stones is the consequence of cholesterol crystallization from gall bladder bile. This is dependent on three factors:

- cholesterol supersaturation of bile
- crystallization-promoting factors within bile
- motility of the gall bladder.

The majority of cholesterol is derived from hepatic uptake from dietary sources. However, hepatic biosynthesis may account for up to 20%. The rate-limiting step in cholesterol synthesis is β-hydroxyβ-methylglutaryl-CoA (HMG-CoA) reductase, which catalyzes the first step: that is, the conversion of acetate to mevalonate.

- The cholesterol formed is cosecreted with phospholipids into the biliary canaliculus as unilamellar vesicles.
- Cholesterol will crystallize into stones only when the bile is supersaturated with cholesterol relative to the bile salt and phospholipid content. This can occur as a consequence of excess cholesterol secretion into bile, which, in some instances, has been shown to be due to an increase in HMG-CoA reductase activity.
- Increased secretion of cholesterol into bile has also been associated with insulin resistance and the metabolic syndrome (see Chapter 37).
- A high-cholesterol diet increases biliary cholesterol secretion and decreases bile salt synthesis and the bile salt pool in cholesterol gallstone subjects but not in controls.

These findings suggest that increased intestinal uptake of cholesterol could play a role in gallstone pathogenesis. In support of this observation, pharmacological inhibition of cholesterol absorption prevents gallstone formation in a mouse model. Ezetimibe is a highly selective intestinal cholesterol absorption inhibitor, suppressing the uptake of dietary and biliary cholesterol across the brush border membrane of the enterocyte. This may offer a potential therapy for the management/prevention of gallstone formation in patients and it remains under investigation.

While cholesterol supersaturation of bile is essential for cholesterol stone formation, many individuals in whom such supersaturation occurs will never develop stones. It is the balance between cholesterol crystallizing and solubilizing factors that determines whether cholesterol will crystallize out of solution. A number of lipoproteins have been reported as putative crystallizing factors.

Statins are effective in the treatment of hypercholesterolaemia by competitively inhibiting HMG-CoA reductase. There is emerging evidence that statins reduce cholesterol secretion into bile and, by doing so, promote gallstone dissolution. The prevalence of gallstone disease in patient groups taking statins for the management of hypercholesterolaemia appears to be reduced.

Leptin (see Chapter 37) has been shown to increase cholesterol secretion into bile. While obesity is a risk factor for cholelithiasis, rapid weight loss (for instance, due to following a very low-calorie diet) also results in an increased risk of cholesterol gallstones, which may relate to *relative* elevations in leptin levels observed in those affected.

An alternative mechanism of supersaturation is a decreased bile salt content, which may occur as a consequence of bile salt loss (e.g., terminal ileal resection or ileal involvement with Crohn's disease).

The ***composition of the bile salt pool*** may also influence the ability to maintain cholesterol in solution. There is evidence that an increased proportion of deoxycholic acid (a hydrophobic bile acid) in the bile acid pool may predispose to cholesterol stone formation. This has been linked with slow colonic transit, during which cholic acid, the primary bile acid, may undergo microbial enzyme metabolism, yielding deoxycholic acid, which is then absorbed back into the bile salt pool (see **Fig. 23.5**).

Evidence from ***epidemiological***, family, and twin studies points to the role of genetic factors in gallstone formation. A number of lithogenic genes have been identified, which may interact with environmental factors. The process of bile formation is maintained by a network of adenosine triphosphate-binding cassette (ABC) transporters in the hepatocyte canalicular membrane, which enables biliary secretion of cholesterol, bile salts, and phospholipids. This process is regulated by the nuclear receptors farnesoid X receptor (FXR) and liver X receptor (LXR). Loss-of-function mutations in specific ABC transporter genes have been associated with cholesterol gallstones secondary to bile salt and phospholipid deficiencies within the nascent bile (**Fig. 24.2**). However, monogenic susceptibility appears uncommon. There are rare cases in which a single missense mutation of the multidrug-resistant *(MDR3)* gene has been associated with extensive intra- and extrahepatic cholelithiasis at an early age (<40 years).

Gall bladder motility represents a further factor that may influence the cholesterol crystallization from supersaturated bile. There is evidence from animal models that gall bladder stasis leads to cholesterol crystallization mediated by hypersecretion of mucin. Abnormalities of gall bladder motility have been suggested as factors in such circumstances as pregnancy, multiparity, and diabetes, as well as octreotide-related gall bladder stones (see Chapter 34). Recognized risk factors for cholesterol gallstones are shown in **Box 24.1**.

Bile pigment stones

The pathogenesis of pigment stones is entirely independent of cholesterol gallstones. There are two main types of pigment gallstone: black and brown.

Black pigment gallstones are composed of calcium bilirubinate and a network of mucin glycoproteins that interlace with salts, such as calcium carbonate and/or calcium phosphate. These stones range in colour from deep black to very dark brown and have a glass-like cross-sectional surface on fracturing. Because hyperbilirubinaemia is the critical risk factor, black stones are associated with all major ***haemolytic anaemias***, such as spherocytosis, sickle cell disease, and thalassaemia, and also with subclinical haemolysis from prosthetic valve replacements, malaria, hypersplenism from hepatic cirrhosis, and foot trauma in long-distance runners. An increased prevalence of black pigment gallstones is also seen

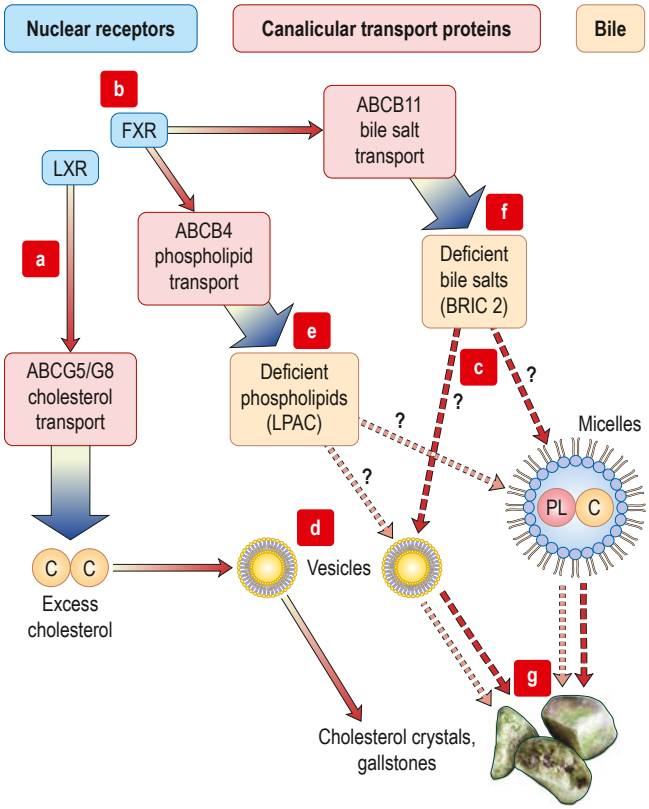


Fig. 24.2 Nascent bile formation at the hepatocytic canalicular membrane, biliary cholesterol solubilization, and gallstone formation. **(a)** ABCG5/G8 transports cholesterol into bile (regulated by nuclear receptor LXR). **(b)** ABCB11 and ABCB4 transport bile salts and phosphatidylcholine into bile (regulated by nuclear receptor FXR). **(c)** Bile cholesterol is solubilized in mixed micelles or contained in cholesterol–phospholipid vesicles. **(d)** Most gallstones are caused by excess biliary cholesterol secretion and subsequent crystallization from supersaturated vesicles (continuous lines). **(e)** Low phospholipid (PL)-associated cholelithiasis (LPAC) occurs if there is deficient phospholipid in bile due to function mutations in gene ABCB4. **(f)** Benign recurrent intrahepatic cholestasis (BRIC) type 2 is characterized by deficient bile salts in bile due to function mutations in gene ABCB11. **(g)** In both LPAC and BRIC type 2 there is an increased risk of gallstone formation.

Box 24.1 Risk factors for cholesterol gallstones

- Increasing age
- Sex (F>M)
- Family history and genetics
- Multiparity
- Obesity ± metabolic syndrome
- Rapid weight loss
- Diet (e.g., high in animal fat/low in fibre)
- Drugs (e.g., contraceptive pill)
- Ileal disease or resection
- Liver cirrhosis
- Spinal cord injury
- Diabetes mellitus
- Acromegaly treated with octreotide
- Total parental nutrition

in Gilbert's syndrome (see Chapter 23), which is associated with enhanced biliary secretion of monoglucuronosyl bilirubin.

There is evidence that bile salt loss into the colon (consequent on ileal resection or ileal disease) promotes solubilization and colonic reabsorption of bilirubin. This enhances the enterohepatic circulation and biliary secretion of bilirubin with the formation of gallstones. Pigment stones have also been linked with bacterial colonization of the biliary tree. Some pigment stones have been shown to contain bacteria, many of which produce glucuronidase and phospholipase, factors that are known to facilitate stone formation. It is speculated that this subclinical bacterial colonization of the bile duct is responsible for pigment stone formation.

Brown pigment stones are usually of a muddy hue and, on cross-section, seem to have alternating brown and tan layers. These stones are composed of calcium salts of fatty acids, as well as calcium bilirubinate. They are almost always found in the presence of bile stasis and/or biliary infection. Brown stones can form in any part of the biliary tree secondary to chronic stasis and the presence of anaerobic bacterial infection.

The ***Oriental hepatolithiasis syndrome (recurrent pyogenic cholangitis)*** is the most serious manifestation of pigment stone disease. Biliary strictures are formed, possibly due to bacterial infection or fluke infestation within the extrahepatic and intrahepatic bile ducts. *Ascaris lumbricoides*, *Clonorchis sinensis*, and *Opisthorchis viverrini* are parasites most commonly associated with this condition. Brown stones may also be the cause of recurrent bile duct stones following cholecystectomy and are also found in the intrahepatic bile ducts in stenosing biliary disease such as Caroli's syndrome and primary sclerosing cholangitis (PSC).

Clinical features

The majority of gallstones are asymptomatic and remain so during a person's lifetime. Gallstones are increasingly detected as an incidental finding at the time of either abdominal radiography or ultrasound scanning (Fig. 24.3). Over a 10- to 15-year period, approximately 20% of these stones will be the cause of symptoms and 10% will involve severe complications. Once gallstones have become symptomatic, there is a strong trend towards recurrent complications, often of increasing severity. Gallstones do not cause dyspepsia, fat intolerance, flatulence, or other vague upper abdominal symptoms.

The clinical syndromes associated with gallstones are shown in Fig. 24.4.

Biliary colic

Biliary colic is the term used for the pain associated with the temporary obstruction of the cystic or CBD by a stone usually migrating from the gall bladder. Despite the term 'colic', the pain of stone-induced ductular obstruction is of sudden onset, severe but constant (not like a 'colic'), and has a crescendo characteristic. Some patients relate the symptoms to over-indulgence with food, particularly when it has a high fat content. The most common time of day for such an episode is in the mid-evening, lasting until the early hours of the morning.

The initial site of pain is usually the epigastrium but there may be a right upper quadrant component. Radiation may occur over the right shoulder and right subscapular region.

Nausea and vomiting frequently accompany more severe attacks. An attack may be terminated spontaneously after a number of hours or by the administration of opiate analgesia. More protracted pain, particularly when associated with fevers and rigors,

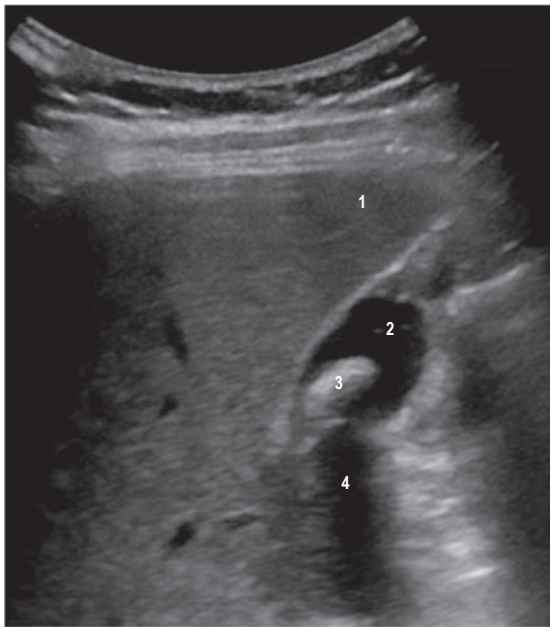


Fig. 24.3 Abdominal ultrasound image demonstrating a gallstone in the gall bladder. This is casting an acoustic shadow, as the ultrasound waves are unable to travel through the gallstone; the structures beneath are therefore not seen. (1) Liver; (2) gall bladder; (3) gallstone; (4) acoustic shadow.

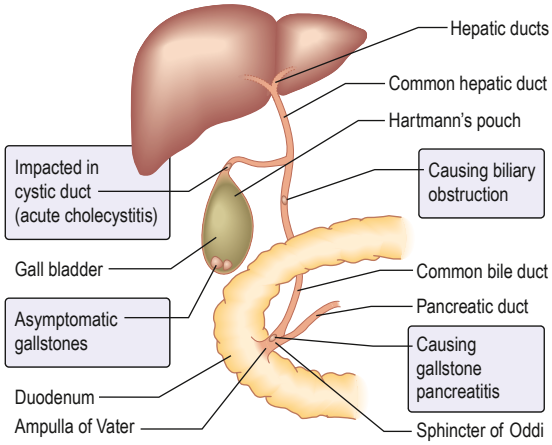


Fig. 24.4 Clinical presentation of gallstones.

suggests secondary complications such as cholecystitis, cholangitis, or gallstone-related pancreatitis (see later).

Acute cholecystitis

The initial event in acute cholecystitis is the obstruction of gall bladder emptying. In 95% of cases a gall bladder stone can be identified as the cause (Fig. 24.5). Such obstruction results in an increase of gall bladder glandular secretion, leading to progressive distension that, in turn, may compromise the vascular supply to the gall bladder.

There is also an inflammatory response secondary to retained bile within the gall bladder. Infection is a secondary phenomenon following this sequence of obstructive, inflammatory, and vascular events.

The initial clinical features of an episode of cholecystitis are similar to those of biliary colic described earlier. Over a number of hours, however, there is progression with severe localized right upper quadrant abdominal pain corresponding to parietal peritoneal involvement in the inflammatory process. The pain is associated with tenderness and muscle guarding or rigidity. This is frequently manifest by Murphy's sign, consisting of pain on taking a deep breath when the examiner's fingers are on the approximate location of the gall bladder. Occasionally, the gall bladder can become distended by pus (an empyema) and, rarely, an acute gangrenous cholecystitis develops, which can perforate, with generalized peritonitis.

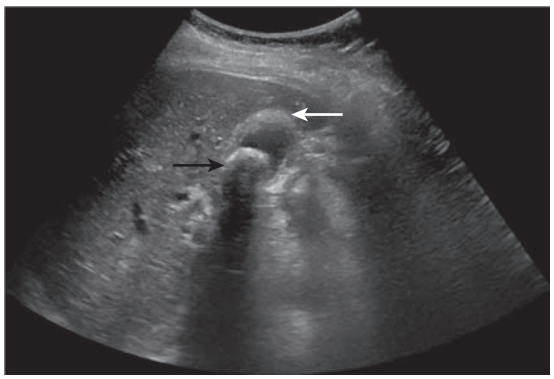


Fig. 24.5 Abdominal ultrasound image demonstrating acute cholecystitis. There is gall bladder wall thickening *(white arrow)* and a large gallstone *(black arrow)*.

Investigations

Biliary colic as a consequence of a stone in the neck of the gall bladder or cystic duct is unlikely to be associated with significant abnormality of laboratory tests. ***Abdominal ultrasound scan*** is the single most useful investigation for the diagnosis of gallstone-related disease (see Fig. 24.3).

Acute cholecystitis is usually associated with a moderate leucocytosis and raised inflammatory markers (e.g., C-reactive protein, CRP). Serum bilirubin, alkaline phosphatase, and aminotransferase levels may be marginally elevated in the presence of cholecystitis alone, even in the absence of bile duct obstruction. More significant elevation of the bilirubin and alkaline phosphatase is in keeping with bile duct obstruction. Abdominal ultrasound has a positive predictive value of 92% and a negative predictive value of 95% in patients with a clinical history in keeping with acute cholecystitis and positive Murphy's sign. Look for:

- ***gallstones*** within the gall bladder, particularly when these are obstructing the gall bladder neck or cystic duct
- ***focal tenderness*** elicited by the ultrasound probe placed over the underlying gall bladder
- ***thickening of the gall bladder wall***, which may also be seen with hypoalbuminaemia, portal hypertension, and acute viral hepatitis.

Gallstones are a common finding in an ageing population, and in the absence of specific symptoms great care should be taken when determining whether the gallstones are responsible for any symptoms.

Differential diagnosis

Typical cases of biliary colic are usually suspected on the clinical history. The differential diagnosis includes irritable bowel syndrome (spasm of the hepatic flexure), carcinoma of the right side of the colon, atypical peptic ulcer disease, renal colic, and pancreatitis.

The differential diagnosis of acute cholecystitis includes a number of other conditions marked by severe right upper quadrant pain

and fever: for example, acute episodes of pancreatitis, perforated peptic ulceration, or an intrahepatic abscess. Conditions above the right diaphragm, such as basal pneumonia, as well as myocardial infarction, may mimic the clinical picture on occasion.

Management

Cholecystectomy

Cholecystectomy is the treatment of choice for virtually all patients with symptomatic gall bladder stones. In patients admitted with specific gallstone–related complications (see later), cholecystectomy should be carried out during the period of that admission to prevent the risk of recurrence. For those presenting with pain alone, an elective procedure can be planned but the waiting time should be minimized to avoid the high risk of recurrent symptoms (approximately 30% over 4 months) and the need for another hospital admission.

Cholecystectomy should not be performed in the absence of typical symptoms just because stones are found on investigation. There is an ongoing debate as to whether prophylactic cholecystectomy is justified in young patients found to have small stones. Such patients have a long period over which they may develop symptomatic disease and small stones are an independent risk factor for the potentially serious complication of gallstone pancreatitis. The most recent guidance available does not recommend prophylactic surgery for asymptomatic gall bladder stones.

The laparotomy approach to cholecystectomy has now been replaced by the laparoscopic technique. Postoperative pain is minimized with only a short period of ileus and the early ability to mobilize the patient. Laparoscopic cholecystectomy can be safely carried out on a day-care basis in an elective setting in otherwise fit patients. This has considerable cost benefits over open cholecystectomy, which is now reserved for a small proportion of patients with contraindications, such as extensive previous upper abdominal surgery, ongoing bile duct obstruction, or portal hypertension.

In approximately 5% of cases, a laparoscopic cholecystectomy is converted to an open operation because of technical difficulties: in particular, adhesions in the right upper quadrant or difficulty in identifying the biliary anatomy.

Specific complications of cholecystectomy

These include a biliary leak, either from the cystic duct or the gall bladder bed. Injury to the bile duct itself occurs in up to 0.5% of laparoscopic operations and may have serious long-term sequelae in the form of a bile duct stricture and secondary biliary liver injury. Injuries to the extrahepatic biliary tree are more likely with variant duct anatomy, such as a low cystic duct insertion or draining of the cystic duct into the right posterior ducts rather than the CBD. There is an overall mortality of 0.2% associated with cholecystectomy.

Acute cholecystitis

The initial management is conservative, consisting of keeping the patient nil by mouth and administering intravenous fluids, opiate analgesia, and intravenous antibiotics. Bacteria that are commonly associated with cholecystitis include *Escherichia coli*, *Bacteroides fragilis*, *Klebsiella*, *Enterococcus*, and *Pseudomonas* species. Antibiotic selection is dictated by local policy and is guided by the severity of any associated sepsis and the result of any blood culture. Options include extended-spectrum cephalosporins (e.g., ceftriaxone) in combination with metronidazole, piperacillin/tazobactam, or imipenem/meropenem.

Cholecystectomy is usually delayed for a few days to allow the symptoms to settle but can then be carried out quite safely in the majority of cases.

When the clinical situation fails to respond to this conservative management, particularly if there is increasing pain and fever, an empyema or gangrene of the gall bladder may have occurred. In this circumstance, urgent imaging (transabdominal ultrasound or CT scan) is required to define the pathology. Surgical intervention is usually needed, although a radiologically placed gall bladder drain can be used as a temporizing measure for the management of an empyema. The placement of a drain is the best option in patients who are considered unfit for early surgery (because of ongoing sepsis or significant co-morbidity). EUS-guided placement of a lumen apposing metal stent (LAMS) between the obstructed gall bladder and the stomach or duodenum has now been adopted by many centres to drain gall bladder empyema when surgery is not feasible.

Stone dissolution and shock wave lithotripsy

These nonsurgical techniques for the management of gall bladder stones are used infrequently but may have a role in a few highly selected patients who may not be fit for cholecystectomy or have declined the surgical option.

Stone dissolution

Pure or near-pure cholesterol stones can be solubilized by increasing the bile salt content of bile. Most experience is with oral ursodeoxycholic acid. The approach requires long-term therapy and the recurrence rate of gallstones is high when therapy is stopped. Additional pharmacological tools for treating cholesterol gallstones include cholesterol-lowering agents that inhibit hepatic cholesterol synthesis (statins) or intestinal cholesterol absorption (ezetimibe), or drugs acting on specific nuclear receptors involved in cholesterol and bile acid homeostasis (see Chapter 23).

Extracorporeal shock wave lithotripsy (ESWL)

A shock wave can be directed either radiologically or by ultrasound on to gall bladder stones. This technique has been shown to have high rates of technical success but only in a restricted patient population. The cystic duct requires patency for stone fragments to pass and rates of post-fragmentation biliary colic are high (~50%). Recurrence rates are high but may be reduced by the pharmacological approaches discussed earlier. The role of extracorporeal shock wave lithotripsy (ESWL) in managing gall bladder stones is therefore not clear.

Post-cholecystectomy syndrome

This refers to right upper quadrant pain, often biliary in type, which occurs a few months after the cholecystectomy but may be delayed for several years. The patients often comment that the pain is identical to that for which the original operation was carried out. In many cases, this syndrome is related to a functional large bowel disorder with colonic spasm at the hepatic flexure (hepatic flexure syndrome). It may be speculated that a functional gut disorder may have been responsible for the original pain that led to the cholecystectomy. In a small proportion of patients the pain is the result of a retained CBD stone; in a further subsection sphincter of Oddi dysfunction (SOD) is a potential cause.

Sphincter of Oddi dysfunction

Sphincter of Oddi dysfunction (SOD) is a clinical syndrome of biliary or pancreatic pain and is caused by either:

- sphincter of Oddi stenosis, for example following prior instrumentation or passage of stones; *or*

- sphincter of Oddi hypertension.

A full history of the type of pain is performed, as well as an assessment of serum liver biochemistry and lipase or amylase during, and between, episodes of pain.

Biliary SOD

In ***type I*** disease, patients have biliary-type pain with *both* raised serum bilirubin, alkaline phosphatase, or aminotransferases during episodes *and* a distended CBD on imaging.

In ***type II*** disease, patients have biliary-type pain and *either* abnormal liver biochemistry during episodes *or* a distended CBD.

In ***type III*** disease, there is recurrent biliary-type pain but no abnormality in the liver biochemistry and a normal calibre CBD.

Management is as follows:

- Type I SOD usually responds to endoscopic sphincterotomy.
- Type II SOD sometimes responds to endoscopic sphincterotomy.
- Type III SOD does not respond to endoscopic sphincterotomy and therefore this should not be performed.

Pancreatic SOD

In ***type I*** disease, patients have recurrent pancreatic pain, a raised serum lipase or amylase, and a dilated pancreatic duct.

In ***type II*** disease, patients have recurrent pancreatic pain with *either* transient elevation of pancreatic enzymes *or* a dilated pancreatic duct.

In ***type III*** disease, patients have pain but neither an elevated lipase or amylase nor a dilated pancreatic duct.

Management is as follows:

- Types I and II SOD may respond to endoscopic sphincterotomy.
- Type III SOD with pain only does not respond to endoscopic sphincterotomy.

Common bile duct stones

The classical features of CBD stones are biliary colic, fever, and jaundice (acute cholangitis). This triad is present only in a minority of patients. Abdominal pain is the most common symptom and has the typical features of biliary colic (see earlier). Jaundice is a variable accompaniment and is almost always preceded by abdominal pain. A patient with bile duct stones may experience sequential episodes of pain, only some of which are accompanied by jaundice. In contrast to malignant bile duct obstruction, the level of jaundice associated with CBD stones characteristically tends to fluctuate. In the elderly or immunocompromised patient, cholangitis may present with very non-specific symptoms, and only associated abnormal liver biochemistry may point to the diagnosis.

Fever is present only in a minority of cases but indicates biliary sepsis and sometimes associated septicaemia. The presence of biliary sepsis is an adverse prognostic factor.

A minority of patients with bile duct stones are discovered incidentally during imaging for gall bladder disease or other intra-abdominal pathology. Some 15% of patients undergoing cholecystectomy will have stones within the bile duct that are detected only at the time of operative cholangiography. The frequency of asymptomatic bile duct stones resulting in complications is not well documented. It is likely that many such stones will pass into the duodenum without causing symptoms. However, the potential for serious complication is well recognized; in most circumstances, incidentally identified bile duct stones are removed (see later).

Examination

If the patient is examined between episodes, there may be no abnormal physical finding. During a symptomatic episode the patient may be jaundiced with a fever and associated tachycardia. There is tenderness in the right upper quadrant, varying from mild to extremely severe.

More widespread abdominal tenderness extending from the epigastrium to the left upper quadrant, associated with distension, may indicate associated stone-related pancreatitis (see later).

Investigations

Laboratory tests

- ***Full blood count*** is usually normal in the presence of uncomplicated bile duct stones.
- ***An elevated neutrophil count*** and raised inflammatory markers (erythrocyte sedimentation rate [ESR] and CRP) are frequent accompaniments of cholangitis.
- ***The raised serum bilirubin*** tends to be mild and often transient. Very high concentrations of bilirubin (>200 μmol/L) almost always reflect complete bile duct obstruction.
- ***Serum alkaline phosphatase and γ-glutamyl transpeptidase*** are similarly elevated in proportion to the degree of hyperbilirubinaemia.
- ***Aminotransferase levels*** are usually mildly elevated, but with complete bile duct obstruction there may be very marked rises of 10 to 15 times the normal value. The alanine aminotransferase is characteristically higher than the aspartate aminotransferase. These high levels may lead to an initial misdiagnosis of a hepatitic process.
- ***Serum lipase or amylase levels*** are often mildly elevated in the presence of bile duct obstruction but are markedly so if stone-related pancreatitis has occurred.
- ***Prothrombin time*** may be prolonged if bile duct obstruction has occurred and is sustained over several days; this reflects decreased absorption of vitamin K.

Imaging

- ***Transabdominal ultrasound*** is the initial imaging technique of choice. ***Bile duct obstruction*** is characterized by dilation of intrahepatic biliary radicles, which are usually readily detected by the ultrasound scan. It may, however, not be possible to identify the cause of obstruction. Stones situated in the distal CBD are poorly visualized by transabdominal ultrasound and up to 50% are missed. The detection of stones within the gall bladder is poorly predictive of the cause of bile duct obstruction. Asymptomatic gallstones are common (up to 15%) in patients who are 65 years and older. Conversely, in 5% to 10% of patients with bile duct stones, no calculi can be seen within the gall bladder.
- ***Magnetic resonance cholangiography (MRC)*** delineates the fluid column within the biliary tree and is a sensitive technique for the detection of CBD stones in the presence of a dilated duct. The technique may be less accurate in the absence of duct dilation (see the role of EUS, later) (Fig. 24.6).
- ***CT scanning*** is an alternative way to detect bile duct dilation. Opaque stones are more readily identifiable within the bile duct than radiolucent cholesterol stones. CT scanning also provides a means of excluding other causes of bile duct obstruction, such as carcinoma of the head of the pancreas.
- ***EUS*** (Fig. 24.7) has enabled high-resolution imaging of the CBD, gall bladder, and pancreas, although, unlike the preceding imaging techniques, it is an invasive procedure. The EUS probe in the duodenum is in close proximity to the distal CBD and hence can identify the large majority of stones at this level. This technique is particularly useful for identifying small calculi (microcalculi) in a non-dilated CBD.

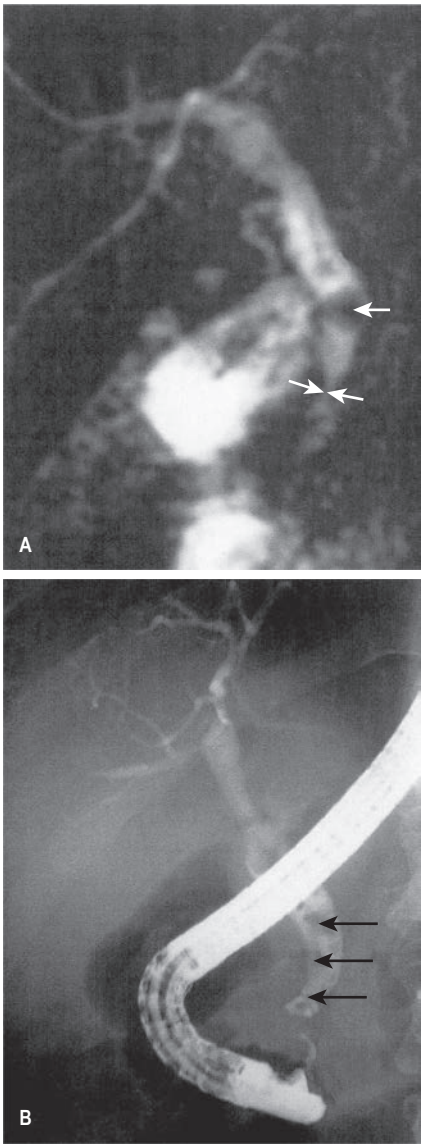


Fig. 24.6 A patient presenting with abdominal pain and jaundice. **(A)** A magnetic resonance cholangiogram (MRC) shows evidence of a distal common bile duct stricture *(bottom arrows)* with a stone in the bile duct proximal to this *(top arrow)*. **(B)** An endoscopic retrograde cholangiopancreatogram (ERCP) in the same patient confirms the details identified in the MRC scan. The stones *(arrowed)* were removed at the time of the ERCP after initial dilation of the stricture.

- ***Endoscopic retrograde cholangiography (ERC).*** In experienced hands, visualization of the CBD will be successful in 98% of cases, providing good documentation of bile duct stones (Fig. 24.8; see also Fig. 24.6) and the facility to immediately proceed to sphincterotomy and stone clearance. However, small stones can still be missed. ERC is an invasive procedure with recognized risks. In almost all circumstances, this is now a therapeutic tool used to remove stones that have been identified by the less invasive investigations described earlier.

Differential diagnosis

Cholangitis may occur independently of gallstones in any condition associated with impaired biliary drainage. It is commonly linked with primary abnormalities of the biliary tree, such as sclerosing

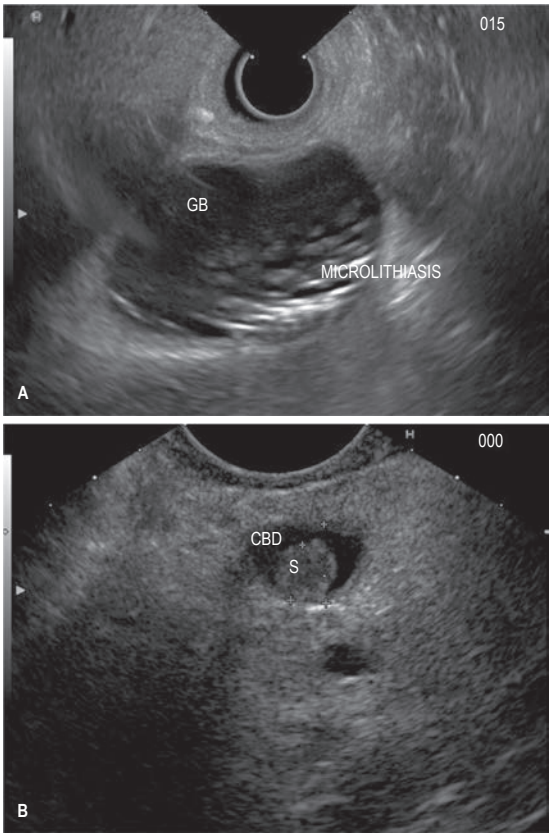


Fig. 24.7 An endoscopic ultrasound (EUS) scan with the probe in the first part of the duodenum. **(A)** The gall bladder (GB) and multiple small, echo-poor stones within (microlithiasis). The patient had presented with recurrent episodes of unexplained abdominal pain. **(B)** The common bile duct (CBD) and a stone (S) clearly identified within the lumen.

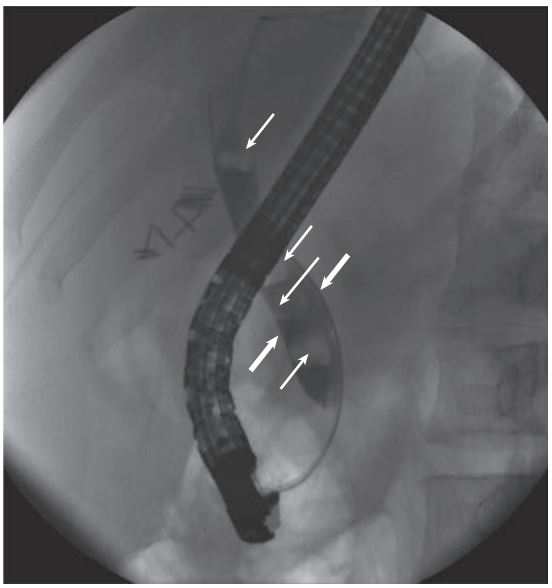


Fig. 24.8 An endoscopic retrograde cholangiopancreatogram (ERCP) carried out in a patient with abdominal pain, fluctuating jaundice, and fever. The cholangiogram shows a dilated common bile duct *(between thick arrows)* and multiple stones within *(arrowed)*. These stones were removed at the time of ERCP by means of balloon and basket retrieval. Note the multiple clips to the left of the upper bile duct, which were placed at a prior laparoscopic cholecystectomy.

cholangitis and Caroli's syndrome (a congenital disorder leading to ectasia/dilation of the intrahepatic bile ducts). Cholangitis may also complicate post-traumatic or surgery-associated bile duct strictures. It is unusual in malignant bile duct obstruction unless there has been prior endoscopic or surgical intervention. Jaundice may also be a feature of acute cholecystitis in the absence of bile duct stones. A stone impacted in the neck of the gall bladder (Hartmann's pouch) or the cystic duct itself may compress and obstruct the bile duct (Mirizzi's syndrome). CBD stones may produce pain, but in the absence of jaundice, the differential diagnosis is that of biliary colic (see earlier).

Management

Acute cholangitis has a high morbidity and mortality, particularly in the elderly or those with serious co-morbidity. Successful management depends on intravenous antibiotics (as for acute cholecystitis) and urgent bile duct drainage. In most circumstances the latter is achieved by the endoscopic retrograde approach (ERC, see earlier). Access to the bile duct is achieved by sphincterotomy and thereafter the stones can be removed by either balloon or basket catheters. In the severely ill patient a plastic stent (a small tube prosthesis) can be inserted into the bile duct to maintain bile drainage without the need to remove the stones, minimizing the time required to complete the procedure. The residual stones can then be cleared and the stent removed endoscopically when the patient has recovered from the acute episode. If endoscopic drainage is not available or is prevented by an inability to access the second part of the duodenum, a radiologically placed percutaneous transhepatic biliary drain (PTBD) represents an alternative management option. Surgical drainage during an acute cholangitic episode has been associated with a high mortality and has been replaced by the endoscopic or percutaneous approach.

Urgent endoscopic bile duct clearance is also indicated in some patients with acute gallstone pancreatitis but only when this coexists with persisting bile duct obstruction (see later). Patients who have retained CBD stones after a previous cholecystectomy are also optimally managed by endoscopic clearance.

Patients shown to have CBD stones as well as gall bladder stones may be treated using different approaches:

- ***At the time of laparoscopic cholecystectomy***, which can also include exploration of the CBD via the cystic duct or via direct choledochotomy. By using these techniques the surgeon can extract stones from the CBD. The skills and time required for bile duct exploration are considerably greater than those for laparoscopic cholecystectomy alone. This single-procedure approach minimizes the length of hospital admission. Concerns around an increased risk of bile duct injury with bile duct exploration have not been confirmed in controlled studies.
- ***An independently timed endoscopic approach***, either immediately before or after the cholecystectomy. Removal of CBD stones by this method is preferred in the UK. Large bile duct stones (>10 mm) may present a significant challenge to endoscopic removal. Mechanical lithotripsy facilitates stone fragmentation and removal into the duodenum. Extracorporeal shock wave stone fragmentation has been used but is not widely available. The recently increased availability of endoscopic cholangioscopy (with direct visualization of the bile duct) has facilitated intraductal lithotripsy, utilizing electrohydraulic or laser probes.
- ***A combined laparoscopic and endoscopic approach***, in which an ERC and duct clearance are carried out under the same anaesthetic used for the laparoscopic cholecystectomy. This approach leads to a shortened hospital stay and the need for only one anaesthetic.

Complications of gallstones

- Acute cholecystitis and acute cholangitis are discussed in the 'Acute cholecystitis' and 'Common bile duct stones' sections.
- Gallstone-related pancreatitis is discussed in the 'Gallstone pancreatitis' section.
- Gallstones can occasionally erode through the wall of the gall bladder into the intestine, giving rise to a biliary enteric fistula. Passage of a gallstone through to the small bowel can give rise to an ileus or true obstruction.
- There is little evidence that gallstones are associated with an increased risk of adenocarcinoma of the gall bladder (see 'Carcinoma of the gall bladder' section).

MISCELLANEOUS CONDITIONS OF THE BILIARY TRACT

Gall bladder

There are a number of non-calculous conditions of the gall bladder, some of which have been associated with symptoms.

Acalculous cholecystitis

Almost 10% of gall bladders removed for biliary symptoms are shown to have chronic inflammation within the wall but an absence of gallstones. Such cases are described as acalculous cholecystitis. In many instances the gall bladder inflammation is minor and of doubtful significance. In some cases, chemical inflammation of the gall bladder may occur from reflux of pancreatic enzymes back into the biliary tree, usually through the common channel at the ampulla of Vater. Bacterial and viral infections of the gall bladder have been recognized as a cause of acalculous cholecystitis. The decision to carry out cholecystectomy in the absence of defined gall bladder stones should be guided by the specific features of the history and by evidence of a diseased gall bladder wall on ultrasound scanning.

A distinct subtype of acalculous cholecystitis is characterized by severe necroinflammation of the gall bladder and generally occurs in an elderly and already critically ill group of patients, usually after major trauma or surgery. Around 20% of these patients develop inflammatory masses with subsequent cholestasis and jaundice; gall bladder perforation is a frequent complication. Morbidity and mortality are very high, and aggressive management of sepsis with antibiotics and cholecystostomy (percutaneous gall bladder drainage) or urgent cholecystectomy is required.

Cholesterolosis of the gall bladder

In cholesterolosis, cholesterol and other lipids are deposited in macrophages within the lamina propria of the gall bladder wall. These may be diffusely situated, giving a granular appearance to the gall bladder wall, or on occasion may be more discrete, giving a polypoid appearance (see later). Cholesterolosis of the gall bladder may coexist with gallstones but also occurs independently. Some degree of gall bladder cholesterolosis is found in up to 26% of laparoscopic cholecystectomy specimens, and as an incidental finding in up to 12% of autopsies in an elderly population. Interestingly, in contrast to cholelithiasis, rates are equal in males and females. It is doubtful whether isolated gall bladder cholesterolosis is a cause of symptoms.

Adenomyomatosis of the gall bladder

Adenomyomatosis is a gall bladder abnormality characterized by hyperplasia of the mucosa, thickening of the muscle wall, and multiple intramural diverticula (the so-called 'Rokitansky–Aschoff sinuses').

The condition is usually detected as an incidental finding during investigation for possible gall stone disease. It has been suggested that this condition is secondary to increased intraluminal gall bladder pressure but this is not proven. Gallstones frequently coexist, particularly when the adenomyomatosis is in a segmental distribution, but there is no evidence to support a direct relationship. It is unlikely that adenomyomatosis alone is a cause of biliary symptoms and it is not considered a pre-malignant state.

Chronic cholecystitis

There are no symptoms or signs that can conclusively be shown to be due to chronic cholecystitis. Symptoms attributed to this condition are vague, such as indigestion, upper abdominal discomfort, or distension. There is no doubt that gall bladders studied histologically can show signs of chronic inflammation and, occasionally, a small, shrunken gall bladder is found either radiologically or on ultrasound examination. However, these findings can be seen in asymptomatic people and therefore this clinical diagnosis should not be made. Most patients with chronic right hypochondrial pain are suffering from functional bowel disease (see Chapter 22).

Extrahepatic biliary tract

Primary sclerosing cholangitis

In up to 40% of patients with primary sclerosing cholangitis (PSC; see also Chapter 23) the clinical course is influenced by a dominant hilar or distal biliary stricture. This is particularly relevant in those patients who do not have established advanced liver involvement, in whom maintenance of bile flow may protect the liver from secondary biliary injury. Drainage with a surgical hepaticojejunostomy has been beneficial in some cases, but outcomes of such restricted surgery in PSC are generally inferior to those of orthotopic liver transplantation. Repeated endoscopic balloon dilation of the dominant stricture, with or without temporary short-term stenting, has been associated with sustained improvement in jaundice and even prolonged transplant-free survival. A significant minority of dominant strictures in PSC, particularly those at the hilum, represent development of an associated cholangiocarcinoma (see later). Development of cholangiocarcinoma in this context carries a very poor prognosis and surgical resection is rarely feasible. A benign dominant stricture in a non-cirrhotic patient, with recurrent cholangitis, refractory jaundice, and pruritus, may be an indication for orthotopic liver transplantation. Superadded cholangiocarcinoma in PSC was previously an absolute contraindication to orthotopic liver transplantation in most healthcare systems. However, liver transplantation is now an option in a small subset of PSC patients who develop hilar cholangiocarcinoma.

Autoimmune cholangitis

Immunoglobulin (Ig) G4–associated cholangitis is the biliary manifestation of a multisystem fibro-inflammatory disorder in which affected organs have a characteristic lymphoplasmacytic infiltrate rich in IgG4-positive cells (see Chapter 6). The original description of this condition was in the context of autoimmune pancreatitis and around 70% of these patients have evidence of IgG4 cholangiopathy (see 'Autoimmune chronic pancreatitis' section). However, IgG4 cholangiopathy can exist in the absence of pancreatic involvement.

The large majority of cases are recognized in middle-aged or elderly men. Presentation is varied, depending on the systems involved, but may include abdominal pain and jaundice. Both intra- and extrahepatic biliary strictures may be seen and the findings may be misinterpreted as representing cholangiocarcinoma or PSC. The diagnosis relies on clinical suspicion, confirmation of an elevated serum IgG4 level, a typical lymphocytoplasmic infiltrate on histological examination of involved tissue, and clinical response to glucocorticosteroid treatment. The condition is almost always responsive to steroids but can lead to hepatic failure.

Biliary cysts (choledochal malformation)

Cystic malformations may occur anywhere in the biliary tree, although they are most commonly extrahepatic. The resulting dilation of the bile duct may be of saccular, diverticular, or fusiform configuration. In many cases there is an associated abnormal pancreatobiliary junction—a congenital malunion where the pancreatic duct drains directly into the CBD. The majority of symptomatic cases present in childhood with features of cholangitis, jaundice, or a palpable mass. The formation of stones and sludge within the cystic segment may predispose to acute relapsing pancreatitis. In adult life, choledochal cysts may be a differential diagnosis in patients presenting with symptoms suggestive of bile duct stones. Extrahepatic bile duct cysts must be fully resected to avoid recurrent biliary sepsis, as well as reduce the risk (approximately 15%) of subsequent cholangiocarcinoma.

Benign bile duct strictures

Benign strictures are a recognized complication of biliary surgery. They may result from inadvertent direct stapling of the duct, or may be a secondary consequence of ischaemic injury (often in association with a bile duct leak). Strictures may also occur at the level of any bile duct anastomosis, either enteric or duct-to-duct. Biliary stricturing is also a rare complication of major trauma to the right upper quadrant. The inflammation and fibrosis of chronic pancreatitis commonly impinge on the intrapancreatic portion of the CBD (see later). This can result in cholestasis, jaundice, and cholangitis.

In most cases, initial therapy includes endoscopic balloon dilation of the stricture and temporary bile duct stenting (see later). This may provide definitive management, but in some cases, surgical intervention with hepaticojejunostomy is required.

Haemobilia

Haemobilia describes blood in the biliary tree. This may be a consequence of liver trauma or a complication of liver surgery. Biopsy of the liver and erosion of a gallstone or hepatobiliary tumour into adjacent vascular structures are also well-recognized causes. The result is a fistula between a blood vessel and a bile duct.

Haemobilia may be a cause of significant gastrointestinal blood loss and should be suspected when melaena is accompanied by right-sided upper abdominal pain and jaundice, particularly in the context of recent hepatobiliary intervention. However, the bleeding may occur without any overt biliary symptoms. If the diagnosis is suspected, bleeding may be managed by occlusion of the feeding blood vessel by radiological embolization. Some patients will require surgery to control the bleeding point.

Tumours of the biliary tract

Gall bladder polyps

Polyps of the gall bladder are a common finding, being seen in approximately 4% of all patients referred for hepatobiliary

ultrasonography. The vast majority of these are small (<5 mm) and non-neoplastic; they are inflammatory in origin or composed of cholesterol deposits (see earlier). Adenomas are the most common benign neoplasm of the gall bladder. Only a proportion of these have cancerous potential. The only reliable predictor of malignant risk is polyp size (>10 mm). Cholecystectomy is recommended for any polyp of 10 mm or larger in diameter.

Carcinoma of the gall bladder

Adenocarcinoma of the gall bladder represents 1% of all cancers. The mean age of occurrence is the early sixties, with a female-to-male ratio of 3:1. Gall bladder stones are often found in association with gall bladder cancer; gallstones have been suggested as an aetiological factor but this relationship remains unproven. Diffuse calcification of the gall bladder (porcelain gall bladder), considered to be the end stage of chronic cholecystitis, has been associated with cancer of the gall bladder and is an indication for early cholecystectomy. Adenomatous polyps of the gall bladder in excess of 10 mm in diameter are also recognized as pre-malignant lesions (see earlier).

Carcinoma of the gall bladder may be detected incidentally at the time of planned cholecystectomy for gallstones; in such circumstances, resection of an early lesion may be curative. Radical surgery with negative resection margins offers the only potential cure. However, early lymphatic spread to the liver and adjacent biliary tract precludes curative resection in more advanced lesions. Palliative chemotherapy treatment may be given, usually with gemcitabine (a nucleoside analogue) and cisplatin (a platinum-containing anticancer drug), with some evidence of a modest improvement in survival which is improved further with the addition of durvalumab (an immune checkpoint inhibitor). A small proportion of cases are sensitive to radiotherapy. However, the generally advanced stage of disease at presentation means that the overall 5-year survival is less than 5%.

Cholangiocarcinoma

Cancers of the biliary tree (see also Chapter 23) are classified as intrahepatic (above the hilum of the liver) or extrahepatic (involving the hilum or bile duct distal to the hilum). The latter are classified either on the TMN classification (see **Box 41.12**) or on the site of the lesion (Bismuth–Corlette classification of biliary strictures). These malignancies represent approximately 1% of all cancers. A number of associations have been identified, such as that with choledochal malformation (see earlier) and that with PSC (see 'Primary sclerosing cholangitis' section). Chronic infection of the biliary tree with parasitic liver flukes, particularly *O. viverrini* or *C. sinensis*, has also been strongly implicated in areas where they are endemic. Extrahepatic bile duct malignancy usually presents with jaundice and may be confirmed on imaging tests, initially ultrasound and thereafter CT and, in particular, magnetic resonance cholangiopancreatography (MRCP; **Fig. 24.9**). Typical findings are of a bile duct stricture with proximal biliary dilation, with or without a visible mass. Histopathological diagnosis often proves difficult because the malignant cells are relatively few and contained within a dense stroma. Endoscopically obtained cytology specimens have only 30% sensitivity. This yield can be enhanced by using additional endoscopic sampling techniques, such as transpapillary biopsy, and analytical enhancements, such as fluorescent *in situ* hybridization. The recent application of direct endoscopic cholangioscopy has enabled direct visualization of biliary lesions and targeted biopsy.

Cholangiocarcinoma is often detected at a late stage and is characterized by early perineural, vascular, and lymphatic spread.

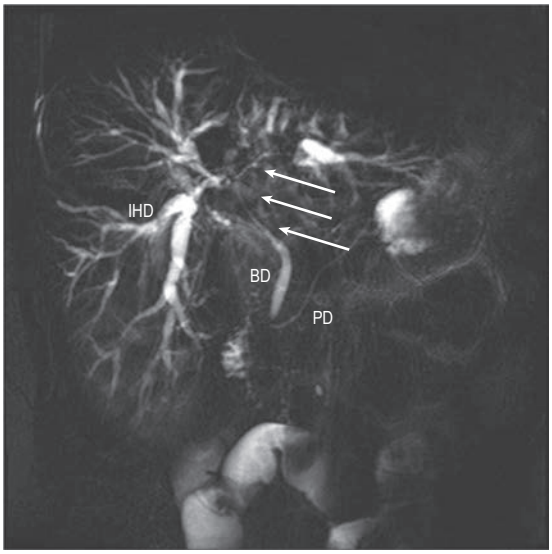


Fig. 24.9 Magnetic resonance cholangiopancreatography (MRCP) image of hilar cholangiocarcinoma extending into the left main duct. There is a normal-calibre distal bile duct (BD) and pancreatic duct (PD). The common hepatic duct and left main duct are strictured *(arrowed)*. There is marked intrahepatic duct (IHD) dilation.

In a minority of cases, complete surgical resection is feasible, offering the only chance of cure. Cholangiocarcinoma of the CBD may be amenable to a limited bile duct resection. Very distal lesions require a pancreatoduodenectomy (Whipple procedure), and perihilar lesions frequently require partial hepatic resection in addition to biliary resection. In an increasing number of centres neoadjuvant chemoradiation therapy followed by orthotopic liver transplantation has been used to cure localized hilar cholangiocarcinoma. Results from this approach show promise, but liver transplantation is an option for only a minority of cholangiocarcinoma patients in healthcare systems where it is available.

The majority of patients with cholangiocarcinoma are treated palliatively with biliary decompression (see later) and gemcitabine- and cisplatin-based chemotherapy regimens, more recently with the addition of durvalumab.

Secondary malignant involvement of the biliary tree

Carcinoma of the head of the pancreas frequently presents with CBD obstruction and jaundice. Metastases to the bile duct from distant cancers are uncommon. Melanoma is the most frequent neoplasm to do so. Infiltration of the bile duct is not uncommon in disseminated lymphoma. Other carcinomas that may give rise to bile duct metastases, in order of frequency, are those arising in the lung, breast, and colon, as well as those from the pancreas (metastatic as compared to direct infiltration).

Management

Palliation of malignant bile duct obstruction

All patients must be fully staged for operability using the imaging techniques described earlier. However, in the greater proportion of patients the treatment is palliative. Relief of bile duct obstruction has been shown to improve quality of life considerably and, with pain control, is the mainstay of effective palliation. Effective biliary decompression is also critical in jaundiced patients who wish to proceed with palliative chemotherapy. ERC permits the insertion of

stents into the biliary tree to re-establish bile flow. The initial use of plastic stents has largely been replaced by self-expanding metal stents that have considerably longer periods of patency (Fig. 24.10). In the small proportion of patients in whom ERC and bile duct drainage is not possible, the percutaneous route offers an alternative method of stent placement under radiological control. The placement of LAMS endoscopically using EUS has now emerged as another safe and effective way of decompressing an obstructed, very dilated CBD into the duodenum, when ERC and stent placement is not feasible.

Further reading

Rushbrook SM, Kendall TJ, Zen Y et al. British Society of Gastroenterology guidelines for the diagnosis and management of cholangiocarcinoma. *Gut* 2024; 73:16–46.
Do-Youn Oh, Aiwu Ruth He, Shukui Qin et al. Durvalumab plus gemcitabine and cisplatin in advanced biliary tract cancer. *NEJM Evid* 2022; 1(8).
Roy LJ, van Wanrooij, Michiel Bronswijk, Rastislav Kunda et al. Therapeutic EUS: European Society of Gastrointestinal Endoscopy (ESGE) technical review. *Endoscopy* 2022; 54:310–332.

PANCREAS

Anatomy and function

Structure

The pancreas extends retroperitoneally across the posterior abdominal wall from the second part of the duodenum to the spleen. Anatomically, it is divided into a head, which rests within the concavity of the duodenum; a body, lying behind the base of the stomach; and a tail, which ends abutting the spleen (Fig 24.11). The neck of the pancreas lies between the body and head, and is in front of the superior mesenteric artery and vein. The head of the pancreas surrounds these two vessels. An uncinate process emerges from the lower part of the head, lying behind the superior mesenteric artery. The pancreas consists of exocrine and endocrine cells. The exocrine pancreas comprises 98% of the parenchyma.

The functional unit of the exocrine pancreas is composed of an acinus and its draining ductule. A ductule from the acinus drains into interlobular (intercalated) ducts, which, in turn, drain into the main pancreatic ductal system. The main pancreatic duct itself joins the CBD to enter the duodenum as a short single duct at the ampulla of Vater.

The endocrine component is scattered throughout the gland in the form of pancreatic islets (of Langerhans).

Exocrine function

The pancreatic acinar cells are responsible for the production of digestive enzymes. These include amylase, lipase, colipase, phospholipase, and the proteases (trypsin and chymotrypsin). These enzymes are stored within the acinar cells in secretory granules and are released by exocytosis (Fig. 24.12). The enzymes released by the acinar cells are transported into the duodenum by a high-volume pancreatic secretion, the majority of which is produced by the ductal cells. This fluid has a high concentration of bicarbonate, which neutralizes the gastric acid that has emptied into the duodenum. The neutralization of gastric acid is essential to facilitate pancreatic enzyme activity, which is pH dependent (requiring a neutral pH).

After ingestion of a meal, pancreatic exocrine secretion is regulated by cephalic, gastric, and intestinal stimuli. The cephalic phase is stimulated by behavioural cues related to the sight, smell, and

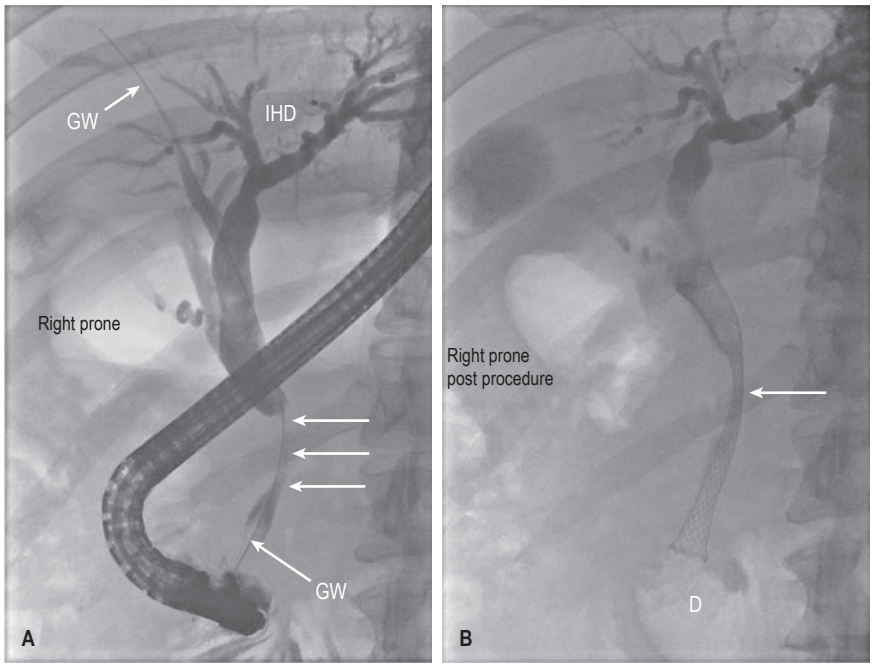


Fig. 24.10 An endoscopic retrograde cholangiopancreatogram in a patient presenting with painless jaundice. **(A)** There is a tight stricture in the mid-common bile duct extending proximally over 4 cm (extent defined by *arrows*). The intrahepatic ducts (IHD) proximally are dilated. A guidewire (GW) has been placed endoscopically across the stricture. **(B)** A self-expanding metal stent has been placed across the stricture and released. The stent is compressed at the level of the stricture *(arrowed)* but will open fully over 24 hours. The contrast in the intrahepatic ducts has largely drained through the stent. The distal margin of the stent is in the duodenum (D).

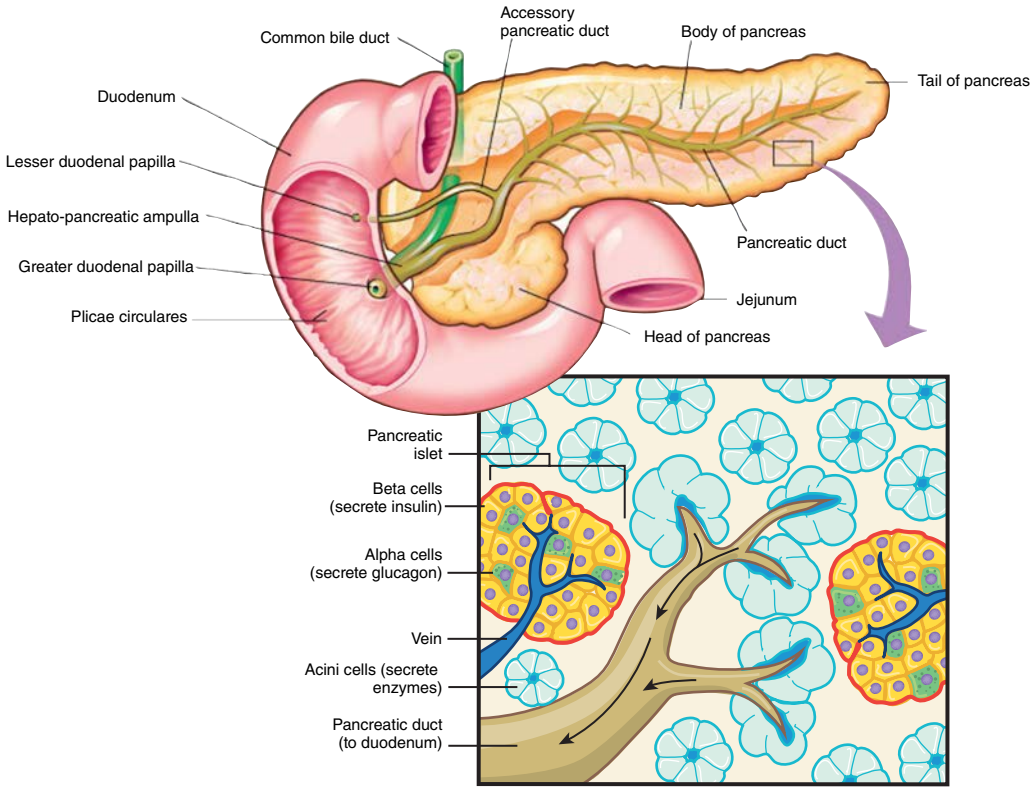


Fig. 24.11 Anatomy and structure of the pancreas. The box at the foot of the image illustrates the microscopic organization of the exocrine (acini) and endocrine (islets) components of the pancreas. The pancreas. *(From Thibodeau GA, Patton KT:* Anatomy & physiology *, ed 5, St. Louis, 2003, Mosby.)*

taste of food. The input from these sensory stimuli is integrated in the central nervous system at the dorsal vagal complex, and the output is transmitted to the exocrine pancreas via the vagus nerve. The gastric phase of pancreatic secretion results from the effects of the meal in the stomach. The major stimulus of pancreatic secretion in the gastric phase is gastric distension, which causes secretion of an enzyme-rich fluid with little secretion of water and bicarbonate. This phase is also under vagal control. The intestinal phase of pancreatic secretion starts when protein, fat, and gastric acid from the stomach enter the duodenum and continues for the duration of the digestive period. It is mediated by both hormones and enteropancreatic vagovagal reflexes. Feedback regulatory events eventually terminate pancreatic secretion.

- ***Cholecystokinin (CCK)*** plays a major role in meal-stimulated digestive enzyme secretion during the intestinal phase of pancreatic secretion. The hormone is produced in specialized gut endocrine cells (I cells) of the mucosa of the small intestine and is secreted in response to intraluminal food. There are no CCK receptors in human pancreatic cells, and CCK acts via receptors on vagal afferent fibres to stimulate pancreatic secretion.
- ***Secretin*** is also released from specialized enteroendocrine cells of the small intestine during a meal and in particular during duodenal acidification. Secretin has a direct effect on the pancreatic acinar cells, as well as the ductal cells. There is also a vagal-mediated secretory response. Secretin action is mediated via G-coupled receptors and calcium-mediated enzyme release. Secretin results in a bicarbonate-rich pancreatic secretion.

The efficiency of the digestive phase of pancreatic secretion requires a negative feedback mechanism to bring the process to a close. Completion of the postprandial secretory phase involves both neural and hormonal control.

- ***Pancreatic polypeptide*** is released from the islet cells of the pancreas in response to a meal and has an inhibitory effect on acinar enzyme secretion, via both a local effect and central receptors.
- ***Somatostatin***, present within the pancreas, stomach, and central nervous system, also has an inhibitory effect. It is released in response to food and its effect is mediated by both direct pancreatic acinar inhibition and a central nervous system inhibitory component.
- ***Peptide YY (PYY)***, contained in endocrine cells of the distal small intestine, is released by nutrients within the ileum and inhibits pancreatic secretion by acting on the acinar cells themselves, as well as centrally via the inhibitory regulation of vagal nerve.

There is also evidence that proteases within the duodenal lumen also have a negative feedback effect on acinar secretion. The gut-related peptides leptin and ghrelin, as well as influencing appetite behaviour, are also regulatory factors in the exocrine function of the pancreas. This effect is believed to occur via hypothalamic centres.

Endocrine pancreas

This consists of hormone-producing cells arranged in nests or islets (islets of Langerhans). The hormones produced are secreted directly into the circulation and there is no access to the pancreatic ductular system. There are ***five main types of islet cell*** corresponding to different secretory components:

- Beta cells are the most common and are responsible for insulin production.

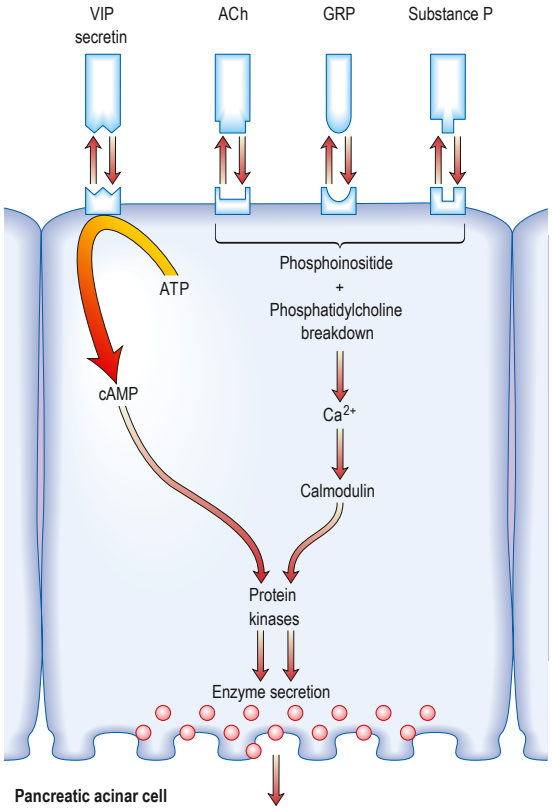


Fig. 24.12 Stimulation of digestive enzyme secretion from the acinar cell. G-protein-coupled receptors have been identified for vasoactive intestinal peptide (VIP), secretin, acetylcholine (ACh), gastrin-releasing polypeptide (GRP), and substance P. There are no cholecystokinin (CCK) receptors in humans. *ATP,* Adenosine triphosphate; *cAMP,* cyclic adenosine monophosphate.

- Alpha cells produce glucagon.
- D cells produce somatostatin.
- PP cells produce pancreatic polypeptide.
- Enterochromaffin cells produce serotonin.

A number of other hormones have been identified within the endocrine pancreas, including gastrin-releasing peptide, neuropeptide Y, and galanin. These are believed to be neurotransmitters active in the neuro-gastrointestinal axis.

Investigations

Assessment of exocrine function

Pancreatic exocrine function is assessed in the investigation of patients with possible chronic pancreatic disease. Clinically evident fat malabsorption does not occur until there has been an 85% to 90% reduction in pancreatic lipase and is therefore a very late manifestation of pancreatic disease.

Direct tests of pancreatic function

These tests rely on the analysis of a duodenal aspirate following pancreatic stimulation, using either naso-duodenal tube or endoscopic aspiration. The original test involved pancreatic stimulation with the oral administration of a specified meal (Lundh meal) then replaced with intravenous secretin and CCK. The aspirate was assessed for pancreatic enzymes and bicarbonate production. These procedures are invasive and require meticulous technique, and so are now rarely used.

Non-invasive indirect tests of pancreatic function

Faecal tests

- ***Faecal elastase*** is an enzyme produced in the pancreas; as it is not degraded in the intestine, it has high concentrations within the faeces. Diminished levels are seen in moderate and severe pancreatic insufficiency. This assay has replaced the faecal chymotrypsin test. It is unaffected by the administration of oral pancreatic enzyme supplements as these do not contain elastase. However, as it relies on the concentration of elastase within the stool sample, it can render false low results in the context of diarrhoea or water contamination during sample collection, due to dilutional effects.
- ***Faecal fat estimation.*** This is rarely performed.

Oral pancreatic function tests

- ***N-benzoyl-L-tyrosyl-p-aminobenzoic acid*** (basis of the PABA test) and ***fluorescein dilaurate*** are oral compounds utilized in pancreatic function tests. They are digested by pancreatic enzymes, releasing substrates that are excreted and measured in the urine. Both tests are commercially available and have good sensitivity in moderate to severe pancreatic exocrine failure.

Clinical application of pancreatic function tests

While the invasive duodenal aspiration tests represent the most sensitive and specific means of assessing pancreatic function, they are very rarely used. The non-invasive tests are widely available but are highly sensitive only in the detection of severe pancreatic insufficiency. The faecal elastase test (in a commercially available form) provides similar sensitivity and specificity, and is the test of choice as a screening tool for pancreatic insufficiency.

Pancreatic imaging

Imaging (see Chapter 23) has a pivotal role in the investigation and management of pancreatic disease, which covers the spectrum of acute, chronic, and malignant conditions.

- ***A plain abdominal X-ray*** may show the calcification associated with chronic pancreatitis, particularly when alcohol is the aetiology.
- ***Transabdominal ultrasound*** of the pancreas usually offers reasonable views of the pancreas; it is a useful screening investigation for inflammation or neoplasia, and is reasonably sensitive for detection of gallstone disease in a patient with pancreatitis. Views may be limited by overlying bowel gas, and ultrasound should not be relied on in the exclusion of pancreatic neoplasia if clinical suspicion exists.
- ***CT scan*** with contrast enhancement and following a specific pancreatic protocol remains the 'gold standard' imaging technique for the investigation of pancreatic disease.
- ***MRI scanning*** represents an alternative to CT. MRCP gives clear definition of the pancreatic duct, as well as the biliary tree. Gallstones may also be identified in the biliary tree using MRI/MRCP.
- ***Endoscopic ultrasound (EUS)*** is very useful for identifying distal CBD stones, which may be the cause of an episode of acute pancreatitis. EUS can identify the early changes of chronic pancreatitis before these are evident with other imaging methods. There is also a role for this technique in staging the operability of pancreatic adenocarcinoma, particularly with respect to vascular invasion. EUS is now considered the imaging technique of choice for investigating cystic lesions of the pancreas (see later). The technique allows fine-needle aspiration or biopsy, as well

as the therapeutic option of cyst drainage using plastic or more recently LAMS. EUS is a sensitive means of detecting small pancreatic tumours, particularly those of neuroendocrine origin.

- ***Endoscopic retrograde cholangiopancreatography (ERCP)*** is an invasive procedure and, while having a good safety profile, is associated with a small risk of serious adverse events: in particular, the precipitation of acute pancreatitis. The availability of MRCP and EUS has provided alternative means of defining pancreatic pathology. ERCP retains a therapeutic role in pancreatic disease (see later).
- ***^{18F}Fluorodeoxyglucose positron emission tomography (FDG-PET)/CT and ^{68}gallium-DOTATATE PET/CT*** are scintigraphic techniques with a high diagnostic accuracy in patients with suspected neuroendocrine tumours, including those of the pancreas. They are also useful in defining the location of a neuroendocrine tumour, identifying multifocal small lesions, evaluating treatment options, and detecting metastatic disease.

In summary, an initial transabdominal ultrasound, supplemented by CT, provides sufficient diagnostic information for most inflammatory and neoplastic conditions of the pancreas. MRI and MRCP are now widely available and provide additional information, particularly with respect to pancreatic ductular and biliary anatomy. EUS is a useful, albeit more invasive, tool for the investigation of both benign and malignant disease of the pancreas, and facilitates fine-needle aspiration and biopsy of targetable lesions.

PANCREATITIS

Classification

Pancreatitis is divided into ***acute*** and ***chronic***. By definition, acute pancreatitis is a process that occurs on the background of a previously normal pancreas and can return to normal after resolution of the episode. Nevertheless, acute alcoholic pancreatitis occurs only in chronic misuse of alcohol. In chronic pancreatitis there is continuing inflammation with irreversible structural changes.

In practice, the differentiation between acute and chronic pancreatitis may be difficult. Any of the causes of acute pancreatitis, if untreated, may result in recurrent episodes, classified as ***acute relapsing pancreatitis*** and eventually bringing about permanent structural changes. In other cases, recurrent episodes of acute pancreatitis may represent exacerbations of an underlying chronic process.

Acute pancreatitis

Acute pancreatitis is a syndrome of inflammation of the pancreatic gland initiated by any acute injury. The causes of acute pancreatitis are listed in **Box 24.2**. In the Western world, gallstones and alcohol account for the vast majority of episodes. Alcohol also causes chronic pancreatitis (see later). The severity of the pancreatitis may range from mild and self-limiting to extremely severe, with extensive pancreatic and peripancreatic necrosis, as well as haemorrhage. In the most severe form (approximately 10% of cases) the mortality is between 40% and 80%.

Pathogenesis

This is still not completely understood. A precipitating event, such as gallstones or alcohol, is thought to induce the acute episode. Whatever the nature of the initiating insult, the pancreatic inflammatory response is secondary to the premature and exaggerated activation of digestive enzymes, principally trypsin, within the pancreas itself. In severe cases the subsequent cascade of autodigestion, microvascular injury, systemic inflammatory response, and bacterial translocation can result in a devastating outcome.

? Box 24.2 Causes of acute pancreatitis

Common
- Gallstones
- Alcohol
- Iatrogenic (post-ERCP, surgery)
- Idiopathic (<10%)

Less common
- Trauma
- Infections, e.g., mumps, Coxsackie B, HIV, adenovirus
- Pancreatic tumours
- Hereditary
- Congenital pancreatic abnormalities, e.g., pancreas divisum
- Metabolic, e.g., hypercalcaemia, hypertriglyceridaemia
- Venom, e.g., scorpion stings, spider
- Sphincter of Oddi dysfunction (see 'Sphincter of Oddi dysfunction' section)
- Miscellaneous
- Drugs (see next section)

Drugs
- Azathioprine/mercaptopurine
- Didanosine
- Oestrogens
- Antibiotics, e.g., tetracycline
- Valproic acid
- Furosemide
- Sulphonamides
- Aminosalicylates
- Corticosteroids
- Metronidazole
- ACE inhibitors
- DPP-4 inhibitors
- Incretin mimetics (GLP-1 agonists)

Gallstone pancreatitis

The inducing effect is obstruction to pancreatic drainage at the ampulla by a stone or associated oedema. In this pathological situation, trypsinogen is cleaved (by cathepsin B) to trypsin, and trypsin degradation by chymotrypsin C is impaired and quickly overwhelmed. Intracellular calcium also increases and may also cause early activation of trypsinogen with upregulation of nuclear factor kappa B, leading to extensive acinar cell damage.

Alcohol-induced pancreatitis

There is evidence that alcohol interferes with calcium homeostasis in pancreatic acinar cells. In addition, activation of pancreatic stellate cells by acetaldehyde occurs, with production of collagen and then matrix proteins.

Clinical features

Acute pancreatitis is a differential diagnosis in any patient with upper abdominal pain. The pain usually begins in the epigastrium, accompanied by nausea and vomiting. As inflammation spreads throughout the peritoneal cavity, the pain becomes more intense. Involvement of the retroperitoneum frequently leads to back pain.

The patient may give a history of previous similar episodes or be known to have gallstones. An attack may follow an alcoholic binge. However, in many cases, there are no obvious aetiological factors.

Physical examination at the time of presentation may show little more than a patient in pain with some upper abdominal tenderness but no systemic abnormalities. In severe disease the patient has a

tachycardia and hypotension, and is oliguric. Abdominal examination may show widespread tenderness with guarding, as well as reduced or absent bowel sounds. Specific clinical signs that support a diagnosis of severe necrotizing pancreatitis include periumbilical ***(Cullen's sign)*** and flank bruising ***(Grey Turner's sign)*** (Fig. 24.13). In patients with gallstone aetiology the clinical picture may also include jaundice or cholangitis.

Diagnosis

Blood tests

- ***Serum amylase*** is a sensitive test to diagnose acute pancreatitis if it is greater than three times the upper limit of normal. It performs best when measured within 24 hours of the onset of pain due to a relatively short half-life. A number of other conditions may occasionally cause a very elevated amylase (Box 24.3). Amylase levels gradually fall back towards normal over the next 3 to 5 days. With a late presentation the serum amylase level may give a false-negative result.
- ***Urinary amylase*** levels may be diagnostic, as they remain elevated over a longer period of time.
- ***Serum lipase*** levels are also raised in acute pancreatitis and remain elevated for a longer period of time than amylase levels, increasing diagnostic sensitivity in patients presenting more than 24 hours after onset of pain. Sensitivity is also increased compared to amylase in alcoholic pancreatitis where repeated episodes of inflammation may result in a progressive depletion of pancreatic enzymes; this is due to the higher content of lipase than amylase in pancreatic acinar tissue. As a result of these perceived advantages, many hospitals have now adopted serum lipase over amylase as their first-line test for acute pancreatitis.
- ***CRP level*** is useful in assessing disease severity and prognosis.
- ***Other baseline investigations*** include a full blood count, serum creatinine and electrolytes, blood glucose, liver biochemistry, plasma calcium, and arterial blood gases. These are documented at presentation and then repeated at 24 and 48 hours, providing a basis for assessing the severity of an attack (see later).

Fig. 24.13 Cullen's sign (periumbilical bruising) and Grey Turner's sign (flank bruising) coexisting in a patient with acute pancreatitis. These signs are observed infrequently in severe pancreatitis as haemorrhagic fluid tracks from the pancreas into the anterior abdominal wall and retroperitoneum. (Coexistence of Cullen's and Grey Turner's signs in Acute Pancreatitis 2009 AJM.)

Box 24.3 Elevation of serum amylase unrelated to pancreatitis

Leakage of upper gastrointestinal contents into the peritoneum
- Upper gastrointestinal perforation
- Biliary peritonitis
- Intestinal infarction

Inherited abnormalities of amylase
- Macroamylasaemia

Reduced clearance
- Acute or chronic renal impairment

Salivary disorders
- E.g., parotitis, salivary gland calculus

Radiology

- An erect ***chest X-ray*** is mandatory to exclude gastroduodenal perforation, which may also raise serum pancreatic enzymes (see Box 24.3). A supine abdominal film is generally not required but may show gallstones (if calcified) or pancreatic calcification.
- An ***abdominal ultrasound scan*** is used as a screening test to identify a possible biliary (gallstone) cause of pancreatitis. Gallstones are difficult to detect in the distal CBD but dilated intrahepatic ducts may occur in the presence of bile duct obstruction. Stones within the gall bladder are not sufficient to justify a diagnosis of gallstone-related pancreatitis. The ultrasound may also demonstrate pancreatic swelling and necrosis, as well as peripancreatic fluid collections if present. In severe pancreatitis the pancreas may be difficult to visualize because of gas-filled loops of bowel.
- ***Contrast-enhanced CT scanning*** is essential in all but the mildest attacks of pancreatitis. It should be performed after 72 hours to assess the extent of pancreatic necrosis. CT provides very valuable prognostic information. Later, repeated CT scans can detect other complications, including fluid collections, abscess formation, and pseudocyst development (Fig. 24.14).
- ***MRI (MRCP)*** assesses the degree of pancreatic damage and identifies gallstones within the biliary tree. MRI is particularly useful to differentiate between fluid and solid inflammatory masses.
- ***ERCP*** is used as a treatment measure to remove bile duct stones in selected cases of gallstone-related pancreatitis (see later).

Assessment of disease severity

The majority of cases of acute pancreatitis are mild and run a short, self-limiting course. However, approximately 25% follow a more complicated course and in 10% this is life-threatening. The revised Atlanta criteria define mild, moderately severe, or severe pancreatitis and are summarized in Box 24.4; they require assessment of organ failure using the Modified Marshall scoring system (Box 24.5). In practice, patients with moderately severe acute pancreatitis should be managed as having severe disease until sustained improvement is seen.

In severe cases mortality may be as high as 80%. The initial clinical course is marked by a systemic inflammatory response syndrome (SIRS), haemodynamic instability, and multiple organ failure. The consequences of SIRS, including respiratory and acute kidney injury, account for most of the mortality within the first 7 days of an episode of severe acute pancreatitis. After 7 days the majority of deaths are linked to delayed complications, such as sepsis due to infected collections or haemorrhage due to inflammatory effects on major vessels.

Accurate identification of patients likely to progress to severe pancreatitis permits appropriate monitoring and intensive care to be put in place. Early clinical assessment has been shown to have poor sensitivity for predicting a severe attack. Similarly, individual laboratory tests have very limited value. Therefore, multiple factors

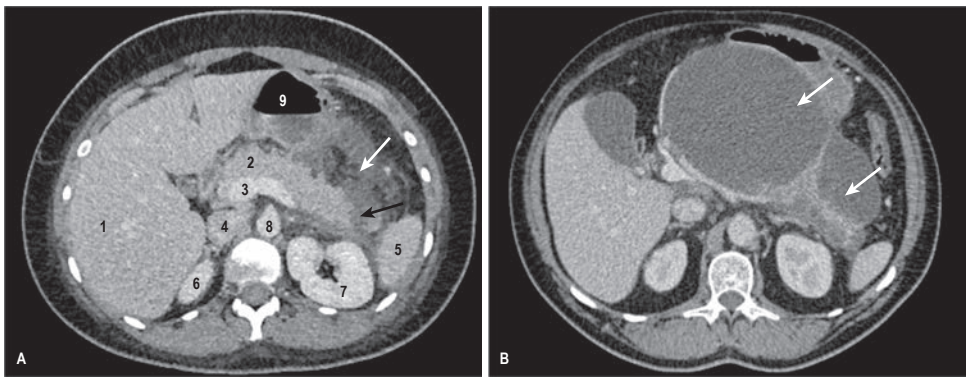


Fig. 24.14 Acute pancreatitis. **(A)** Abdominal computed tomography (CT) demonstrating features of acute pancreatitis. There is marked inflammatory change/free fluid surrounding the pancreas *(white arrow)* and a focal area of necrosis in the pancreatic tail *(black arrow)*. (1) Liver; (2) pancreas; (3) portal vein; (4) inferior vena cava; (5) spleen; (6) right kidney; (7) left kidney; (8) aorta; (9) stomach. **(B)** Abdominal CT in a patient who presented with acute pancreatitis a few weeks prior to the scan demonstrating large pseudocysts *(arrowed)*.

Box 24.4 Severity of acute pancreatitis: revised Atlanta criteria 2013

Mild
- Absence of organ failure
- Absence of local complications

Moderate/severe
- Local complication and/or transient organ failure (<48 hours)

Severe
- Persistent organ failure (>48 hours), defined by the modified Marshall score for the evolution of organ failure (see **Box 24.5**)

have been used to develop clinical scoring systems. The ***Glasgow and Ranson scoring systems*** (**Boxes 24.6 and 24.7**) are based on such parameters and have been shown to have 80% sensitivity for predicting a severe attack, although only after 48 hours from presentation. The ***Acute Physiology and Chronic Health Evaluation (APACHE) II score*** has been extensively adopted as a means of assessing the severity of a wide spectrum of illness. It is based on common physiological and laboratory values, patient age, and the presence or absence of other chronic health problems, including obesity (**Box 24.8**). This scoring system appears to have a high sensitivity and can be applied as early as 24 hours after onset of symptoms. There is evidence that obesity affects the outcome of an episode of pancreatitis, perhaps because adipose tissue is a substrate for activated enzyme activity, worsening the inflammatory reaction. All scoring systems for predicting the severity of acute pancreatitis have proven to have modest utility in clinical practice. However, systematic classification tools can guide consistent clinical assessment and are certainly necessary in clinical studies and audit of outcomes.

Management

The initial management of acute pancreatitis is similar for all causes. A multiple factor scoring system (ideally, APACHE II with a modification for obesity) should be used at the end of the first 24 hours after presentation to allow identification of patients with a predicted severe attack. This should be repeated at 48 hours to identify a further subgroup that appears to be moving into the severe category. These patients should be managed in a high-dependency or intensive care unit and the case should be discussed with a specialist pancreatic unit at an early stage. Even patients outside the severe category may require considerable supportive care.

Early fluid losses in acute pancreatitis may be large, requiring well-maintained intravenous access and a urinary catheter to monitor circulating volume and renal function. Arterial and central venous catheterization may be necessary to facilitate close monitoring of haemodynamic status and permit repeat sampling of arterial blood gases. An initial moderate fluid resuscitation regimen is now preferred to an aggressive fluid resuscitation regimen (due to heightened risks of fluid overload), but careful monitoring of haemodynamic parameters remains essential to avoid under-resuscitation of patients with moderate or severe pancreatitis.

- ***Nasogastric suction.*** This prevents abdominal distension and vomitus, and hence the risk of aspiration pneumonia.
- ***Baseline arterial blood gases.*** These are a key predictive factor for severity of an episode and determine the need for continuous oxygen administration.
- ***Prophylactic antibiotics.*** Controlled data for the use of antibiotics are available but the results are not uniform in showing benefit, particularly in demonstrating improved mortality. There is some evidence that the β-lactam imipenem reduces the incidence of infected pancreatic necrosis. Antibiotics should usually be reserved for cases where cholangitis or infected necrotic tissue is strongly suspected and, if possible, the choice should be guided by blood culture results.
- ***Analgesia requirements.*** Opiates are usually required for immediate post-presentation pain control. Unless there is prompt control of pain, a patient-controlled system of administration is indicated to provide continuous and adequate pain relief. Fentanyl has been used widely for this application. There is a theoretical risk that morphine and diamorphine might exacerbate pancreatic ductular hypertension by causing sphincter of Oddi contraction. Although there are no clinical data to support this risk, some clinicians avoid these drugs in acute pancreatitis.
- ***Feeding.*** In a severe episode there is little likelihood of oral nutrition for a number of weeks. Total parenteral nutrition has been associated with a high risk of infection and so enteral nutrition is preferred. In the absence of gastroparesis, most patients will tolerate nasogastric administration of feed without exacerbation of pain. In those with gastroparesis or poorly tolerated nasogastric feeding (exacerbation of pain or precipitation of nausea and

Box 24.5 Modified Marshall scoring system[a]

Score system	0	1	2	3	4
Respiratory PO_2/FiO_2	>400	301–400	201–300	101–200	≤101
Renal Serum creatinine (μmol/L)	134	134–169	170–310	311–439	>439
Cardiovascular Systolic blood pressure (mmHg)	>90	<90 Fluid-responsive	<90 Not fluid-responsive	<90 pH <7.3	<90 pH <7.2

[a]A score of ≥2 in any system defines the presence of organ failure.

Box 24.6 Glasgow prognostic criteria in acute pancreatitis[a]

Criteria	Measurement
Age	>55 years
White blood cell count	>15 × 10^9/L
Blood glucose	>10 mmol/L
Serum urea	>16 mmol/L
Serum albumin	<30 g/L
Serum aminotransferase	>200 U/L
Serum calcium	<2 mmol/L
Serum lactate dehydrogenase	>600 U/L
P_aO_2	<8.0 kPa (60 mmHg)

[a]More than three positive factors during the first 48 hours suggest severe pancreatitis and a poorer prognosis.

Box 24.7 Ranson criteria in pancreatitis

Criteria	Measurement
At admission	
Age	>55 years
White blood cell count	>16 × 10^9/L
Blood glucose	>11 mmol/L
Serum lactate dehydrogenase	>600 U/L
Serum aminotransferase	>250 U/L
Within 48 hours	
Haematocrit fall	>10%
Serum aminotransferase	>200 U/L
Serum calcium	<2 mmol/L
Serum urea increase	>1.8 mmol/L
Base deficit	>4 Meq/L
P_aO_2	<8.0 kPa (60 mmHg)

vomiting), post-pyloric feeding should be instituted by the endoscopic placement of a nasojejunal tube.

- ***Anticoagulation.*** Low-molecular-weight heparin should be given for prophylaxis of deep vein thrombosis.

In a small proportion of patients, multiorgan failure will develop in the first few days after presentation, reflecting the extent of SIRS. Such patients may require positive-pressure ventilation and renal support (see Chapter 13).

Gallstone-related pancreatitis

In patients with gallstone-related pancreatitis and associated bile duct obstruction (particularly when complicated by cholangitis), endoscopic intervention with sphincterotomy and stone extraction is the treatment of choice. In the absence of bile duct obstruction, sphincterotomy and stone extraction are of proven benefit only when the episode of pancreatitis is predicted to be severe. In less severe cases of gallstone-related pancreatitis the presence of residual bile duct stones can be assessed electively by MRCP or EUS in the recovery phase of the acute episode; if stones are present, they can be removed by ERCP. To prevent a recurrent episode of gallstone pancreatitis, cholecystectomy should be carried out as soon as feasible after the acute episode has resolved.

Complications

These are listed in **Box 24.9**. After the first 7 days, the prognosis of acute pancreatitis is most closely related to the extent of pancreatic necrosis. This can be most accurately assessed by contrast-enhanced CT, which should be carried out in all patients with severe disease after the first week. Extensive necrosis (>50% of

Box 24.8 Acute Physiology and Chronic Health Evaluation (APACHE) II scoring system[a]

Physiological

- Temperature
- Heart rate
- Respiratory rate
- Mean arterial pressure
- Glasgow Coma Scale score
- Presence of acute kidney injury
- Age
- Organ insufficiency
- Immunocompromise

Laboratory

- Oxygenation (P_aO_2)
- Arterial pH
- Serum:
  - Sodium
  - Potassium
  - Creatinine
  - Haematocrit
  - White blood cell count

[a]APACHE is applied within 24 hours of admission. Score can range from 0 to 71 (normal to abnormal). Higher scores correspond to more severe disease and a higher risk of death. Body mass index (BMI) is an additional parameter that can be added specifically in assessing severity of acute pancreatitis.

Box 24.9 Complications of acute pancreatitis

Systemic
- Systemic inflammatory response syndrome (SIRS)
- Multiorgan dysfunction

Pancreas
- Pancreatic fluid collections
- Necrosis±infection
- Pancreatic abscess
- Pancreatic pseudocyst (after 4–6 weeks)

Lungs
- Pleural effusion
- Acute respiratory distress syndrome (ARDS)—hypoxia
- Pneumonia

Kidney
- Acute kidney injury

Gastrointestinal tract
- Gastrointestinal bleeding from gastric or duodenal erosions
- Paralytic ileus

Hepatobiliary
- Jaundice
- Common bile duct obstruction
- Portal vein thrombosis

Metabolic
- Hypoglycaemia
- Hyperglycaemia
- Hypocalcaemia

Haematological
- Disseminated intravascular coagulation (DIC)

the pancreas) is associated with a high risk of further complicated disease, frequently requiring radiological, endoscopic, or surgical intervention.

In this minority of patients with extensive necrosis of the pancreatic and peripancreatic tissues, superimposed infection is associated with a greatly increased risk of mortality. A 'step-up', minimally invasive approach is now preferred, with open surgical debridement avoided if possible. Long courses of antibiotics are given, and percutaneous drainage and EUS-guided endoscopic LAMS placement with necrosectomy are used to clear infected collections. The aims of such interventions are to control infection, evacuate devitalized tissues (the culture medium for invasive infection), and promote conditions for healing.

The best outcomes from intervention are achieved when debridement is delayed until at least 4 weeks after the onset of pancreatitis. When the damaged area has been walled off and liquefaction has begun, a pseudocyst develops.

Prognosis

The vast majority of patients with a mild to moderate episode of acute pancreatitis will make a full recovery with no long-term sequelae. Recurrent episodes of pancreatitis may occur, particularly if there has been any long-term pancreatic ductular damage. Patients with more severe acute pancreatitis may develop pancreatic insufficiency with respect to both exocrine (malabsorption) and endocrine function (diabetes); both carry their own significant, lifelong morbidity.

Chronic pancreatitis

Aetiology

In high-income countries, alcohol is reported to be the only aetiological factor in 60% to 80% of cases. There is a sizeable list of other reported aetiological factors, which can be categorized into toxic–metabolic, genetic, autoimmune, recurrent acute or severe acute pancreatitis, obstructive and idiopathic causes (**Box 24.10**).

Box 24.10 Causes of chronic pancreatitis

- Alcohol
- Chronic kidney disease
- Hereditary:
  - Cystic fibrosis
  - Trypsinogen and inhibitory protein defects
- Tropical chronic pancreatitis
- Obstructive
- Trauma
- Idiopathic
- Hypercalcaemia
- Recurrent acute pancreatitis
- Autoimmune (immunoglobulin G4)

Pathogenesis

There is increasing evidence that an increase in activated trypsin within the pancreas is a common pathway for the development of chronic pancreatitis. This may occur because of premature activation of trypsinogen to trypsin, or of impaired inactivation/clearance of the activated enzyme from the pancreas. It is believed that the increased or prolonged intrapancreatic enzyme activity leads to the precipitation of proteins within the duct lumen in the shape of plugs. These then form a nidus for calcification and cause ductal obstruction, leading to ductal hypertension and further pancreatic damage (**Fig. 24.15**).

In the case of alcohol-related chronic pancreatitis, ethanol increases intracellular calcium, promoting trypsinogen activation to trypsin, as well as diminishing the protective inactivation pathway. The observation that the majority of people drinking excess alcohol do not develop pancreatitis suggests that the disease process is a complex interaction of different mechanisms. It is proposed that the alcohol is only one factor that interacts with other environmental and genetic influences (see later).

Genetic aspects of chronic pancreatitis

Several genetic factors have been identified that influence the process of trypsin activation and inactivation. Cationic trypsinogen is the major form of trypsinogen produced in the pancreas and is encoded by the *PRSS1* gene (**Fig. 24.16**). Gain-of-function mutations of this gene are recognized as the major factor in hereditary pancreatitis, an autosomal dominant condition with high penetrance.

Calcium levels within the pancreas have a role in the process of activation and inactivation of trypsinogen/trypsin and are, in part, modulated by the calcium-sensing receptor (CASR). Mutations coding for this receptor have been associated with pancreatitis and are believed to facilitate the damaging effects of alcohol on the pancreas.

The serine protease inhibitor Kazal type 1 *(SPINK-1)* is a specific trypsin inhibitor and is co-secreted with trypsinogen by the acinar cells. Loss-of-function mutations of the *SPINK-1* gene have been associated with the development of chronic pancreatitis and identified as a factor in the development of tropical pancreatitis (almost certainly interacting with environmental triggers).

Chymotrypsin C is produced in trace amounts by the acinar cells and has also been shown to have a role in trypsin degradation. Loss-of-function mutations of the encoding gene have been identified in patients with chronic pancreatitis.

The cystic fibrosis transmembrane conductance regulator *(CFTR)* is expressed on the apical surface of the acinar cells and

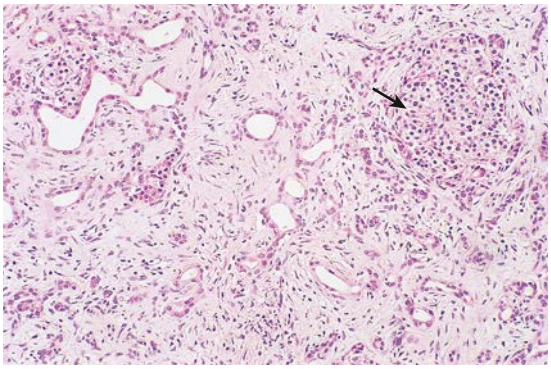


Fig. 24.15 Histology of chronic pancreatitis. There is considerable loss of acini and replacement by fibrosis. Inflammatory cells are relatively inconspicuous at this late stage. Islets of Langerhans (one is *arrowed*) sometimes escape destruction but their loss can result in diabetes mellitus. *(From Underwood JC [ed.]. General and Systematic Pathology, 4th edn. Edinburgh: Churchill Livingstone; 2004, with permission.)*

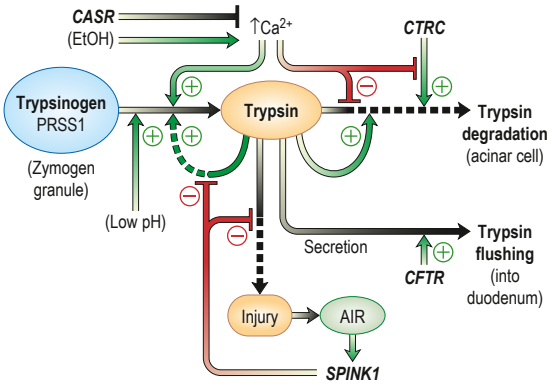


Fig. 24.16 Genetic factors in trypsinogen activation and trypsin inactivation. Early activation of trypsinogen to trypsin within the pancreas is crucial in the development of pancreatitis. Trypsinogen activation is promoted by cationic trypsinogen mutations (*PRSS1+*), active trypsin, high calcium (Ca^{2+}), and low pH. Calcium levels are regulated in part by calcium-sensing receptor (CASR) and dysregulated by ethanol (EtOH). Active trypsin degradation is facilitated by CTRC and by other active trypsin molecules, but blocked by high calcium. Active trypsin within the pancreas leads to pancreatic injury, which causes an acute inflammatory response (AIR). AIR upregulates expression of serine protease inhibitor Kazal 1 (SPINK-1), which blocks active trypsin and therefore prevents further activation of trypsinogens and limits further tissue injury. Cystic fibrosis transmembrane conductance regulator (CFTR) represents an extra-acinar cell mechanism to eliminate trypsin by flushing it out of the pancreatic duct and into the duodenum. *(Adapted from Whitcomb DC. Genetic aspects of pancreatitis.* Ann Rev Med *2010; 61:413–424.)*

is responsible for maintaining a high-volume, bicarbonate-rich pancreatic secretion. This high-volume secretion is responsible for flushing the activated trypsin into the duodenum. The homozygote or complex heterozygote *CFTR* mutations associated with the cystic fibrosis disease state are almost always manifest by perinatal/early pancreatic exocrine failure (see 'Cystic fibrosis' section). An increased frequency of a single *CFTR* mutation in patients with idiopathic chronic pancreatitis has been identified.

The identification of a genetic component to the development of chronic pancreatitis has led to speculation that, in many cases, the evolution of the disease is dependent on the complex interaction of gene–gene and gene–environmental factors.

Autoimmune chronic pancreatitis

Two types of autoimmune chronic pancreatitis (ACP) have been identified. The ***most common variant (type 1)*** is seen predominantly in middle-aged men and is associated with raised serum and tissue levels of IgG4. Other autoantibodies, including those directed towards nuclear and smooth muscle antigens, are also observed. Extrapancreatic tissue involvement is common, including the biliary tree (autoimmune cholangitis; see earlier), as well as thyroid, salivary gland, and renal tissue (see Chapter 6). In all these disorders there is a raised serum IgG4 level and, pathologically, there is a dense lymphoplasmacytic infiltrate with many IgG4-positive plasma cells, a mild to moderate eosinophil infiltrate, and an obliterative phlebitis in some organs. ACP is one of the few settings in which the pathogenesis of the disease may be independent of the activated trypsin pathways.

The ***second variant (type 2)*** tends to occur in early midlife with an equal sex distribution and does not have the autoimmune markers or IgG4-positive cells. Some 30% of cases are associated with inflammatory bowel disease. Disease is much more likely to be restricted to the pancreas and lacks the associations with other organ/tissue involvement seen with the type 1 variant. The hallmark of both types of autoimmune pancreatitis is evidence of responsiveness to steroids (see later).

The presentation of autoimmune pancreatitis is varied, particularly in type 1, in which extrapancreatic disease may predominate. Abdominal pain and weight loss are common features; jaundice may be an early symptom, both secondary to bile duct obstruction by the inflamed head of pancreas, and a manifestation of the cholangitis seen in type 1 cases.

Clinical features

Epigastric pain is the most common presentation of chronic pancreatitis, often radiating into the back. The pattern of pain may be episodic, with short periods of severe pain, or is chronic and unremitting. For those with an alcohol-related aetiology, exacerbations of the pain may follow further alcohol excess, although this is not a uniform relationship. Alcohol excess or meals with a high fat content may also lead to exacerbations of pain, independent of the aetiology.

During periods of abdominal pain, anorexia is common and weight loss may be severe. This is particularly so in those patients with chronic, unremitting symptoms. Exocrine insufficiency may develop at any time and occasionally malabsorption is the presenting feature in the absence of abdominal pain. Diabetes occurs in approximately 30% of cases and is usually a late event in the disease process, almost always following the development of exocrine insufficiency. Jaundice secondary to obstruction of the CBD during its course through the fibrosed head of the pancreas may also occur and may be a presenting feature in a small proportion of patients.

Investigations

The extent to which investigations are required is dependent on the clinical setting.

- ***Serum amylase and lipase*** levels may be elevated, but in advanced disease there may not be sufficient residual acinar tissue to produce this elevation.
- ***Serum IgG4 levels*** should be measured in those cases with suspected autoimmune pancreatitis.

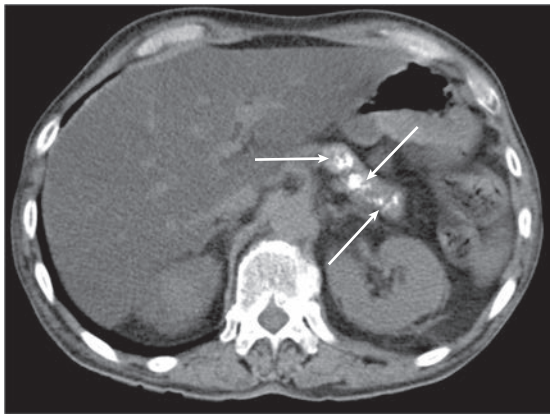


Fig. 24.17 Contrast-enhanced CT scan in a patient with chronic pancreatitis. Multiple calcific densities *(arrowed)* are demonstrated along the line of the main pancreatic duct.

- ***Faecal elastase*** level will be low in the majority of patients with moderate to severe pancreatic disease.
- ***Gene mutation analysis*** should be carried out in selected cases when the aetiology is uncertain. This is most relevant in patients presenting below the age of 40. Common mutations of the *PRSS1*, *SPINK-1*, and *CFTR* encoding genes are available via reference centres.
- ***Transabdominal ultrasound scan*** is frequently used for initial assessment.
- ***Contrast-enhanced CT scanning*** provides a more detailed assessment. In the presence of pancreatic calcification and a dilated pancreatic duct, the diagnosis of chronic pancreatitis can be readily established (Fig. 24.17). This may be much more difficult when these features are not present.
- ***MRI with MRCP*** is utilized to define more subtle abnormalities of the pancreatic duct, which may be seen in non-dilated chronic pancreatitis. Administration of intravenous secretin during an MRCP can afford dynamic images of pancreatic duct distension that offer an indirect measure of pancreatic exocrine function.
- ***MRCP*** has replaced diagnostic ERCP in the diagnosis of chronic pancreatitis. It has particular utility in defining ductal anatomy, strictures, and leaks.
- ***EUS*** is used increasingly when doubt about the diagnosis remains after non-invasive imaging, or specifically when complications of chronic pancreatitis, including pseudocyst formation and the possible development of malignancy, need assessment.

Differential diagnosis

The differential diagnosis is that of pancreatic malignancy. Carcinoma of the pancreas can reproduce many of the symptoms and imaging abnormalities that are commonly seen with chronic pancreatitis. The diagnosis of malignancy should be considered in patients with a short history when there is a localized pancreatic mass. Considerable difficulties may arise when a malignancy develops on the background of established chronic pancreatitis (the latter being a recognized pre-malignant lesion).

High-quality imaging is able to define malignant features with a localized mass lesion, local invasion, and lymph node enlargement. EUS with fine-needle biopsy provides the most accurate diagnosis of a potential mass lesion.

Management

In patients with alcohol-related chronic pancreatitis, long-term abstinence is likely to be of benefit, although this has been difficult to prove. Tobacco smoking is also an independent and strong risk factor for chronic pancreatitis and pancreatic cancer, and smoking cessation is strongly recommended. Autoimmune pancreatitis is steroid-responsive and failure to respond would put the diagnosis in question. Relapse is common when the steroids are withdrawn, and long-term immunomodulators (e.g., azathioprine) may be required.

Abdominal pain

For short-term flare-ups of pain, a combination of a non-steroidal anti-inflammatory drug (NSAID) and an opiate (tramadol) is usually sufficient for symptomatic relief. In patients with chronic unremitting pain, this may be inadequate and also risks opiate dependence.

Tricyclic antidepressants (e.g., amitriptyline) and membrane-stabilizing agents (e.g., pregabalin) are used for chronic pain and reduce the need for opiates. Coeliac axis nerve block may produce good pain relief but is unreliable in its efficacy and duration of action. In the majority of patients, some spontaneous improvement in pain control occurs with time. After a 6- to 10-year period, some 60% of patients will become pain-free. For recurrent severe or debilitating chronic pain, both endoscopic and surgical interventions have been used. Such intervention is indicated in patients with a dilated pancreatic duct upstream of a pancreatic duct stricture or stone. The endoscopic approach has centred on improving duct drainage by removing intraductal stones and duct stenting to maintain patency. ESWL has been used to fragment stones within the head of the pancreas.

Surgical intervention usually involves a duct drainage procedure, which can be combined with partial resection of the diseased pancreas. Trials have reported improved pain control following surgical intervention, as compared with the endoscopic approach. However, many patients have a high level of debility (and often continued alcohol excess) and are unsuitable for major surgery due to co-morbidity or technical factors such as the presence of varices within the operative field. In such circumstances, endoscopic therapy is justified as a first measure and there is no evidence that this adversely influences subsequent surgery.

Malabsorption

The steatorrhoea associated with pancreatic exocrine insufficiency may be intrusive, with up to 30 mmol of fat lost per 24 hours. This will usually improve with pancreatic enzyme supplements. Current preparations are presented in the form of microspheres, which reduce the problems of acid degradation in the stomach. An acid suppressor (H_2-receptor antagonist or proton pump inhibitor) is also given. Despite this, a proportion of patients continue to suffer malabsorption, usually reflecting the inadequate mixing of the pancreatic supplements with the food, as well as the low pH in the duodenum secondary to inadequate pancreatic bicarbonate production. There is no justification to reduce fat intake below the recommended levels of a normal diet, as this will contribute to the malnutrition seen in patients with chronic pancreatitis. Dietetic input is valuable for the management of malabsorption and also for monitoring and alleviating the malnutrition that is often seen secondary to abdominal pain and secondary food aversion.

Diabetes

Diabetes associated with pancreatic endocrine failure is termed type 3C diabetes and may be difficult to control, with a rapid progression

from oral hypoglycaemic agents to an insulin requirement. Labile blood sugar levels are a common problem secondary to inadequate glucagon production by the damaged pancreas.

Autoimmune pancreatitis

Most patients respond to glucocorticoid therapy, for example prednisolone 40 mg daily for 4 to 6 weeks. Relapses are treated with azathioprine or with rituximab.

Specific complications

The most common structural complication of chronic pancreatitis is a ***pancreatic pseudocyst***, a fluid collection surrounded by granulation tissue. These usually occur in relationship to a period of enhanced inflammatory activity within the pancreas giving abdominal pain but may develop silently during what would appear to be a stable phase. Intra- or retroperitoneal rupture, bleeding, or cyst infection may occur. The larger cysts may occlude nearby structures, including the duodenum and the bile duct. In pseudocysts less than 6 cm in diameter, spontaneous resolution can be anticipated. In larger cysts that have been present for a period longer than 6 weeks, resolution is less common and a long-term complication rate of approximately 30% can be anticipated. Many pseudocysts are closely opposed to the posterior wall of the stomach or duodenum and can be successfully drained endoscopically using EUS. A direct fistula is created between the pseudocyst lumen and the gastric or duodenal lumen, which is then kept patent by the insertion of removable stents. This approach will be successful in approximately 75% of cases. Surgical drainage is required for failures of endoscopic therapy or in circumstances where the pseudocyst anatomy does not allow endoscopic access.

Ascites and, occasionally, ***pleural effusions*** can be a direct consequence of chronic pancreatitis when there has been disruption of the main pancreatic duct. A high amylase level in the ascites or pleural fluid confirms the aetiology. Such disruptions of the main pancreatic duct require surgical intervention or ERCP and pancreatic duct stent placement.

There is an increased risk of ***pancreatic cancer*** in patients with chronic pancreatitis. The risk of malignancy is closely related to the duration of the inflammatory process. The highest incidence has been reported in hereditary pancreatitis, with a 50-fold increase and a lifetime risk as high as 40%. This reflects the early onset (in childhood) of the disease. Increases of 20- to 30-fold have been described in patients carrying other gene mutations and early onset of disease. The lifetime risk of malignancy in other causes of chronic pancreatitis, such as alcohol, which develop much later, is 10% to 15%. Cancer surveillance programmes have been proposed for the very high-risk groups (hereditary pancreatitis and other causes of early-onset disease), usually starting around the age of 40 years and relying on yearly imaging and tumour marker measurement.

Cystic fibrosis

Some 85% of people with cystic fibrosis (see Chapter 20) will have pancreatic exocrine failure, and in the majority of these this will develop *in utero* or the perinatal period. Malabsorption and failure to thrive are common presentations in the perinatal period and first year of life, and diabetes will develop subsequently in 30% of these patients during their lifetime.

In the remaining 15% of cases there is sufficient pancreatic exocrine secretion, which may persist throughout the patient's lifetime. This reflects residual *CFTR* function and is associated with class IV and V mutations. These patients tend to present later in life and have less severe pulmonary involvement, although this is not uniform. Some of these cases will progress to pancreatic exocrine insufficiency at variable times in the course of their disease. A small proportion develop symptomatic pancreatitis as part of this process.

The ***management*** of pancreatic exocrine insufficiency in cystic fibrosis is necessary to optimize growth and overall nutrition. ***Pancreatic enzyme supplements*** are closely titrated against the level of steatorrhoea. ***Fat intake*** should be maintained to avoid nutritional deficit. A daily ***lipase intake*** of up to 10 000 units/kg body weight is required. The efficacy of the supplements may be improved by the use of a ***proton pump inhibitor***. These drugs reduce the risk of acid denaturation of the enzymes. Despite optimization of the use of enzyme supplements, a degree of fat malabsorption may persist. This reflects other luminal abnormalities of fat absorption in cystic fibrosis that are related to viscous mucus, small bowel bacterial colonization, and poor mixing of the food with bile.

Further reading

de-Madaria E, Buxbaum JL, Maisonneuve P et al. Aggressive or moderate fluid resuscitation in acute pancreatitis. *N Engl J Med* 2022; 387:989–1000.
Kamisawa T, Zen Y, Pillai J et al. IgG4-related disease. *Lancet* 2015; 385:1460–1471.
Lankisch PG, Apte M, Banks PA. Acute pancreatitis. *Lancet* 2015; 386:85–96.
National Institute for Health and Care Excellence. *NICE Guideline NG104: Pancreatitis*. NICE 2018; https://www.nice.org.uk/guidance/ng104/evidence/full-guideline-pdf-6535536157.
Santhi Swaroop Vege, MD and Suresh T. Chari, MD Chronic Pancreatitis. *N Engl J Med* 2022; 386:869–878.

PANCREATIC CANCER

The many types of pancreatic cancer can be divided into two main groupings. The vast majority of cases occur in the exocrine component of the pancreas. There are several subtypes of exocrine pancreatic cancer but their diagnosis and treatment have much in common. A small minority of pancreatic cancers arise in the endocrine tissue of the pancreas and have different clinical characteristics.

Pancreatic adenocarcinoma

The incidence of pancreatic cancer in the West has been estimated at approximately 10 cases per 100 000, with a slight increase in incidence over the last 10 years. The diagnosis is rarely made in individuals younger than 40 years of age, and the median age at diagnosis is 71 years. Approximately 60% of patients with this condition are male. It is the eighth leading cause of death from cancer in males and the ninth leading cause of death from cancer in females worldwide. Some 96% of pancreatic cancers are adenocarcinoma in type and the large majority are of ductal origin.

Aetiology

Smoking is associated with a twofold increase. Excessive intake of alcohol or coffee and excessive use of aspirin have also been implicated. There is an increased incidence of pancreatic cancer among patients with a history of diabetes and chronic pancreatitis. The risk of developing pancreatic cancer is most marked in those patients with a genetic mutation predisposing to chronic pancreatitis (there is a 50 times increased risk in the presence of a *PRSS-1* mutation; see **Fig. 24.16**). Although it is estimated that 5% to 10% of pancreatic cancers have an inherited component, the genetic basis for familial

aggregation has not been identified in most cases. A subgroup of such high-risk kindreds carry germline mutations of DNA repair genes, such as *BRCA2* and the partner and localizer of *BRCA2*. Among people with a known family history of pancreatic cancer in a first-degree relative, the relative risk of pancreatic cancer developing is increased by a factor of 2, 6, and 30 in people with one, two, and three affected family members, respectively (Box 24.11).

Pathogenesis

Data suggest that pancreatic cancer results from the successive accumulation of gene mutations (Fig. 24.18). The cancer originates in the ductal epithelium and evolves from pre-malignant lesions to fully invasive cancer. The lesion called ***pancreatic intraepithelial neoplasia (PanIN)*** is the best-characterized histological precursor of pancreatic cancer. The progression from minimally dysplastic epithelium (PanIN grades 1A and 1B) to more severe dysplasia (grades 2 and 3) and finally to invasive carcinoma is paralleled by the successive accumulation of mutations. More than 90% of cases of PanIN of all grades have *KRAS* mutations. The mutational inactivation of the *CDKN2A*, *p53*, and *SMAD* family member *(SMAD4)* tumour suppressors is detected with increasing frequency in type 2 and type 3 lesions of PanIN, suggesting that they are rate-limiting events for tumour progression. Exome-sequencing studies have identified additional loss-of-function mutations encoding components of the SWI/SNF nucleosome remodelling complex, which are cumulatively detected in approximately 10% to 15% of pancreatic adenocarcinomas, as well as other, less frequent alterations.

Box 24.11 Relative risks of pancreatic cancer in patients with a family history or associated gene mutations

Risk factor	Relative risk
Familial pancreatic cancer	
Two first-degree relatives affected	18
Three first-degree relatives affected	57
Hereditary pancreatic cancer syndromes	
BRCA2 mutation	5.9
Familial atypical multiple mole melanoma (FAMMM)	16
Peutz–Jeghers's syndrome	36
Hereditary pancreatitis	50

A small percentage of pancreatic adenocarcinomas arise from cystic lesions, including ***intraductal papillary mucinous neoplasm (IPMN)*** and ***mucinous cystic neoplasia*** (see later). There is evidence that these cystic neoplasms demonstrate a similar multistep genetic and histological progression to invasive adenocarcinoma, although there are recognized differences in the mutational events that occur.

Clinical features

Approximately two-thirds of pancreatic cancers are located in the head of the pancreas, and the remainder in the body and tail (Fig. 24.19). Patients with pancreatic cancer most commonly present with abdominal pain, anorexia, and weight loss. Many have experienced low-grade symptoms for a number of months before they present for investigation. New onset of depressive symptoms has also been identified as an early manifestation. In many cases the onset of these non-specific symptoms is at a stage when there is already advanced local or metastatic disease. The pain associated with pancreatic cancer frequently radiates through into the back and, in some cases, is partially relieved by leaning forwards. Jaundice is a common, and maybe an early, manifestation of tumours of the ampulla and pancreatic head. This reflects the occlusion or compression of the distal CBD as it traverses the head of pancreas before entering the duodenum at the level of the ampulla.

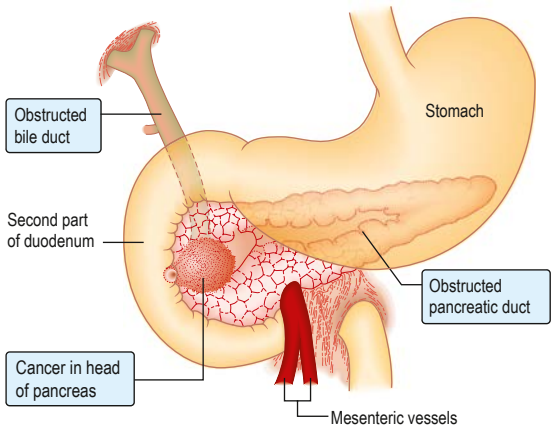


Fig. 24.19 Carcinoma of the head of the pancreas. The close relationship of a carcinoma of the head of pancreas to the surrounding structures.

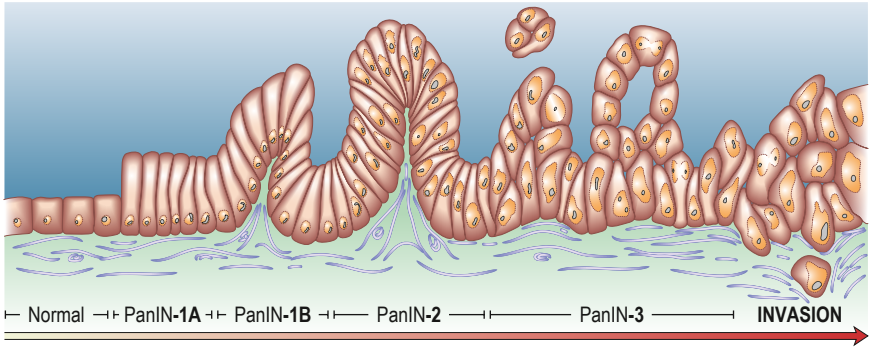


Fig. 24.18 Genetic model for the development of pancreatic ductal adenocarcinoma. The stages indicate histological progression with specific gene alterations. *PanIN,* Pancreatic intraepithelial neoplasia. *(Modified from Maitra A, Adsay NV, Argani P et al. Multicomponent analysis of the pancreatic adenocarcinoma progression model using a pancreatic intraepithelial neoplasia tissue microarray.* Mod Pathol *2003;16:902–912, with permission from the Nature Publishing Group.)*

The patient may have noticed pale stools, dark urine, and itching associated with bile duct obstruction in the absence (or prior to the onset) of detectable jaundice. Diabetes is present in at least 50% of patients with pancreatic cancer and may predate other manifestations of disease. The obstruction of the pancreatic duct by the cancer may also lead to symptoms of malabsorption and overt steatorrhoea. An episode of symptomatic pancreatitis may, on occasion, be the presenting clinical picture. Other unusual presenting features can include thromboembolic phenomena, polyarthritis, and skin nodules. These manifestations, distant to the tumour itself, have not been fully explained but may precede the development of a detectable pancreatic mass lesion by many months.

Signs

At the time of first presentation, there may be an absence of physical signs. Evidence of weight loss is common. Jaundice may be present, with associated excoriations prompted by pruritus. In a proportion of cases the gall bladder will be palpable ***(Courvoisier's sign)*** and secondary to an obstructed bile duct (or cystic duct). A palpable epigastric or central abdominal mass may be present as a reflection of advanced local disease. Liver metastases may be reflected in hepatomegaly.

Diagnosis and investigations

Patients with pancreatic adenocarcinoma frequently present for investigation at a late stage of tumour development, precluding curative surgery. In many cases, symptoms have been present for a number of months but have not led to investigation that is specifically aimed at confirming or excluding a pancreatic cancer. To improve outcomes from this cancer there is increasing emphasis on earlier diagnosis. This is focused on targeted investigation in patients recognized as high-risk (see earlier). More emphasis is also placed on earlier referral for investigation in patients over the age of 40 who have low-grade symptoms that persist after initial assessment and empirical therapy. The upper gut symptoms frequently prompt referral for an upper gastrointestinal endoscopy. In most cases this will not add diagnostic information and it is essential that a negative endoscopy does not delay further investigation—in particular, CT imaging. Increasing efforts have been made to detect possible early pancreatic cancer in patients newly diagnosed with diabetes. This includes NICE guidance on cancer screening that GPs consider an urgent CT scan, or an urgent ultrasound scan if CT is not available, to assess for pancreatic cancer in people aged 60 and over with weight loss and new-onset diabetes. However, it is important to note that less than 1% of patients newly diagnosed with diabetes will be diagnosed with pancreatic cancer within three years of that diabetes diagnosis.

- ***Transabdominal ultrasound*** is the initial imaging investigation in the majority of patients. In the presence of bile duct obstruction this will confirm dilated intrahepatic bile ducts, as well as a mass in the head of the pancreas. Ultrasound is less reliable when the cancer is found in the body and tail of the pancreas because of overlying bowel gas, and has a sensitivity of detection of 60%. A negative transabdominal ultrasound scan should not be considered as excluding the diagnosis of pancreatic cancer.
- ***Contrast-enhanced CT scan*** should confirm the presence of a mass lesion in most cases of pancreatic adenocarcinoma (Fig. 24.20). Lymph node involvement and metastatic disease will also be identified. If there is a high index of suspicion, a dual-phase pancreatic protocol CT scan should be requested from the outset. A high-quality pancreatic protocol CT is required as a staging procedure prior to planned curative surgery. Extending the CT to the chest will exclude pulmonary metastases.
- ***EUS*** is the most sensitive (>85%) non-surgical procedure for the detection of pancreatic cancer (Fig. 24.21). In most cases, it does not add to the diagnostic and staging information provided by the CT scan. However, the procedure is valuable for the definition of small (<2 cm) lesions of the pancreas, which may be missed on CT. EUS is now the approach of choice for obtaining histological confirmation of the underlying malignancy. A histological/cytological diagnosis is essential prior to planned chemotherapy. There are concerns that needle sampling of the tumour prior to a planned curative resection might lead to cancer cell seeding. This has been reported as a rare event (1%–2%).
- ***MRI and PET-CT scanning*** are useful techniques in patients when the local tumour or possibility of metastases has not been adequately defined. PET-CT is generally required in patients going forward for curative resection.

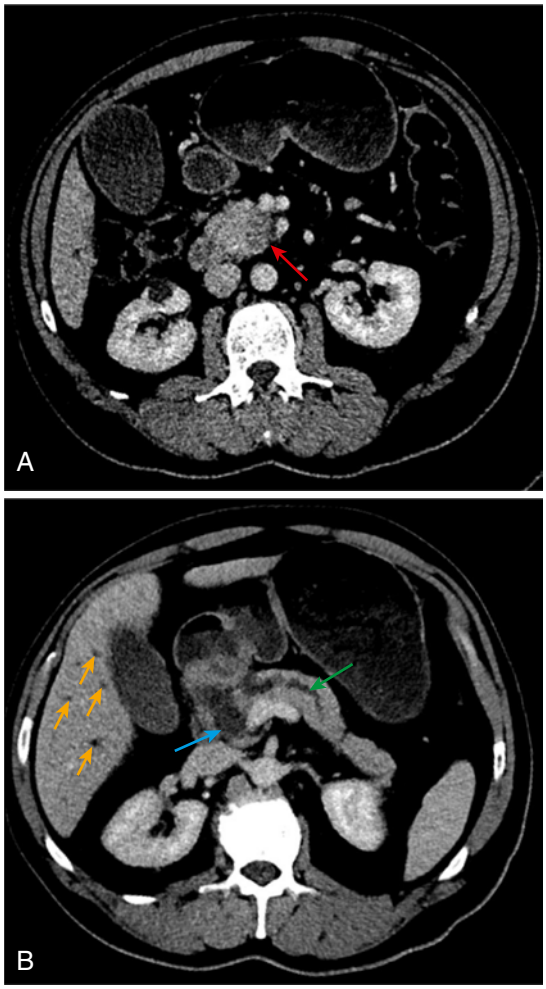


Fig. 24.20 Axial CT images demonstrating pancreatic cancer. **(A)** An image through the pancreatic head demonstrates a low-attenuation lesion in the head of the pancreas that is contacting the superior mesenteric artery and superior mesenteric vein *(arrowed)*. **(B)** A more superior image demonstrates dilation of both the pancreatic duct *(green arrow)* and common bile duct *(blue arrow)* (known as the double duct sign) due to compression of the ducts by the mass. Dilated intrahepatic ducts are also visible in the imaged portion of the liver *(yellow arrows)*.

- ***Several tumour markers*** have been evaluated for the diagnosis and monitoring of pancreatic cancer. The CA19–9 has reasonable sensitivity (80%) but a high false-positive rate. In individual patients, single values of these tumour markers may be of little help but a progressive elevation over time is often diagnostic; in such circumstances, tumour marker levels can be used to monitor response to treatment.

Differential diagnosis

The diagnosis should not be difficult in the presence of painless jaundice or epigastric pain radiating into the back with progressive weight loss. Unfortunately, many patients present with less specific symptoms including lethargy, poorly localized pain, change in bowel habit, and weight loss. Imaging, particularly abdominal CT, should be performed if pancreatic cancer is suspected. IgG4-related autoimmune pancreatitis is now recognized as a differential diagnosis in patients presenting with abdominal pain, jaundice, and an abnormal pancreas on imaging (localized or diffuse enlargement; see earlier). Pancreatic cancer may rarely present with recurrent episodes of typical acute pancreatitis.

Management

Please see Chapter 41.

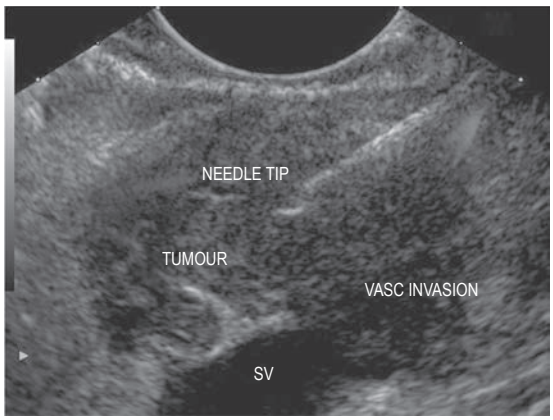


Fig. 24.21 An EUS scan with the probe in the distal stomach in a patient with pancreatic cancer. The patient presented with intractable periumbilical pain. A mass lesion of the pancreatic body is defined (tumour). There is invasion by the lesion into the splenic vein (SV, vasc invasion). The line of a transendoscopic needle is seen (needle tip), which has been placed under the guidance of the ultrasound probe to obtain a needle aspirate for cytological diagnosis.

Pancreatic cystic neoplasms

Cystic lesions of the pancreas are common. Around 75% will be pseudocysts (see 'Pancreatitis/Specific complications') but, of the remainder, the majority are true cystic neoplasms. Careful characterization of lesions with CT and MRI and discussion in a specialist pancreatic multidisciplinary team meeting are crucial. There is a high potential for the development of malignancy in large, true cystic neoplasms and therefore resection is generally recommended. The decision to operate may be difficult in patients with small (<3 cm) lesions of the head of the pancreas (in the absence of confirmed malignancy at presentation) and in those with significant co-morbidity. An initially conservative approach with follow-up imaging is then justified. Differentiation between pseudocysts and true cystic neoplasms may be difficult, even with multiple imaging techniques. EUS and fine-needle aspiration are frequently helpful in characterizing cystic lesions of the pancreas. EUS appearance (**Fig. 24.22**), cytology, and the measurement of cyst fluid carcinoembryonic antigen, glucose, and amylase may help to categorize cystic lesions and identify malignant change.

Serous cystadenomas are composed of multiple small, cystic cavities lined by cuboidal, glycogen-rich, mucin-poor cells. These lesions tend to occur in an elderly age group and are often an asymptomatic finding. Malignant transformation in a serous cystadenoma is extremely rare and so surgical resection is rarely required. Larger serous cystadenomas may cause local compressive complications, including pain, which may warrant surgical resection.

Solid pseudopapillary neoplasms are rare lesions, usually found in females in their fourth decade; they may occur anywhere in the pancreas. Resection is required, as around 20% show malignant features on histological examination.

Mucinous cystadenomas are almost exclusively found in females, usually in the fifth or sixth decade, and are sited in the pancreatic body and tail. Multilocular cysts are lined by tall, mucin-synthesizing cells. Around 20% of these lesions are malignant at the time of presentation and the majority appear to have malignant potential. Surgical resection is generally recommended.

IPMN is a pancreatic cystic neoplasm that can arise in either the main pancreatic duct (main duct IPMN) or its side branches (branch duct IPMN). The majority are found in males between the ages of 60 and 70. Presentation is usually with pancreatic pain but branch duct IPMN is often an incidental finding. IPMNs are slowly progressive but carry a significant malignant potential. Generally, all main duct IPMNs and large branch duct IPMNs should be resected. Small branch duct IPMNs have a much lower malignant potential and are generally kept under radiological surveillance.

Cystic degeneration of a pancreatic adenocarcinoma or pancreatic neuroendocrine tumour can occur.

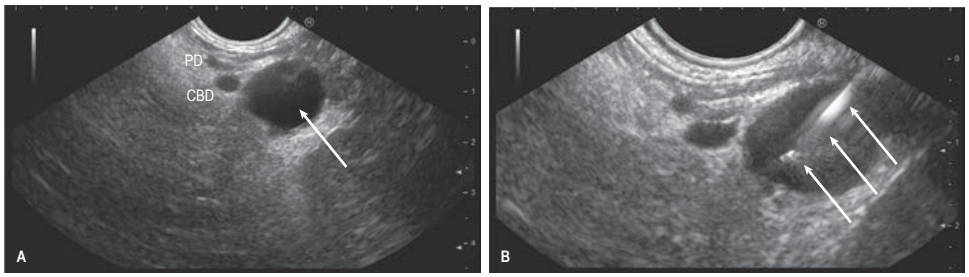


Fig. 24.22 Linear EUS images. (A) A cystic lesion *(arrowed)* in the head of the pancreas; the common bile duct (CBD) and pancreatic duct (PD) are also visible. The lesion is an intraductal papillary mucinous neoplasm. **(B)** The same lesion being sampled by fine needle aspiration for cytology, fluid amylase, and fluid carcinoembryonic antigen estimation. The needle is *arrowed*.

Pancreatic neuroendocrine tumours

Pancreatic neuroendocrine tumours (NETs) are thought to arise from the cells of the islets of Langerhans and can secrete a variety of active hormones, most frequently insulin. Pancreatic NETs have increased in incidence over the last two decades, reaching an incidence of 4–5/1 000 000 population. They represent a heterogeneous group of tumours with varying symptomology, tumour biology, and prognosis. Between 25% and 50% of the patients present with symptoms related to hormones released from the tumour. The remainder has 'non-functioning' tumours that may be an incidental finding or present with symptoms related to tumour bulk, such as obstruction, jaundice, bleeding, or abdominal mass. The increasing incidence of pancreatic NETs has been attributed in part to the widespread use of high-resolution cross-sectional imaging, leading to incidental diagnosis of asymptomatic lesions.

In addition to insulin and glucagon, pancreatic NETs may synthesize ectopic hormones that are not usually found in the pancreas, such as gastrin, adrenocorticotrophin, vasoactive intestinal peptide, and growth hormone. While many pancreatic endocrine tumours are multihormonal, one peptide tends to predominate and is responsible for any clinical syndrome (see later). Non-functioning pancreatic NETs often contain peptide hormone on immunostaining but this remains functionally inactive. The majority of endocrine pancreatic tumours are malignant in their behaviour. Between 10% and 15% of pancreatic NETs are linked to an inherited syndrome, such as the multiple endocrine neoplasia type 1 (MEN-1) (see Chapter 34) or von Hippel–Lindau (VHL) syndrome (see Chapter 26).

Clinical syndromes

Insulinoma is described in Chapter 34.

A ***gastrinoma*** accounts for approximately 1 in 1000 cases of duodenal ulcer disease. The ulceration results from hypersecretion of gastric acid secondary in response to high serum gastrin concentrations, secreted by the pancreatic tumour (Zollinger–Ellison's syndrome). Recurrent severe duodenal ulceration occurs, with only a partial response to acid suppression. The diagnosis is confirmed by an elevated gastrin level. High-dose proton pump inhibitors are used to suppress symptoms.

A ***VIPoma*** is an endocrine pancreatic tumour producing vasoactive intestinal polypeptide (VIP). This causes secretory diarrhoea secondary to the stimulation of adenyl cyclase within the enterocyte (Verner–Morrison or watery diarrhoea syndrome). The clinical syndrome is one of profuse watery diarrhoea, hypokalaemia, and a metabolic acidosis. To produce the syndrome, the tumours are usually in excess of 3 cm in diameter.

Glucagonomas are rare α-cell tumours that are responsible for a syndrome of migratory necrolytic dermatitis, weight loss, diabetes mellitus, deep vein thrombosis, anaemia, and hypoalbuminaemia. The diagnosis is made by measuring pancreatic glucagon in the serum.

Somatostatinomas are rare malignant D-cell tumours of the pancreas. These tumours cause diabetes mellitus, gallstones, and diarrhoea/steatorrhoea. They can be diagnosed by high serum somatostatin levels.

Investigations

Diagnosis is based on a combination of biochemical and histopathological markers. The biochemical diagnosis includes measurement of circulatory chromogranin A or specific hormones such as gastrin, insulin, glucagon, and VIP. The histopathology includes features such as positive staining for chromogranin A and specific hormones such as gastrin, pro-insulin, and glucagon.

Tumour localization depends on cross-sectional imaging, including CT and MRI scanning. The majority of NETs express somatostatin receptors, and a radiolabelled somatostatin analogue (such as octreotide) provides a means of tumour localization using scintigraphy. PET scanning with ^{18}F-FDG-PET/CT or ^{68}Ga-DOTATATE PET/CT is now widely used in the localization of primary disease and any metastases.

Identification of the primary and possibly metastatic lesions may be difficult despite multiple imaging techniques. EUS is the most sensitive means of detecting small primary NETs in the pancreas. It also offers the opportunity to undertake fine needle biopsy of lesions, affording a histological diagnosis (**Fig. 24.23**).

Management

Treatment of pancreatic NETs requires a multidisciplinary approach and depends on the presence or absence of metastatic (usually hepatic) disease. Where possible, surgical resection of the primary lesion is the optimal management of pancreatic endocrine tumours, as it offers the only possible cure. The propensity of many endocrine tumours to metastasize early precludes a cure in many cases. Debulking of the tumour, including liver metastases, is frequently carried out to facilitate systemic treatment.

Somatostatin analogues, such as octreotide and lanreotide, are used to control hormonal-related symptoms and also have a tumour-modulating effect. Radionuclide therapy using somatostatin analogues (peptide receptor radionuclide therapy [PRRT] with ^{177}Lutetium-DOTATATE) has proven benefit in patients with tumours that express a high content of somatostatin receptors.

Newer ***targeted and immunological systemic treatments*** used in the management of advanced pancreatic NET include everolimus (an mTOR inhibitor) and sunitinib (a tyrosine kinase inhibitor). Pancreatic NETs show a very high degree of vascularization, as well as abundant production and secretion of growth factors. There is evidence of benefit from anti-angiogenesis therapy utilizing vascular endothelial growth factor (VEGF) antagonists.

The ***chemotherapeutic agents*** temozolomide in combination with capecitabine or streptozotocin + 5-fluorouracil are the first-line chemotherapy regimens used to treat non-resectable pancreatic NET.

In patients with extensive liver metastases, occlusion of the arterial blood flow by hepatic arterial embolization may control hormone-related symptoms. In many cases the tumours are slowly progressive and may allow a reasonable quality of life for many years.

Further reading

Del Chiaro M, Sugawara T, Karam SD et al. Advances in the management of pancreatic cancer. *BMJ* 2023; 383.

Kos-Kudła B, Castaño JP, Denecke T et al. European Neuroendocrine Tumour Society (ENETS) 2023 guidance paper for nonfunctioning pancreatic neuroendocrine tumours.

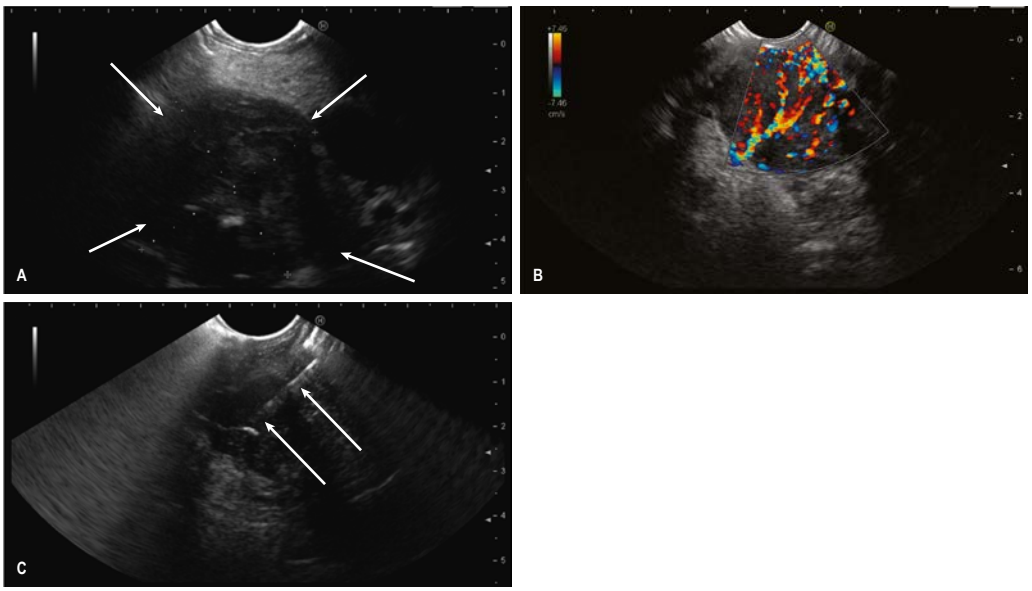


Fig. 24.23 Linear EUS images. **(A)** A heterogeneous solid mass *(arrowed)* in the head of the pancreas with some central calcification. **(B)** Doppler study demonstrating the vascularity of the lesion. **(C)** Endoscopic ultrasound-guided fine needle aspiration *(arrowed)* of the same lesion, which confirmed the presence of a neuroendocrine tumour.

Significant websites

https://www.bsg.org.uk *British Society of Gastroenterology guidelines.*
https://www.emedicine.medscape.com *Cholecystitis and cholelithiasis.*
https://www.gastrohep.com *Resources for gastroenterology, hepatology, and endoscopy.*
https://www.gi.org *American College of Gastroenterology guidelines.*
https://www.nice.org.uk *Guidelines and service standards.*
https://www.ueg.eu/education *Online learning in gastroenterology.*

Questions for this chapter can be found on Elsevier eBooks+.

25 Nutrition

Shameer J. Mehta and Sarah Faloon

Clinical skills 762
Assessment and diagnosis of malnutrition 762
Management 763
Malnutrition: introduction and background 763
Mechanisms in under-nutrition 764
Dietary requirements 764
Starvation-related under-nutrition 769
Re-feeding syndrome 773
Nutrition and ageing 774
Micronutrients and deficiencies 774
Fat-soluble vitamins 775
Water-soluble vitamins 776
Minerals 777
Food fortification 780
Nutritional support 780
Principles of nutritional support 780
Nutritional requirements for adults 781
Enteral nutrition 781
Parenteral nutrition 782
Intestinal failure 783
Functional classification of intestinal failure 784
Short bowel syndrome 784
Chronic intestinal failure: alternatives to home parenteral nutrition 786
Food intolerances 787

CORE SKILLS AND KNOWLEDGE

It is vital that healthcare professionals consider nutrition as a contributor to health. Patients should be screened for under-nutrition on admission to hospital and referred to dieticians and nutrition teams for assessment. This is usually followed by the setting of nutritional goals and an individualized nutritional care plan. This is especially relevant where patients cannot swallow (such as for patients who have suffered a stroke) or in those who are hypercatabolic (e.g., during critical illness). Clinical nutrition and intestinal failure lie within the specialism of gastroenterology.

Key skills in nutrition include:

- Developing an awareness of patients at risk of malnutrition and performing a nutritional assessment
- Understanding the role of dietary supplements and clinically assisted nutrition
- Being aware of syndromes associated with specific micronutrient deficiencies
- Understanding the classification of intestinal failure and the treatment principles.

Opportunities for learning about nutrition include shadowing dieticians reviewing patients in the community or as inpatients, attending specialist nutrition ward rounds led by gastroenterologists, or attending multidisciplinary meetings. The nutrition multidisciplinary team usually consists of a mix of gastroenterologists, dieticians, nutrition clinical nurse specialists, pharmacists, and surgeons with an interest in intestinal failure if reconstructive surgery can be performed to restore nutritional autonomy.

CLINICAL SKILLS

Assessment and diagnosis of malnutrition

Screening

Detection of malnutrition begins with screening using validated tools. Early identification of individuals at risk is essential. The Malnutrition Universal Screening Tool (MUST, see **Fig. 25.1**) is the most commonly used screening tool in the UK. It consists of three parts:

1. Body mass index (BMI)
2. History of recent unintentional weight loss
3. Acute disease effect

Composite scoring results in the stratification of patients into low-, medium-, or high-risk categories for malnutrition and guides the user to develop an individualized treatment plan. BMI is calculated using the following equation: BMI = mass (kg)/height^2 (m). A normal BMI is considered to fall between 18.5 and 24.9 kg/m^2, with <18.5 kg/m^2 considered underweight, ≥25 kg/m^2 considered overweight, and ≥30 kg/m^2 classed as obese. For black, Asian, and some other minority ethnic groups, the normal BMI range is 18 to 23 kg/m^2, owing to the increased risk of developing conditions such as type 2 diabetes mellitus with a lower BMI than observed in white populations.

Another screening tool, the Nutritional Risk Score 2002 (NRS-2002), is validated among hospital inpatient populations and is used widely across Europe. In addition to nutritional components, severity of disease is also incorporated into the score to reflect the expected increase in nutritional requirements.

The aim of the screening process is to identify patients who require a more detailed nutritional assessment and formulation of a targeted individualized management plan. Re-screening of inpatients at 7-day intervals is recommended.

Nutritional assessment

Once patients have been identified as being at risk of under-nutrition, physiological and nutritional measures are used to confirm and quantify this risk. This is usually undertaken by a dietician who

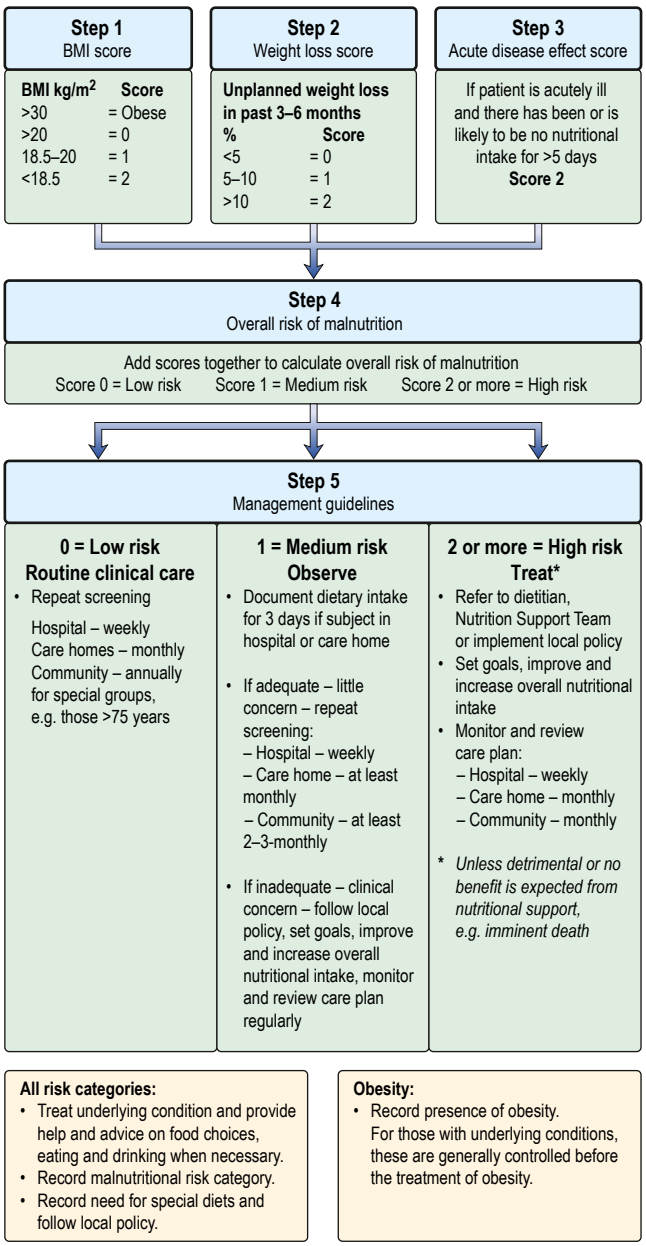


Fig. 25.1 Malnutrition Universal Screening Tool (MUST). *BMI,* Body mass index. *(With permission from the British Association for Parenteral and Enteral Nutrition (BAPEN), at: http://www.bapen.org.uk.)*

will perform a thorough assessment looking at the patient holistically. Assessment should take an 'ABCDE' approach (see **Fig. 25.2**).

Clinical examination

Bedside clinical examination can elicit the signs of malnutrition and the underlying driving pathology. A collection of specific signs may indicate particular nutritional deficiencies and subsequent complications. **Fig. 25.3** illustrates clinical features of malnutrition and how they relate to particular conditions.

Global assessment tools

Global assessment tools are then used to confirm formally a diagnosis of malnutrition and establish its severity in order to inform nutritional intervention. One such tool is the Subjective Global Assessment (SGA, **Fig. 25.4**), which refers to an overall evaluation of a patient's history and physical examination and uses structured clinical parameters to diagnose malnutrition. The SGA has been validated in multiple clinical scenarios and can predict morbidity and mortality associated with malnutrition.

Recently, the Global Leadership Initiative on Malnutrition (GLIM) published updated international criteria on the diagnosis of under-nutrition. These are divided into two categories: phenotypic criteria and aetiological criteria. The phenotypic criteria are (1) non-volitional weight loss, (2) low BMI, and (3) reduced muscle mass. The aetiological criteria are (1) reduced food intake or assimilation and (2) disease burden or inflammation. A diagnosis of malnutrition can be made when an individual meets at least one phenotypic and one aetiological criterion. The phenotypic criteria are well-established manifestations of under-nutrition, whereas the aetiological criteria help to determine treatment intervention. The phenotypic criteria also determine the degree of severity of malnutrition. Features such as >10% weight loss within the past 6 months and/or a BMI < 18.5 and/or severe muscle mass deficit are considered severe.

Management

Once under-nutrition has been confirmed using the above measures, a management plan with nutritional intervention can be developed (**Fig. 25.5**). This must take into consideration the above factors, goals of treatment (i.e., weight gain, improving or restoring function), and an assessment of the risk of re-feeding (see 'Re-feeding syndrome' section).

Further reading

Allen B, Saunders J. Malnutrition and undernutrition: causes, consequences, assessment and management. *Medicine* 2023.
Cederholm T, Jensen GL, Correia MI et al. GLIM criteria for the diagnosis of malnutrition—a consensus report from the global clinical nutrition community. *JCSM* 2019; 10(1):207–17.
Duerksen DR, Laporte M, Jeejeebhoy K. Evaluation of nutrition status using the subjective global assessment: malnutrition, cachexia, and sarcopenia. *Nutr Clin Pract* 2021; 36(5):942–56.
Stratton R, Smith T, Gabe S. *Managing Malnutrition to Improve Lives and Save Money*. On behalf of BAPEN (British Society of Enteral and Parenteral Nutrition). 2018.

MALNUTRITION: INTRODUCTION AND BACKGROUND

Nutrition is essential for health. A number of acute and chronic illnesses negatively impact on food intake and metabolism, resulting in under-nutrition and consequent increased morbidity and mortality. Additionally, food intake determines health and plays a major role in conditions such as cardiovascular disease, diabetes, and cancer.

The term malnutrition is used to describe a deficiency, excess, or imbalance of nutrients, which can lead to adverse effects on body composition, function, and clinical outcome. The focus of this chapter will be on under-nutrition. Overnutrition and obesity are covered in Chapter 37. Under-nutrition remains a major clinical and public health concern worldwide.

Globally, malnutrition affects approximately 1 billion people. In low- and middle-income countries, the lack of food and poor usage of available food, often in association with coexistent inflammatory processes, often result in under-nutrition and micronutrient deficiencies. Worldwide, in 2011, 165 million children under 5 years of age were affected by growth stunting, and at least 52 million by muscle wasting. Under-nutrition is also common in the UK, affecting over 3 million people. This can be disease related or a

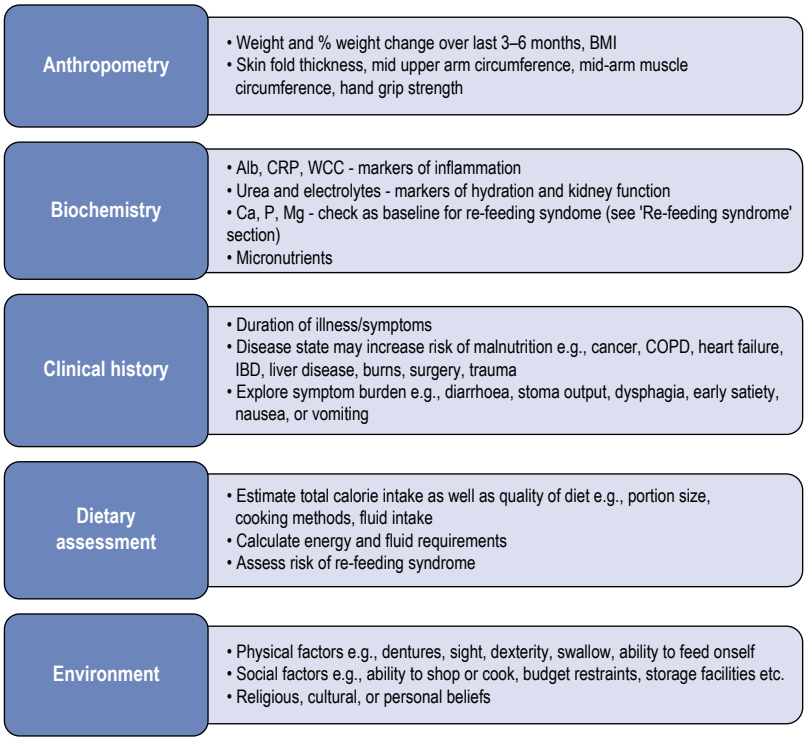


Fig. 25.2 The ABCDE approach to nutritional assessment. *Alb,* Albumin; *BMI,* body mass index; *Ca,* calcium; *CRP,* C-reactive protein; *Mg,* magnesium; *P,* phosphate; *WCC,* white cell count.

consequence of psychosocial circumstances (**Box 25.1**). According to one report by The British Association of Parenteral and Enteral Nutrition (BAPEN), 25% to 34% of patients admitted to hospital and 30% to 42% of elderly people admitted to care homes are at risk of malnutrition.

Malnutrition has negative effect on the function and recovery of all body systems (**Box 25.2**). In the hospital setting, malnutrition is associated with longer length of stay, increased morbidity and mortality. Crucially, unlike other risk factors, it is also modifiable with cheap and safe interventions. It therefore presents an ideal opportunity for intervention and it is vital that healthcare professionals are trained to assess for, diagnose, and treat malnutrition.

MECHANISMS IN UNDER-NUTRITION

Dietary requirements

In the UK, dietary reference values for food, energy, and nutrients are stated as ***reference nutrient intakes*** (RNIs), on the basis of data from the Food and Agriculture Organization (FAO-WHO), United Nations University (UNU) expert committee, and elsewhere. The RNI is sufficient, or more than sufficient, to meet the nutritional needs of 97.5% of healthy people in a population. Most people's daily requirements are less than this, and so an ***estimated average requirement (EAR)*** is also given, which will certainly be adequate for most. A lower reference nutrient intake (LRNI), which fails to meet the requirements of 97.5% of the population, is also given. The RNI figures quoted in this chapter are for the age group 19 to 50 years. These represent values for healthy subjects and are not always appropriate for patients with disease.

Water and electrolyte balance

Water and electrolyte balance is dealt with fully in Chapter 27. About 1 L of water is required in the daily diet to balance insensible losses but much more is usually drunk, the kidneys being able to excrete large quantities. The daily RNI for sodium is 70 mmol (1.6 g) but daily sodium intake varies in the range 90 to 440 mmol (2–10 g). These are needlessly high intakes of sodium, which are thought to play a role in causing hypertension (see Chapter 18). The World Health Organization (WHO) also recommends at least 90 mmol (3.5 g) of potassium, which would reduce blood pressure and cardiovascular risk of stroke and coronary artery disease.

Energy

Food is necessary to provide the body with energy (**Fig. 25.6**). The SI unit of energy is the joule (J), and 1 kJ = 0.239 kcal. The conversion factor of 4.2 kJ, equivalent to 1.00 kcal, is used in clinical nutrition.

Energy balance

Energy balance is the difference between energy intake and energy expenditure. Weight gain or loss is a simple but accurate way of indicating differences in energy balance.

There are two approaches to assessing energy requirements for subjects who are weight-stable and close to energy balance:

- assessment of energy intake
- assessment of total energy expenditure.

Energy intake

Energy intake can be estimated from dietary surveys and, in the past, this has been used to decide daily energy requirements. However, measurement of energy expenditure gives a more accurate assessment of requirements.

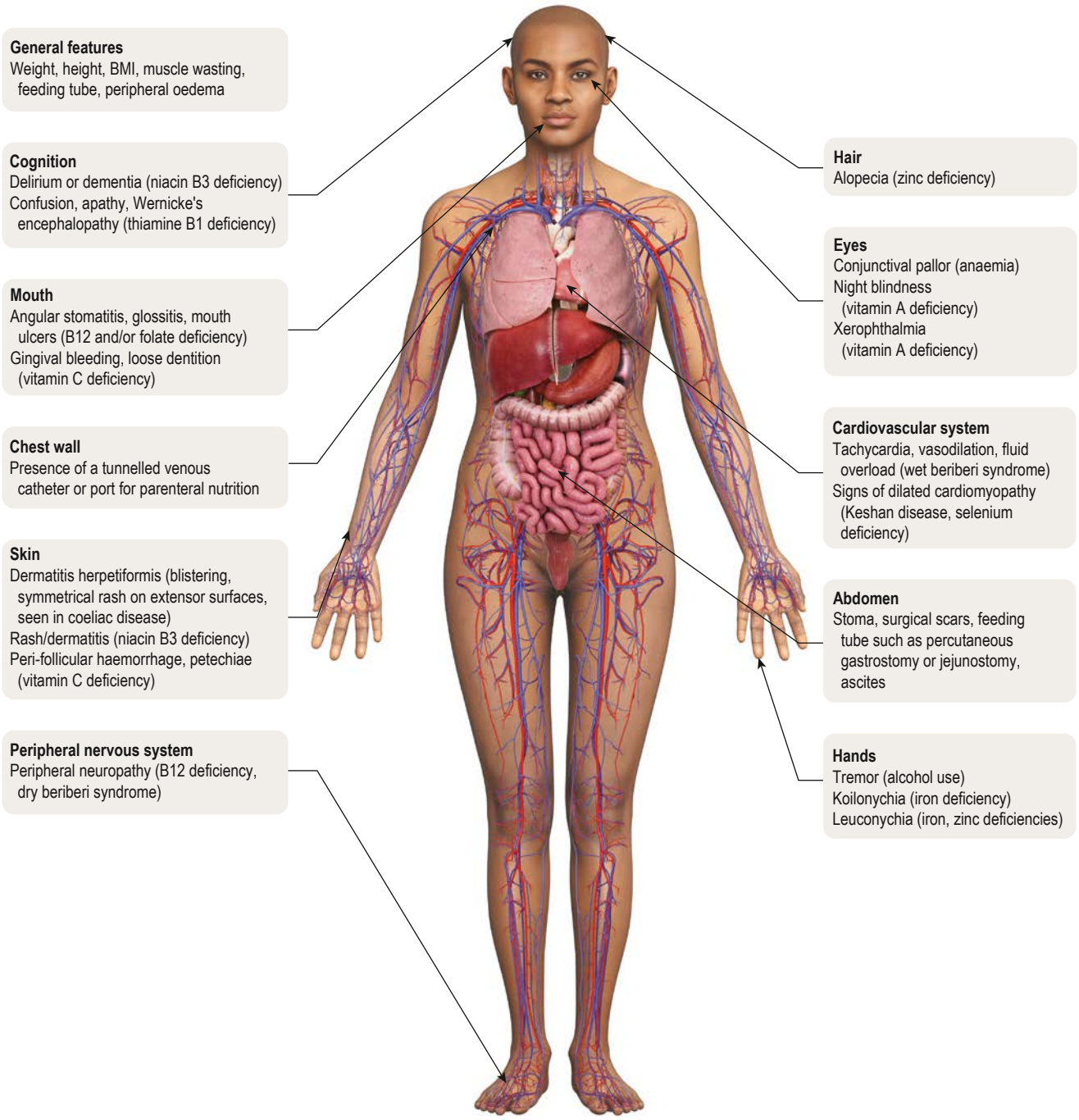


Fig. 25.3 **Examination: clinical features of malnutrition.**

Energy expenditure

Daily energy expenditure (Fig. 25.7) is the sum of:

- the basal metabolic rate (BMR)
- the thermic effect of food eaten
- occupational activities
- non-occupational activities.

Total energy expenditure can be measured using a double-labelled water technique. Water containing the stable isotopes ^{2}H and ^{18}O is given orally. As energy is expended, carbon dioxide and water are produced. The difference between the rates of loss of the two isotopes is used to calculate the carbon dioxide production, which is then used to calculate energy expenditure. This can be done on urine samples over a 2- to 3-week period with the subject ambulatory. The technique is accurate, but it is expensive and requires the availability of a mass spectrometer. An alternative tracer technique for measuring total energy expenditure is to estimate CO_2 production by isotopic dilution. A subcutaneous infusion of labelled bicarbonate is administered continuously by a minipump, and urine is collected to measure isotopic dilution by urea, which is formed from CO_2. Other methods for estimating energy expenditure, such as heart rate monitors or activity monitors, are also available but are less accurate.

SUBJECTIVE GLOBAL ASSESSMENT RATING FORM	
Patient name: ID #: Date:	
HISTORY	
WEIGHT/WEIGHT CHANGE: ***(Included in K/DOQI SGA)*** **1. Baseline Wt:** _____(Dry weight from 6 months ago) **Current Wt:** _____(Dry weight today) **Actual Wt loss/past 6 mo:** _____ **% loss:** _____(actual loss from baseline or last SGA) **2. Weight change over past two weeks:** _____No change _____Increase _____Decrease	**Rate 1–7**
DIETARY INTAKE No change_____(Adequate) No change_____(Inadequate) 1. Change: Sub optimal intake:_____Protein_____Kcal_____Duration_____ Full liquid:_____Hypocaloric liquid_____Starvation_____	
GASTROINTESTINAL SYMPTOMS ***(Included in K/DOQI SGA-anorexia or causes of anorexia)*** **Symptom:** **Frequency:*** **Duration:+** _____ None _____ _____ _____ Anorexia _____ _____ _____ Nausea _____ _____ _____ Vomiting _____ _____ _____ Diarrhea _____ _____ Never, daily, 2–3 times/wk, 1–2 times/wk > 2 weeks, < 2 weeks	
FUNCTIONAL CAPACITY **Description** **Duration:** _____ No dysfunction _____ _____ Change in function _____ _____ Difficulty with ambulation _____ _____ Difficulty with activity (Patient specific "normal") _____ _____ Light activity _____ _____ Bed/chair ridden with little or no activity _____ _____ Improvement in function _____	b
DISEASE STATE/COMORBIDITIES AS RELATED TO NUTRITIONAL NEEDS Primary diagnosis_____Comorbidities_____ Normal requirements ____Increased requirements ____Decreased requirements _____ Acute metabolic stress: _____ None _____Low _____Moderate _____High	
PHYSICAL EXAM	
_____ Loss of subcutaneous fat (Below eye, triceps, biceps, chest) ***(Included in K/DOQI SGA)*** ____Some areas _____All areas _____ Muscle wasting (Temple, clavicle, scapula, ribs, quadriceps, calf, knee, interosseous ***(Included in K/DOQI SGA)*** ____Some areas _____All areas _____ Edema (Related to under-nutrition/use to evaluate weight change)	
OVERALL SGA RATING	
Very mild risk to well-nourished = 6 or 7 most categories or significant, continued improvement. **Mild–moderate** = 3, 4, or 5 ratings. No clear sign of normal status or severe malnutrition. **Severely malnourished** = 1 or 2 ratings in most categories/significant physical signs of malnutrition.	

Fig. 25.4 Subjective Global Assessment tool.

BMR can be calculated by measuring oxygen consumption and CO_2 production, but it is more usually taken from standardized tables (Box 25.3) that only require knowledge of the subject's age, weight, and sex.

The ***physical activity ratio*** (PAR) is expressed as multiples of the BMR for both occupational and non-occupational activities of varying intensities (Box 25.4).

In the UK, the estimated 'average' daily requirement (EAR) for energy is:

- for a 55-year-old female—8700 kJ (2079 kcal)
- for a 55-year-old male—10 800 kJ (2581 kcal).

Energy requirements increase during the growing period, with pregnancy and lactation, and sometimes following infection or trauma. For example, the EAR for pregnant women in the third trimester increases by 800 kJ per day (200 kcal/day). In general, the increased BMR associated with inflammatory or traumatic conditions is counteracted, or more than counteracted, by a decrease in physical activity, so that total energy requirements are not increased.

In the basal state, energy demands for resting muscle are 20% of the total energy required, those for abdominal viscera 35% to 40%, those for brain 20% and those for heart 10%. There can be more than a 50-fold increase in muscle energy demands during exercise.

Energy stores

Although virtually all body fat and glycogen are available for oxidation, less than half the protein is available for oxidation. Fig. 25.8 shows that fat accounts for the largest reserves of energy in both

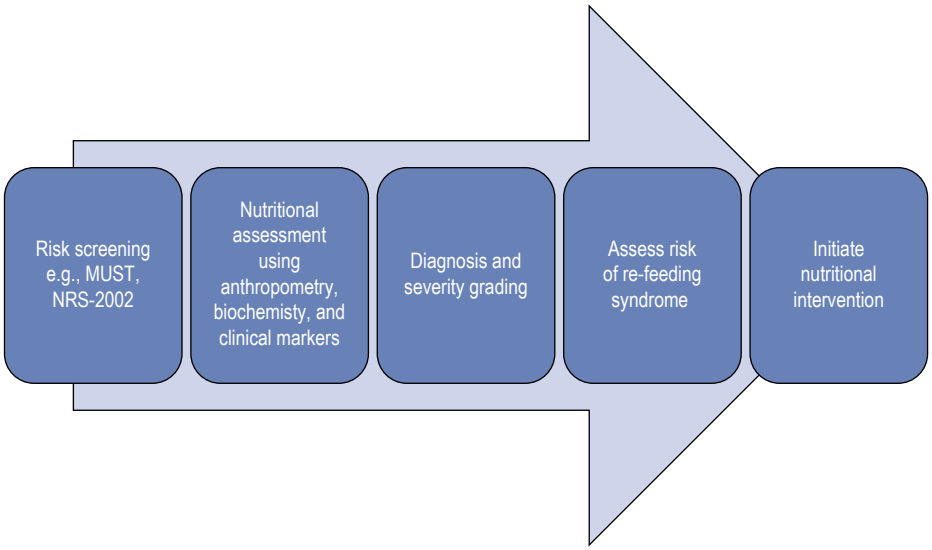


Fig. 25.5 **Process of nutritional screening and assessment.**

Box 25.1 Who is at risk of under-nutrition?

- Older patients, particularly if they are living in a care home or nursing home
- People with chronic illnesses such as chronic inflammatory conditions, chronic liver disease, cancer
- People with chronic progressive conditions such as degenerative neuromuscular conditions or dementia
- People who abuse alcohol or drugs
- Patients affected by social deprivation such as poverty or social isolation, limited mobility, inability to feed self, or lack of transport
- Physical factors such as poor dentition

Box 25.2 Consequences of under-nutrition

- Reduced muscle mass and strength (sarcopenia)
- Impaired gut integrity and immunity
- Decreased immune function with increased susceptibility to infections and slow wound healing
- Depression and/or apathy
- Reduced cardiac output
- Impaired liver function and fatty change

lean and obese subjects. The size of the stores determines survival during starvation.

Body weight

Body weight depends on energy balance. Intake depends not only on food availability but also on a number of complex interrelationships that include the stimulus of appetizing food, the role of hunger, metabolic changes (e.g., hypoglycaemia), and the pleasure and habit of eating. Some people are able to keep their body weight constant within a few kilograms for many years, but most gradually increase their weight owing to a small but continuous increase of intake over expenditure. A gain or loss of energy of 25 to 29 MJ (6000–7000 kcal) would, respectively, increase or decrease body weight by approximately 1 kg.

Protein

In the UK, the adult daily RNI for protein is 0.75 g/kg, with protein representing at least 10% of the total energy intake. Most affluent people eat more than this, consuming 80 to 100 g of protein per day. The total amount of nitrogen excreted in the urine represents the balance between protein breakdown and synthesis. In order to maintain nitrogen balance, at least 40 to 50 g of dietary protein is needed. The amount of protein oxidized can be calculated from the amount of nitrogen excreted in the urine over 24 hours using the following equation:

$$\text{Grams of protein required} = \text{urinary nitrogen} \times 6.25$$

(most proteins contain about 16% of nitrogen)

In healthy individuals, urinary nitrogen excretion reflects protein intake. However, excretion does not match intake in catabolic conditions (negative N balance), or during growth or repletion following an illness (positive N balance).

Protein contains many ***amino acids:***

- ***Indispensable (essential).*** There are nine amino acids that cannot be synthesized and must be provided in the diet: tryptophan, histidine, methionine, threonine, isoleucine, valine, phenylalanine, lysine, and leucine.
- ***Dispensable (non-essential).*** These are amino acids that can be synthesized in the body (some may still be needed in the diet unless adequate amounts of their precursors are available).

Animal proteins (e.g., in milk, meat, and eggs) contain a good balance of all the indispensable amino acids, but many proteins from vegetables are deficient in at least one indispensable amino acid. In low- and middle-income countries, protein intake derives mainly from vegetable proteins.

Loss of protein from the body (negative N balance) occurs not only because of inadequate protein intake but also because of inadequate energy intake. When there is loss of energy from the body, more protein is directed towards oxidative pathways and, eventually, gluconeogenesis for energy.

Role of amino acids

- ***Glutamine*** is quantitatively the most significant amino acid in the circulation and in inter-organ exchange.
- ***Alanine*** is released from muscle; it is deaminated and converted into pyruvic acid before entering the citric acid cycle.
- ***Homocysteine*** is a sulphur-containing amino acid that is derived from methionine in the diet. A raised plasma concentration is an independent risk factor for vascular disease.

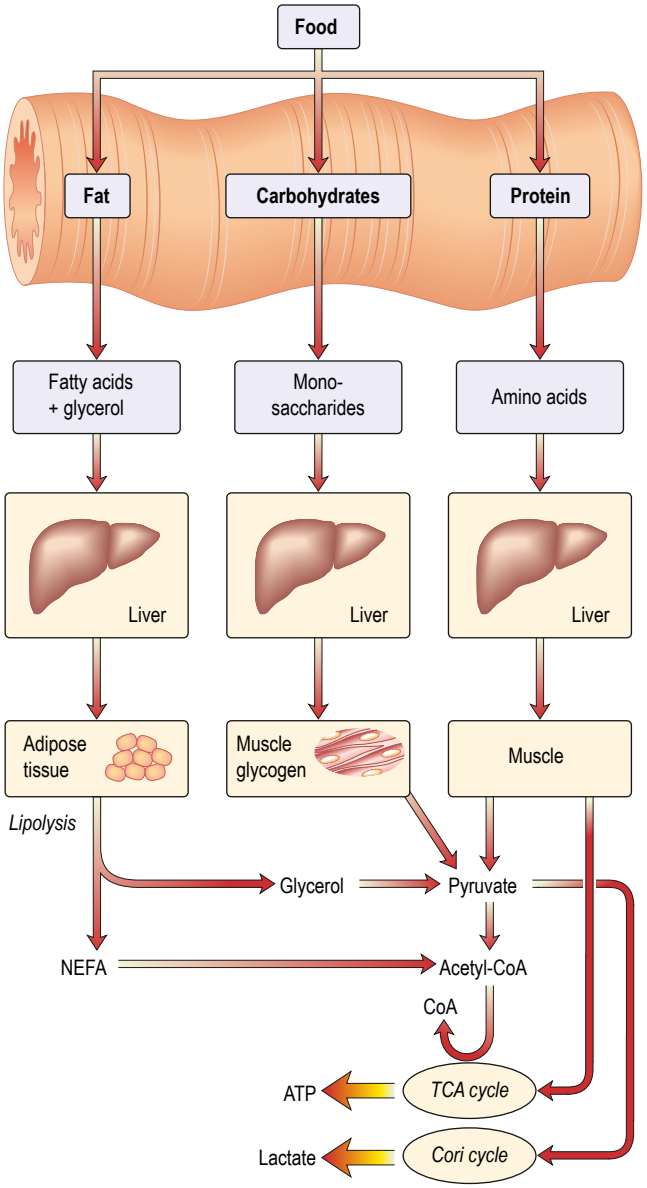


Fig. 25.6 The production of energy from the main constituents of food. Alcohol produces up to 5% of total calories but the variation between individuals is wide. One mole of glucose produces 36 moles of adenosine triphosphate (ATP). *CoA,* Coenzyme A; *NEFA,* non-esterified fatty acids; *TCA,* tricarboxylic acid.

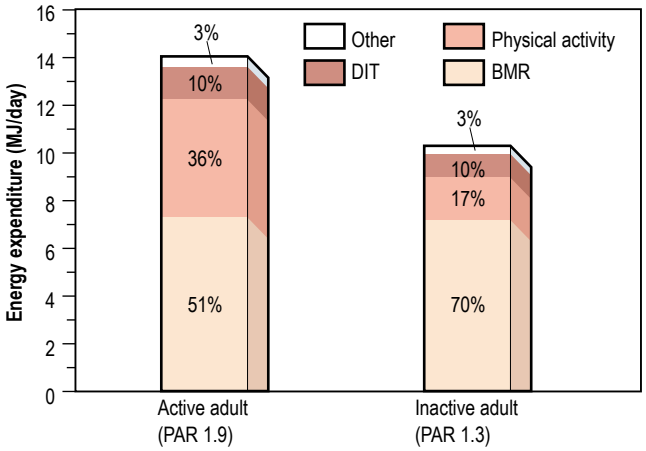


Fig. 25.7 Daily energy expenditure in an active and a sedentary 70-kg adult. *BMR,* Basal metabolic rate; *DIT,* dietary-induced thermogenesis; *PAR,* physical activity ratio.

Box 25.3 Equations for the prediction of basal metabolic rate (BMR) (in MJ/day)

Age range (years)	Equation for predicting BMR[a]	95% confidence limits
Men		
10–17	0.0740×(wt)+2.754	±0.88
18–29	0.0630×(wt)+2.896	±1.28
30–59	0.0480×(wt)+3.653	±1.40
60–74	0.0499×(wt)+2.930	N/A
75+	0.0350×(wt)+3.434	N/A
Women		
10–17	0.0560×(wt)+2.898	±0.94
18–29	0.0620×(wt)+2.036	±1.00
30–59	0.0340×(wt)+3.538	±0.94
60–74	0.0386×(wt)+2.875	N/A
75+	0.0410×(wt)+2.610	N/A

[a]Body weight (wt) in kg.
(Data from Department of Health, 1991.)

Box 25.4 Physical activity ratio (PAR) for various activities (expressed as multiples of basal metabolic rate)

Activity	PAR
Office worker	1.7
Domestic helper/salesperson	2.7
Labourer	3.0
Reading/eating	1.2
Household/cooking	2.1
Gardening/golf	3.7
Jogging/swimming/football	6.9

Amino acids are utilized to synthesize products other than protein or urea. For example:

- ***Glycine*** is required for haem production.
- ***Tyrosine*** is required for melanin and thyroid hormones.
- ***Glutamine, aspartate, and glycine*** are required for nucleic acid bases.
- ***Glutamate, cysteine, and glycine*** are required for glutathione, which is part of the defence system against free radicals.

Fat

Fat is required as part of a healthy and balanced diet; it contains essential fatty acids (EFAs) and carries the fat-soluble vitamins (A, D, E, and K). Dietary fat is chiefly in the form of triglycerides, which are esters of glycerol and free fatty acids. Fatty acids vary in chain length and in saturation.

The ***essential fatty acids*** are linoleic and α-linolenic acid, both of which are precursors of prostaglandins. Eicosapentaenoic and docosahexaenoic acid are also necessary, but can be manufactured to a limited extent in the tissues from linoleic and linolenic acid, and thus a dietary supply is not essential.

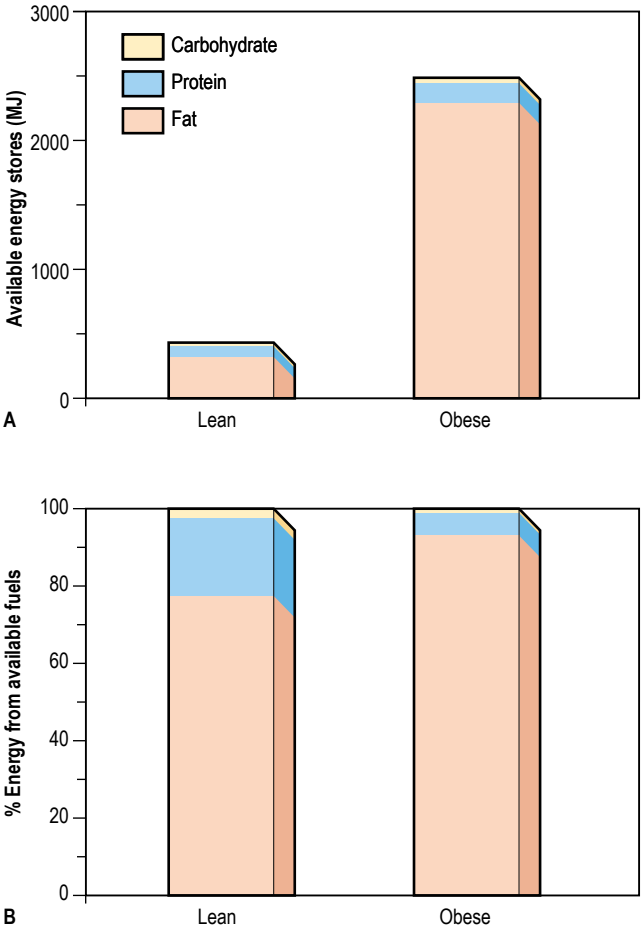


Fig. 25.8 Available energy reserves from different fuels. Values are shown for a lean hypothetical 70-kg adult and an obese 140-kg adult. **(A)** Expressed in MJ. **(B)** Expressed as a percentage of the total.

EFA deficiency can occur in starvation-related malnutrition (see below), and it has been clearly defined as a clinical entity in patients on long-term parenteral nutrition who receive glucose and protein, with no fat. Alopecia, thrombocytopenia, anaemia, and dermatitis can occur within weeks.

Synthesis of triglycerides, sterols, and phospholipids is very efficient. Even with low-fat diets, subcutaneous fat stores can be normal. ***Dietary fat*** provides 37 kJ (9 kcal) of energy per gram. A high-fat intake has been associated with a number of conditions including obesity, cardiovascular disease, type 2 diabetes, and steatotic liver disease. Dietary fats can be broken down into saturated and trans (the unhealthy fats) and monounsaturated and polyunsaturated fats (the healthy fats). Omega-3 is found within polyunsaturated fats, such as oily fish, sunflower oil, and seeds.

Carbohydrate

Carbohydrates are readily available in the diet, providing 17 kJ (4 kcal) per gram of energy (15.7 kJ (3.75 kcal) per gram monosaccharide equivalent). Carbohydrate intake comprises:

- polysaccharides (starch, fibre)
- disaccharides (mainly sucrose)
- monosaccharides (glucose and fructose).

Carbohydrate is cheap, compared with other foodstuffs; a great deal is therefore eaten, usually more than required. WHO suggests that sugar intake should be less than 5% of a person's total energy consumption.

Dietary fibre

Dietary fibre, is often removed in the processing of food. Types of fibre include cellulose, hemicelluloses, pectins, and gums. These are not digested by gut enzymes and therefore pass into the large intestine, where they are partly broken down by colonic bacteria, producing gas and volatile fatty acids, such as butyrate. Foods rich in dietary fibre include wholemeal bread, wholegrain cereals, brown pasta or rice, fruit, and potato skins. Fibre increases faecal bulk and is therefore helpful in the treatment of constipation. The recommended daily intake of fibre for adults in the UK is 30 g, but on average this is much less. Each extra gram of fibre daily adds approximately 3 to 5 g to the daily stool weight.

The healthy diet

Globally, examples of healthy diets include the Mediterranean or Nordic diets. These diets tend to be low in saturated fats but higher in fibre, oily fish, vegetable oils, and spreads which contain unsaturated fats. These diets also tend to be lower in processed meats, salt, and refined sugars.

Starvation-related under-nutrition

Starvation uncomplicated by disease is relatively uncommon in high-income countries, although some degree of under-nutrition is seen in socially deprived areas. Starvation remains a global issue. The 2021 United Nations Report stated that 828 million people are affected by hunger and the number continues to rise. Events such as conflicts and wars, the COVID-19 global pandemic and frequent extreme climate events are resulting in rising rates of food insecurity, hunger, and under-nourishment worldwide. The next section will focus on the mechanisms of starvation and its clinical manifestations, which are most often seen in less economically developed countries.

Pathophysiology of starvation

In the first 24 hours following low dietary intake, the body relies on the breakdown of hepatic glycogen to glucose as an energy source (Fig. 25.9). Hepatic glycogen stores are limited and therefore gluconeogenesis is soon necessary to maintain glucose levels. Gluconeogenesis takes place mainly from pyruvate, lactate, glycerol, and amino acids, especially alanine and glutamine. The majority of protein breakdown takes place in muscle, with eventual loss of muscle bulk.

Lipolysis, the breakdown of the body's fat stores, also occurs. This process is inhibited by insulin, although serum levels of this hormone diminish as starvation continues. Stored triglyceride is hydrolyzed by lipase to glycerol, which is used for gluconeogenesis, and also to non-esterified fatty acids, which can be used directly as a fuel or oxidized in the liver to ketone bodies.

Adaptive processes take place as starvation continues, to prevent available protein from being completely utilized. There follows a decrease in metabolic rate and total body energy expenditure. Central nervous metabolism changes from glucose as a substrate to ketone bodies. Gluconeogenesis in the liver decreases, as does protein breakdown in muscle, and both of these processes are inhibited directly by ketone bodies. Most of the energy at this stage comes from adipose tissue, with some gluconeogenesis from amino acids, particularly from alanine in the liver and glutamine in the kidney.

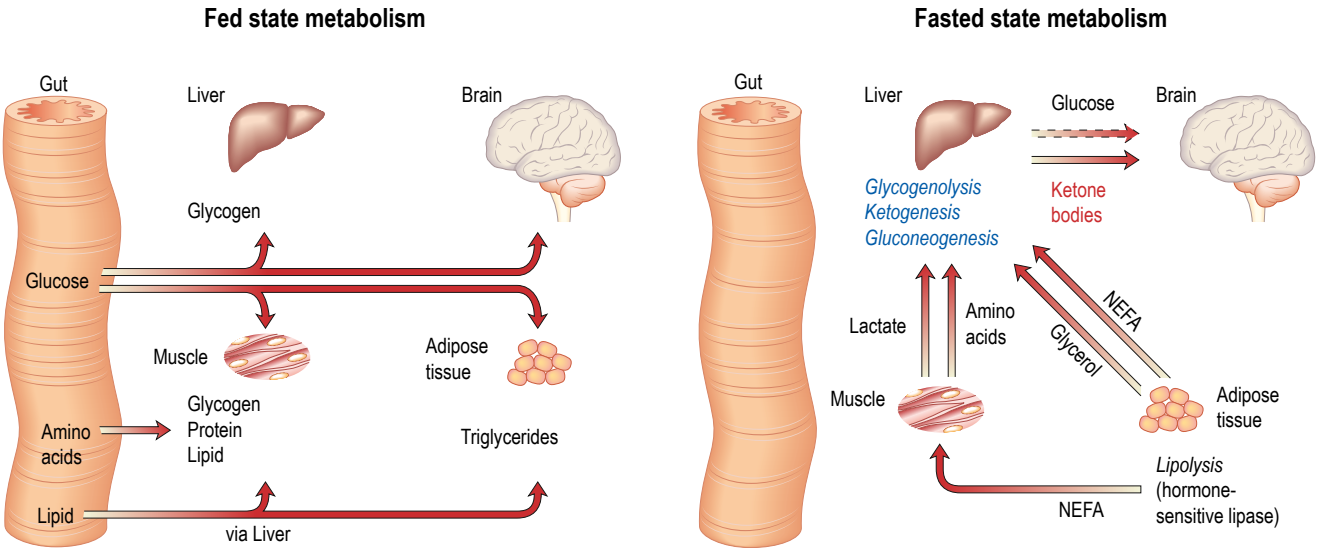


Fig. 25.9 Metabolism in the fed and the fasted state. *NEFA,* Non-esterified fatty acids.

The *metabolic response to prolonged starvation* differs between lean and obese individuals. One of the major differences concerns the proportion of energy derived from protein oxidation, which determines the proportion of weight loss from lean tissues. This proportion may be up to three times smaller in obese subjects than in lean subjects. It can be regarded as an adaptation that depends on the composition of the initial reserves (see Fig. 25.8). This means that deterioration in body function is more rapid in lean subjects, and survival time is shorter in lean subjects (approximately 2 months) compared with the obese (in whom it can be at least several months).

Following trauma or shock, some of the adaptive changes do not take place. Glucocorticoids and cytokines (see below) stimulate the ubiquitin–proteasome pathway in muscle, which is responsible for accelerated proteolysis in muscle in many catabolic illnesses. In starvation, there is a decrease in BMR, while in inflammatory and traumatic disease, there is often an increase in BMR. These changes all result in continuing gluconeogenesis with massive muscle breakdown, and further reduction in survival time.

Regulation of metabolism

Tissue metabolism is regulated by multiple coordinated processes. Some are rapid, involving nerves, while others are slower, involving circulating substrates and hormones. Factors include:

- *Circulating substrate concentrations.* The uptake and metabolism of ketone bodies, which serve as the major fuel for the brain during prolonged starvation, are primarily determined by the circulatory concentration, which can increase up to 5 mmol/L or more. The liver is responsible for the production of ketone bodies, which is, in turn, controlled by the availability of fatty acids derived from adipose tissue. Substrates may also compete with each other for metabolism; for example glucose competes with non-esterified fatty acids for uptake and metabolism in muscle and heart (the glucose–fatty acid cycle), and this is independent of hormones.
- *Blood flow.* The delivery of substrates to tissues depends not only on their circulating concentration but also on the blood flow to tissues. In many tissues, there is coupling between metabolic activity and blood flow, with arterioles regulating blood flow to the tissue according to demand; for example blood flow to muscle increases during exercise.
- *Signals.* Hormones and other signals, such as cytokines (see below), regulate intracellular metabolism.

Insulin/glucagon ratios in the fed and fasted state

In the fed state, insulin/glucagon ratios are high. Insulin promotes synthesis of glycogen, protein, and fat, and inhibits lipolysis and gluconeogenesis.

In the fasted state, insulin/glucagon ratios are low. Glucagon acts mainly on the liver and has no action on muscle. It increases glycogenolysis and gluconeogenesis, as well as increasing ketone body production from fatty acids. It also stimulates lipolysis in adipose tissue. Catecholamines have a similar action to glucagon but also affect muscle metabolism. These agents both act via cyclic adenosine monophosphate (cAMP) to stimulate lipolysis, producing free fatty acids that can then act as a major source of energy.

Proportion of lean-to-fat tissue

During weight loss uncomplicated by disease, the proportion of lean-to-fat tissue loss (or proportion of energy derived from protein metabolism) is greater in lean than overweight/obese individuals.

During acute disease, loss of lean tissue, which is associated with protein oxidation, can be particularly rapid. Hormones such as corticosteroids, pro-inflammatory cytokines, and insulin resistance are all involved.

Role of cytokines

The metabolic response to trauma, injury, and inflammation depends on the balance between *pro*-inflammatory (e.g., tumour necrosis factor, TNF; interleukin-2, IL-2) and *anti*-inflammatory cytokines (e.g., IL-10), and the production of many of these cytokines is influenced by genetic polymorphisms. Since many chronic diseases have an inflammatory component, these changes have wide-reaching metabolic implications. Cytokines such as IL-1, IL-6, and TNF play a significant role in regulating metabolism. In acute diseases, they contribute to the catabolic process, glycogenolysis, and acute phase protein synthesis. TNF, which inhibits lipoprotein lipase, is one of a number of 'cachexia factors' in patients with cancer. It is unclear how these cytokines interact with central feeding

Fig. 25.10 Malnourished children. **(A)** Marasmus. **(B)** Kwashiorkor. *(Courtesy of Dr Paul Kelly.)*

pathways to cause anorexia. However, in animal models of both cancer and inflammatory bowel disease, many peripheral and central mediators of appetite are involved. For example, neuropeptide Y levels in the hypothalamus are often inappropriately low, leading to a reduced drive to feed.

Starvation-related malnutrition leads to a depression of immunological defences, resulting in an increased susceptibility to infection. This also detrimentally affects muscle strength and fatigue, reproductive function (e.g., in anorexia nervosa, see Chapter 30), wound healing, and psychological function (depression, anxiety, loss of libido).

In children, growth failure is a key element in the diagnosis of starvation-related malnutrition. WHO standards for optimal growth in children of 0 to 4 years have been adopted by low- and middle-income and high-income countries. They aim to reflect optimal rather than prevailing growth in both high-income and low- and middle-income countries, since they describe children born after a healthy pregnancy to non-smoking, relatively affluent mothers who breast-fed their children exclusively or predominantly for the first 6 months of life. The general principles of management are similar in high-income and low- and middle-income countries but resources are required to manage the problems once they have been identified. When malnutrition is apparent and the underlying aetiology cannot be corrected immediately some form of nutritional support is necessary (see later).

Clinical features

Severe acute malnutrition (SAM) as a result of near-total starvation produces unique clinical appearances in children that are virtually never seen in high-income countries. Seventeen million children in 2017 had SAM and this is 2.5% of all the children under 5 in the world. Children under 5 years may present with the following:

- *Kwashiorkor* occurs typically in a young child displaced from breast-feeding by a new baby. It is often precipitated by infections such as measles, malaria, and diarrhoeal illnesses. The child is apathetic and lethargic with severe anorexia. There is generalized oedema with skin pigmentation and thickening (**Fig. 25.10B**). The hair is dry and sparse, and may become reddish or yellow in colour. The abdomen is distended owing to hepatomegaly and/or ascites. The serum albumin is always low. The exact cause is unknown, but theories related to diet (low in protein and high in carbohydrate) and free radical damage in the presence of inadequate antioxidant defences have been proposed.
- *Marasmus* is the childhood form of starvation, which is associated with obvious wasting. The child looks emaciated, and there is obvious muscle wasting and loss of body fat. There is no oedema. The hair is thin and dry (**Fig. 25.10A**). The child is not as apathetic or anorexic as with kwashiorkor. Diarrhoea is frequently present and signs of infection must be looked for carefully.

The *WHO classification of severe malnutrition* (**Box 25.5**) makes no distinction between kwashiorkor and marasmus because the approach to treatment is similar in both. The WHO classification of chronic under-nutrition in children is based on standard deviation (SD) scores. Thus, children with an SD score between –2 and –3 (between 3 and 2 SD scores below the median, corresponding to a value between 0.13 and 2.3 centiles) can be regarded as being at moderate risk of under-nutrition; below an SD score of –3, there is a risk of severe malnutrition. A low weight-for-height is a measure of thinness (wasting when pathological) and a low height-for-age is a measure of shortness (stunting when pathological). Those with oedema and clinical signs of severe malnutrition are classified as having oedematous malnutrition.

Box 25.5 Classification of childhood malnutrition

	Moderate malnutrition	Severe malnutrition[a]
Symmetrical oedema	No	Yes: oedematous malnutrition[b]
Weight-for-height SD score	–3 to –2 (70–79%)[c]	<–3 (<70%)[c] (severe wasting)[d]
Height-for-age SD score	–3 to –2 (85–89%)[c]	<–3 (<85%)[c] (severe stunting)

[a]The diagnoses are not mutually exclusive.
[b]Older classifications use the terms kwashiorkor and marasmic kwashiorkor instead.
[c]Percentage of the median National Centre for Health Statistics/WHO reference.
[d]Called marasmus (without oedema) in the Wellcome classification and grade II in the Gomez classification.

Starvation in adults may lead to extreme loss of weight, depending on the severity and duration. Patients may crave food, be apathetic, or complain of cold and weakness with a loss of subcutaneous fat and muscle wasting. The WHO classification is based on BMI, with a value less than 18.5 kg/m^2 indicating malnutrition (severe malnutrition if <16.0 kg/m^2).

Severely malnourished adults and children are very susceptible to respiratory and gastrointestinal infections, leading to an increased mortality in these groups.

Investigations

Investigations to consider, although not always practical in certain settings in the developing world, include:

- *Blood tests* such as full blood count, iron studies, urea and electrolytes, haematinics, serology testing, and blood film. They may show the following results:
  - Anaemia due to folate and iron deficiency is often present, but the haematocrit may be high owing to dehydration
  - Eosinophilia may suggest parasitic infection
  - Electrolyte disturbances
  - Malarial parasites
  - Acute or chronic viral infection, such as human immunodeficiency virus (HIV)
- *Stools* for ova, cysts, and parasites
- *Chest radiography* may reveal pulmonary tuberculosis.

Management

Management must involve the provision of protein and energy supplements and the control of infection. The approach to the treatment of children is described below. Adults do not usually suffer from such severe malnutrition but the same general principles of treatment should be followed.

Resuscitation and stabilization

The severely ill child will require:

- correction of fluid and electrolyte abnormalities
- treatment of shock
- treatment of hypoglycaemia (blood glucose <3 mmol/L), hypothermia (reduce heat loss, and provide additional heat if necessary), and infection (antibiotics). These often coexist.

The standard WHO *oral hydration solution* has a high sodium and low potassium content and is not suitable for severely malnourished children. Instead, the rehydration solution for malnutrition (ReSoMal) is recommended. It is commercially available but can also be produced by modification of the standard WHO oral hydration solution.

Infection is common (**Box 25.6**). Diarrhoea is often due to bacterial or protozoal overgrowth; metronidazole is very effective and is often given routinely. Parasites are also common and, as facilities for stool examination are usually not available, oral mebendazole 100 mg twice daily should be given for 3 days. In high-risk areas, antimalarial therapy is given. A previous study in Malawi suggested that the addition of antibiotics to therapeutic regimens for uncomplicated SAM in children less than 5 years of age is associated with a significant improvement in recovery and mortality.

Large doses of vitamin A are also given because deficiency of this vitamin is common. After the initial resuscitation, further stabilization over the next few days is undertaken, as indicated in **Box 25.7**.

Re-feeding

During the initial management of the acute situation, a balanced diet with sufficient protein and energy is given to maintain a steady state. Large increases in energy can lead to heart failure, circulatory collapse, and death (see later section on *Re-feeding syndrome*). If necessary, fluids and food should be given by nasogastric tube. All severely malnourished children have vitamin and mineral deficiencies and they should receive multivitamin supplementation with folic acid, zinc, and copper.

Box 25.6 Infections seen in severe acute malnutrition in low- and middle-income countries

- Diarrhoea
  - Bacteria
  - Protozoa
  - Helminths
- Malaria
- Tuberculosis
- HIV infection
- Measles
- Respiratory infections

Rehabilitation

Gradually, as the child improves, more energy can be given and during rehabilitation, weight gain is achieved by providing extra energy and protein ('catch-up weight gain'). Children who have been severely ill need constant attention throughout the convalescent period, as home conditions are often poor and feeds are refused. Sensory stimulation and emotional support are major components of management during both the stabilization and rehabilitation phases. The treatment of underlying chronic infective conditions, such as HIV, malaria, and tuberculosis, is also necessary.

Care setting

In some parts of the world where malnutrition is widespread, insufficient hospital facilities emphasize the need for outpatient and community-based programmes, although these require investment, political will, and time to build to full capacity. Programmes may involve the use of ready-to-use therapeutic foods, such as energy-dense pastes with minerals and vitamins, without the need to add water, which could potentially contaminate the food.

Prognosis

Children with extreme malnutrition have a mortality of over 50%. By careful management, this can be reduced significantly to less than 10%, depending on the availability of facilities and trained staff. Treatment of underlying disease is essential. Neural development takes place in the first years of life, a time when severe starvation–associated malnutrition is frequently seen. There is evidence that intellectual impairment and behavioural abnormalities occur in severely affected children. Physical growth is also impaired. These effects may be limited by improved standards of living, balanced diets, and freedom from infection over prolonged periods.

Prevention

Prevention depends not only on the availability of adequate nutrients but also on the education of both governments and individuals in the importance of good nutrition and immunization (**Box 25.8**). Short-term programmes are useful for acute shortages of food, but long-term programmes involving improved agriculture are equally necessary. Food supplements (and additional vitamins) should be given to 'at-risk' groups by adding high-energy food (e.g., milk powder, meat concentrates) to the diet. Pregnancy and lactation

Box 25.7 Timeframe for the management of the child with severe malnutrition (10-step approach recommended by WHO)

Management	Stabilization		Rehabilitation	Follow-up
	Days 1–2	*Days 3–7*	*Weeks 2–6*	*Weeks 7–26*
1. Treat or prevent hypoglycaemia	→			
2. Treat or prevent hypothermia	→			
3. Treat or prevent dehydration	→			
4. Correct electrolyte imbalance	→	→	→	
5. Treat infection	→	→		
6. Correct micronutrient deficiencies	Without iron		With iron	
	→	→	→	
7. Begin feeding	→	→		
8. Increase feeding to recover lost weight			→	→
9. Stimulate emotional and sensorial development	→	→	→	→
10. Prepare for discharge			→	

Box 25.8 Prevention of severe acute malnutrition—GOBIF (a WHO priority programme)

- **G**rowth monitoring: the WHO has a simple growth chart that the mother keeps
- **O**ral rehydration, particularly for diarrhoea
- **B**reast-feeding supplemented by food after 6 months
- **I**mmunization: against measles, tetanus, pertussis, diphtheria, polio, and tuberculosis
- **F**amily planning

are times of high energy requirement and supplements have been shown to be beneficial.

Re-feeding syndrome

While there is no clear internationally defined definition of re-feeding syndrome, it can be described as severe fluid and electrolyte shifts related to metabolic complications in malnourished patients after the re-introduction of nutrition. It was first described following World War II, when released prisoners of war developed cardiac failure on introduction of nutrition following a prolonged period of starvation.

Pathophysiology

During starvation, reduced carbohydrate intake leads to reduced insulin secretion and increased glucagon secretion resulting in utilization of glycogen stores. Glycogen stores within the liver and skeletal muscle are metabolized into glucose, to provide energy for the body's essential processes. Once glycogen stores are consumed, gluconeogenesis from other substrates is relied on, in particular fat followed by protein sources. This results in protein, fat, electrolyte, and vitamin depletion and intolerance to sodium and water. To conserve energy expenditure, the action of cellular pumps is reduced, with electrolytes (phosphate, potassium, and magnesium) able to leak across the cell membrane. Thiamine deficiency occurs due to its water-solubility and inability to be stored.

When feeding resumes, the reintroduction of carbohydrate results in a surge of insulin release. This stimulates glycogen, fat, and protein synthesis. This process relies on phosphate and magnesium, together with co-factors such as thiamine. Insulin also drives potassium into cells via the sodium-potassium adenosine triphosphatase (ATPase) pump, which also transports glucose. The overall effect results in hypokalaemia, hypomagnesemia, hypophosphatemia, thiamine deficiency, and salt and water retention.

Ultimately this can cause cardiac, respiratory, and neuromuscular complications including altered myocardial function and respiratory distress. Thiamine deficiency can result in Wernicke-Korsakoff syndrome. The clinical effects and consequences of re-feeding syndrome are outlined in **Box 25.9**.

Risk factors

The UK National Institute of Health and Care Excellence (NICE) guidance states that patients should be considered at risk of re-feeding syndrome if one or more of the following factors are present:

- BMI < 16 kg/m^2
- Unintentional weight loss >15% in past 3 to 6 months
- Little or no nutritional intake for >10 days
- Low serum levels of potassium, phosphate, or magnesium before feeding.

Additionally, patients should be considered at risk of re-feeding syndrome if two or more of the following are present:

- BMI < 18.5 kg/m^2
- Unintentional weight loss >10% in past 3 to 6 months
- Little or no nutritional intake for >5 days
- History of alcohol misuse or drugs including insulin, chemotherapy, antacids, or diuretics.

Conditions associated with a greater risk of re-feeding syndrome include chronic alcohol use disorder, cancer, and anorexia nervosa.

Management and prevention of re-feeding syndrome

If a patient is considered to be at risk of re-feeding syndrome, the following steps should be taken:

- nutritional support should be commenced at a maximum of 10 to 20 kcal/kg/day, with levels increased slowly to meet needs over 4 to 7 days guided by dieticians and clinical judgement
- restoration of circulating volume with strict and regular fluid balance assessment to minimize risk of sodium and fluid overload
- daily monitoring of serum potassium, phosphate, and magnesium levels and proactive replacement
- provision of thiamine intravenously, for example Pabrinex Intravenous High Potency solution which contains water-soluble vitamins B_1, B_2, B_3, B_6, and vitamin C.

Box 25.9 Clinical effects and consequences of re-feeding syndrome

Electrolyte disturbance	CVS effects	Respiratory effects	CNS effects
Hypophosphataemia	Arrhythmias Altered myocardial function Cardiac failure	Acute respiratory failure	Weakness/paralysis Paraesthesia Confusion/delirium Seizures Coma Rhabdomyolysis
Hypokalaemia	Arrhythmias Cardiac arrest	Respiratory depression	Paralysis Paraesthesia Rhabdomyolysis
Hypomagnesaemia	Arrhythmias	Respiratory depression	Ataxia Seizures Coma Muscle tremors Paraesthesia Tetany
Low-serum thiamine levels	Cardiac failure Cardiomegaly	Pulmonary oedema Pleural effusions	Ataxia Coma Wernicke's encephalopathy

CNS, Central nervous system; *CVS*, cardiovascular system.

Nutrition and ageing

Most populations in high-income countries are ageing, with the current average life expectancy in the UK standing at 80.9 years. Furthermore, populations are becoming increasingly multimorbid. More than half of the population aged over 60 have multimorbid conditions. Ageing is a degenerative process that, in combination with disease, is linked to anabolic resistance and increased catabolism. This places older men and women at high risk of developing pathological weight loss and eventually under-nutrition. The process of ageing is also considered to be a low-level inflammatory state. This combination of catabolic and inflammatory processes interacts with psychosocial factors such as cognitive impairment, social isolation, and bereavement resulting in malnutrition. The prevalence of under-nutrition in older adults is significant, affecting over 10% of people over the age of 65. Under-nutrition is consistently associated with increased morbidity and mortality, with large-scale prospective studies demonstrating that being underweight is more strongly associated with death than obesity in older adults. The immune system, in particular, is sensitive to nutrient depletion leading to increased susceptibility to infection, poor wound healing, and pressure sore development.

Sarcopenia is a progressive and generalized skeletal muscle disorder that is associated with increased likelihood of adverse outcomes including falls, fractures, physical disability, and mortality. Sarcopenia was previously thought to be solely a disorder of older people and associated with ageing. However, it is now recognized to begin earlier in life and have a causal link with chronic inflammatory conditions. Low muscle strength is the principal determinant.

Maintenance of physical activity continues to be necessary for overall health, regardless of age. However, energy expenditure among older people is reduced, meaning a lower energy requirement. For people aged 60 and above, irrespective of age, the daily energy requirement has been set at approximately 1.5×BMR. Owing to reduced fat-free mass (from an average of 60 kg to 50 kg in men, and from 40 kg to 35 kg in women), overall BMR is reduced.

Nutritional deficits in the elderly are common and may be due to many factors, such as dental problems, depression, and lack of motivation. Significant malnourishment in high-income countries is usually secondary to social problems or disease. In elderly people resident in nursing or care homes, multiple nutrient deficiencies are common. Vitamin D supplementation (colecalciferol 20 μg [800 units] daily) may be required because of poor diet or inadequate sunlight exposure. Owing to the high prevalence of osteoporosis in elderly people, an increased daily calcium intake (1–1.5 g/day) is often recommended.

Further reading

Cruz-Jentoft AJ, Bahat G, Bauer J et al. Sarcopenia: revised European consensus on definition and diagnosis. *Age Ageing* 2019; 48(1):16–31.
Culkin A, White R. Refeeding syndrome. *Medicine* 2023.
Elia M. *The Cost of Disease-Related Malnutrition in the UK and Economic Considerations for the use of Oral Nutritional Supplements (ONS) in Adults.* Health Economic Report on Malnutrition in the UK. BAPEN Reports 2012.
Kelly P. Starvation and its effects on the gut. *Adv Nutr* 2021 May; 12(3):897–903.
Morton RW, Traylor DA, Weijs PJM, Phillips SM. Defining anabolic resistance: implications for delivery of clinical care nutrition. *Curr Opin Crit Care* 2018; 24(2):124–130.
National Collaborating Centre for Acute Care UK. *Nutrition Support for Adults: Oral Nutrition Support, Enteral Tube Feeding and Parenteral Nutrition*. London: National Institute for Health and Care Excellence (NICE). 2006.
Scientific Advisory Committee on Nutrition. *Dietary Reference Values for Energy*. The Stationery Office. London; 2011 https://www.gov.uk/government/publications/sacndietary-reference-values-for-energy.
Scientific Advisory Committee on Nutrition. *Carbohydrates and Health*. The Stationery Office. London; 2015 https://www.gov.uk/government/publications/sacncarbohydrates-and-health-report.
World Health Organization. *Guideline: Updates on the Management of Severe Acute Malnutrition in Infants and Children*. WHO: Geneva; 2013.

MICRONUTRIENTS AND DEFICIENCIES

The term micronutrient relates to vitamins and minerals essential to the maintenance of tissue function and homeostasis. In the majority of individuals, a healthy balanced diet should provide adequate intake. Deficiency can result in clinical symptoms and metabolic consequences. There are several factors that can contribute to micronutrient deficiency. These include:

- Environment, poverty, and social deprivation
- Chronic gastrointestinal conditions and malabsorption
- Obesity or post-bariatric surgery
- Chronic alcohol use

- Medications such as metformin, diuretics, and proton pump inhibitors
- Restrictive diets.

Micronutrient replacement depends on whether the deficiency occurs at an individual or population level, with provision of micronutrients considered a priority for the WHO. While sole preparations can be administered, it is worthwhile considering multivitamin replacements as deficiencies can often coexist.

Fat-soluble vitamins

Fat-soluble vitamins (A, D, E, K) are absorbed in the small intestine and follow lipid metabolism pathways. Patients at risk of deficiency include those with pancreatic insufficiency, post-cholecystectomy, inflammatory bowel disease, chronic liver disease, and intestinal failure.

Vitamin A

Vitamin A (retinol) is part of the family of retinoids, which is present in food and the body as esters combined with long-chain fatty acids. The richest food source is liver, but it is also found in milk, butter, cheese, egg yolks, and fish oils. Retinol or carotene is added to margarine in the UK and other countries. β-carotene is the main carotenoid found in green vegetables, carrots, and other yellow and red fruits. β-carotene is cleaved in the intestinal mucosa by carotene dioxygenase, yielding retinaldehyde, which can be reduced to retinol. Retinol is stored in the liver and is transported in plasma bound to an α-globulin, retinol-binding protein (RBP).

Function

Vitamin A has several metabolic roles, most importantly:

- Retinaldehyde in its *cis* form is found in the opsin proteins in the rods (rhodopsin) and cones (iodopsin) of the retina (see Chapter 31). Light causes retinaldehyde to change to its *trans* isomer, and this leads to changes in membrane potentials that are transmitted to the brain.
- Retinol and retinoic acid are involved in the control of cell proliferation and differentiation.

Vitamin A deficiency

Worldwide, vitamin A deficiency and xerophthalmia (see below) are major causes of blindness in young children, despite intensive preventative programmes.

Xerophthalmia has been classified by the WHO (**Box 25.10**). Impaired adaptation, followed by night blindness, is the first effect. There is dryness and thickening of the conjunctiva and the cornea (xerophthalmia occurs as a result of keratinization). Bitot's spots—white plaques of keratinized epithelial cells—are found on the conjunctiva of young children with vitamin A deficiency. These spots can, however, be seen without vitamin A deficiency, possibly caused by exposure. Corneal softening, ulceration, and dissolution (keratomalacia) eventually occur, superimposed infection is a frequent accompaniment and both lead to blindness (**Fig. 25.11**).

Diagnosis

In parts of the world where deficiency is common, diagnosis is made on the basis of the clinical features, and deficiency should always be suspected if any degree of malnutrition is present. Blood levels of vitamin A will usually be low, but the best guide to diagnosis is a response to replacement therapy.

Box 25.10 Classification of xerophthalmia by ocular signs

Ocular signs	Classification
Night blindness	XN
Conjunctival xerosis	XIA
Bitot's spot	X2
Corneal xerosis	X2
Corneal ulceration/keratomalacia <1/3 corneal surface	X3A
Corneal ulceration/keratomalacia >1/3 corneal surface	X3B
Corneal scar	XS
Xerophthalmic fundus	XF

(From WHO/UNICEF/IVACG 1988.)

Fig. 25.11 Vitamin A deficiency. Blindness due to vitamin A deficiency and corneal scarring. *(Courtesy of Ehrnstrom P. In: Bolognia J, Jorrizzo J, Rapini R [eds].* Dermatology, *2nd edn. St Louis: Mosby; 2008: Fig. 51.6.)*

Management

Urgent treatment with retinol palmitate 100 000 units orally (30 mg equivalent) should be given. In the presence of vomiting and diarrhoea, 30 mg of vitamin A is given intramuscularly. Referral for specialist ophthalmic treatment is necessary in severe cases.

Prevention

Most Western diets contain sufficient amounts of dairy products and green vegetables, but vitamin A is added to foodstuffs (e.g., margarine) in some countries. Vitamin A is not destroyed by cooking.

In some low- and middle-income countries, vitamin A supplements are given at the time a child attends for measles vaccination. Food fortification programmes are another approach. Education of the population is necessary and in particular, pregnant women and children should be encouraged to eat green vegetables and yellow fruits.

Possible adverse effects of vitamin A excess

- ***High intakes of vitamin A.*** Chronic ingestion of retinol can cause liver and bone damage, hair loss, double vision, vomiting, headaches, and other abnormalities.
- ***Retinol is teratogenic.*** The incidence of birth defects in infants is high with vitamin A intakes of more than 3 mg a day during pregnancy. In pregnancy, extra vitamin A or consumption of liver is not recommended in the UK. However, β-carotene is not toxic.

Vitamin D

Vitamin D is discussed in more detail in Chapter 33, where the most common manifestations of deficiency are outlined (rickets

and osteomalacia). Vitamin D status is assessed by measurement of 25-hydroxyvitamin D in the serum. Vitamin D receptors are distributed widely in human tissues but their function in many non-musculoskeletal tissues still remains poorly understood. Vitamin D status has been linked to a wide range of diseases, including:

- cardiovascular disorders (ischaemic heart disease, heart failure, hypertension)
- respiratory disorders (chest infections)
- renal disorders (progression of renal disease)
- endocrine disorders (type 1 and type 2 diabetes)
- neuropsychiatric disorders (depression, cognitive deficits)
- cancer (e.g., prostate, breast, colon) and mortality from various causes
- multiple sclerosis (see Chapter 28).

It has therefore been suggested that vitamin D may have a role in global health, and not just the health of the musculoskeletal system. Reduced circulating concentrations of vitamin D can result not only from lack of exposure to sunlight and a poor diet but also from inflammation, smoking, and obesity.

Vitamin E

Vitamin E is group of vitamins rich in antioxidant properties that offer protection from harmful free radicals linked to cancer and cardiovascular disease. Existing in eight forms, α-tocopherol is the naturally occurring bioavailable form absorbed in the small intestine. Vegetables and seed oils, including soya bean, saffron, sunflower, cereals, and nuts, are the main sources. Animal products are poor sources of the vitamin. Vitamin E is absorbed with fat and transported in the blood largely in LDLs. Deficiency is rare and often asymptomatic, and can occur in severe malnutrition.

Function

The biological activity of vitamin E results principally from its antioxidant properties. In biological membranes, it contributes to membrane stability. It protects cellular structures against damage from a number of highly reactive oxygen species, including hydrogen peroxide, superoxide, and other oxygen radicals. Vitamin E may also affect cell proliferation and growth.

Vitamin E deficiency

Deficiency is seen only in children with abetalipoproteinaemia (see Chapter 22) and in patients on long-term parenteral nutrition. The severe neurological deficit (gross ataxia) can be prevented with vitamin E injections.

Vitamin K

Vitamin K is found as phylloquinone (vitamin K_1) in green, leafy vegetables, dairy products, and rapeseed and soya bean oils. Intestinal bacteria can synthesize the other major form of vitamin K, menaquinone (vitamin K_2), in the terminal ileum and colon. Vitamin K is absorbed in a similar manner to other fat-soluble substances in the upper small gut. Some menaquinones must also be absorbed, as this is the major form found in the human liver.

Function

Vitamin K is a co-factor that is necessary for the production not only of blood clotting factors (II, VII, IX, and X, and other proteins involved in coagulation) but also of proteins that are necessary for the formation of bone. Vitamin K is a co-factor for the post-translational carboxylation of specific protein-bound glutamate residues in γ-carboxyglutamate (Gla). Gla residues bind calcium ions to phospholipid templates, and this action on factors II, VII, IX, and X, and on proteins C and S, is required for coagulation to take place. Bone osteoblasts contain three vitamin K–dependent proteins: osteocalcin, matrix Gla protein, and protein S, which have a role in bone matrix formation. Osteocalcin contains three Gla residues, which bind tightly to the hydroxyapatite matrix, depending on the degree of carboxylation; this leads to bone mineralization. There is, however, no convincing evidence that vitamin K deficiency or antagonism affects bone other than rapidly growing bone.

Vitamin K deficiency

Vitamin K deficiency results in inadequate synthesis of clotting factors, which leads to an increase in the prothrombin time and haemorrhage. It has also been linked to bone health, but there is inadequate evidence to support the routine use of vitamin K to prevent osteoporosis and reduce fracture risk. Deficiency occurs in the newborn, in cholestatic jaundice, and with concomitant use of vitamin K antagonists.

The newborn

Deficiency occurs in the newborn owing to:

- poor placental transfer of vitamin K
- modest quantities of vitamin K in breast milk
- the lack of hepatic stores of menaquinone (due to a lack of intestinal bacteria in the neonate).

Deficiency leads to a haemorrhagic disease of the newborn, which can be prevented with prophylactic vitamin K. Vitamin K (phytomenadione 1 mg, intramuscularly) is administered to all neonates after the risks have been discussed with parents and consent has been obtained.

Cholestatic jaundice

When bile flow into the intestine is interrupted, malabsorption of vitamin K occurs as no bile salts are available to facilitate absorption and the prothrombin time increases. This can be corrected by giving 10 mg of phytomenadione intramuscularly or intravenously. (Note that an increased prothrombin time caused by liver disease does not respond to vitamin K injection, there being no shortage of vitamin K—just reduced synthesis of clotting factors.) In patients with chronic cholestasis (e.g., primary biliary cholangitis), oral therapy with a water-soluble preparation, menadiol sodium phosphate 10 mg daily, is used.

Concomitant vitamin K antagonists

Oral anticoagulants, such as warfarin, antagonize vitamin K. Antibacterial drugs also interfere with the bacterial synthesis of vitamin K.

Water-soluble vitamins

Vitamin C (ascorbic acid)

Humans are unable to synthesize ascorbic acid from glucose, meaning that adequate dietary intake is essential to achieve normal concentrations. Vitamin C is present in citrus fruits, potatoes, and vegetables. Ascorbic acid has many important actions including the promotion of iron absorption and formation of cross-linking in connective tissues. Most absorption occurs within the jejunum and terminal ileum.

Vitamin C deficiency (scurvy)

The clinical syndrome associated with vitamin C deficiency is known as scurvy. Patients at risk of this and in whom deficiency should be suspected include the elderly and those who are socially deprived

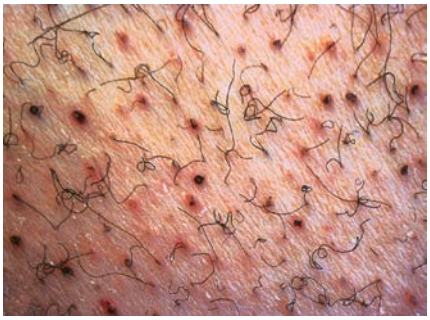


Fig. 25.12 Scurvy. Corkscrew hairs and perifollicular haemorrhage on the lower extremity. *(From Bolognia JL, Jorrizzo JL, Schaffer JV [eds].* Dermatology, *3rd edn. St Louis: Mosby; 2012, with permission.)*

or have alcohol use disorder. Deficiency results in defective collagen synthesis leading to the following symptoms:

- Keratosis of hair follicles with 'corkscrew' hair (Fig. 25.12)
- Perifollicular haemorrhages
- Swollen, spongy gums with bleeding and loosening of teeth
- Spontaneous bruising and haemorrhage
- Failure of wound healing
- Anaemia.

Plasma levels and leucocyte-platelet layer can be used to make a laboratory-based diagnosis; however, this is rarely done in clinical practice. Treatment is with 250 mg of ascorbic acid in daily divided doses and symptoms improve dramatically with treatment.

B vitamins

B vitamins are water-soluble vitamins that act as coenzymes and precursors essential for metabolism. They are generally absorbed in the small intestine. Box 25.11 outlines the essential functions of B vitamins, dietary sources, clinical features of deficiency, and in whom to suspect deficiency.

Clinical conditions as a manifestation of B vitamin deficiencies

Pellagra

The clinical syndrome associated with niacin (vitamin B_3) deficiency is known as pellagra. Pellagra is characterized by the '4 Ds': (1) dermatitis (Fig. 25.13), (2) delirium, (3) diarrhoea, and (4) death. The skin manifestation is a photosensitive, blistering, and erythematous rash which affects the sun-exposed areas. One should suspect this deficiency in patients with malnutrition, alcohol use disorder, and/or homelessness. The presentation can often mimic alcohol withdrawal or intoxication or mental illness. Treatment with intravenous B vitamin supplementation (Pabrinex) is indicated, especially as other water-soluble vitamin deficiencies are likely to coexist.

Beriberi

This clinical condition arising as a result of thiamine (vitamin B_1) deficiency is now confined to the poorest areas of South-east Asia. There are two main clinical types of beriberi, which, surprisingly, rarely occur together.

- ***Dry beriberi*** usually presents insidiously with a symmetrical polyneuropathy. The initial symptoms are heaviness and stiffness of the legs, followed by weakness, numbness, and pins and needles. The ankle jerk reflexes are lost and eventually signs of polyneuropathy may progress to involve the trunk and arms. Cerebral involvement occurs, producing the picture of the Wernicke–Korsakoff syndrome (Chapter 28).
- ***Wet beriberi*** causes oedema. Initially, this is of the legs, but it can extend to involve the whole body, with ascites and pleural effusions. The peripheral oedema may mask the accompanying features of dry beriberi. Thiamine deficiency impairs pyruvate dehydrogenase with accumulation of lactate and pyruvate, producing peripheral vasodilatation and eventually oedema. The heart muscle is also affected and heart failure occurs, causing a further increase in the oedema. Initially, there are warm extremities, a full, fast, bounding pulse, and a raised venous pressure ('high-output state'), but eventually heart failure advances and a poor cardiac output ensues. The electrocardiogram may show conduction defects.

Infantile beriberi occurs, usually acutely, in breast-fed babies at approximately 3 months of age. The mothers show no signs of thiamine deficiency but presumably their body stores must be virtually nil. The infant becomes anorexic, develops oedema, and has some degree of aphonia. Tachycardia and tachypnoea develop and, unless treatment is instituted, death occurs quickly.

Both dry and wet forms of beriberi respond well to thiamine supplementation if diagnosed and treated early. For wet beriberi, a rapid disappearance of oedema after thiamine is diagnostic. Other causes of oedema must be considered (e.g., renal or liver disease), while the polyneuropathy of dry beriberi is indistinguishable from that due to other causes, which thus require careful exclusion. The diagnosis is confirmed by measurement of the circulating thiamine concentration or transketolase activity in red cells using fresh heparinized blood.

Wernicke-Korsakoff syndrome

In the developed world, thiamine deficiency is usually seen in the context of alcohol dependence and re-feeding syndrome. Rarely, wet beriberi syndrome can develop and this will need to be distinguished from alcohol-related cardiomyopathy. Usually however, thiamine deficiency will manifest with Wernicke-Korsakoff syndrome. This syndrome consists of dementia, ataxia, ophthalmoplegia, and nystagmus. If treated promptly it can be reversible. Treatment is with intravenous thiamine replacement for 2 weeks.

Minerals

Calcium

Essential for bone and tooth formation. Dairy foods are rich sources of calcium. Postmenopausal women and those following a lactose-free diet are at risk of deficiency. Calcium absorption from the gastrointestinal tract is vitamin D dependent. Increased calcium is required in pregnancy and lactation, when dietary intake must be increased. Calcium deficiency is usually due to vitamin D deficiency.

Chloride

Essential for gastric hydrochloric acid production and therefore the process of digestion. Renal disease can put patients at risk of deficiency.

Chromium

Essential for lipid and carbohydrate metabolism and insulin action. Short bowel syndrome is a risk factor.

Copper

Required for nervous system and immune system health, and for the production of caeruloplasmin. It is found in nuts and shellfish. Deficiency can present with cardiac arrhythmias and myeloneuropathy. Menkes' disease is a rare sex-linked recessive condition, with

Box 25.11 B vitamins: functions, sources, and deficiency and replacement

B vitamin	Actions	Dietary source	Deficiency	Who to suspect in	Replacement
B_1 thiamine	Essential for energy release from carbohydrates DNA and RNA production	Fish Fortified flour	Wernicke's encephalopathy Beriberi syndrome (see text)	Chronic alcohol use Liver disease Re-feeding	i.v. Pabrinex or oral thiamine
B_2 riboflavin	Co-factor for energy production, cellular function, and fat metabolism	Leafy vegetables Additional riboflavin is synthesized by colonic microbiota	Rare Symptoms include dry skin, angular stomatitis, and thyroid problems		Vitamin B compound strong 1–2 tablets tds
B_3 niacin	Energy production	Fortified processed foods Poultry	Rare Pellagra (see text)	Elderly Homeless Chronic alcohol use	i.v. Pabrinex or 300 mg oral nicotinamide daily
B_5 pantothenic acid	Synthesizes co-factor A, catabolic and metabolic reactions	Meat Dairy produce	Fatigue Anorexia		Oral pantothenic acid 5 mg daily
B_6 pyridoxine	Group of 6 compounds required for protein metabolism	Meat Potatoes	Microcytic anaemia Depression Dry skin Paraesthesia	Chronic alcohol use Dialysis Isoniazid use (tuberculosis)	Oral pyridoxine 20–50 mg 1–3 times per day
B_7 biotin	Regulates gene expression Immune response	Fish Eggs	Dry skin Brittle nails Depression	Chronic alcohol use Short bowel syndrome	Oral biotin 10 mg daily
B_9 folate	Essential coenzyme for RNA and DNA transfer and cell division	Leafy green vegetables Nuts	Megaloblastic anaemia Pancytopenia	Chronic alcohol use Drugs such as methotrexate, sulfasalazine, phenytoin, trimethoprim can cause folate deficiency	Folic acid 5 mg od Preconception folic acid (400 mcg) is required to reduce risk of neural tube defects
B_{12} vitamins	Blood cell production CNS function	Fish Dairy foods	Macrocytic anaemia Glossitis Peripheral neuropathy	Pernicious anaemia Post-gastrectomy Metformin use Proton pump inhibitor use e.g., omeprazole, lansoprazole Nitrous oxide abuse	IM hydroxocobalamin 1 mg 3 times per week for 2 weeks then 1 mg every 3 months or PO cyanocobalamin 10–150 mcg daily

a mutation in the copper-transporting gene *ATP7A*. Infants with this abnormality develop growth failure, intellectual disabilities, bone lesions, and brittle hair. Copper toxicity occurs in Wilson's disease (see Chapter 23).

Fluoride

Essential for dental health, with fluoridated toothpastes, mouthwashes, and fluoridated water supplies needed to maintain concentrations. In areas where the level of fluoride in drinking water is less than 1 part per million (0.7–1.2 mg/L), dental caries are relatively more prevalent. Fluoridation of the water provides 1 to 2 mg daily, resulting in a reduction of tooth decay in children by about 50%. There is little fluoride in food. Fluoride-containing toothpaste may add up to 2 mg/day.

Iodine

Essential for the production of thyroid hormones. Many areas throughout the world lack iodine in the soil, and so iodine deficiency, which impairs brain development, is a WHO priority. Two billion people worldwide (one-third of whom are children) have insufficient iodine intake. Endemic goitre (Fig. 25.14) occurs in remote areas where the daily intake is below 70 μg, and in those parts 1% to 5% of babies are born with congenital hypothyroidism (with severe stunting, learning difficulty, and a goitre). In these areas, iodized oil should be given intramuscularly to all reproductive women every 3 to 5 years. Salt iodization is now practised in many countries and is a simple, cost-effective way to prevent deficiency.

Iron

The most abundant trace element within the human body. It is required to maintain normal physiological processes and for energy metabolism. The main function of iron is as a functional component of haem, participating in oxygen binding and transport. Iron is also involved in cellular processes such as DNA synthesis, cell proliferation, and gene regulation. Dietary sources include red meat and fortified grain products. Worldwide, iron deficiency is the most common nutritional deficiency. Patients with underlying gastrointestinal

Fig. 25.13 Pellagra. Hyperpigmentation with desquamation of the dorsal aspects of the hands and forearms. *(From Bolognia JL, Jorrizzo JL, Schaffer JV [eds].* Dermatology, *3rd edn. St Louis: Mosby; 2012, with permission.)*

Fig. 25.14 Goitre due to iodine deficiency. *(From Dhillon RS, East CA. Ear,* Nose and Throat and Head and Neck Surgery, *4th edn. Elsevier; 2013, with permission.)*

cancers and menstruating and pregnant women are at risk of deficiency, which can present with a microcytic anaemia.

Magnesium

Essential to maintain muscle, nerve, and cardiovascular function and as a co-factor for many chemical reactions in the body. Dietary sources include leafy vegetables, nuts, and seeds. Malabsorptive gastrointestinal diseases and diuretic use predispose to deficiency. Magnesium deficiency can present with myopathy, fatigue, cardiac arrhythmias, and seizures. Biochemically, hypokalaemia frequently coexists due to the role of magnesium in cellular potassium uptake.

Manganese

Essential co-factor needed for amino acid, cholesterol, glucose, and carbohydrate metabolism. Dietary sources include leafy vegetables, nuts, and grains. Deficiency is rare. Toxicity is more of an area for concern, which can cause neurological problems such as tremor, gait disturbance, and facial spasms. Patients on long-term parenteral nutrition, and also patients with chronic liver disease are at risk of toxicity.

Molybdenum

Essential for co-factor production needed for drug and toxin metabolism. Deficiency is rare as high amounts are present in many foods, including grains and legumes.

Phosphate

A component of phospholipids found in RNA, DNA, and ATP. Absorption occurs in the first half of the small bowel, from foods including dairy produce and fish. Deficiency can present with bone pain, weakness, and paraesthesia, which can develop acutely in refeeding syndrome. Hyperphosphataemia is seen in renal failure as a result of reduced excretion.

Potassium

An intracellular cation that forms part of the sodium–potassium ATP transporter channel which maintains adequate fluid balance. Diarrhoea and vomiting increase losses, with diuretics also reducing serum concentrations. Concentrations also fall rapidly in refeeding syndrome. Dietary sources include bananas, potatoes, and legumes. Symptoms of deficiency include fatigue and muscle weakness, while cardiac arrhythmias may also occur.

Sodium

Essential for regulation of fluid and electrolyte balance, muscle contraction, and nerve conduction. Dietary sources include processed meats, cheeses, and sauce. Hyponatraemia, a low serum sodium which may reflect either sodium deficiency or water excess, is discussed in Chapter 27.

Selenium

A trace element important for thyroid hormone synthesis, DNA production, and protection from oxidative damage. Dietary sources include bread and poultry. Clinical deficiency is rare, but seen in areas of China where Keshan disease, a selenium-responsive cardiomyopathy, occurs. Deficiency may also cause a skeletal myopathy.

Zinc

Involved in many metabolic pathways and is essential for the synthesis of RNA and DNA. Acrodermatitis enteropathica is an inherited disorder caused by malabsorption of zinc. Infants develop growth retardation, severe diarrhoea, hair loss, and a skin rash, which can occur anywhere on the body but is most often found around the mouth, genitalia, and hands (**Fig. 25.15**). A similar rash occurs in adults suffering from zinc deficiency due to other causes; see below. This condition provides a model for zinc deficiency. Zinc supplementation results in a complete cure. Zinc deficiency probably also plays a role in protein-energy under-nutrition and in many diseases in children in the developing world. Zinc supplementation has been demonstrated as being of some benefit in, for example, the prevention of diarrhoeal diseases and acute respiratory infections; it also improves growth. Levels of zinc have also been shown to be low in some patients with malabsorption or skin disease, and in patients with acquired immunodeficiency syndrome (AIDS), but its exact role in these situations is disputed. Excessive amounts of zinc carry a low risk of toxicity, but high levels can interfere with iron and copper absorption. Conversely, administration of copper or iron to treat deficiencies such as iron deficiency anaemia can precipitate zinc deficiency. Wound healing is impaired with moderate zinc deficiency and is improved by zinc supplements. Impaired taste and smell, hair loss, and night blindness are also features of severe zinc deficiency.

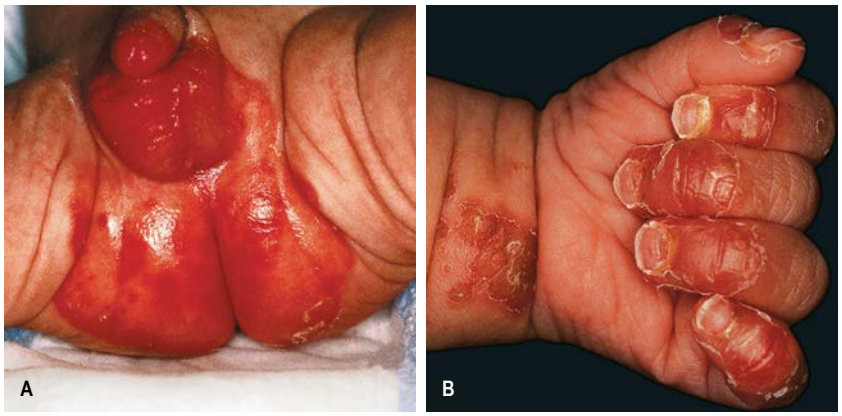


Fig. 25.15 Zinc deficiency (genetic). **(A)** Erythema and erosions. **(B)** Crusting and desquamation. *(From Bolognia JL, Jorrizzo JL, Schaffer JV [eds].* Dermatology, *3rd edn. St Louis: Mosby; 2012, with permission.)*

Box 25.12 Disease-specific risks of micronutrient depletion or deficiency

Disease	Deficiency favouring disease development	Deficiency as a result of disease
Alcohol use disorder		A, D, E, K, B_1, B_2, B_3, B_6, B_7, B_9, B_{12}, C, zinc
Alcohol-related hepatitis	B_6, zinc	
Anaemia	B_1, B_6, B_9, B_{12}, iron, copper, cobalt	
Cardiomyopathies Heart failure	B_1, B_6, D, selenium, iron	
Chronic obstructive pulmonary disease	D, copper, selenium, manganese, zinc	
Chronic (type 3) intestinal failure		B_2, B_7, B_9, B_{12}, A, D, E, K, copper, iron, zinc
Chronic atrophic gastritis		B_9, B_{12}, C, D, iron
Diabetes mellitus	B_9, chromium	
Inflammatory bowel disease		B_1, B_6, B_{12}, A, D, E, K, iron, selenium, zinc
Multiple sclerosis	B_7	
Obesity post-bariatric surgery		A, D, E, K, B_1, B_9, B_{12}, C, copper, iron, zinc
Osteoporosis	B_{12}, D, K, copper, iron, zinc, manganese, fluoride	
Chronic kidney disease		B_1, B_6, B_9, K, D, copper, selenium, zinc

Adapted from Mette M Berger et al. ESPEN micronutrient guideline. Clin Nutr. *2022 Jun;41(6):1357–1424. doi: 10.1016/j.clnu.2022.02.015. Epub 2022 Feb 26.*

A list of common disorders and micronutrient deficiencies both predisposing to and resulting from these is found in **Box 25.12**.

Food fortification

Fortified food is a food product to which vitamins, minerals, energy, protein, or other nutrients, or a combination of them, have been added to increase energy or nutrient density. It can play an important role in improving micronutrient intake of the population as a whole, particularly when targeted appropriately.

As defined by the WHO and the Food and Agricultural Organization of the United Nations (FAO), fortification refers to 'the practice of deliberately increasing the content of an essential micronutrient, that is vitamins and minerals (including trace elements) in a food, to improve the nutritional quality of the food supply and to provide a public health benefit with minimal risk to health'.

Fortification of foods with specific nutrients is common. In the UK, margarine and milk are fortified with vitamins A and D, flour with calcium, iron, thiamine, and niacin, and breakfast cereals with several vitamins and iron.

Further reading

Awasthi S, Peto R, Read S et al. Vitamin A supplementation every 6 months with retinol in 1 million pre-school children in north India: DEVTA, a cluster-randomised trial. *Lancet* 2013; 381:1469–1477.
Berger MM, Shenkin A, Schweinlin A et al. ESPEN micronutrient guideline. *Clin Nutr* 2022; 41(6):1357–424.
Granger M, Eck P. Dietary vitamin C in human health. *Adv Food Nutr Res* 2018; 83:281–310.

NUTRITIONAL SUPPORT

Principles of nutritional support

The goals of nutritional support are to treat or prevent malnutrition and to improve patient-related outcomes. All patients should be screened for malnutrition on admission and the findings linked to a care plan, preferably under the supervision of a trained multidisciplinary team, including a dietician. There are many different means of providing nutritional support in both the community and hospital settings. Plans are discussed with patients and consent is taken for any invasive procedure (e.g., nasogastric tube, parenteral

nutrition). If the patient is unable to give consent, the healthcare team should act in the patient's best interest, considering any previously expressed wishes of the patient and views of the family.

The delivery of nutritional support can be described as either enteral or parenteral. Enteral nutrition refers to artificial feeding, which uses the gastrointestinal tract as a means of achieving calorie requirements. This incorporates oral nutritional supplements, nasogastric or percutaneous gastrostomy tubes, and post-pyloric feeding with nasojejunal or jejunostomy tubes.

Parenteral nutrition (PN) refers to intravenous nutrition and can be total or supplementary. PN is only indicated when it is not possible to supply nutrition using the gastrointestinal tract, that is due to intestinal failure or intestinal insufficiency (see later).

Nutritional requirements for adults

The exact nutrient requirements in many disease states are not clearly defined and vary with the stage and severity of disease, as well as nutritional status. There are different nutritional guidelines for patients with liver cirrhosis, cancer, chronic kidney disease, inflammatory bowel disease, obesity, and critical illness. However, generally, the following guidance applies to adults:

Fluid. Fluid requirements are generally calculated using 25 to 35 mL/kg/day (>60 years = 30 mL/kg, 18–60 years = 35 mL/kg). However, various other factors need to be taken into consideration. For instance, gastrointestinal losses (e.g., high-output jejunostomy or diarrhoea), fever and insensible losses, and also fluid-overloaded states such as renal failure, congestive cardiac failure, and oedema.

Energy. As previously discussed, typical energy requirements are approximately 7.5 to 10.0 MJ/day (1800–2400 kcal/day). When aiming for weight gain, an additional 400 to 600 kcal/day can be considered in metabolically stable patients. Disease increases resting energy expenditure but may also decrease physical activity. When energy requirements are calculated, other parameters affecting energy expenditure should also be considered. For example, fever, shivering, the work of breathing, and inflammation can all increase requirements. Conversely, muscle relaxants, analgesia, hypothermia, and adapted ventilation can all reduce requirements. Baseline energy requirements can be determined using a weight-based calculation, for example 25 to 35 kcal/kg/day. Predictive equations such as the Penn State equation are used to calculate energy requirements in critically ill patients who are ventilated on intensive care.

Carbohydrate as an energy source is provided in the form of glucose. It is generally provided in amounts of up to 50% to 70% of non-nitrogen energy. Glucose is essential for central nervous system functions, however exceeding the safe level can result in hyperglycaemia, hepatic steatosis, and impaired respiratory function with increased CO_2 production. The safe level can be determined using the glucose oxidation rate (4–7 mg/kg/min/day); when this has reached its maximum, excessive glucose can no longer be oxidized and it will be converted to fat or carbon dioxide.

Lipid in PN solutions provides a concentrated form of energy reducing the need for infusion of large amounts of glucose. It also prevents EFA deficiency. Lipid provides <50% of total non-nitrogen energy.

Protein. Protein is provided in the form of nitrogen, which is one of the main components of amino acids. Most stable patients require 0.8 to 1.4 g/kg/day of protein, where 1 g of nitrogen is equivalent to 6.25 g of protein. Requirements depend on the degree of stress and catabolism present. Most patients receive at least 11 to 15 g of nitrogen per day. Extra protein may be needed in severely catabolic conditions, such as extensive burns, sepsis, and major trauma.

Electrolyte requirements. The electrolyte requirements in an adult are outlined in **Box 25.13**.

Requirements are increased in patients with gastrointestinal effluents, for example high sodium and potassium losses from surgical drains, jejunostomies, ileostomies, and diarrhoea. The excretion of these minerals in various effluents can provide an indication of the additional requirements (see **Box 27.11**). Requirements may be lower in patients with fluid overload or if baseline levels are deranged (e.g., those with hypernatraemia or hyperkalaemia). The requirements for calcium and magnesium are higher in enteral than in parenteral nutrition because only a proportion of these minerals are absorbed by the gut.

Trace elements. For trace elements such as iodine, fluoride, and selenium, which are well absorbed, the requirements are similar in enteral and parenteral nutrition. For other trace elements, such as iron, zinc, manganese, and chromium, the requirements in parenteral nutrition are substantially lower than in enteral nutrition.

Vitamins. Many vitamins are given in greater quantities in patients receiving PN than in those receiving enteral nutrition. This is because patients on PN may have increased requirements, partly because of severe disease, partly because they may already have depleted pools of vitamins, and partly because some vitamins degrade during storage. Vitamin K is usually absent from parenteral nutrition regimens and therefore it may need to be administered separately.

Enteral nutrition

Enteral nutrition is indicated in patients who cannot meet their nutritional requirements by their standard oral dietary intake and have a functioning and accessible gastrointestinal tract. Clinical situations where enteral nutrition via a feeding tube may be indicated include a reduced conscious level such as in hepatic encephalopathy, when a patient has an unsafe swallow secondary to neurological disorders such as a stroke, or when a patient is critically ill in the intensive care unit.

Nasogastric feeding is a generally well-tolerated and safe method of nutritional therapy. Some common problems include compliance and tube displacement (especially if a patient is confused or agitated), nausea/vomiting, diarrhoea, aspiration, and refeeding syndrome. Decisions regarding nasal versus percutaneous feeding tube will be determined by factors such as reversibility of acute pathology, prognosis, and patient choice. When long-term enteral nutrition greater than 4 to 6 weeks is anticipated, then

Box 25.13 Electrolyte requirements for adults

Electrolyte	Na^+	K^+	Ca^{2+}	Mg^{2+}	PO_4^{3-}	Cl^-
Requirement (mmol/kg/day)	1.0–1.5	1.0–1.5	0.1–0.15	0.1–0.2	0.5–0.7	1.0–1.5

Box 25.14 Indications for PEG insertion

Unsafe swallow	Failure of feeding
• Acute ischaemic or haemorrhagic stroke • Progressive neuromuscular disease e.g., motor neurone disease, Huntington's disease, Parkinson's disease, bulbar and pseudobulbar palsies	• Cystic fibrosis • Head and neck cancers

insertion of a percutaneous endoscopically placed gastrostomy (PEG) should be considered.

Percutaneous enteral feeding tube placement is performed to facilitate direct enteral feeding for patients whose oral intake is insufficient to meet their nutritional requirements or when anatomical abnormalities or neurological function precludes the normal process of swallowing. Common indications are outlined in **Box 25.14**. This list is not exhaustive and it should be noted that in clinical practice, there may be other situations where PEG placement is suitable after assessment by a wider multidisciplinary nutrition team.

The procedure takes place endoscopically (**Fig. 25.16**) and should be performed under full aseptic non-touch technique and with antibiotic prophylaxis to prevent skin site infection and clinically relevant bacteraemia. The surface marking for PEG placement requires endoscopic visualization with finger indentation and transillumination (when infrared light emitted from the endoscope into the gastric wall is visualized externally). Where endoscopic access is not possible, gastrostomies can be placed radiologically or surgically.

Complications arising from PEG insertion may be minor or major and can develop immediately after tube placement or later. Immediate and early risks from PEG placement include respiratory compromise, aspiration, bleeding, and peritonitis. Later complications include infection, leakage of gastric contents, and buried bumper syndrome, where the internal bumper of the PEG is pulled up too tightly against the gastric mucosa and over time erodes into it, becoming 'buried'. This can lead to obstruction of the tube with feed leaking back onto the skin. This should be suspected if the PEG tube does not push in easily. It can be prevented by regular rotation and gently pushing the tube in on a weekly basis.

Post-pyloric feeding

Post-pyloric feeding, that is feeding beyond the stomach, can be provided by a nasojejunal tube or jejunostomy. This method of feeding is indicated when there is pathology such as gastric outlet obstruction (GOO), or an anatomical reason that precludes gastric feeding such as large hiatus hernia or gastrectomy. GOO is a clinical syndrome characterized by abdominal pain and postprandial vomiting. Causes of GOO can be benign or malignant. Pancreatic cancer with local invasion and distal gastric cancer are common causes. Nasojejunal tubes require endoscopic placement, although bedside placement devices are also available. For long-term post-pyloric feeding, jejunostomies can be placed endoscopically, through an extension of an existing PEG, radiologically or surgically.

Feeding regimen

Enteral feeds are sterile nutritional liquids which are often nutritionally complete. Feeds are available in different preparations and volumes. Nutritional assessment helps to inform which type of feed is most suitable for the patient. The feeding regimen should be tailored to meet the patient's individual requirements and target goals of

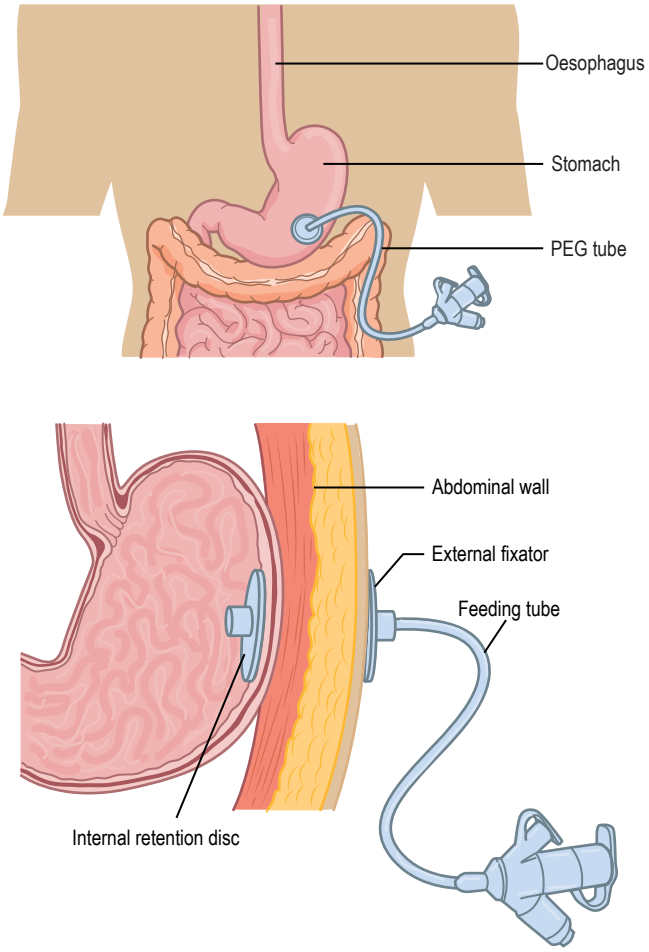


Fig. 25.16 Percutaneous endoscopic gastrostomy.

nutritional intervention. Most feeds provide 1 to 2 kcal/mL. Standard polymeric feeds contain whole protein and long-chain triglycerides (i.e., in the form in which they are within a normal diet), and are nutritionally complete. Some patients with severely impaired gastrointestinal function and malabsorption will require a pre-digested feed (i.e., semi-elemental/elemental). These feeds contain protein as low molecular weight short peptides or free amino acids. Carbohydrate provides the majority of the energy and energy is also provided from fat as either medium- or long-chain triglycerides. The aim is to improve absorption of nutrients. Choice of feed, rate, and volume will be guided by dietetic assessment, underlying clinical condition, and risk of re-feeding. Patients receiving enteral feeding require close monitoring to ensure nutritional requirements are being met, electrolyte and metabolic disturbances are reduced, and complications are avoided. Monitoring includes a food chart, fluid balance, weight, biochemistry, observations, bowel chart, nausea or vomiting, and gastric residual volumes to help determine tolerance of feeding and absorption.

Parenteral nutrition

Parenteral nutrition (PN) is the term to describe intravenous support provided to achieve normal fluid and nutritional balance and is indicated in the management of intestinal failure. PN can provide sufficient nitrogen for protein synthesis, calories for energy and electrolytes, vitamins, and trace elements to meet the patient's nutritional requirements. To establish a PN regimen, the patient

Box 25.15 Common types of vascular access for parenteral nutrition

Type of access	Disadvantages
Peripherally inserted central catheters (PICC)	Risk of infection, thrombosis Difficult to self-access if administering home PN (though mitigated by extension sets)
Tunnelled central venous catheters e.g., Hickman line	Risk of infection, thrombosis Requires surgical removal
Implanted port	Not commonly used therefore not all healthcare staff familiar with accessing Port infections generally cannot be treated with antibiotics alone and require port removal

needs to be assessed by a dietician and undergo a nutritional assessment. As discussed earlier, there are many factors that need to be considered as part of a nutritional assessment. These include anthropometry, weight, BMI, recent weight loss, biochemistry, and any changes to nutritional intake, which can indicate whether the patient is malnourished. Medical history is important as certain co-morbidities might influence patients' requirements.

Vascular access for parenteral nutrition

When it comes to delivery of PN, different types of venous catheters exist, some of which are more suitable in the short term and others in the long term (Box 25.15). Tunnelled lines are traditionally preferred for long-term PN administration. These are commonly known as Hickman lines, whereby catheters enter a central vein and are tunnelled through the skin to exit over the chest wall. Alternatively, peripherally inserted central catheters (PICCs) can be used for short- or medium-term PN. With either type of venous access, the risk of venous thrombosis is reduced when the catheter tip is located at the superior vena cava–right atrium junction. This can be useful when interpreting chest radiographs for venous line position.

Complications of parenteral nutrition

Complications can be related to vascular access or the parenteral nutrition itself. The most serious complication is infection, of which there are different types. An exit site infection refers to inflammation or purulence within 2 cm of an exit site. A tunnel infection refers to inflammation along a catheter greater than 2 cm from the exit site, and a pocket infection refers to inflammation or necrosis overlying a port, which is used to deliver PN in some centres.

Catheter-related bloodstream infection (CRBSI) refers to a bacteraemia that has originated from an intravenous catheter. Risk factors include the type of catheter used, the number of infusions per week, and whether patients are self-caring or nursed. Factors shown to reduce infection rates include strict asepsis during catheter handling, the use of 2% chlorhexidine during catheter insertion, the absence of a stoma, a single-lumen catheter (compared to a multi-lumen catheter), and the antimicrobial line lock agent taurolidine.

Initial diagnostic methods without catheter removal include paired quantitative blood cultures using pour plates. A CRBSI is diagnosed if the same organism is isolated from both blood obtained via the central catheter and a peripheral vein, and the colony count of microbes grown from blood obtained through the catheter hub is at least 3-fold greater than the colony count from blood obtained from a peripheral vein. This is considered gold standard but is not practiced universally. An alternative is paired *qualitative* blood cultures with a differential time to positivity of greater than 120 minutes. However, this requires standardization of factors such as volume of blood in each sample and delivery to labs within 30 minutes.

Other complications include thrombosis and intestinal failure-associated liver disease, or IFALD. IFALD is defined as abnormal liver function with evidence of radiological or histological liver abnormalities occurring in a patient with intestinal failure, in the absence of alternative liver pathology.

Box 25.16 Pathophysiological classification of intestinal failure

Mechanism	Potential causes
Reduced absorption	
Short bowel syndrome	Extensive intestinal resection after infarction Crohn's disease Volvulus Trauma
Intestinal fistula	Post-abdominal operation Crohn's disease Malignancy Radiation enteropathy
Extensive small bowel disease	Crohn's disease Radiation enteropathy Intestinal lymphangiectasia Chemotherapy-induced enteritits
Restriction of oral/enteral intake	
Intestinal dysmotility	Postoperative ileus Chronic intestinal pseudo-obstruction
Intestinal mechanical obstruction	Malignancy Crohn's disease Abdominal adhesions Desmoid tumours

Further reading

Burch N. Artificial nutrition support. *Medicine*. 2023.
Pittiruti M, Hamilton H, Biffi R et al. ESPEN Guidelines on Parenteral Nutrition: central venous catheters (access, care, diagnosis and therapy of complications). *Clin Nutr* 2009; 28: 365–377.
Staun M et al. *ESPEN Guidelines on Parenteral Nutrition: Home Parenteral Nutrition (HPN) in Nutrition Patients* 2009; 28: 467–479.
Stroud M, Duncan H, Nightingale J. Guidelines for enteral feeding in adult hospital patients. *Gut* 2003; 52(suppl 7):vii1–2.
Westaby D, Young A, O'Toole P, Smith G, Sanders DS. The provision of a percutaneously placed enteral tube feeding service. *Gut* 2010; 59(12):1592–605.

INTESTINAL FAILURE

Intestinal failure (IF) is the rarest organ failure, although the prevalence is increasing. Intestinal failure is defined by the European Society for Clinical Nutrition and Metabolism (ESPEN) as the reduction of gut function below the minimum necessary for the absorption of macronutrients and/or water and electrolytes, such that intravenous supplementation is required to maintain health and/or growth. The reduction of gut absorptive function that does not require intravenous supplementation can be considered as intestinal insufficiency.

There are five major pathophysiological mechanisms that cause intestinal failure (Box 25.16).

Where short bowel syndrome, intestinal fistulae, and extensive small bowel disease are concerned, the primary mechanism

of intestinal failure is reduced intestinal absorptive capacity, either due to reduction of mucosal surface area or intestinal bypass. Conversely, with intestinal dysmotility and mechanical obstruction, the primary mechanism leading to IF is restricted oral or enteral nutrition. This can occur either as a result of feeding-related exacerbation of GI symptoms or due to mechanical intestinal obstruction.

Furthermore, there are many secondary pathophysiological mechanisms that can contribute to the development of IF. These include:

- Accelerated gastrointestinal transit time
- Increased intestinal secretion of fluids and electrolytes within dilated or obstructed bowel segments
- Loss of fluids/electrolytes through high output stomas, gastric drainage, vomiting
- Disease-related hypophagia
- Small bacterial overgrowth
- Increased metabolic demand due to concomitant infection, inflammation, or sepsis.

Functional classification of intestinal failure

The most useful classification in clinical practice is the functional classification, which describes types 1, 2, and 3 IF. In all IF types, a multidisciplinary and multispecialty approach is critical, ideally delivered by a team with expertise and in a centre with the correct infrastructure and services to deliver optimal care. Patient choice is clearly also vital and should underpin every decision made.

Type 1 intestinal failure

This describes an acute, and often self-limiting, type of IF diagnosed in approximately 15% of patients after abdominal surgery, or in association with critical illness. This usually resolves within days to weeks. Patients with type 1 IF generally have excellent outcomes with no long-term sequelae.

Type 2 intestinal failure

Type 2 IF is usually reversible but lasts for more than 28 days. It commonly arises as a consequence of abdominal sepsis and enterocutaneous fistula(e), bowel obstruction, or ileus and may be associated with metabolic consequences.

In type 2 IF, where PN or fluid is required for months and when patients are metabolically unstable in the initial period, the 'SNAP' mnemonic is useful to guide management. SNAP stands for Sepsis, Nutrition, Anatomy, and planned Procedure. Sepsis control is the first priority, as this is the leading cause of death in type 2 IF. It includes prompt diagnosis, radiological or surgical drainage of abdominal collections, and tailored antibiotic therapy. Alternative sources of sepsis should also be considered, such as urinary or central venous catheters. Fungal infection in prolonged sepsis and antibiotic exposure should also be borne in mind.

Normalization of hydration status, electrolyte levels, and acid–base balance are also critical during this period, after which delivery of appropriate PN is necessary. This can be challenging—this is a highly catabolic state and not only does protein balance need addressing but also optimization of tissue function while avoiding overfeeding, which can impair liver function. Enteral nutrition is rarely sufficient because of the loss of function of the gastrointestinal tract. Therefore, parenteral nutrition often represents the main option, alone or in combination with enteral nutrition. It is also crucial to take steps to prevent, monitor, and manage re-feeding syndrome where necessary.

Unfortunately, the abdominal cavity is hostile for several weeks or months after an initial laparotomy, meaning any reconstructive surgery would risk fistulation. Therefore, early surgery should be limited to sepsis control. Once the initial period of stabilization has been negotiated, delineation of intestinal anatomy should be undertaken with appropriate imaging modalities, which subsequently informs planning of definitive surgery. It is usual to wait for between 3 and 6 months, after sepsis has been cleared and the inflammatory response has subsided, before definitive surgery is undertaken.

The most frequent outcomes of patients with type 2 IF are full intestinal rehabilitation (about 40%) and long-term enteral tube nutrition or transition to type 3 IF status requiring prolonged home PN (together about 50%). In-hospital mortality of type 2 IF has been reported to be as high as over 10%.

Type 3 intestinal failure

Type 3 IF describes chronic and usually irreversible intestinal failure for which long-term parenteral nutrition delivered at home is the first-line treatment. These patients are usually metabolically stable. Lifelong dependence is typical, unless patients are considered for alternatives such as intestinal transplant, glucagon-like peptide (GLP) analogue therapy or bowel lengthening procedures (see later).

For type 3 IF, weaning from home PN (HPN) within 2 years of starting may occur in 20% to 50% of patients, due to intestinal adaptation. Patients with benign disease have a high probability of long-term survival on HPN: 1-, 5-, 10-, and 20-year survival has been reported as 93%, 71%, 59%, and 28%, respectively. However, type 3 IF may also be associated with life-threatening complications and the condition itself may be highly disabling and impair the patient's quality of life. A summary of the clinical features and management of the three types of intestinal failure can be found in **Box 25.17**.

Short bowel syndrome

Short bowel syndrome, or SBS, is defined as the combination of clinical features associated with a remaining small bowel length of less than 200 cm. The normal small intestinal length, measuring from the duodenal jejunal flexure, varies from about 275 to 850 cm depending on sex and height. SBS is a major cause of chronic intestinal failure—indeed it is responsible for three-quarters of IF due to benign causes in adults. Depending on the anatomy of the remnant bowel, four categories of short bowel are identified: end-jejunostomy with no colon in continuity, end-ileostomy with no colon in continuity, jejunocolic anastomosis with no ileocaecal valve and a part of the colon in continuity, and jejunoileal anastomosis with both the ileocaecal valve and the entire colon in continuity (**Fig. 25.17**).

Adaptation

During the acute phase following bowel resection, the absorptive capacity of the residual small bowel is limited by increased transit times, venous congestion, loss of surface area, and bacterial overgrowth.

Intestinal adaptation is a natural and spontaneous compensatory mechanism following bowel resection whereby structural and functional changes in the remnant bowel attempt to maximize nutrient and fluid absorption. The ileum has greater adaptive potential than the jejunum. Adaptation usually takes place over 1 to 2 years post resection and occurs partly by increasing the absorptive area but also the slowing of gut transit. The process is promoted by the

Box 25.17 Summary of functional classification of intestinal failure (IF)

	Type 1 IF	Type 2 IF	Type 3 IF
Features	Acute, self-limiting Often post operative	Prolonged condition (>28 days) Potentially reversible Metabolically unstable Sepsis is common Often related to complex abdominal surgery and its complications e.g., fistula development	Chronic condition Usually irreversible Stable patient
Management	Short-term PN and i.v. fluid Consider and treat reversible causes e.g., ileus	PN/i.v. fluid for months SNAP approach	Home PN/i.v. fluid GLP2 analogues Bowel lengthening surgery Intestinal transplant
Prognosis	Excellent	Intestinal rehabilitation 40% Long-term enteral or parenteral nutrition 50% Inpatient mortality 10%	5-year survival 71% 10-year survival 59%

presence of nutrients in the gastrointestinal lumen, gut hormones, and pancreatico-biliary secretions.

Causes of short bowel syndrome

The causes of SBS can be thought of as being anatomical or functional (**Box 25.18**). Congenital causes or paediatric short bowel syndrome will not be discussed here.

Anatomical causes include extensive bowel resection, a proximal jejunostomy, a proximal fistula (e.g., enterocutaneous, entero-enteric, or entero-colic), or multiple bowel resections over time. Functional causes include extensive mucosal disease impairing normal function, due to inflammation (e.g., Crohn's disease), following abdomino-pelvic radiation or, less commonly, small intestinal bacterial overgrowth. Causes may coexist, for example a patient with Crohn's disease and previous bowel resections but also active current inflammation, or a patient with a previous irradiated bowel, malignant bowel obstruction, and a proximal jejunostomy.

Pathophysiology of short bowel syndrome

The total normal daily volumes of gastrointestinal secretions in a healthy adult are as follows:

- Food and drink intake 1500 mL
- Saliva 750 mL
- Gastric secretions 1250 mL
- Biliary secretions 1000 mL
- Pancreatic secretions 1000 mL
- Jejunal secretions 2500 mL

This totals approximately 8 L per day. However, normal daily stool volumes are around 150 mL. Therefore, in short bowel syndrome, where the distal intestine is lost or out of continuity, intestinal effluent volumes will be much higher.

All fluid entering the gut originates proximal to the ileum, and reabsorption occurs distal to the ileum. A healthy individual is considered to be a 'net absorber', meaning that increased fluid consumption will lead to greater absolute intestinal absorption. The proximal intestinal secretory response occurs up to a sodium level of 90 to 100 mmol/L. Given that the sodium content of almost all drinks is below this, drinking will lead to a further secretory effect. This is overcome in a healthy individual by the absorptive capacity of the distal small bowel and colon to ensure net absorption. However, in a patient with SBS, their absorptive capacity is lost, meaning that additional drinking drives secretion but without reabsorption, resulting in paradoxical further dehydration.

Management of short bowel syndrome

Given the above underlying mechanism of dehydration in SBS, the aims of management are to reduce gastrointestinal secretions, slow gut transit, correct fluid and electrolyte abnormalities, optimize nutritional status, and manage symptoms of thirst. One such management strategy is fluid restriction of hypotonic fluid with intake of hypertonic drinks. Hypertonic drinks should have a sodium content of greater than 90 mmol/L, such as St Mark's solution (**Box 25.19**). Other medical treatments include acid suppression with proton pump inhibitors such as omeprazole, to reduce gastric secretions. Loperamide and codeine can be used as antimotility agents.

With regards to nutrition, enteral nutrition is important for the integrity of the gut barrier and immune function. The need for parenteral support in the form of intravenous fluids and/or nutrition is determined largely by the underlying anatomy and bowel length (**Box 25.20**).

Complications of short bowel syndrome

The following complications are usually seen in patients with SBS and presence of colon in continuity.

D-lactic acidosis

Short bowel syndrome is a risk factor for D-lactic acidosis. This is a rare clinical condition and patients typically present with clinical features of confusion, ataxia, and slurred speech. Blood gas analysis classically demonstrates a raised anion gap metabolic acidosis with a normal L-lactate level. A D-lactate level of above 3 mmol/L is considered to be significant. In SBS, there is an absence or reduction of small bowel area, which is usually responsible for the absorption of simple carbohydrates. The unabsorbed carbohydrate then passes into the colon, where it is fermented by colonic microbiota to form D-lactate. Treatment of D-lactic acidosis is with intravenous fluids and sodium bicarbonate. Patients should be advised to reduce the amount of carbohydrates in their diet.

Renal stones

Oxalates are natural compounds found within the diet (spinach, almonds, beetroot, blackberries, blueberries, and raspberries) and are also produced within the body as a waste product. Oxalate is absorbed in the colon and precipitates within the renal tubule, leading to damage. In a healthy individual, calcium binds to oxalate within the intestinal lumen reducing the amount available for absorption and therefore preventing renal damage. However, in a

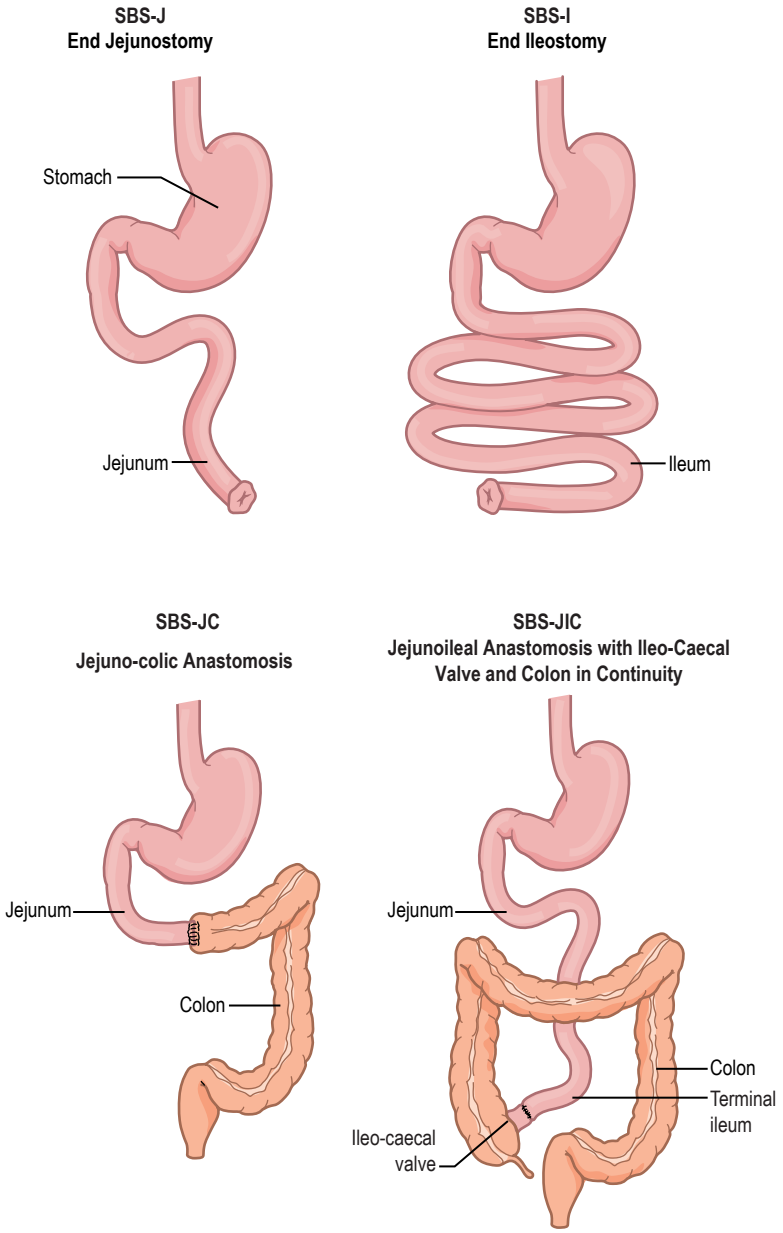


Fig. 25.17 **Short bowel syndrome.**

Box 25.18 Causes of short bowel syndrome

Anatomical causes	Functional causes
• Extensive bowel resection e.g., mesenteric ischaemia • Proximal jejunostomy • Proximal fistula • Multiple bowel resections	• Mucosal inflammation e.g., Crohn's disease • Radiation enteropathy • Small intestinal bacterial overgrowth

Box 25.19 St Mark's solution

Ingredients for 1 L batch of St Mark's solution
Glucose powder 20 g (6 teaspoons) Table salt (sodium chloride) 3.5 g (1 level 5 mL teaspoon) Sodium bicarbonate 2.5 g (heaped 2.5 mL teaspoon)

patient with SBS, unabsorbed long-chain fatty acids compete for calcium binding. The resultant effect is more free oxalate available for absorption and an increased risk of oxalate renal stone formation. This problem occurs in up to a quarter of patients with SBS and colon in continuity. General measures to reduce this risk include adequate hydration, reduced dietary fat, increased dietary calcium, and to adopt a low oxalate diet.

Chronic intestinal failure: alternatives to home parenteral nutrition

Chronic intestinal failure should be managed by a multidisciplinary team at a specialist IF centre. The goals of care should be to maximize enteral function, while attempting to wean PN requirements, avoid or minimize any PN-related complications, and improve quality of life.

Box 25.20 Nutritional and fluid support in short bowel syndrome

Small bowel length(from DJ flexure)	Jejunostomy	Jejunocolic anastomosis
0–50 cm	Parenteral nutrition + intravenous fluids	Parenteral nutrition
51–100 cm	Parenteral nutrition +/– intravenous fluids	Oral nutritional support
101–150 cm	Oral nutrition + oral glucose solution	No support required
151–200 cm	Oral glucose solution	No support required

Intestinal transplant

An alternative to home parenteral nutrition (HPN) is intestinal transplantation. This is a rare procedure and only occurs at specialist centres. There are three different types of intestinal transplantation—isolated small bowel transplants, liver-containing grafts, and multi-visceral transplants, which include the stomach and/or pancreas.

The indications for intestinal transplant are outlined in **Box 25.21**. The main indications are due to complications related to PN: progressive IFALD, recurrent life-threatening catheter-related sepsis, and loss of central venous access. Referral for consideration of transplant should be considered before life-threatening complications develop. Poor quality of life and acceptance of PN can be considered alongside other factors, and this could be an emerging indication over the next decades as post-transplant outcomes continue to improve. There are also non-intestinal failure-related indications for transplant. Multi-visceral transplant can be considered in patients who require surgical evisceration for treatment of invasive desmoid tumours or pseudomyxoma peritonei. Another emergency indication is mesenteric vascular catastrophe.

The survival rates for post-transplant patients are improving but remain inferior to those on HPN, with 1- and 5-year figures of 88.2% and 72.6%, respectively. Outcomes are better with non–liver-containing grafts versus liver-containing grafts. Complications of intestinal transplant include rejection, graft versus host disease, infection, and post-transplant lymphoproliferative disorder. Currently, intestinal transplantation remains a rescue procedure for life-threatening complications because of the superior survival figures with HPN. However, as outcomes improve, transplantation may start to be considered earlier for patients on HPN.

Bowel lengthening procedures

A variety of surgical techniques to improve bowel length have been described. These procedures are not performed routinely and require specialist centres. The Bianchi procedure involves a longitudinal incision of dilated bowel and an end-to-end anastomosis, resulting in increased intestinal length. The serial transverse enteroplasty, or STEP procedure, involves diagonal breaks in dilated bowel to increase length and slow transit time. Lastly, in the spiral intestinal lengthening and tailoring (SILT) procedure, a spiral incision is made in dilated bowel, which then allows resewing and intestinal extension.

GLP2 analogues

Endogenous glucagon-like peptide 2 (GLP2) is a gut hormone produced by the L cells in the small intestine in response to carbohydrate ingestion. The recombinant GLP2 analogue teduglutide, containing a substitution of the amino acid alanine with glycine at position 2 of the polypeptide chain, is used in clinical practice to promote intestinal absorption, having been originally found to increase mesenteric blood supply and mucosal growth, while reducing gastric emptying and gastric secretions in animal studies.

Further reading

Eliasson J, Hvistendahl MK, Freund N, Bolognani F, Meyer C, Jeppesen PB. Apraglutide, a novel glucagon-like peptide-2 analog, improves fluid absorption in patients with short bowel syndrome intestinal failure: findings from a placebo-controlled, randomized phase 2 trial. *JPEN J Parenter Enteral Nutr* 2022; 46(4):896–904.
Elsabbagh A, Hawksworth J, Khan K et al. Long-term survival in visceral transplant recipients in the new era: a single-center experience. *Am J Transplant* 2019; 19:2077–2091.
Jeppesen PB, Pertkiewicz M, Messing B et al. Teduglutide reduces need for parenteral support among patients with short bowel syndrome with intestinal failure. *Gastroenterology* 2012; 143:1473–1481.e3.
Kowlgi NG, Chhabra L. D-lactic acidosis: an underrecognized complication of short bowel syndrome. *Gastroenterol Res Pract*. 2015; 2015.
Nightingale JM. How to manage a high-output stoma. *Frontline Gastroenterol* 2022; 13(2):140–51.
Pironi L, Arends J, Baxter J et al. ESPEN endorsed recommendations. Definition and classification of intestinal failure in adults. *Clin Nutr* 2015; 34: 171–180.
Pironi L, Arends J, Bozzetti F et al. ESPEN guidelines on chronic intestinal failure in adults. *Clin Nutr* 2016; 35:247–307.
Rutter CS, Sharkey L. Intestinal transplantation. *Medicine*. 2023.
Yannam GR, Sudan DL, Grant W et al. Intestinal lengthening in adult patients with short bowel syndrome. *J Gastrointest Surg* 2010; 14:1931–1936.

FOOD INTOLERANCES

Many people ascribe various gastrointestinal symptoms to food, and many such sufferers are seen and started on exclusion diets. Adverse reactions to food do exist. These can be divided into food allergy and food intolerance. Food allergy incorporates immunological reactions—both IgE and non-IgE mediated. Common triggers of food allergies include cows' milk, eggs, peanuts, tree nuts, wheat, and shellfish. These should be managed by an immunologist. Food intolerance describes the occurrence of similar symptoms on a non-immune basis and will be the focus of this section.

Lactose intolerance

This is probably the most common form of food intolerance, with lactose malabsorption reported as affecting up to 60% to 70% of the world's population. Loss of intestinal lactase expression in the first two decades of life (lactase non-persistence) is the dominant phenotype worldwide, with highest rates found in East Asia and West Africa (70%–100%). Lactase persistence is the dominant phenotype in Western countries, with Northern Europeans least likely to be affected (rates of <5%). Patients develop symptoms of abdominal bloating and diarrhoea following ingestion of lactose, which is a sugar present in dairy products such as milk, butter, and cheese. It is defined as symptoms on lactose exposure in individuals with lactose malabsorption. Tests for lactose intolerance are usually not necessary. Most patients are advised to trial a lactose elimination diet, where lactose-containing foods are avoided and symptoms reassessed. Alternatively, a hydrogen breath test can be performed although this is not routinely done and there is variable reliability of these tests.

For quality of life, patients are often advised that they may continue to eat lactose up to their individual point of tolerance, rather than completely eliminating lactose from the diet. Harder cheeses contain less lactose than soft cheeses as the lactose sugar is spent

Box 25.21 Indications for intestinal transplant

Complications or failure of PN	Treatment of underlying condition	Poor patient acceptance
• Liver failure • Central venous thrombosis of two or more central veins • Frequent catheter-related bloodstream infections • Frequent episodes of severe dehydration	• Invasive desmoid tumours • Congenital mucosal disorders • Ultra-short bowel syndrome	• Frequent hospitalizations • Inability to function • Unwillingness to accept long-term home PN

during the fermentation of cheese. Lactose-free or alternative milks tend to be fortified with calcium. Most individuals with lactose malabsorption can tolerate up to 12 g of lactose, which is equivalent to a large glass of milk, without experiencing symptoms.

FODMAPs

FODMAPS (fermentable oligosaccharides, disaccharides, monosaccharides, and polyols) are fermentable carbohydrates which are found in many food groups and can trigger symptoms of irritable bowel syndrome (IBS); such as abdominal cramping, bloating, and loose stool. Unlike other carbohydrates, FODMAPs are not easily absorbed within the small intestine and therefore pass into the colon where they get fermented by bacteria. This leads to production of gas and fluid shifts within the large bowel resulting in the symptoms mentioned above. Therefore, it is sometimes recommended to follow a low FODMAP diet under the guidance of a specialist dietician for a limited period for symptom relief and identification of specific triggers.

Non-coeliac gluten sensitivity

Non-coeliac gluten sensitivity (NCGS) describes a syndrome of symptomatic response to gluten ingestion in patients with no serologic or histologic evidence of coeliac disease (see Chapter 22). Common symptoms include abdominal pain, bloating, and diarrhoea. Extraintestinal manifestations such as headache and 'brain fog' have also been reported. In patients presenting with symptoms that they attribute to gluten, it is important to exclude coeliac disease and IgE-mediated wheat allergy. However, there are no tests that distinguish NCGS from IBS. The clinical response to a gluten-free diet may be caused by a variety of mechanisms, including placebo effect and reduction in FODMAPs (particularly fructans), as well as by true gluten sensitivity. Patients should see a specialist dietician before embarking on a gluten-elimination diet and counselled on potential nutritional deficiencies such as calcium, vitamin D, iron, and zinc.

Further reading

Dale HF, Biesiekierski JR, Lied GA. Non-coeliac gluten sensitivity and the spectrum of gluten-related disorders: An updated overview. *Nutr Res Rev* 2019; 32(1):28–37.

Misselwitz B, Butter M, Verbeke K, Fox MR. Update on lactose malabsorption and intolerance: pathogenesis, diagnosis and clinical management. *Gut* 2019; 68(11):2080–91.

Whelan K, Staudacher H. Low FODMAP diet in irritable bowel syndrome: a review of recent clinical trials and meta-analyses. *Curr Opin Clin Nutr Metab Care* 2022; 25(5):341–7.

Questions for this chapter can be found on Elsevier eBooks+.

26 Kidney and urinary tract disease

M. Magdi Yaqoob and Neil Ashman

Clinical skills 790
History 790
Quantification of proteinuria 790
Examination 790
A diagnostic approach to new kidney disease 790
Background 792
Anatomy and physiology 792
Functional anatomy 792
Physiology 794
Glomerular filtration rate 795
Drugs, toxins, proteins, and the kidney 796
Endocrine function 796
Autocrine function 797
Investigation of kidney and urinary tract disease 797
Examination of the urine 798
Blood tests 799
Imaging techniques 799
Transcutaneous kidney biopsy 800
The glomerulus and glomerular disease 800
Nephrotic syndrome 802
Non-nephrotic range proteinuria 803
Causes of proteinuria and nephrotic syndrome 803
Glomerulonephritis 808
Mixed nephritic and nephrotic syndrome 813
Other glomerular disorders 816
Kidney involvement in other diseases 817
Hypertension and the kidney 819
Renovascular disease 820
Kidney calculi and nephrocalcinosis 821
Urinary tract obstruction 825
Drugs and the kidney 828
Urinary tract infection 829
Tubolointerstitial nephritis 833
Acute kidney injury 835
Renal replacement therapy in AKI 837
Chronic kidney disease 837
Complications of CKD 838
Management of CKD 843
Kidney replacement therapy 843
Kidney transplantation 846
Cystic kidney disease 849
Tumours of the kidney and genitourinary tract 851
Diseases of the prostate gland 852
The urinary tract in the elderly 852

CORE SKILLS AND KNOWLEDGE

Kidney disease is common: acute kidney injury (AKI) affects around 20% of hospitalized patients, and 10% to 15% of the general population meet diagnostic criteria for chronic kidney disease (CKD). The kidney is involved in systemic conditions including diabetes, hypertension, autoimmune disorders, haematological malignancies, and infections including human immunodeficiency virus (HIV) and viral hepatitis.

Kidney physicians work in lifetime partnership with patients, their families, and a multidisciplinary team including nurses, therapists, counsellors, nutritionists, technicians, and pharmacists. Most patients with kidney disease are managed in outpatient settings, although patients with AKI may require inpatient investigation and management. Specialist kidney expertise including transplantation and dialysis are offered in tertiary units. Kidney physicians manage immunosuppression during and after kidney transplantation, and supervise peritoneal and haemodialysis. Nephrologists need to be general physicians, comfortable with complexity and co-morbidity.

Key skills in kidney medicine include:

- Diagnosing AKI and managing its complications
- Understanding CKD causes, complications, and progression
- The principles of dialysis and kidney transplantation.

Opportunities for learning kidney medicine include attending outpatient clinics, assessing patients with AKI presenting to emergency or acute admissions departments, and spending time in specialist centres with patients receiving kidney replacement therapies.

CLINICAL SKILLS

History

In its early stages, kidney disease may be asymptomatic and diagnosed incidentally on blood or urine tests. History-taking in kidney medicine (Box 26.1) is tailored to the clinical context of the individual patient.

Quantification of proteinuria

Quantifying the concentration of urinary protein is necessary in kidney disease. Testing is done initially using a urine dipstick and then quantified using one of the measures shown in Box 26.2. There is a rough equivalence between methods if proteinuria is non-selective (i.e., if protein is excreted in the urine in the same proportions as in the blood). In healthy people there should be virtually no urinary albumin (uACR <2.5 mg/mmol for males, <3.5 mg/mmol for females); even very low concentrations above this threshold (not detectable by dipstick testing) confer increased cardiovascular risk. High concentrations of proteinuria generally indicate glomerular disease, and uPCR > 350 mg/mmol indicates nephrotic syndrome if hypoalbuminuria and oedema are also present.

Examination

Except for palpable polycystic kidneys, there may be few physical signs specific to kidney disease. Signs of systemic disease with kidney involvement may be evident, for example, systemic lupus erythematosus, systemic sclerosis, diabetic microvascular disease, cardiovascular disease. The figure on the next page shows potential signs of kidney disease in the **grey boxes**. The **yellow boxes** indicate signs related to renal replacement in patients with kidney failure or due to previous treatment including physical signs of peritoneal dialysis, haemodialysis, and transplantation.

A diagnostic approach to new kidney disease

A diagnostic approach to kidney dysfunction is presented in Box 26.3. This is relevant in AKI and CKD with superimposed AKI or a rapid deterioration in kidney function.

Box 26.2 Quantifying the concentration of urinary protein

24-hour urine protein collection (g/24 h)	Urine protein:creatinine ratio (uPCR, mg/mmol)	Urinary albumin:creatinine ratio (uACR, mg/mmol)
1	100	70
2	200	150
3	300	200
4+	400+	300+

Box 26.1 History-taking in kidney and urinary tract disease

Assessing the likely duration of kidney dysfunction

- Duration of symptoms
- Previous occasions on which urinalysis or measurement of serum creatinine might have been performed, e.g., pre-employment health check, new patient check, pregnancy.
- Urine volume:
  - Oliguria is a symptom of acute kidney injury
  - Slowly progressive kidney disease can cause an increase in urine volume when a reduction in tubular function leads to a salt- and water-wasting. Urine volume increases as tubular concentrating ability reduces meaning that polyuria/nocturia may suggest the length of time for which CKD has been present.

Symptoms of uraemia

Symptoms are common when the serum urea concentration exceeds 40 mmol/L, though many patients develop uraemic symptoms at lower concentrations. Symptoms are generally non-specific but may include:

- malaise, loss of energy
- loss of appetite and weight loss
- insomnia
- nocturia/polyuria
- itch
- nausea/vomiting
- 'restless legs' syndrome
- bone pain
- peripheral oedema
- symptoms due to anaemia
- amenorrhoea
- erectile dysfunction
- CNS symptoms in advanced uraemia: mental slowing, clouding of consciousness, myoclonic twitching, seizures

History that may reveal the aetiology of kidney disease:

- ***Medical history*** focusing on diseases that commonly affect the kidney: diabetes, hypertension, systemic inflammatory diseases, blood-borne viruses, myeloma
- ***Recent ill health***—e.g., sepsis or dehydration causing kidney hypoperfusion and pre-renal AKI
- Symptoms of ***urinary tract obstruction*** (see Chapter 47)
- Symptoms of ***systemic inflammatory or malignant conditions:*** fever, malaise, joint pain, rash, eye inflammation, hair fall, nasal discharge, haemoptysis
- ***Medications*** use now and in the past. Remember known nephrotoxic agents (such as non-steroidal anti-inflammatory drugs, which the patient may have obtained over-the-counter and therefore may forget to disclose), and traditional and herbal remedies, which are implicated in kidney disease
- ***Family history*** of kidney disease or kidney replacement therapy

In advanced CKD

- Do they understand their diagnosis?
- What form of renal replacement offers the best quality/quantity of life?
- Do they wish to consider dialysis or prefer a supportive/palliative approach?

In patients receiving dialysis

- How are they managing their treatment?
- Are there any symptoms relating to anaemia or bone disease?
- Is transplantation medically achievable and feasible?

Kidney transplant recipients

- Any there any symptoms to suggest opportunistic infections?
- How are they managing with their immunosuppressant medications? Are strategies needed to aid adherence?
- Have they noticed any skin lesions/changes?
- Can their lifestyle be optimized to increase the survival of their transplant?

SIGNS OF KIDNEY DISEASE

General features
Short stature (childhood kidney disease)
Pallor (anaemia)
Photosensitive pigmentation
Yellowish complexion

Observation chart
Hypoxia and hypertension suggest fluid overload

Fundoscopy
Signs of diabetic or hypertensive eye disease

Mouth
Dry mouth suggests intravascular fluid depletion

Neck
Jugular venous pressure reflects fluid status

Chest
Basal crackles suggest fluid overload

Heart
Pericardial friction rub (uraemia)

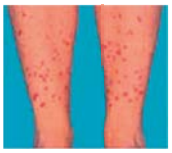
Skin
Scratch marks due to uraemic pruritus

Abdomen
Grossly enlarged kidneys may be palpated in polycystic kidney disease

Peripheral pulses
Reduced in atherosclerotic disease

Hands
Brown discoloration of the nails

Legs
Signs of fluid overload
Purpuric rash suggests an underlying vasculitis with potential kidney involvement



EVIDENCE OF RENAL REPLACEMENT THERAPY

General features
Tremor may suggest toxicity from calcineurin inhibitor immunosuppression

Skin
Basal cell carcinomas and other malignant or pre-malignant change with long-term immunosuppressant use

Mouth
Gingival hypertrophy with some immunosuppressant medications

Neck
Scars from previous venous dialysis catheters

Chest
Tunnelled dialysis catheter

Forearms
Evidence of arteriovenous fistulae – check for a thrill to confirm patency

Abdomen
Tenckhoff peritoneal dialysis catheter
Kidney transplant: mass with overlying scar in the iliac fossa

Box 26.3 Causes of kidney dysfunction

Pre-renal:
- Intercurrent illness including systemic infection or sepsis
- Vomiting, diarrhoea, and dehydration
- Concurrent use of antihypertensives or diuretics in intercurrent illness, e.g., angiotensin-converting enzyme (ACE) inhibitors, angiotensin receptor blockers

Intrarenal:
- Disease with kidney involvement, e.g., cast nephropathy in myeloma
- Nephrotoxic medication
- Symptoms suggestive of systemic inflammatory disease
- Proteinuria/haematuria

Post-renal:
- Symptoms of urinary tract obstruction
- Anuria
- A condition associated with urinary tract obstruction: prostate disease, bladder or pelvic cancer
- Bladder distension or hydronephrosis on ultrasound

BACKGROUND

Nephrology is a major medical specialty; over 10% of the world's population are thought to be living with kidney disease.

The kidneys maintain our internal environment and influence other systems when they fail. Glomerular pathology, an array of local and systemic diseases causing kidney injury, and the complex consequences of fluid and electrolyte shifts make nephrology broad and challenging. Chronic kidney disease is often a life-long journey, encompassing different treatment modalities and necessitating an appreciation of and support for the physical, psychological social adaptations required when living with CKD.

ANATOMY AND PHYSIOLOGY

Functional anatomy

The ***kidneys*** are paired organs, which in adults are 11 to 14 cm in length, 5 to 6 cm in width and 3 to 4 cm in depth. They lie in the retroperitoneum, on either side of the vertebral column at the level of T12–L3 (the right kidney lies lower than the left, pushed down by the liver), enclosed in a fibrous capsule. Kidneys have an outer cortex and an inner medulla (Fig. 26.1). There are about 1 million ***nephrons*** in each kidney. Each nephron contains a glomerulus, proximal tubule, loop of Henle, distal tubule, and collecting duct. All glomeruli lie in the cortex. The tubules dip in and out of the medulla, where the collecting ducts merge to form the ducts of Bellini, emptying at a papilla at the apex of the renal pyramid into a calyx. Urine flows through merging calyces into the renal pelvis, ureters, and bladder.

The renal arteries branch off the abdominal aorta, dividing until arterial blood reaches the glomerulus. About 25% of people have dual or multiple renal arteries on one or both sides. ***Afferent glomerular arterioles*** arise from interlobular branch arteries to supply the glomerular capillary tuft, which drains into an ***efferent glomerular arteriole***. Efferent arterioles from (outer) cortical glomeruli drain into a peritubular capillary network within the cortex and then into increasingly large proximal branches of the renal vein. In contrast, blood from the (inner) juxtamedullary glomeruli passes via vasa recta in the medulla and returns via the cortex to renal veins that drain into the inferior vena cava.

The capsule and ureters are innervated via T10–12 and L1 nerve roots, and pain related to the kidney is felt over the corresponding dermatomes.

Nephron

The glomerular tuft is a ball of capillaries enclosed by Bowman's capsule, a chamber lined with specialized ***parietal epithelial cells*** that marks the origin of the tubule. The tuft, held together and regulated by ***mesangial cells*** serves as a filtration barrier, allowing filtrate from plasma to move into the urinary (Bowman's) space (Fig. 26.2). The rate of glomerular filtration is influenced by changes in the contractile tone of either afferent or efferent arterioles; for example, efferent vasoconstriction increases transglomerular capillary pressure and filtration. Glomerular capillaries are lined with ***endothelial cells***, fenestrated with 60 to 80 nm pores and covered with charged glycocalyx. The glomerular basement membrane (GBM), about 300 nm thick and made of type IV collagen, laminin, and heparin sulphate, separates the endothelium from podocytes. Podocytes anchor to the GBM by an extensive trabecular network of foot processes, and hang into Bowman's space. The interdigitating foot processes of podocytes form the 40 nm filtration slit, a narrow potential space traversed by a protein 'zipper' that may prevent the passage of larger molecules (such as albumin) into the urinary space, and regulates the architecture and function of podocytes.

Mesangial cells (thought to be related to macrophages) sit within the tuft, able to contract and relax to control blood flow and the filtration surface area along the glomerular capillaries in response to a host of mediators. They also secrete the ***mesangial matrix***, which provides the scaffold for glomerular capillaries.

The ***kidney tubules*** are lined by epithelial cells, which alter the composition of filtrate that eventually forms urine. Proximal tubular cells have a luminal brush border to increase (by thirtyfold) the surface area exposed to filtrate, rich in transporters and channels. The loop of Henle is lined with squamous cells, which are more permeable to water than solute, and cuboidal epithelium, which is more permeable to solute. The distal tubule regulates electrolytes and pH through cuboidal epithelium, and the cortical portion of the collecting ducts contains two cell types with different functions: principal cells (sodium, potassium, and water homeostasis) and intercalated cells (acid–base homeostasis). Finally, interstitial fibroblast-like cells in the cortex produce erythropoietin in response to hypoxia.

Juxtaglomerular apparatus

The juxtaglomerular apparatus sits between the afferent arteriole and the distal convoluted tubule of the same nephron (Fig. 26.3). It is critical to regulation of blood flow and glomerular filtration rate via tubuloglomerular feedback. It includes the macula densa: modified epithelial cells in the distal convoluted tubule. Columnar epithelial cells in the macula densa sense the concentration of tubular fluid sodium (higher filtrate flow means more delivered sodium), triggering adenosine-mediated vasoconstriction of the afferent arteriole to reduce glomerular filtration. Juxtaglomerular cells secrete renin, able to induce aldosterone release, allowing retention of salt to maintain fluid balance in response to decreased glomerular filtration when needed.

Further reading

Zehra T, Cupples WA, Branko B. Tubuloglomerular feedback synchronization in nephrovascular networks. *JASN* 2021; 32(6):1293–1304.

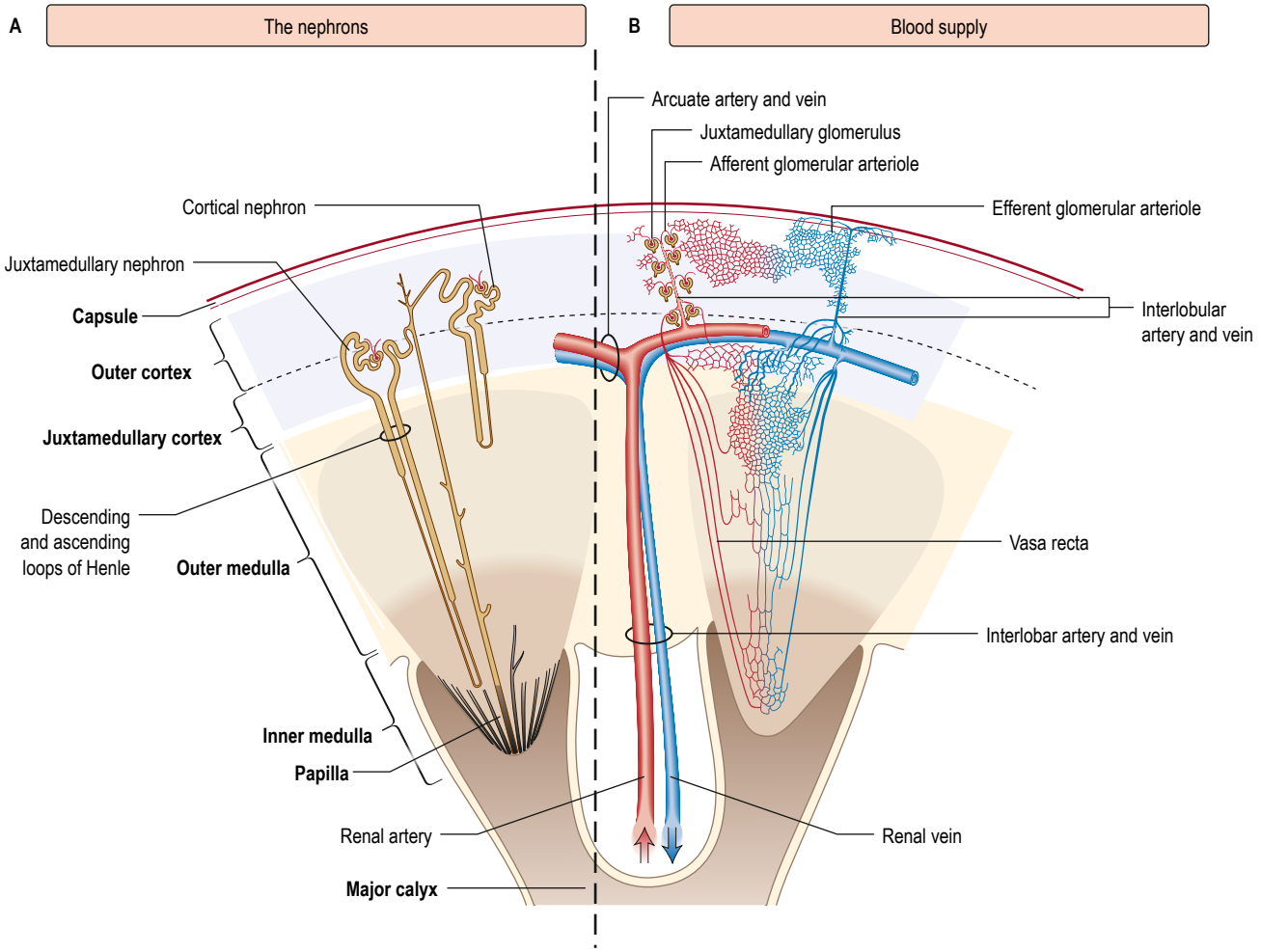


Fig. 26.1 Functional anatomy of the kidney. **(A)** The nephrons. **(B)** Arterial and venous supply. *(After Standring S (ed.).* Gray's Anatomy, *41st edn. London: Elsevier; 2016.)*

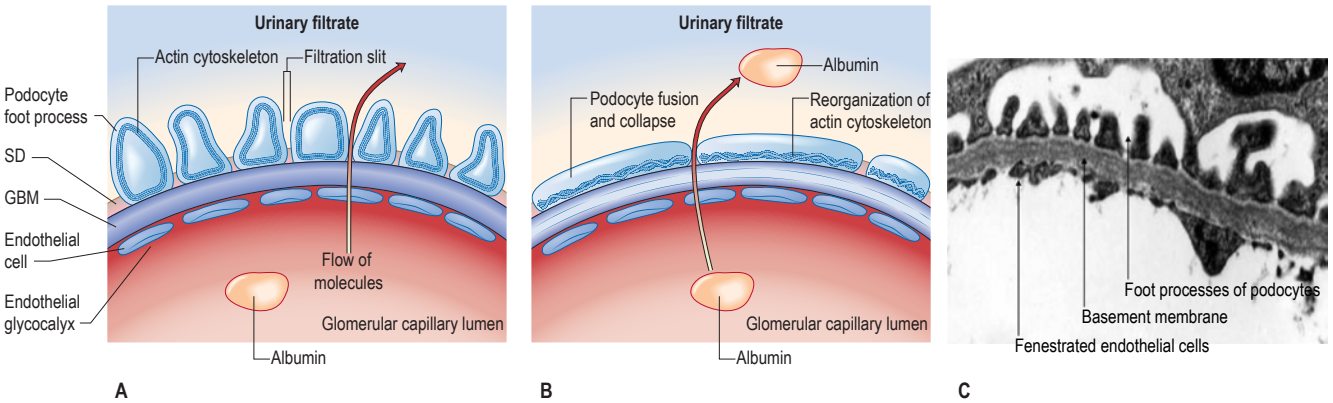


Fig. 26.2 The glomerular filtration barrier. **(A)** Blood enters the glomerular capillaries and is filtered across the endothelium and the glomerular basement membrane (GBM), through the filtration slits between podocyte foot processes to produce the primary urine filtrate. In healthy glomeruli, this barrier restricts the passage of macromolecules. The proteins that form the slit diaphragm (SD) are essential for the normal functioning of the filtration barrier. **(B)** Genetic or acquired loss of slit diaphragm proteins leads to foot process effacement, breakdown of the barrier, and leakage of albumin into the urine (central to the pathophysiology of glomerular diseases and nephrotic syndrome). **(C)** Electron micrograph of the normal filtration barrier. *(A, B, Adapted from Quaggin S. Sizing up sialic acid in glomerular disease.* J Clin Invest *2007; 117:1480–1483.)*

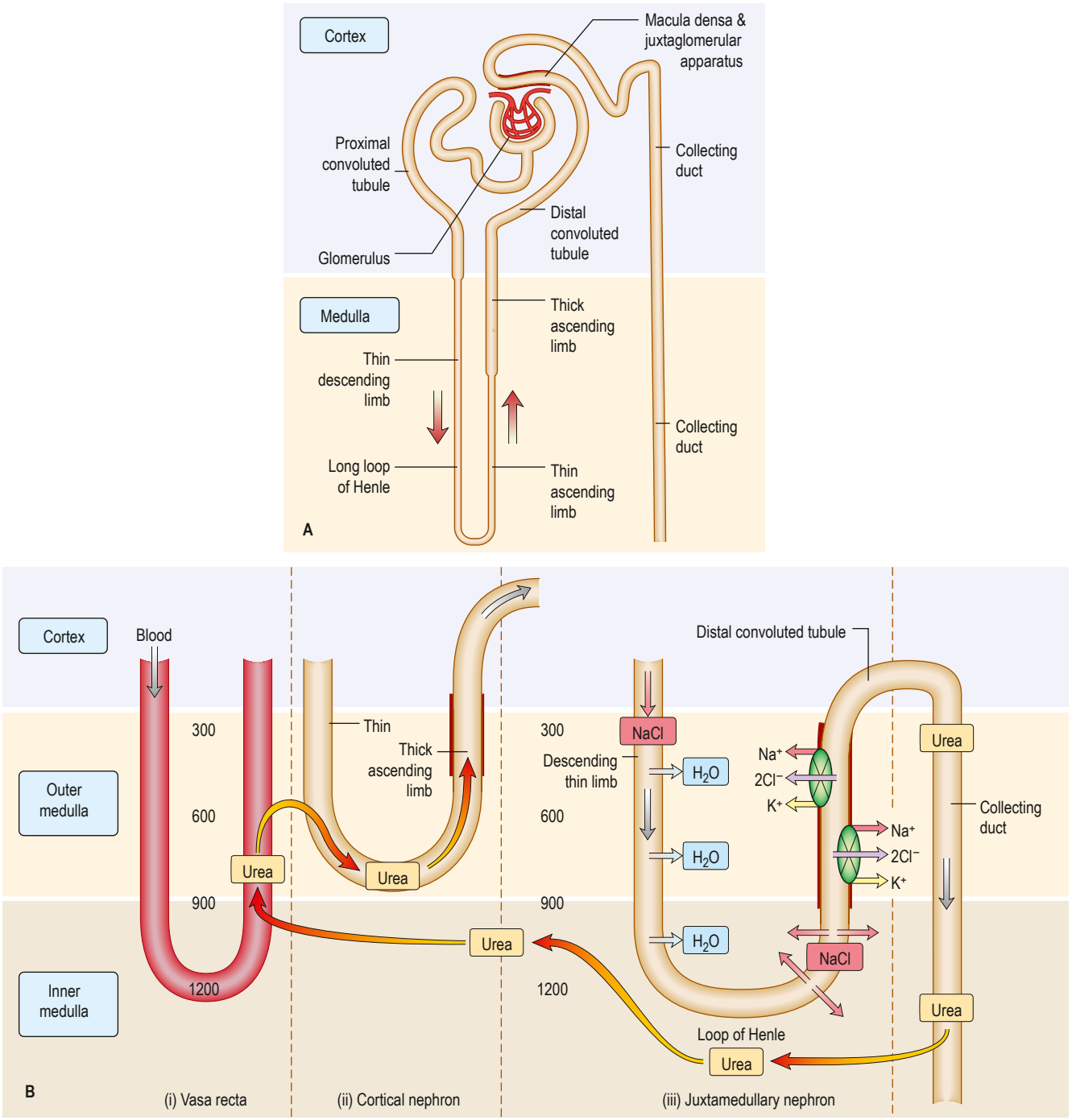


Fig. 26.3 Functional anatomy of the nephron. **(A)** Principal parts of the nephron. The distal tubule near its own glomerulus is termed the juxtaglomerular apparatus. It contains the macula densa. **(B)** The countercurrent system (i) Vasa recta: vessels descend from the cortex into the medulla and then turn back towards the cortex. (ii) The descending limb of the cortical nephron extends into the outer medulla. (iii) Juxtamedullary nephron: the descending limb dips into the hypertonic inner medulla. Numbers indicate approximate osmolalities.

Physiology

A ***hydrostatic pressure gradient*** of approximately 10 mmHg (capillary pressure of 45 mmHg minus 10 mmHg of pressure within Bowman's space and 25 mmHg of plasma oncotic pressure) provides the driving force for ultrafiltration of virtually protein-free and fat-free fluid across the glomerular filter into Bowman's space and the renal tubule (Fig. 26.4).

The ***glomerular filtration rate*** (GFR) varies with age and sex but is approximately 120 to 130 mL/min/1.73 m^2 surface area in adults. This means that ultrafiltration of 170 to 180 L of water and unbound small-molecular-weight constituents of blood occurs each day. If these volumes of ultrafiltrate were excreted unchanged as urine, it would be necessary to drink huge amounts of water and salts to maintain balance. Instead, reabsorption occurs.

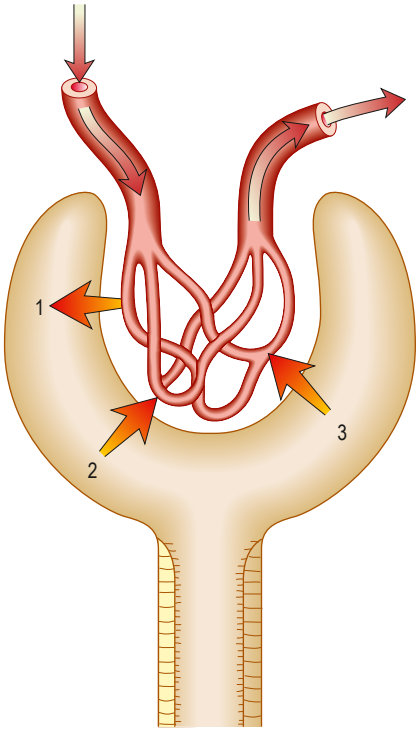


Fig. 26.4 Pressures controlling glomerular filtration. (1) Capillary hydrostatic pressure (45 mmHg). (2) Hydrostatic pressure in Bowman's space (10 mmHg). (3) Plasma protein oncotic pressure (25 mmHg). *Arrows* (1, 2, 3) indicate the direction of the pressure gradient.

Absorption of solutes

Essential electrolytes and other blood constituents, such as glucose and amino acids, are absorbed from filtrate along the course of the nephron (see Fig. 26.3).

Sodium filters freely, and 60% to 80% of filtered sodium (and water) is reabsorbed in the late proximal tubule, where the apical membrane Na^+/H^+ exchanger (NHE3) trades hydrogen ions into the lumen for absorbed sodium, with anionic chloride (Cl^-) accompanying sodium to maintain electric neutrality. Secreted H^+ allows bicarbonate (HCO_3^-), formed from water and CO_2 by cellular and luminal carbonic anhydrase, to exit the basolateral membrane with absorbed Na^+. Recently, sodium–glucose co-transporters 1 and 2 (SGLT-1 and -2) have generated a lot of interest because of the use of SGLT-2 inhibitors in the treatment of diabetes. SGLT-2 is a predominant proximal tubular transporter that absorbs glucose and sodium. SGLT-2 inhibitors promote glycosuria and natriuresis.

Around 25% of Na^+ is absorbed in the thick ascending limb of the loop of Henle, as NaCl by the $Na^+/K^+/2Cl^-$ co-transporter (NKCC). The remaining 5% is absorbed by the thiazide-sensitive NaCl co-transporter and the epithelial sodium channel (ENaC) in the principal cell of the collecting duct. Of the recommended daily intake of up to 6 g per day of ***salt*** (equivalent to around 2 g Na^+), only around 5% to 10% is excreted in urine, stool, and sweat.

Potassium is freely filtered at the glomerulus, largely reabsorbed in the proximal tubule, and secreted in the distal tubule and collecting ducts. A clinical consequence of this is that elimination of unwanted potassium is less dependent on GFR than elimination of urea or creatinine.

Virtually all bicarbonate, glucose, and amino acids are absorbed in the proximal tubule though there is a maximal absorptive capacity for many compounds. For example, if blood glucose concentrations are above the upper reference limit, the amount filtered (filtered load = GFR × plasma concentration) may exceed maximal absorptive capacity and glucose will 'spill over' as glycosuria. Inherited or acquired defects in tubular function may lead to incomplete absorption of a *normal* filtered load, and substances such as glucose, amino acids, phosphate, sodium, potassium, and calcium appear in the urine, either singly or in combination. Examples include cystinuria (a defect of a specific amino acid transporter) or Fanconi's syndrome (a proximal tubule disorder leading to phosphaturia, aminoaciduria, and glycosuria). Other compounds filtered and reabsorbed or secreted to a variable extent include urate, many organic acids, and many drugs or their metabolic breakdown products. The more tubular secretion of a compound that occurs, the less dependent elimination is on GFR.

Absorption of water

Urine is concentrated by the countercurrent multiplier mechanism in the loop of Henle, the medullary interstitium, medullary blood vessels (vasa recta), and the collecting ducts (Fig. 26.3). Sodium and urea in water flow as filtrate into the descending loop, which is permeable to water and impermeable to sodium. In the thick ascending limb, which is impermeable to water, active absorption of sodium by the $Na^+/K^+/2Cl^-$ co-transporter into the interstitium increases tissue osmolarity. Because of the hairpin nature of the loop, water from the permeable descending limb enters the interstitium by osmosis, where the vasa recta return fluid to the circulation. Constant active absorption of sodium in the ascending loop multiplies this process over time, and the solute concentration at the tip of the loop is fourfold that of extracellular fluid (Fig. 26.3B). By this mechanism, salt and water are returned to the circulation. Filtrate entering the collecting duct is increasingly concentrated by the action of arginine vasopressin (AVP, or vasopressin), leading to the release of intracellular aquaporin water channels, which insert themselves across the apical membrane allowing water to be absorbed along an osmotic gradient. When the effect of AVP wears off, water channels return to the cell cytoplasm. The final urine volume is 1 to 2 L daily.

Glomerular filtration rate

In health, the GFR remains constant owing to intrarenal regulatory mechanisms. In disease (e.g., a reduction in intrarenal blood flow, damage to or loss of glomeruli, obstruction to the free flow of ultrafiltrate along the tubule), the GFR will decrease and the ability to eliminate waste and regulate the volume and composition of body fluid declines. This is measured as a rise in the plasma urea or creatinine, or as a reduction in measured GFR.

The concentration of urea or creatinine in serum represents the dynamic equilibrium between production and elimination. In healthy subjects there is an enormous reserve of excretory function, and serum urea and creatinine do not rise above the normal range until there is a reduction of 50% to 60% in the GFR (Fig. 26.5). Thereafter, the concentration of urea depends on both the GFR and urea production rate (Box 26.4). The latter is heavily influenced by protein intake and tissue catabolism. The concentration of creatinine is much less dependent on diet but can vary with age, sex, and muscle mass. Once it is elevated, serum creatinine is a better guide to GFR than urea and is widely used to monitor further deterioration in the GFR.

Measuring or estimating the GFR

If a substance is filtered and unmodified by the tubules as it passes along the nephron, the concentration of that substance per mL will

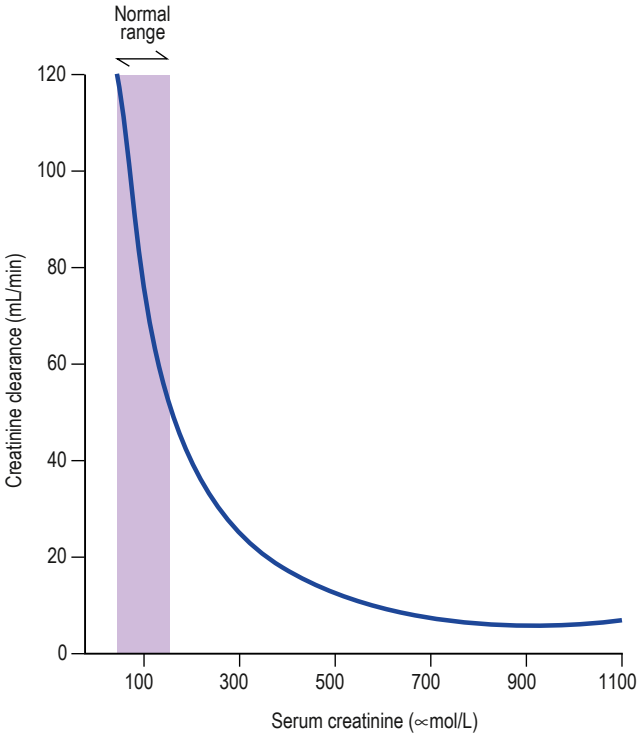


Fig. 26.5 Creatinine clearance versus serum creatinine. Serum creatinine does not rise above the upper reference limit until there is a 50%–60% reduction in glomerular filtration (creatinine clearance).

Box 26.4 Factors influencing serum urea concentrations

Increased serum urea	Increased production: • High-protein diet • Increased catabolism • Surgery • Infection • Trauma • Corticosteroid therapy • Upper gastrointestinal bleeding • Cancer	Reduced elimination: • Glomerular disease • Reduced renal blood flow • Hypotension • Dehydration • Urinary obstruction • Tubulointerstitial nephritis
Decreased serum urea	Decreased production: • Low-protein diet • Reduced catabolism, e.g., old age • Liver failure	Increased elimination • Pregnancy

be the same in blood and urine. Accurate assessment of GFR, particularly when urea and creatinine may be in the normal range or near normal, can be done via cystatin C, which is a freely filtered low-molecular-weight protein and more accurate marker of kidney function than creatinine, though measurement has not widely entered clinical practice.

Daily production of creatinine (principally from muscle cells) is fairly constant and little affected by protein intake. Serum creatinine and excreted (urinary) creatinine vary little throughout the day, allowing measurement of ***creatinine clearance***. Urine is collected over 24 hours for measurement of total urinary creatinine, paired with a single plasma creatinine measured during the 24-hour period. Creatinine clearance is calculated as $U \times VP$, where U = urine concentration of creatinine, V = rate of urine flow in mL/min, and P = plasma concentration of creatinine. Normal ranges are 90 to 140 mL/min in males and 80 to 125 mL/min in females.

Alternatives for quantifying true GFR include measuring iohexol or inulin clearance, or radio-isotope (^{51}Cr-EDTA) GFR. Neither is practical for clinical practice.

Estimated GFR (eGFR)

Measurement of GFR is cumbersome, time-consuming, and may be inaccurate if 24-hour urine collections are incomplete. Several formulae estimate GFR from serum creatinine and patient characteristics, often derived from large trials. Variables include age, weight, and sex. For patients with chronic kidney disease, the convenience and ease of eGFR monitoring led to its widespread adoption. Commonly used equations have their shortcomings and are less accurate the closer GFR is to normal. CKD-EPI is most reliable for predicting eGFR above 60 mL/min/1.73 m^2. Race adjustment of equations has been abandoned due to overestimation compared to measured GFR and the systemic disadvantage of patients of black ethnicity. A new equation using cystatin C (eGFRcys) does not include adjustment for race. Race-free eGFR equations represent an important step forwards, and validation across the full range of kidney disease, body mass index (BMI), and ethnicities continues.

Further reading

Inker LA, Eneanya ND, Coresh J et al. New creatinine- and cystatin c-based equations to estimate gfr without race. *N Engl J Med* 2021; 385(19):1737–1749.
Tsai JW, Cerdeña JP, Goedel WC et al. Evaluating the impact and rationale of race-specific estimations of kidney function: Estimations from U.S. NHANES, 2015-2018. *EClinicalMedicine* 2021; 42:101197.

Drugs, toxins, proteins, and the kidney

A substantial fraction of prescription drugs is handled and eliminated by the kidney. Many medications (e.g., penicillins, cephalosporins, diuretics, NSAIDs, antivirals) circulate in the plasma as small organic anions, and are eliminated in the proximal tubule by a specific organic anion transporter (OAT) system. The kidney is also a major site for the catabolism (and therefore elimination) of many small-molecular-weight proteins and polypeptides, including hormones such as insulin, parathyroid hormone (PTH), and calcitonin, by endocytosis carried out by the megalin–cubilin complex in the brush border of proximal tubular cells. In CKD, clearance is reduced and half-life is prolonged. This accounts, for example, for the reduced insulin requirements of patients with diabetes as kidney function declines.

Endocrine function

Renin–angiotensin system

Pro-renin, synthesized by specialized arteriolar smooth muscle cells in the JGA, is cleaved into the active proteolytic enzyme, renin. Active renin is stored in the JGA and released in response to a falling intravascular volume, pressure, or increased fluid losses via the kidney. These are 'sensed' as:

- pressure changes in the afferent arteriole or changes in sympathetic tone
- chloride and osmotic concentration in the distal tubule via the macula densa
- local prostaglandin and nitric oxide release.

In the blood, renin converts angiotensinogen to inactive angiotensin I, which is further cleaved by angiotensin-converting enzyme (ACE) in lung and vascular endothelium into active ***angiotensin II (AII)***, which acts via two receptor types, AT_1 and AT_2. Binding of AII to AT_1 in the heart, blood vessels, kidney, adrenal cortex, lung, and brain leads to:

- rapid, powerful vasoconstriction
- stimulation of the adrenal zona glomerulosa to increase aldosterone production (over hours or days) leading to sodium (and water) absorption in the collecting duct
- vasoconstriction of efferent > afferent renal arterioles increasing glomerular capillary pressure to maintain GFR.

The renin–angiotensin system can be blocked at several points with renin inhibitors, ACE inhibitors, and angiotensin II receptor antagonists (AII-RAs). These are useful agents in the treatment of chronic kidney disease, hypertension, and heart failure. There are differences in action: AII-RAs are specific for AT_1 receptors, but ACE inhibitors also block kinin breakdown leading to an accumulation of kinins, causing a cough in some patients. In diabetes mellitus, absorption of excessively filtered glucose and sodium via SGLT-1 and SGLT-2 in the proximal tubules reduces sodium delivery at the JGA, which leads to activation of the renin–angiotensin axis and causes increased GFR. This is the hyperfiltration of diabetes, with an emerging therapeutic role for SGLT-2 inhibitors.

Erythropoietin

Erythropoietin (EPO) is the major stimulus for erythropoiesis, the synthesis of red cells. It is a glycoprotein produced principally by fibroblast-like cells in the renal interstitium. In ***hypoxia***, the α and the β subunits of hypoxia-inducible factor 2 (HIF-2) are expressed, forming a heterodimer causing erythropoietin gene transcription. Once formed, erythropoietin binds to its receptors on erythroid precursor cells, to maintain normal red cell synthesis. Under ***normal oxygen*** conditions, only the HIF-2-β subunit is constitutively expressed. The α subunit undergoes proline hydroxylation in the presence of iron and oxygen. The hydroxylated HIF-2-α subunit binds to von Hippel–Lindau protein, with subsequent degradation via proteasomes, so no erythropoietin is transcribed. This and other hydroxylation steps have an absolute requirement for oxygen and form the basis of oxygen sensing.

Loss of kidney substance with decreased erythropoietin production results in a normochromic, normocytic anaemia. Conversely, erythropoietin secretion may be increased in polycystic kidney disease, benign kidney cysts, or renal cell carcinoma leading to polycythaemia.

Recombinant human erythropoietin is used to treat anaemia in chronic kidney disease. In addition, there is emerging use of oral HIF stabilizers to promote endogenous erythropoietin production.

Vitamin D metabolism

Naturally occurring vitamin D (cholecalciferol) requires hydroxylation in the liver at position 25 and again by 1α-hydroxylase in the kidney to produce metabolically active 1,25-dihydroxycholecalciferol (1,25-$(OH)_2D_3$). Activity of 1α-hydroxylase is increased by:

- high plasma concentrations of PTH
- low concentrations of serum phosphate
- low concentrations of 1,25-$(OH)_2D_3$.

1α-hydroxylase activity is decreased by fibroblast growth factor 23 (FGF23), which is produced by osteocytes in response to high phosphate in CKD. (FGF23 increases phosphate excretion in the urine by downregulation of the sodium–phosphate co-transporter in the proximal tubule and decreases intestinal phosphate absorption). FGF23 is increasingly recognized as a possible mediator of bone mineral disease in CKD.

Both 1,25-dihydroxycholecalciferol and 25-hydroxycholecalciferol are degraded in part by being hydroxylated at position 24 by 24-hydroxylase. The activity of this enzyme is reduced by PTH and increased by 1,25-$(OH)_2D_3$ (which therefore promotes its own inactivation).

Reduced 1α-hydroxylase activity in CKD leads to relative deficiency of 1,25-$(OH)_2D_3$. As a result, gastrointestinal calcium and, to a lesser extent, phosphate absorption are reduced, and bone mineralization is impaired. Reduced gut calcium absorption leads to hypocalcaemia sensed by a specific calcium-sensing receptor (CaSR) on parathyroid glands, which in turn induces release of PTH. Reduced receptor binding alters the set-point for release of PTH when serum calcium falls. This combination contributes to the (common) hyperparathyroidism seen in patients with CKD.

Autocrine function

Prostaglandins

Prostaglandins are unsaturated, oxygenated fatty acids, derived from the enzymatic metabolism of arachidonic acid, mainly by constitutively expressed cyclo-oxygenase-1 (COX-1) or inducible COX-2. COX-1 is highly expressed in the collecting duct, while COX-2 expression is restricted to the macula densa. Both COX isoforms convert arachidonic acid into prostaglandin E_2 (PGE_2) in the collecting duct (leading to a natriuresis/diuresis); prostacyclin (PGI_2), and thromboxane A_2, a vasoconstrictor, are mainly synthesized in the glomerulus. These all maintain kidney blood flow and GFR in the face of vasoconstrictors such as angiotensin II, catecholamines, and α-adrenergic stimulation. Inhibition of prostaglandin synthesis by NSAIDs results in a fall in GFR, sometimes sufficient to cause AKI, particularly in the elderly, and in volume depletion, particularly when ACE inhibitors or AII-RAs are also used.

Nitric oxide

Nitric oxide, a molecular gas, is formed by the action of three isoforms of nitric oxide synthase (NOS), all of which are expressed in the kidney: eNOS is found in vessels, nNOS mainly in the macula densa and inner medullary collecting duct, and iNOS in several tubule segments. Nitric oxide binds soluble guanylate cyclase, enhancing the synthesis of cyclic guanosine monophosphate (cGMP) from guanosine triphosphate (GTP), and mediates the following physiological actions in the kidney:

- regulation of kidney perfusion and glomerular pressure
- natriuresis, by inhibiting Na^+/K^+-ATPase and NHE3
- antagonism of AVP
- modulation of tubuloglomerular feedback.

Further reading

O'Hara DV, Lam CSP, McMurray JJV et al. Applications of SGLT2 inhibitors beyond glycaemic control. *Nat Rev Nephrol* 2024; 20:513–529.
Takkavatakarn K, Thammathiwat T, Phannajit J et al. The impacts of hypoxia-inducible factor stabilizers on laboratory parameters and clinical outcomes in chronic kidney disease patients with renal anemia: a systematic review and meta-analysis. *Clin Kidney J* 2023; 16(5):845–858.

INVESTIGATION OF KIDNEY AND URINARY TRACT DISEASE

The kidneys have a great deal of reserve; as a result, early stages of CKD are often asymptomatic, even as renally cleared metabolites begin to accumulate. It is not clear which metabolites

lead to specific symptoms, but products of protein catabolism (nitrogenous waste products) are thought to be important. These metabolites, also called uraemic toxins, must have a small molecular size since haemodialysis, which clears only relatively small molecules, improves symptoms. Little else is known with certainty.

Serum urea and creatinine concentrations are measured in CKD as surrogates of accumulating uraemic toxins because measurement is easy, and there is a rough correlation between urea and creatinine concentrations and symptoms. However, these substances are not particularly toxic in themselves.

Examination of the urine

Unless urine is visibly bloody (or cola-coloured, in cases of myoglobinuria or haemoglobinuria), macroscopic inspection of the urine is not helpful. Cloudy, offensive-smelling urine may denote infection, but this is not specific or sensitive and should be confirmed on culture.

Volume

Healthy adults will pass between 800 and 2500 mL of urine per 24 hours (roughly 1 mL/kg per hour). Usually, the minimum urine output capable of maintaining solute excretion for health (without accumulation) is around 650 mL per day. If the kidney loses concentrating capacity (as occurs in CKD or diabetes insipidus), higher urine volumes are needed for the same daily solute output, and urine outputs may increase well above 2500 mL. High daily volumes are also seen with glycosuria or increased protein catabolism following surgery, as the solute load requiring excretion is higher. ***Nocturia*** is a symptom of kidney disease as patients without the ability to concentrate solute into a small urine volume will wake needing to pass urine during the night.

Dipsticks (chemical testing)

Dipsticks are cheap and hugely helpful in the initial investigation of suspected kidney disease, using reagents fixed on pads that change colour on reacting with specific elements in urine.

Specific gravity and osmolality

Urine specific gravity (SG, where <1.008 is dilute and >1.020 is concentrated) is a measure of the weight of dissolved particles in urine, reflecting urine osmolality. Measurement of urine SG or osmolality can be helpful in confirming loss of concentrating ability. It can also be helpful in oliguric patients, when a high SG might suggest prerenal AKI, as opposed to established acute tubular necrosis (ATN) (see Chapter 27).

Urinary pH

Measurement of urinary pH (usually 5.5–6.5) is helpful in the investigation and treatment of renal tubular acidosis (see Chapter 27, 'Renal tubular acidosis' section).

Blood

Haematuria may come from any site within the urinary tract (Fig. 26.6):

- ***Overt bleeding from the urethra*** is suggested when blood is seen at the start of voiding and then the urine becomes clear.
- ***Blood diffusely present*** throughout the urine usually comes from the bladder or above.
- ***Blood only at the end of micturition*** suggests bleeding from the prostate or bladder base.

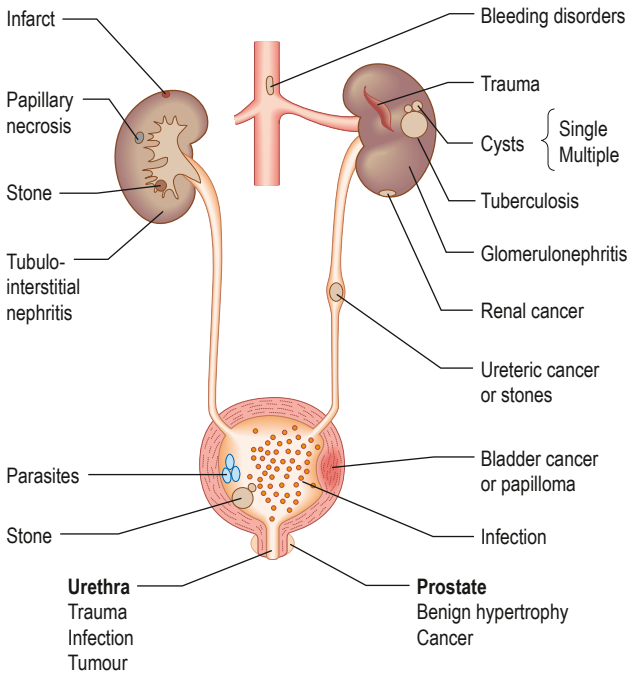


Fig. 26.6 Sites and causes of haematuria.

Females will commonly have dipstick-positive haematuria during a period; it is usually worth repeating testing after menstruation.

Once it is found on dipstick, a mid-stream sample can be spun at 2000 rpm and examined for red cells or casts on microscopy. Red cells (usually 1–2 per high-power field) are dysmorphic (altered in shape) if they originate from the early nephron (glomerular bleeding) and may be accompanied by red-cell casts. Casts are cylindrical bodies, moulded in the shape of the distal tubular lumen. Red-cell casts—even if only single—are suggestive of glomerulonephritis. A dipstick that is strongly positive for haematuria with no red cells seen on microscopy suggests haemoglobinuria or myoglobinuria.

Protein

Proteinuria is a common sign of kidney disease; normal excretion is less than 150 mg per day. Dipsticks will detect proteinuria from around 50 to 150 mg/L. If two separate urine samples find proteinuria on dipstick testing, a random urine protein:creatinine ratio (uPCR) should be measured. This convenient test has replaced timed 24-hour urine collections for proteinuria. An alternative is the albumin:creatinine ratio (uACR). As albumin is usually the dominant protein lost in the urine, patients who have albuminuria always have proteinuria (and uPCR will be higher than uACR). For people with diabetes, the urine ACR has prognostic significance and is the preferred screening test for ***microalbuminuria*** in early diabetic kidney disease. Both tests are expressed as mg protein per mmol creatinine. Generally, an uACR of 2.5 to 20 mg/mmol corresponds to albuminuria of 30 to 300 mg daily (crudely, multiply the result by 15 to approximate excretion in 24 h). Similarly, uPCR of 3 to 30 mg/mmol corresponds to urine protein excretion of 30 to 300 mg per day (multiply by a factor of 10). The uPCR assay is relatively cheap and identifies patients with both tubular and glomerular proteinuria. uACR or uPCR concentrations independently predict all-cause and cardiovascular mortality in the general population and allow risk stratification of patients with CKD for future kidney outcomes. Pyrexia, exercise, and an upright posture ***(postural proteinuria)*** all

increase urinary protein output but are benign. ***Urine electrophoresis and immunofixation*** are needed to detect light chains, which can be present in myeloma without a detectable serum paraprotein.

Electrolytes

Urinary electrolytes are unhelpful in CKD. The use of urinary sodium concentration in the distinction between pre-renal and intrinsic kidney disease is discussed in Chapter 27 ('Decreased extracellular volume' section).

Glucose

Glycosuria is uncommon, so a positive test for glucose should trigger diagnostic testing for diabetes mellitus.

Bacteria and pus cells

Dipstick tests for bacteriuria are based on the detection of ***nitrite*** produced from the reduction of urinary nitrate by bacteria, and on the detection of leucocyte esterase, an enzyme specific for neutrophils. Both tests have limitations. When positive, a mid-stream sample should be sent for microscopy, culture, and sensitivity (MC&S).

White blood cells (WBCs) and bacteria may be seen on microscopy. Urine tract infection (UTI) is suggested by 10 WBCs/mL or more in fresh mid-stream urine sample. However, not all pyuria is a UTI, and WBCs are also seen with kidney stones, tubulointerstitial nephritis, papillary necrosis, tuberculosis, and interstitial cystitis. White cell casts may occur.

Blood tests

The use of serum urea, creatinine, and GFR as measures of kidney function is discussed earlier (see 'Glomerular filtration rate' section). A general assessment of kidney disease might include the following:

Serum biochemistry

- Serum urea, electrolytes, and bicarbonate.
- Serum creatinine, with calculation of eGFR.
- Electrophoresis and immunofixation with serum-free light chains for myeloma.
- Calcium, phosphate, alkaline phosphatase, PTH, liver function, creatine kinase, lactate dehydrogenase.
- Blood glucose and HbA1C.

Haematology

- Full blood count, blood film, and ESR.
- Tests for sickle cell disease if relevant.

Immunology

- ***Complement components:*** low in active kidney disease due to systemic lupus erythematosus (SLE), mesangiocapillary glomerulonephritis, post-streptococcal glomerulonephritis, and cryoglobulinaemia.
- ***Autoantibody screening:*** SLE (e.g., ANA, anti-dsDNA), scleroderma (e.g., anti-Scl-70, anti-centromere, anti-RNA polymerase III), granulomatosis with polyangiitis and microscopic polyangiitis (e.g., c-ANCA, p-ANCA), Goodpasture's syndrome (anti-GBM). Anti-phospholipase A2 receptor antibodies in membranous nephropathy.
- ***Cryoglobulins*** in mesangiocapillary glomerulonephritis.
- ***Antibodies to streptococcal antigens*** (antistreptolysin O titre (ASOT), anti-DNAse B) if post-streptococcal glomerulonephritis (PSGN) is possible.
- ***Antibodies to hepatitis B and C*** may point to polyarteritis or membranous nephropathy (hepatitis B) or cryoglobulinaemic kidney disease (hepatitis C).
- ***Antibodies to HIV*** in HIV-associated kidney disease.

Imaging techniques

Ultrasonography

Ultrasound of the kidneys and bladder is safe and non-invasive, avoiding ionizing radiation and intravascular contrast medium. It is the imaging method of choice for:

- Assessing kidney size and symmetry. Normal-sized kidneys with abnormal function may suggest an acute cause. Kidneys scar as they fail and may shrink in length and volume.
- Ruling out obstruction, either of the bladder (bilateral hydronephrosis), ureters (unilateral hydronephrosis), or the kidney itself (pelvicalyceal dilation due to high ureteric or renal pelvis disease).
- Characterizing kidney masses as cystic (simple cysts, polycystic kidneys) or complex/solid (benign/malignant kidney tumours, infected collections).
- Guiding interventions aimed at relieving obstruction (percutaneous nephrostomy).
- ***Doppler ultrasonography (duplex)*** for confirming renal vein patency. Doppler for arterial disease in native kidneys is often unreliable.
- Bladder assessment: tumours, stones, post-micturition volume to assess bladder emptying.
- Contrast-enhanced ultrasound using gas-filled microbubbles may offer better visualization of organs and blood vessels.

Computed tomography

Unenhanced computed tomography (CT) is the first-line investigation for ***ureteric colic*** and suspected kidney calculi. It has superseded excretion urography (intravenous urography [IVU]/intravenous pyelography [IVP]). It is also used to:

- characterize kidney masses that are indeterminate on ultrasound imaging
- stage kidney and bladder tumours
- evaluate the retroperitoneum for tumours, retroperitoneal fibrosis (peri-aortitis), and other causes of ureteric obstruction
- assess kidney trauma
- visualize the renal arteries and veins with angiography.

Positron emission tomography (PET), using ^{18}F-fluorodeoxyglucose (FDG), is useful for detection of infection (e.g., in a cyst), inflammation, or tumours, and is often used in conjunction with CT (PET-CT).

Magnetic resonance imaging

Magnetic resonance imaging (MRI) is used as an alternative to CT with no irradiation:

- to stage prostate (and renal and bladder) cancer
- to reconstruct the anatomy of the renal arteries using magnetic resonance ***angiography*** (MRA) with gadolinium contrast

Care is advised with gadolinium in patients with kidney insufficiency (eGFR < 30/mL) because of the risk of nephrogenic systemic fibrosis. However, with newer gadolinium-based contrast agents (e.g., gadobenate dimeglumine, gadoteridol, gadoterate meglumine, and gadobutrol), the risk for NSF is exceedingly low and the potential harm of withholding indicated diagnostic investigation may outweigh risk.

Plain radiography (X-ray)

A plain radiograph of the abdomen may be useful to identify kidney calcification or radiodense calculi in the kidney, renal pelvis, and the line of the ureters or bladder (Fig. 26.7).

Micturating cystourethrography

After bladder catheterization, contrast is instilled into the bladder. The catheter is then removed, and the patient screened during voiding to check for vesico-ureteric reflux and to study urethra and bladder emptying. It may be useful in children with recurrent urinary tract infection.

Aortography or renal arteriography

Conventional or digital subtraction angiography (DSA) has largely been replaced by CT or MR angiography. It may be used if fibromuscular dysplasia (FMD) suspected, to allow simultaneous therapeutic intervention.

Renal scintigraphy

Isotope studies are helpful for dynamic or static investigation of perfusion or excretion. Following venous injection of a bolus of tracer, emissions from the kidney are recorded by gamma camera. Technetium-labelled diethylenetriaminepenta-acetic acid (^{99m}Tc-DTPA) is excreted by glomerular filtration and can be used to confirm kidney perfusion. Dimercaptosuccinic acid labelled with technetium (^{99m}Tc-DMSA) is filtered by the glomerulus and then binds to proximal tubular cells. It is used in static studies and is useful to assess the relative contribution of function from asymmetric kidneys, and to highlight 'photon-deficient' areas (isotope not seen) suggestive of scars or infarction. Mercaptoacetyltriglycine (MAG3) labelled with technetium (^{99m}Tc) is excreted by tubular secretion, so resistance to flow in the pelvis or ureter (with obstruction) prolongs parenchymal transit. On whole-kidney renograms, the time–activity curve fails to fall after an initial peak, or continues to rise (Fig. 26.8), confirming hold-up to flow. Furosemide may be given to exaggerate urine output and emphasize delay to aid diagnosis.

Transcutaneous kidney biopsy

Kidney biopsy (Box 26.5) is carried out under ultrasound guidance and requires interpretation by a specialist pathologist. It is helpful in the investigation of the nephritic and nephrotic syndromes, acute and chronic kidney disease, haematuria after normal urological investigations, and kidney transplant dysfunction. Native kidney biopsy material is examined by conventional histochemical staining, by electron microscopy, and by immunoperoxidase or immunofluorescence. Techniques like *in situ* hybridization and polymerase chain reaction (PCR) analysis are also widely used. The complications of transcutaneous kidney biopsy are shown in Box 26.6.

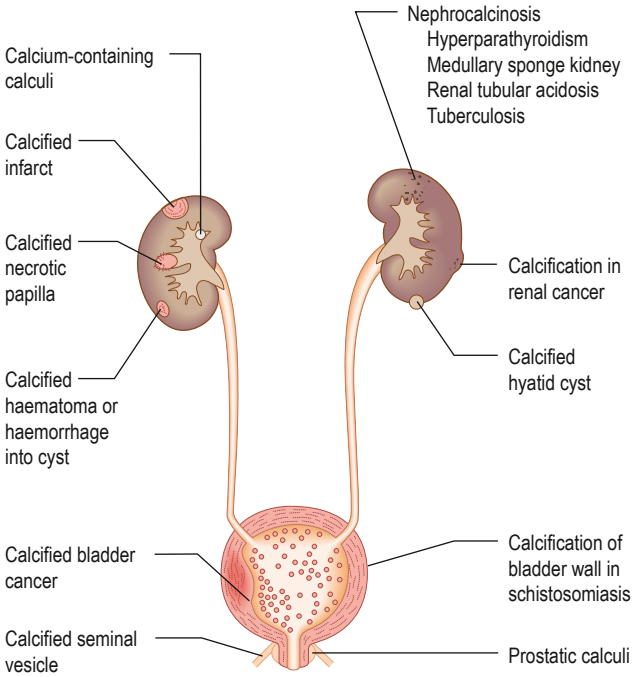


Fig. 26.7 Calcification in the kidney and urinary tract. Calculi can occur at any site.

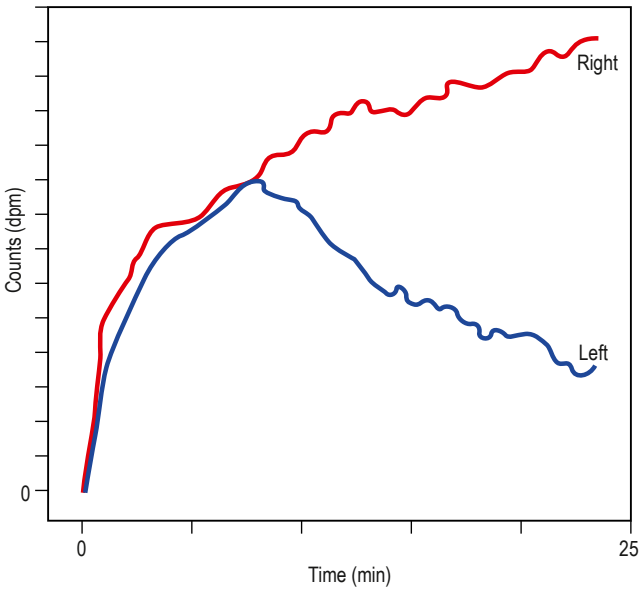


Fig. 26.8 Dynamic scintigram. Note the progressive rise of the right kidney curve to a plateau (in contrast to the normal left kidney curve) due to right-sided urinary tract obstruction.

THE GLOMERULUS AND GLOMERULAR DISEASE

The glomerular filtration barrier allows or prevents filtration in the glomerulus. It consists of the fenestrated endothelium, the GBM, and the terminally differentiated visceral epithelial cells known as podocytes (Figs. 26.2 and 26.9). Podocytes attach to the GBM by foot processes via adhesion molecules, such as $\alpha_3\beta_1$ and dystroglycans. Adjacent podocytes are joined laterally via their foot process by slit diaphragms. The various proteins comprising the slit diaphragm include nephrin, CD2-associated protein (CD2AP), canonical transient receptor potential channel 6 (TRPC6), podocin, P-cadherin, α- and β-catenin, and zonula occludens-1 (ZO-1). They co-localize within the subcellular domain to function as a molecular sieve. These proteins, in addition to providing structural support to the cytoskeletal proteins like filamentous actin, also provide signalling to maintain normal podocyte function. Abnormalities in any of these proteins result in the breakdown of the filtration barrier and consequent leak of macromolecules.

Box 26.5 Transcutaneous kidney biopsy

Before biopsy

- Check blood pressure, platelet count, and clotting are safe for biopsy
- Group and save for crossmatch if needed.
- Informed consent including benefits, risks, and other options if biopsy not performed.

Biopsy

- Patient lies prone with a pillow under the abdomen.
- Localize the kidney by ultrasound.
- Inject local anaesthetic along the planned biopsy track.
- Instruct the patient to hold a breath (prevent kidney movement) when biopsy is performed

After biopsy

- Apply dressing to the biopsy site.
- Bedrest after procedure, usually for at least 6 h.
- Monitor heart rate and blood pressure regularly.
- Advise the patient to represent for urgent care if macroscopic haematuria or symptoms of hypotension.

Box 26.6 Complications of transcutaneous kidney biopsy

- Macroscopic haematuria (up to 10%)
- Flank pain, can refer to shoulder tip
- Perirenal haematoma
- Arteriovenous aneurysm formation (approximately 20%), almost always of no clinical significance
- Profuse haematuria requiring transfusion (1%–3%)
- Profuse haematuria requiring radiological (embolization) or surgical (nephrectomy) intervention (approximately 1 in 400)
- Infection
- Mortality (approximately 0.1%)

Podocyte changes

The podocyte structure is maintained by actin, which supports the cytoskeleton. A rearrangement of the actin cytoskeleton leads to foot process effacement (flattening) and albumin leaks into the urine (Fig. 26.9). Recovery of the cytoskeleton reverses proteinuria. The cytoskeleton can be altered by:

- abnormalities of cytoskeletal proteins, e.g., α-actinin-4 mutations in hereditary focal segmental glomerular sclerosis
- injury to or abnormalities of slit diaphragm proteins
- changes in the GBM
- direct injury to podocytes by viral infection, drugs, toxins, or local activation of renin–angiotensin.

Glomerular disease

Glomerular disease is the third most common cause (after diabetes and hypertension) of kidney disease requiring renal replacement (end-stage kidney disease [ESKD]) in Europe and the USA, accounting for 10% to 15% of such patients. Glomerular diseases are characterized by:

- either immunologically mediated inflammatory glomerular injury or structural/functional glomerular damage without inflammation
- associated interstitial damage
- symmetrical kidney involvement
- secondary mechanisms of glomerular injury including fibrin deposition, platelet aggregation, neutrophil infiltration, and free radical-induced damage
- haemodynamic consequences that further disturb glomerular function.

Describing glomerular disease

The nomenclature relating to glomerular disease can be confusing, as descriptive histological terms overlap with clinical syndromes and more recent molecular insights into pathogenesis. Glomerular disease is usually described by kidney biopsy findings. Commonly used terms are shown in Box 26.7. If there is predominant inflammation, glomerular disease is described as a ***glomerulonephritis***. If inflammation is absent, the term ***glomerulopathy*** is used. There remains much overlap and the terms are often (wrongly) used interchangeably. It may be better to think about glomerular disease in terms of the component of the GBM that is predominantly affected:

- **Podocytes** (in the urinary compartment) are principally involved in glomerular diseases (usually glomerulopathies) that present as the ***nephrotic syndrome*****.**
- **Endothelial and mesangial cells** (in the endocapillary compartment) are principally involved in glomerular disease presenting as ***nephritis*** (glomerulonephritis), where haematuria, proteinuria, and often hypertension are evident.
- Podocytes, endothelial, and mesangial cells may be equally involved if a glomerulonephritis causes nephrosis.

Clinical classification of glomerular disease is also often used, although there is no complete correlation between histopathological types and clinical features. Four major glomerular syndromes are described:

- ***Nephrotic syndrome:*** proteinuria (>3.5 g/day), hypoalbuminaemia, oedema, lipiduria, and hyperlipidaemia. Podocyte malfunction or injury is often causative.
- ***Glomerulonephritis (nephritic syndrome)*** presents as one of:
  - ***Acute glomerulonephritis:*** abrupt onset glomerular haematuria (red blood cell casts or dysmorphic red blood cells), non-nephrotic-range proteinuria, oedema, hypertension, and transient kidney impairment
  - ***Rapidly progressive glomerulonephritis:*** features of acute nephritis, focal necrosis with or without crescents, and rapidly progressive kidney failure over weeks
- ***Mixed nephritic/nephrotic presentations:*** where glomerulonephritis is part of a systemic disease (e.g., lupus nephritis, cryoglobulinaemia, Henoch–Schönlein purpura), a nephritic syndrome is often associated with the nephrotic syndrome.
- ***Asymptomatic haematuria, proteinuria***, or both.

Investigation of glomerular diseases is shown in Box 26.8.

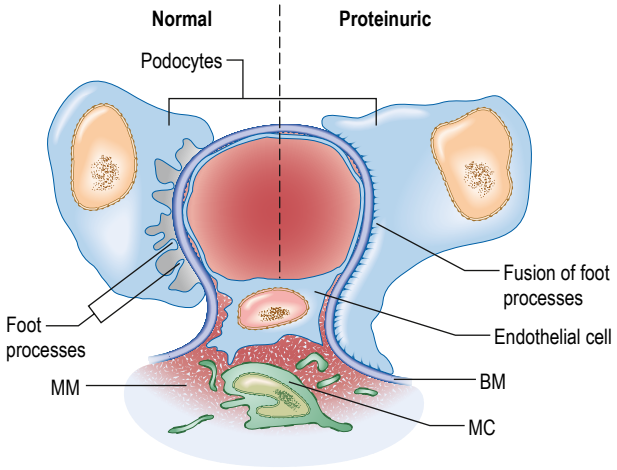


Fig. 26.9 A normal *(left)* and a proteinuric *(right)* glomerulus. A capillary loop shows normal glomerular morphology *(left)* including podocytes with pseudopodia (foot processes). In proteinuria *(right)*, fusion of foot processes is characteristic of many diseases. *BM,* Basement membrane; *MC,* mesangial cell; *MM,* mesangial matrix.

Box 26.7 Glomerular disease: commonly used terms

- ***Focal:*** some, but not all, glomeruli have the lesion
- ***Diffuse*** (global): most glomeruli (>75%) have the lesion
- ***Segmental:*** part, not all, of the glomerulus is affected (most focal lesions are also segmental, e.g., focal segmental glomerulosclerosis)
- ***Global:*** all glomeruli are symmetrically involved
- ***Proliferative:*** an increase in cell numbers due to hyperplasia of one or more of the resident glomerular cells with or without inflammation
- ***Membrane alterations:*** capillary wall thickening due to deposition of immune deposits or alterations in basement membrane
- ***Crescent formation:*** epithelial cell proliferation with mononuclear cell infiltration in Bowman's space

Box 26.8 Investigation of glomerular diseases

Investigations	Positive findings
Urine microscopy	Red cells, red-cell casts
Urinary protein	Nephrotic or sub–nephrotic-range proteinuria quantified by uACR or uPCR
Serum urea	May be increased
Serum creatinine	May be increased
Culture (throat swab, discharge from ear, swab from inflamed skin)	Nephritogenic organism (not always detected)
Antistreptolysin-O titre	Elevated in post-streptococcal nephritis
C3 and C4 concentrations	May be reduced
Antinuclear antibody (ANA)	Significant titre in systemic lupus erythematosus
Antinuclear cytoplasmic antibody (ANCA)	Positive in some vasculitides
Anti-glomerular basement membrane (anti-GBM)	Positive in anti-GBM disease (Goodpasture's syndrome)
Cryoglobulins	Increased in cryoglobulinaemia
Chest radiograph	Cardiomegaly, pulmonary oedema, pulmonary haemorrhage
Kidney imaging	Usually structurally normal renal tract
Kidney biopsy	Diagnosis of glomerulopathy

Nephrotic syndrome

Nephrotic syndrome is characterized by:

- hypoalbuminaemia
- >3.5 g proteinuria/day
- hyperlipidaemia
- salt and water retention, leading to oedema.

Pathophysiology

Hypoalbuminaemia

Loss of 3.5 g or more of urinary protein (largely albumin) per day leads to hypoalbuminaemia. Normal dietary protein intake in the UK is around 70 g daily and the normal liver can synthesize albumin at a rate of 10 to 12 g daily. How, then, does a daily urinary protein loss of 3.5 g result in hypoalbuminaemia? This can be partly explained by increased catabolism of reabsorbed protein, largely albumin, in the proximal tubules, even though the rate of albumin synthesis is increased.

Proteinuria

Proteinuria occurs due to the pathology of the glomerular barrier (podocytes, basement membrane, fenestrated endothelium, endothelial charge) allowing passage of more and larger molecules.

Hyperlipidaemia

Increased synthesis of lipoproteins (such as apolipoprotein B, C-III lipoprotein (a)) occurs as a direct consequence of a low plasma albumin. Low-density lipoprotein (LDL) increases, partly due to upregulation of a liver serine protease, ***pro-protein convertase subtilisin kexin-9 (PCSK9)***, which causes internalization of LDL receptors. There is no change (or a decrease) in high-density lipoprotein (HDL), meaning an increased LDL:HDL ratio. Very-low-density lipoprotein (VLDL) and/or intermediate-density lipoprotein (IDL) fractions also increase. Reduced clearance of the principal triglycerides bearing lipoprotein (chylomicrons and VLDL), as high plasma concentrations of free fatty acid (FFA) trigger release of appropriately sialylated angiopoietin-like 4 (ANGPTL4) protein from adipose tissue, heart, and skeletal muscles, inhibits lipoprotein lipase and results in hypertriglyceridaemia.

Oedema in hypoalbuminaemia

See Chapter 27 ('Increased extracellular volume' section).

Management

General measures

- ***Initial management*** should be with dietary sodium restriction and a loop diuretic (e.g., furosemide or bumetanide). There is little consensus on diuretic use in the nephrotic syndrome, but by convention unresponsive patients require the addition of a synergetic thiazide-like diuretic (such as metolazone) or an ENaC inhibitor such as amiloride. Nephrotic patients may malabsorb diuretics (as well as other drugs) owing to gut mucosal oedema, and intravenous administration may be needed. Patients are sometimes hypovolaemic, and moderate oedema may have to be accepted to avoid postural hypotension.
- ***Normal protein intake*** is advisable. A high-protein diet (80–90 g protein daily) increases proteinuria, which may be harmful in the long term.
- ***Hypercoagulable states*** predispose to venous thrombosis. The hypercoagulable state is due to an imbalance between naturally occurring pro-coagulant/pro-thrombotic factors (accelerated activated intrinsic pathway, fibrinogen, platelet function, and fibrin–platelet interaction) and anticoagulant/antithrombotic factors (reduced activity of antithrombin III, protein S and C) which promotes in situ thrombosis in deep veins or arteries. Prolonged bed rest should be avoided, as thromboembolism is common (particularly in membranous nephropathy). Long-term prophylactic anticoagulation may be indicated, and if renal vein thrombosis occurs, treatment dose anticoagulation is required.
- ***Sepsis*** is a major cause of death in nephrotic patients. Increased susceptibility to infection is partly due to loss of immunoglobulin in the urine. Pneumococcal infections are particularly common and pneumococcal vaccine should be given. Early detection and aggressive treatment of infections, rather than long-term antibiotic prophylaxis, constitute the best approach.
- ***Lipid abnormalities*** are responsible for an increase in the risk of cardiovascular disease in patients with proteinuria. Treatment of hypercholesterolaemia starts with an HMG-CoA reductase inhibitor ('statin').
- ***ACE inhibitors and/or angiotensin II receptor antagonists*** (AII-RAs) are indicated for their antiproteinuric properties in all

types of glomerulonephropathy. They reduce proteinuria by lowering glomerular capillary filtration pressure (a fall in efferent tone decreases the transglomerular capillary pressure) and therefore protein loss into the urinary space.

- ***SGLT-2 inhibitors*** may delay CKD progression and reduce proteinuria in kidney conditions associated with nephrotic-range proteinuria, with emerging evidence of disease-specific benefit.

Specific measures

Treat the underlying cause of urinary protein leak. **Box 26.9** shows the glomerular lesions commonly associated with the nephrotic syndrome.

Non-nephrotic range proteinuria

In asymptomatic patients, isolated proteinuria without haematuria is often an incidental finding. It is usually less than 1 g/day with normal kidney function. Over 50% of these patients have (benign) postural proteinuria, although it may be an early sign of a serious glomerular lesion such as membranous glomerulopathy, IgA nephropathy, FSGS, diabetic nephropathy or amyloidosis (see later). Mild and transient proteinuria may also accompany a febrile illness, congestive heart failure, or infectious diseases with no clinical significance.

Causes of proteinuria and nephrotic syndrome

Minimal-change nephropathy (minimal-change disease)

The name is derived from the fact that glomeruli appear normal on light microscopy (**Fig. 26.10**). However, on electron microscopy there is fusion of podocyte foot processes. Neither immune complexes nor anti-GBM antibody can be demonstrated by immunofluorescence.

Underlying pathophysiological mechanisms include:

- Immature differentiating CD34 stem cells (rather than mature T lymphocytes)
- Interleukin-13 (IL-13) affecting podocyte function
- Vascular endothelial growth factor (VEGF) production
- Upregulation of vascular hyposialylated-angiopoietin-like 4 (ANGPTL4)
- Atopy is present in 30% of cases of MCN, and allergic reactions can trigger nephrosis
- Drugs: NSAIDs, lithium, antibiotics (cephalosporins, rifampicin, ampicillin), bisphosphonates, and sulfasalazine
- Infection: hepatitis C virus, HIV, tuberculosis

Clinical features

MCN is responsible for most nephrotic syndrome in childhood. Proteinuria is usually 'highly selective': albumin is lost in the urine but not higher-molecular-weight proteins such as immunoglobulins. Oedema is usual and may present periorbitally. The condition accounts for 20% to 25% of cases of adult nephrotic syndrome. It is often regarded as a condition that does not lead to CKD (but see focal segmental glomerulosclerosis later).

Management

- Manage symptoms with general measures (see earlier).
- High-dose corticosteroid therapy with prednisolone 60 mg/m^2 daily (up to a maximum of 80 mg/day) for a maximum of 4 to 6 weeks, followed by 40 mg/m^2 every other day for a further 4 to 6 weeks reverses proteinuria in more than 95% of children. Response rates in adults are lower and response may occur only after many months. Spontaneous remission also occurs, and steroid therapy is usually withdrawn/withheld if urinary protein loss is insufficient to cause hypoalbuminaemia and oedema. In both children and adults, if remission lasts for 4 years after discontinuing steroid therapy, further relapse is rare.
- Two-thirds of children relapse after steroid therapy and further courses of corticosteroids are required. One-third of these patients regularly relapse on steroid withdrawal, so that a second-line agent should be added after repeat induction with steroids.
- Ciclosporin or tacrolimus may be effective (with or without steroids) but must be continued long-term to prevent relapse on stopping treatment. The antiproteinuric effect of these calcineurin inhibitors is normally attributed to its immunosuppressive action but may result from the stabilization of the actin cytoskeleton in kidney podocytes. Ciclosporin inhibits the calcineurin-mediated dephosphorylation of synaptopodin (a regulator of actin cytoskeleton) and protects it from cathepsin L-mediated degradation. These results have shed new light on the role of calcineurin signalling in proteinuric kidney diseases. Excretory function and ciclosporin and tacrolimus blood concentrations must be monitored as both drugs are potentially nephrotoxic.

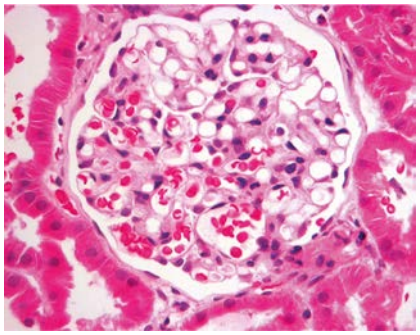


Fig. 26.10 Normal glomerulus on light microscopy in minimal-change nephropathy.

Box 26.9 Glomerulopathies associated with nephrotic syndrome

Clinical syndrome	Primary (kidney) disease	Secondary (systemic) disease
Nephrotic syndrome	Minimal-change nephropathy Congenital nephrotic syndrome Focal segmental glomerular sclerosis Membranous nephropathy	Amyloidosis Diabetic nephropathy
Mixed nephrotic/nephritic	Mesangiocapillary glomerulonephritis Mesangial proliferative glomerulonephritis	Systemic lupus erythematosus Cryoglobulinaemic disease Henoch–Schönlein syndrome Idiopathic fibrillary glomerulopathy Immunotactoid glomerulopathy Fibronectin glomerulonephropathy

- Rituximab, a depleting monoclonal antibody directed against CD20 on B lymphocytes, is increasingly used in steroid-unresponsive and frequently relapsing disease.
- Cyclophosphamide 1.5 to 2.0 mg/kg/day for 8 to 12 weeks with prednisolone 7.5 to 15 mg/day increases the likelihood of long-term remission. Steroid-unresponsive patients may respond. No more than two courses of cyclophosphamide should be prescribed in children because of the risk of side-effects, which include future infertility (azoospermia and premature ovarian failure).
- In corticosteroid-dependent children, the anthelminthic agent levamisole 2.5 mg/kg to a maximum of 150 mg on alternate days is also useful in maintenance of remission. Levamisole's mode of action is attributable to its direct effects on podocytes by inducing expression of glucocorticoid receptor (GR) and activated GR signalling in podocytes.

Congenital nephrotic syndrome

Congenital nephrotic syndrome (Finnish type) is an autosomal recessively inherited disorder due to mutations in the gene coding for a transmembrane protein, nephrin. Incidence is 1 in 8200 in Finland. Nephrin is a critical element of the filtration slit, and its loss of function results in massive proteinuria shortly after birth. The disorder can be diagnosed *in utero*, as increased α-fetoprotein in amniotic fluid is a common feature. This condition is characterized by relentless progression to ESKD. Other inherited nephrotic syndromes involve mutations in genes that encode other podocyte proteins, such as podocin, α-actinin 4, and Wilms' tumour suppressor gene. Congenital nephrotic syndrome patients are steroid-resistant. Mutation analysis in steroid-resistant nephrotic syndrome (SRNS) has identified more than 30 recessive or dominant genes, revealing the encoded proteins that are essential for glomerular function. Mutation analysis could be offered to all individuals who manifest with SRNS before the age of 25 years, for diagnosis and for exploration of possible therapies, as it may permit personalized treatment options based on genetic causation by way of 'precision medicine'.

Focal segmental glomerulosclerosis

Focal segmental glomerulosclerosis (FSGS) describes a sclerotic glomerular lesion that affects some but not all glomeruli (focal), and some but not all segments of each tuft (segmental).

- ***Primary FSGS*** is an unusual primary cause of the nephrotic syndrome.
- ***Secondary FSGS*** looks similar on light microscopy. Although proteinuria may be heavy, hypoalbuminaemia is *unusual.*

Primary focal segmental glomerulosclerosis

This disease usually presents with massive proteinuria (usually non-selective), haematuria, hypertension, and kidney impairment. The associated nephrotic syndrome is often resistant to steroid therapy. All age groups are affected. It usually recurs in transplanted kidneys, sometimes within days of transplantation, and particularly in patients with aggressive primary kidney disease.

Aetiology

A circulating permeability factor causes the increased protein leak: plasma from patients increases membrane permeability in isolated glomeruli. Kidneys transplanted into murine models of FSGS develop the lesion, but kidneys from FSGS-prone mice transplanted to a normal strain are protected. Removal of this factor by plasmapheresis results in transient amelioration of proteinuria. The identity of the permeability factor remains unknown but recent findings suggest that cardiotrophin-like cytokine 1 is a likely candidate. Soluble urokinase-like plasminogen activator receptor (SuPAR) was initially thought to be involved but now appears less likely based on recent experimental and clinical evidence.

Pathology

Segmental glomerulosclerosis is seen on light microscopy, which later progresses to global sclerosis. The deep glomeruli at the corticomedullary junction are affected first. These may be missed on transcutaneous biopsy, leading to a mistaken diagnosis of MCN (see earlier). A pathogenetic link may exist between MCN and FSGS, as a proportion of cases classified as having the former condition develop progressive CKD, which is unusual. Electron microscopy demonstrates primarily foot process effacement, occasionally in a patchy distribution. The degree of podocyte foot process effacement on electron microscopy can help distinguish between 'primary' and 'secondary' FSGS. If severe foot process effacement is present in normal and sclerosed glomeruli on electron microscopy, primary FSGS is more likely (this is not the case if foot process effacement is largely localized to sclerosed glomeruli alone).

Five histological variants of FSGS exist:

- ***Classic FSGS*** (Fig. 26.11A): the involved glomeruli show sclerotic segments in any location of the glomerulus.
- ***Tip FSGS/glomerular tip lesion*** (Fig. 26.11B): segmental sclerosis at the tubular pole of all affected glomeruli at a very early stage. These patients have a more favourable response to steroids and the disease runs a more benign course.
- ***Collapsing FSGS*** (Fig. 26.11C): podocytes are usually enlarged and coarsely vacuolated with wrinkled and collapsed capillary walls. Commonly seen in HIV infection when it is termed HIV-associated nephropathy (HIVAN; see later).
- ***Perihilar variant*** (Fig. 26.11D): perihilar sclerosis and hyalinosis in more than 50% of segmentally sclerotic glomeruli. Frequently observed with secondary FSGS.
- ***Cellular variant*** (Fig. 26.11E): at least one glomerulus with segmental hypercellularity (proliferation) that occludes the capillary lumen.

Management

- ***Prednisolone*** 0.5 to 2 mg/kg per day in most patients and continued for 6 months before the patient is considered resistant to therapy, which is common.
- ***Ciclosporin or tacrolimus*** may be effective in reducing or stopping urinary protein excretion. Relapse is very common, so long-term use is commonly required.
- ***Rituximab:*** B lymphocyte depletion therapy is an additional strategy used particularly when response to steroids and calcineurin inhibitors is poor. It induces partial and or complete remission in 30% to 50% of cases.
- ***Cyclophosphamide***, ***chlorambucil***, or ***azathioprine.*** In patients with mesangial hypercellularity and tip lesion, cyclophosphamide 1 to 1.5 mg/kg per day with 60 mg of prednisolone for 3 to 6 months, followed by maintenance prednisolone and azathioprine can be used.

About 50% progress to ESKD within 10 years of diagnosis, particularly if resistant to therapy. Recurrence following kidney transplantation is high with a poor prognosis. Plasmapheresis or immunoabsorption has been the mainstay of treatment in patients with post-transplant recurrence with modest results. Anti-CD80/86 (abatacept) has been tried with some success in patients with podocyte upregulation of CD80. Adrenocorticotrophic hormone (ACTH) gel applied topically has been used successfully in refractory cases of post-transplant FSGS.

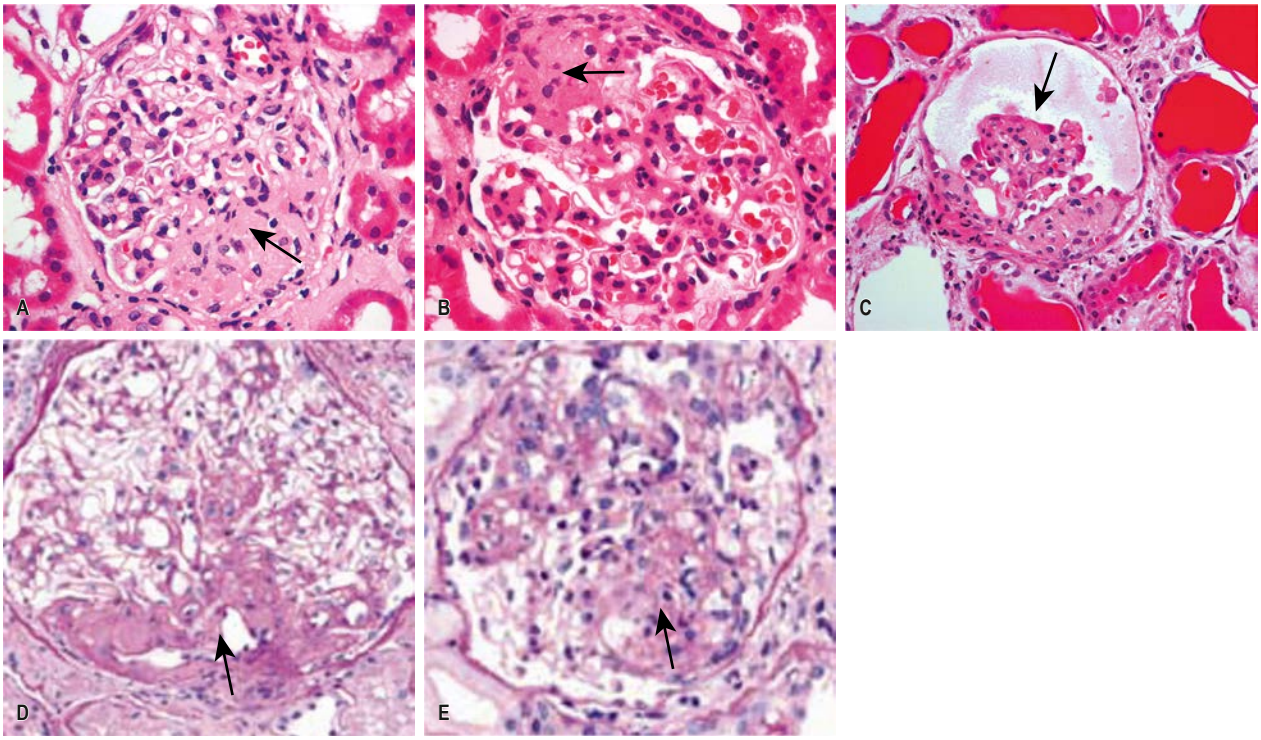


Fig. 26.11 Focal segmental glomerulosclerosis (FSGS). **(A)** Classic FSGS showing sclerotic segments in the glomerulus. **(B)** Glomerular tip lesions showing segmental sclerosis at the pole of the glomerulus. **(C)** Collapsing FSGS. **(D)** FSGS perihilar variant. **(E)** FSGS cellular variant. Arrows show the sclerotic segment in each light micrograph.

Secondary FSGS

Secondary FSGS with similar glomerular changes is seen as a ***secondary phenomenon*** when it occurs due to a reduction in functioning nephron number. In this context, FSGS represents a common final glomerular lesion in response to haemodynamic glomerular strain. As nephrons fail, increased flow through the remaining nephrons leads to glomerular hypertrophy, hyperfiltration, and hydraulic injury with secondary changes of FSGS. Associations include:

- ***reduced nephron number:*** nephrectomy, hypertension, obesity, ischaemia, sickle nephropathy, reflux nephropathy, chronic allograft nephropathy, IgA nephropathy, scarring following renal vasculitis
- ***mutations*** in specific podocyte genes
- ***viruses:*** HIV type 1, erythrovirus B19, cytomegalovirus, Epstein–Barr virus, simian virus 40
- ***drugs:*** heroin, all interferons, anabolic steroids, lithium, sirolimus, pamidronate, calcineurin inhibitors.

APOL1 gene mutations on chromosome 22 (selected due to protection against African trypanosomiasis in patients of African ancestry) are associated with podocyte injury hypothesized due to cationic action in response to insults such as hypertension, SLE, sickle cell trait or disease, HIV, and SARS-CoV-2. In a recent phase 2a clinical study, targeted inhibition of APOL1 channel function with oral inaxaplin reduced proteinuria by 47% in 13 patients with an APOL1 disease allele and FSGS. Results of ongoing trials are awaited.

HIV-associated nephropathy

Glomeruli are characteristically 'collapsed' on light microscopy (Fig. 26.11C): podocytes are enlarged, hyperplastic, and coarsely vacuolated, containing protein absorption droplets and overlying capillaries with varying degrees of wrinkling and collapse of the walls. Direct podocyte HIV-1 infection is associated with loss of podocyte-specific markers such as Wilms' tumour factor and synaptopodin. HIVAN presents with nephrotic-range proteinuria, oedema, and CKD, which can be rapid in progression. Anti-retroviral therapy (ART) may reverse the lesions and restores kidney function if treatment is commenced early.

Membranous glomerulopathy

Autoimmune membranous glomerulopathy occurs mainly in adults. It presents with asymptomatic proteinuria or nephrotic syndrome. Microscopic haematuria, hypertension, and/or kidney impairment may accompany the nephrotic syndrome. In membranous glomerulopathy, one-third of patients undergo spontaneous or therapy-related remission but, as in other glomerular diseases, hypertension and a greater degree of kidney impairment are poor prognostic signs. About 40% develop CKD, usually in association with persistent nephrotic-range proteinuria. Younger people, females, and those with asymptomatic moderate proteinuria at the time of presentation do best.

Pathogenesis

Most patients (75%) with idiopathic or autoimmune membranous nephropathy have been found to have under galactosylated immunoglobulin G4 (IgG4)-type autoantibodies against M type phospholipase A_2 receptor (PLA_2R), a glycoprotein constituent of normal glomeruli. PLA_2R is present in normal human podocytes and in immune deposits in patients with idiopathic membranous nephropathy, indicating that it could be a major autoantigen in this disease. It is linked to human leucocyte antigen (HLA)-DQA1. Other target antigens include:

- thrombospondin type 1 domain-containing 7A (THSD7A) in 3% of PLA_2R-positive and 10% of PLA_2R-negative patients
- exostosin ½ associated with membranous lupus nephritis
- neural epidermal growth factor-like 1 (NELL-1) associated with malignancy and leads to IgG1 co-dominant antibody deposition in the kidney
- serine protease HTRA1
- semaphorin 3B
- protocadherin 7
- netrin G1
- contactin 1
- neural cell adhesion molecule-1
- transforming growth factor beta receptor 3
- neural-derived neurotropic factor
- proprotein convertase subtilisin/kexin type 6 associated with NSAID use.

In secondary membranous nephropathy, where no detectable anti-PLA_2R or glomerular antigen staining can be found, glomerular histology is identical to that seen in autoimmune disease. Causes include:

- drugs, e.g., penicillamine, gold, NSAIDs, probenecid, mercury, captopril
- autoimmune disease, e.g., SLE, thyroiditis
- infection, e.g., hepatitis B, hepatitis C, schistosomiasis, *Plasmodium malariae*
- cancer, e.g., lung, colon, stomach, breast, lymphoma
- other, e.g., sarcoidosis, sickle cell disease.

On light microscopy, capillary loops appear thick. Using a periodic acid–Schiff or silver stain (which highlights basement membrane), 'spikes' of basement membrane are visible. On electron microscopy small, electron-dense deposits in the subepithelial aspects of the capillary walls are seen, encircled by perpendicular basement membrane spikes. Uniform granular capillary wall deposits of PLA_2R antigen and IgG subclasses (IgG4 is predominant in idiopathic membranous nephropathy), as well as complement C3, are seen on immunofluorescence. Late in the disease, deposits are surrounded by basement membrane and are undergoing resorption, which appears as uniform thickening of the capillary basement membrane on light microscopy (Fig. 26.12).

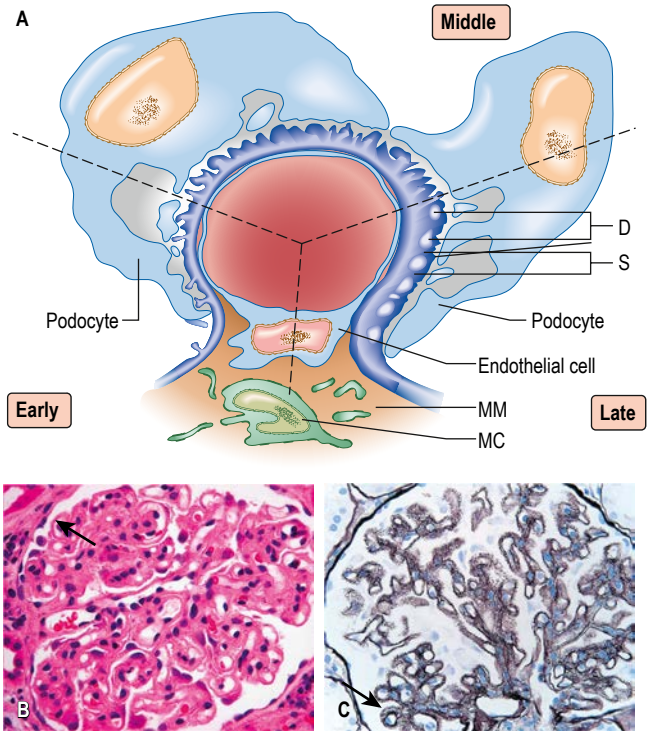


Fig. 26.12 Membranous glomerulopathy. (A) Membranous nephropathy showing early, middle, and late stages. *D,* Subepithelial deposits; *MC,* mesangial cell; *MM,* mesangial matrix; *S,* spikes. **(B)** Light microscopy of membranous nephropathy showing thickened basement membrane *(arrowed)*. **(C)** Silver stain showing spikes *(arrowed)*. *(A, After Marsh FP.* Postgraduate Nephrology. *Oxford: Butterworth Heinemann; 1985.)*

Management

As many as one-third or more of patients will undergo spontaneous remission within 6 to 12 months, particularly if kidney function is normal, anti-PLA_2R antibodies are low in titre, and proteinuria is modest. Patients with heavier proteinuria, progressive kidney dysfunction, and a high titre of anti-PLA_2R antibodies are considered for early treatment. Management includes:

- ACE inhibitors or AII-RAs at the maximum tolerated dose.
- If indicated: anticoagulation to prevent venous thromboembolism, diuretics, statin.
- Rituximab (anti-CD20): effective in autoimmune disease at inducing remission, maintaining or improving kidney function, and reducing proteinuria.
- Cyclophosphamide (1.5–2.5 mg/kg per day for 6–12 months with 1 mg/kg per day of oral prednisolone on alternate days for the first 2 months) OR chlorambucil (0.2 mg/kg per day in months 2, 4, and 6, alternating with oral prednisolone 0.4 mg/kg per day in months 1, 3, and 5).
- Ciclosporin or tacrolimus: treatment courses are longer and relapse is common. In recent trials, rituximab was superior to calcineurin therapy.
- Oral corticosteroids are of no benefit in isolation but may have an additive effect.

Amyloidosis

Amyloidosis is a systemic disorder of protein folding, in which normally soluble proteins or fragments are deposited extracellularly as abnormal, insoluble fibrils (usually β-pleated sheets that are resistant to proteolysis), causing progressive organ dysfunction and death. The disease may be acquired or inherited. Classification is based on the nature of the precursor plasma proteins (at least 20) that form the fibrillar deposits. The most common forms are AL amyloidosis (where abnormal protein may be derived from light chains or immunoglobulin) and AA amyloidosis (where deposits form from serum amyloid A protein). The kidney consequences are similar, even if systemic features differ.

Pathophysiology

AL amyloidosis (immunoglobulin light chain associated)

This is a plasma cell dyscrasia, related to multiple myeloma, in which clonal plasma cells in the bone marrow produce immunoglobulins that are amyloidogenic. This may be the outcome of destabilization of light chains owing to substitution of amino acids into the light chain variable region. There is a clonal dominance of amyloid light (AL) chains—either the dominant κ or γ isotype—which are excreted in the urine (as Bence Jones protein). This type of amyloid may be seen in association with other lymphoproliferative disorders, such as myeloma, Waldenström's macroglobulinaemia, or non-Hodgkin lymphoma. It rarely occurs before age 40 years.

Clinical features are related to the organs involved. As well as nephrotic syndrome, the heart may be affected (presenting with heart failure), and autonomic and sensory neuropathies are relatively common. Carpal tunnel syndrome with weakness and paraesthesia of the hands may be an early feature.

On examination, hepatomegaly and, rarely, splenomegaly, cardiomyopathy, polyneuropathy, and bruising may be seen. Macroglossia occurs in about 10% of cases and periorbital purpura in 15%.

Reactive systemic (secondary AA) amyloidosis

Serum amyloid A (SAA) is an acute-phase protein that rises in response to inflammatory stimuli in a similar way to C-reactive protein (CRP). Deposition is therefore related to chronic inflammatory disorders or chronic infection.

Clinical features depend on the nature of the underlying disorder. Chronic inflammatory disorders include rheumatoid arthritis, inflammatory bowel disease, untreated familial Mediterranean fever, and other rare autoinflammatory conditions such as cryopyrin-associated periodic syndromes, TNF receptor–associated periodic fever syndrome, mevalonate kinase deficiency/hyper-IgD and periodic fever syndrome, and deficiency of adenosine deaminase 2. In low- and middle-income countries, it is associated with infections including tuberculosis, bronchiectasis, and osteomyelitis. In high-income countries, it is seen in persons who inject drugs who may have chronic skin abscesses and sinus formation related to injection sites. AA amyloidosis often presents with nephrotic syndrome, CKD, hepatomegaly, and splenomegaly. Macroglossia is not a feature and cardiac involvement is rare.

Familial amyloidoses

These are autosomal dominant diseases where a mutant protein forms amyloid fibrils, usually starting in middle age. The most common form is due to a mutant transthyretin, leading to transthyretin-associated amyloidosis (ATTR). Transthyretin is a tetrameric protein with four identical subunits, which functions as a transport protein for thyroxine and retinol-binding protein and is mainly synthesized in the liver. Over 80 amino acid substitutions have been described which destabilize the protein, which precipitates following stimulation. Though it can cause nephrotic syndrome, kidney disease is less prevalent than with AL amyloidosis. Clinically, peripheral sensorimotor and autonomic neuropathy is common, with symptoms of autonomic dysfunction, diarrhoea, and weight loss. Cardiac problems are usually those of conduction. There may be a family history of unidentified neurological disease. Macroglossia does not occur.

Diagnosis

Diagnosis can be made clinically when features of amyloidosis are present elsewhere. On imaging, kidneys may be large. Scintigraphy with radiolabelled serum amyloid P (SAP) is a technique for quantitatively imaging amyloid deposits *in vivo* and is used to detect rate of regression or progression of amyloidosis over time.

Kidney biopsy is indicated in suspected kidney involvement. Widespread eosinophilic deposits are seen in the mesangium, capillary loops, and arteriolar walls. Deposits stain pink, with green birefringence under polarized light with Congo red (Fig. 26.13). On electron microscopy, characteristic fibrils of amyloid can be seen. Amyloid consisting of immunoglobulin light chains (AL amyloid) can be identified by immunohistochemistry in only 40% of cases, compared to almost 100% of patients with secondary amyloid (AA amyloid). If the primary presentation of amyloidosis is with non-renal disease, or if kidney histology cannot be obtained, amyloid deposits can be seen in tissues obtained from the rectum, gums, abdominal fat, or myocardium.

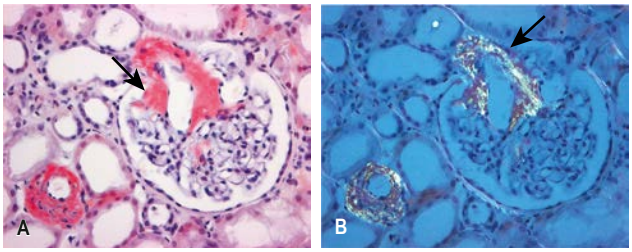


Fig. 26.13 Amyloid. **(A)** Light microscopy of eosinophilic amyloid deposits *(arrowed)*. **(B)** Congo red stain under polarized light showing apple green refringence.

Paraproteinaemia and light chains in the urine may be seen in AL amyloidosis, along with an increase in plasma cells on bone marrow biopsy. In secondary or reactive amyloidosis, there will be an underlying inflammatory disorder. Genetic testing for inherited amyloidosis should be undertaken where there is a relevant family history.

Management

Treatments that reduce production of the amyloidogenic protein can improve organ function and survival:

- AL amyloidosis: myeloma-directed therapy (e.g., bortezomib, rituximab), with or without stem cell transplantation.
- AA amyloidosis: treatment of the underlying inflammatory condition, colchicine in familial Mediterranean fever.
- ATTR amyloidosis: liver transplantation leads to complete disappearance of the mutant protein (transthyretin is predominantly synthesized in the liver) and is considered definitive therapy. However, new targeted therapies including Tafamidis and Diflunisal (tetramer stabilizers) and Patisiran (short interfering RNA inhibiting ATTR synthesis) are being used with some success.

Renoprotective measures should be started (see Box 26.10), as should relevant treatment for nephrotic syndrome and heart failure. If kidney failure progresses, the success of dialysis and kidney transplantation depends on the extent of amyloid deposition in extrarenal sites, especially the heart.

Diabetic nephropathy

Diabetic kidney disease is the leading cause of ESKD in high-income countries, arising largely as a complication of type 2 diabetes mellitus. Diabetic kidney disease occurs in about 20% to 30% of patients with both type 1 and type 2 diabetes. The natural history is similar from the onset of proteinuria, and the histological lesion is the same. Risk factors for nephropathy include poor glycaemic control, hypertension, male sex, ethnicity, and social deprivation.

Pathology

Glomerular hyperfiltration (GFR > 150 mL/min/1.73 m^2) and initial enlargement of kidney volume occur as local vasoactive factors increase flow. The GBM thickens and the mesangium expands. Progressive depletion of podocytes from the filtration barrier through apoptosis or detachment results in podocyturia early in the disease. Proteinuria evolves as filtration pressures rise and the filter is compromised. Later, glomerulosclerosis develops with nodules (Kimmelstiel–Wilson lesion) and hyaline deposits in glomerular arterioles (Fig. 26.14). Mesangial expansion and hyalinosis are partly due to amylin (β-islet-specific amyloid protein) deposits, with increasing proteinuria.

Box 26.10 Renoprotection

Goals of treatment

- Blood pressure <130/80 mmHg
- Proteinuria <0.3 g/24 h

Treatment

Patients with chronic kidney disease and proteinuria >1 g/24 h:

- Angiotensin-converting enzyme inhibitor increasing to maximum dose
- Angiotensin receptor antagonist if goals are not achieved[a]
- Finerenone
- Addition of diuretic to prevent hyperkalaemia and help to control blood pressure
- Addition of calcium-channel blocker (verapamil or diltiazem) if goals are not achieved

Patients with eGFR 20 ml/min/1.73 m^2 or more:

- SGLT-2 inhibitor
- GLP-1 agonist in type II diabetes

Additional management

- Statins to lower cholesterol to <4.5 mmol/L
- Smoking cessation (threefold higher rate of deterioration in chronic kidney disease)
- Treatment of diabetes (HbA_{1c} < 53 mmol/mol)
- Normal protein diet (0.8–1 g/kg body weight)

[a]In type 2 diabetes, start with angiotensin receptor antagonist.

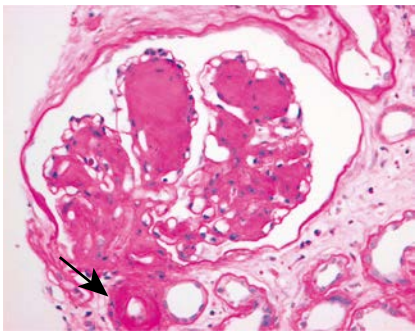


Fig. 26.14 Diabetes mellitus. Advanced diabetic glomerulopathy and arteriolar sclerosis *(arrowed)*.

The Renal Pathology Society has developed a consensus classification combining type 1 and type 2 diabetic nephropathies (Box 26.11). This discriminates lesions by various degrees of severity for international clinical practice.

Management

As in other kidney diseases, nearly the entire course of kidney disease in diabetes is clinically silent. The aim of medical intervention during this silent phase is to slow GFR decline over time (Box 26.10). Despite intensified metabolic and blood pressure control, many still go on to develop ESKD.

Management includes:

- ***Optimization of blood glucose concentrations:*** reduces the incidence of ESKD and other microvascular complications (the so-called 'legacy effect' of both type 1 and type 2 diabetes mellitus).
- ***SGLT-2 inhibitors:*** independent of benefits on glucose control, SGLT-2i (e.g., dapagliflozin, canagliflozin, empagliflozin) reduce CKD progression and reduce cardiovascular events (particularly heart failure) in those with type 2 diabetes and a GFR of ≥20 mL/min/1.73 m^2.

Box 26.11 Renal Pathology Society classification of types 1 and 2 diabetic nephropathy

Class	Name
I	• Isolated glomerular basement membrane thickening (>395 nm in females, >430 nm in males). • No mesangial expansion or mesangial matrix increase • Global glomerulosclerosis involving <50% of glomeruli
IIa	• Mild mesangial expansion
IIb	• Severe mesangial expansion (>25% of total mesangium contains areas of expansion larger than mean area of a capillary lumen)
III	• Nodular intercapillary glomerulosclerosis (≥1 Kimmelstiel–Wilson lesion) • <50% global glomerulosclerosis
IV	• >50% global glomerulosclerosis

- ***Glucagon-like peptide-1 agonists:*** independent of glucose control, GLP-1 (e.g., liraglutide, semaglutide, dulaglutide) offer renoprotection and enhanced weight loss, as well as cardiovascular protection in those with type 2 diabetes and CKD, if individualized glucose control is not optimized on other agents. Combination therapy with SGLT-2i appears to offer additive cardiovascular benefit.
- ***Blood pressure:*** <130/80 mmHg with ACE inhibitors or AII-RA.
- ***Renin-angiotensin blockade:*** ACE inhibitors or AII-RA should be used once microalbuminuria develops, even if blood pressure control is good.
- ***Treatment of dyslipidaemia.***
- ***Lifestyle:*** cessation of smoking, attention to salt intake, weight, and exercise optimization.
- ***Atrasentan:*** a selective endothelin A receptor (ET_AR) antagonist combined with renin–angiotensin system inhibitor is safe and effective in reducing albuminuria. Recent phase three trial data showed a reduced risk of renal events.

Glomerulonephritis

Glomerulonephritis (GN) is immunologically mediated, with involvement of cellular immunity (T lymphocytes, macrophages/dendritic cells), humoral immunity (antibodies, immune complexes, complement), and other inflammatory mediators (cytokines, chemokines, the coagulation cascade). The immune response can be directed against known target antigens, particularly when GN complicates infections, cancers, or drugs. The underlying antigenic target is more often unknown. Primary GN may occur in genetically susceptible individuals (associated with major histocompatibility complex (MHC) genes including *HLA-A1*, *B8*, *DR2*, and *DR3*), following environmental insults. Circulating autoantibodies and/or abnormalities in serum complement, and glomerular deposition of antibodies, immune complexes, complement, and fibrin characterize the condition. GN may present as:

- asymptomatic urinary abnormalities
- acute nephritis (nephritic syndrome)
- rapidly progressive glomerulonephritis.

The same histology can present in any of these ways, and these should be seen as clinical syndromes on a spectrum, rather than as distinct diseases.

Asymptomatic urinary abnormalities. Haematuria with or without sub–nephrotic-range proteinuria in an asymptomatic patient

may lead to the early discovery of potentially serious glomerular disease including SLE, Henoch–Schönlein purpura, post-infectious GN, and idiopathic hypercalciuria in children.

Acute nephritis (nephritic syndrome) classically presents as

- haematuria (macroscopic or microscopic), with red-cell casts on urine microscopy
- proteinuria
- hypertension
- oedema (periorbital, leg, sacral)
- temporary oliguria and uraemia.

Nephritis can present indolently or incidentally (Box 26.12). It is usually distinguished from rapidly progressive glomerulonephritis by the lack of cellular necrosis and crescent formation in the glomeruli seen on biopsy, and the rate at which kidney function declines.

Rapidly progressive glomerulonephritis is a syndrome which includes

- glomerular haematuria: red blood cell casts or dysmorphic red blood cells
- rapidly developing acute kidney failure over weeks to months
- focal glomerular necrosis with or without glomerular crescent development on kidney biopsy (Fig. 26.15).

The 'crescent' is an aggregate of macrophages and epithelial cells in Bowman's space. RPGN can develop with immune deposits (anti-GBM or immune complex type, e.g., SLE) or without immune deposits (pauci-immune, e.g., anti-proteinase 3 [PR3] or anti-myeloperoxidase–antineutrophil cytoplasmic antibodies [MPO-ANCA]-positive vasculitides). It can also develop as an idiopathic primary glomerular disease, or can be superimposed on secondary glomerular diseases such as IgA nephropathy, membranous GN, and post-infective GN. It can be classified based on the pattern of immune complex deposition in glomeruli (seen on immunofluorescence): linear, granular, and negative immunofluorescent patterns (Box 26.13).

Post-streptococcal glomerulonephritis

Acute nephritis 1 to 3 weeks after a streptococcal infection; for example, streptococcal throat infection, otitis media, cellulitis. The infecting organism is nephritogenic Lancefield group A β-haemolytic streptococcus. The latent interval between infection and the development of symptoms and signs of kidney involvement reflects the time taken for immune complex formation, deposition, and glomerular injury to occur. PSGN is now rare in high-income countries. Kidney biopsy shows diffuse, florid, acute inflammation in the glomerulus (without necrosis but occasionally with cellular crescents), with neutrophils and deposition of IgG and complement (Fig. 26.16).

Management

Includes blood pressure control, diuretics and salt restriction for oedema, and dialysis if indicated. If recovery is slow, corticosteroids may be helpful. The prognosis is usually good in children. A small number of adults develop hypertension and/or CKD. Evidence for long-term penicillin prophylaxis after the development of GN is lacking.

Non-streptococcal infective glomerulonephritis

GN occurs rarely in patients with infective endocarditis or infected ventriculoperitoneal shunts ***(shunt nephritis)***. Histological appearances resemble those of post-infectious GN but lesions are usually focal and segmental. Crescentic GN with AKI has been described, particularly with *Staphylococcus aureus* infection. Appropriate antibiotic therapy or surgical eradication of infection in fulminant cases usually results in a return of normal kidney function.

GN also occurs in association with visceral abscesses (mainly pulmonary). Complement concentrations are normal and immune deposits are absent on biopsy. Antibiotic therapy and surgical drainage of the abscess result in complete recovery of kidney function in approximately 50% of patients.

Box 26.12 Diseases commonly associated with acute nephritic syndrome

- Post-streptococcal glomerulonephritis
- Systemic lupus erythematosus (see Box 26.15)
- Henoch–Schönlein syndrome
- Non–streptococcal post-infectious glomerulonephritis, e.g., *Staphylococcus*, pneumococcus, *Legionella*, syphilis, mumps, varicella, hepatitis B and C, echovirus, Epstein–Barr virus, toxoplasmosis, malaria, schistosomiasis, trichinosis
- Infective endocarditis
- Shunt nephritis
- Visceral abscess
- Cryoglobulinaemia

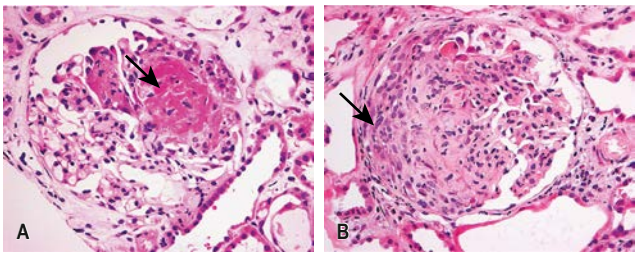


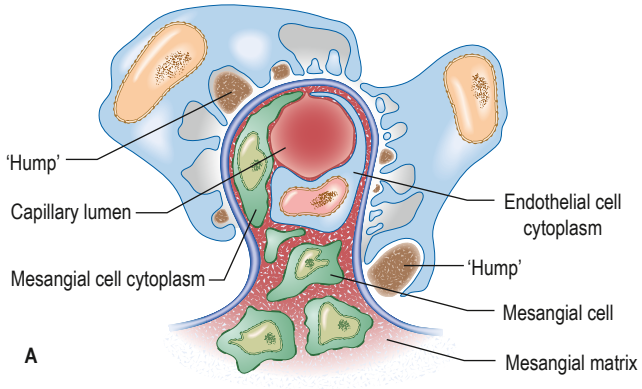
Fig. 26.15 Rapidly progressive glomerulonephritis (RPGN). *Arrows* show 'crescents' with aggregates of macrophages and epithelial cells in Bowman's space. **(A)** Focal necrotizing glomerulonephritis. **(B)** Crescentic glomerulonephritis.

Box 26.13 Types of rapidly progressive glomerulonephritis (RPGN)

Linear immunofluorescent pattern (see Fig. 26.18A)
- Anti-GBM antibody-mediated RPGN
- Goodpasture's syndrome

Granular immunofluorescent pattern (immune complex-mediated RPGN; see Fig. 26.18B)
- Immune complex-mediated RPGN
- Associated with other primary GN:
  - Mesangiocapillary GN (type II > type I)
  - IgA nephropathy
  - Membranous glomerulopathy
- Associated with secondary GN:
  - Post-infectious GN
  - Systemic lupus erythematosus
  - Henoch–Schönlein syndrome
  - Cryoglobulinaemia

Negative immunofluorescent pattern (pauci-immune RPGN)
- ANCA-associated systemic vasculitides



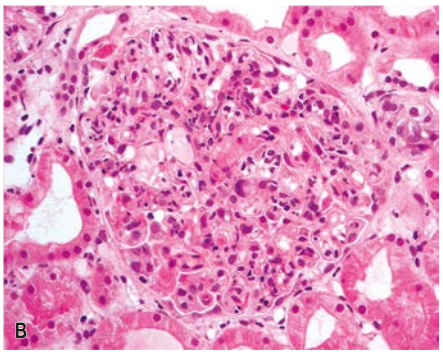


Fig. 26.16 Post-streptococcal glomerulonephritis. **(A)** Large aggregates of immune material ('humps') in the extracapillary area. There is an increase in mesangial matrix and mesangial cells, with occlusion of the capillary lumen by endothelial cell cytoplasm, leucocytes, and mesangial cell cytoplasm. **(B)** Light microscopy showing acute inflammation of the glomerulus with neutrophils. *(After Marsh FP.* Postgraduate Nephrology. *Oxford: Butterworth Heinemann; 1985.)*

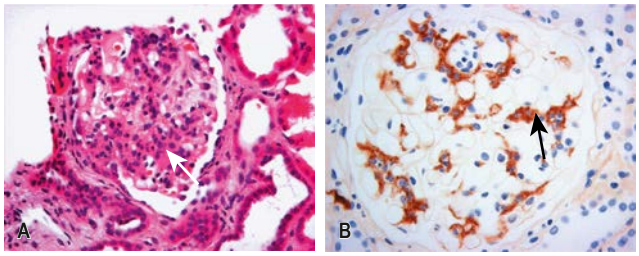


Fig. 26.17 IgA nephropathy. **(A)** Light microscopy showing mesangial cell proliferation *(arrowed)* and increased matrix. **(B)** IgA deposits *(arrowed)* on immunoperoxidase staining.

Box 26.14 Oxford histological classification of IgA nephropathy

Histological variable	Description	Class
Mesangial hypercellularity	Average mesangial hypercellularity[a] >0.5	M1
	Average mesangial hypercellularity <0.5	M0
Segmental glomerulosclerosis	Part of glomerular tuft involved in sclerosis	S1
	No segmental glomerulosclerosis	S0
Endocapillary hypercellularity	Hypercellularity present with luminal narrowing	E1
	No hypercellularity	E0
Tubular atrophy/ interstitial fibrosis	>50% cortex	T2
	26%–50% cortex	T1
	0%–25% cortex	T0

[a]Mesangial hypercellularity is scored 0 for glomeruli with <4 mesangial cells per mesangial area; 1 for those with 4–5 cells; 2 for 6–7 cells; and 3 for ≥8 cells. The mean score for all glomeruli is obtained.

IgA nephropathy

IgA nephropathy (Fig. 26.17) is the most common form of GN worldwide. Demographic and family studies support a genetic contribution to pathogenesis but candidate gene studies are inconsistent. A genome-wide analysis study in European patients showed a strong association with the MHC/DQ and HLA-B loci on chromosome 6p.

Histology

There is a focal and segmental proliferative GN with mesangial deposits of polymeric IgA1. In some cases, IgG, IgM, and C3 are also seen in the glomerular mesangium. Superimposed crescent formation is frequent, particularly following macroscopic haematuria associated with upper respiratory tract infection (synpharyngitic).

An Oxford histological classification for IgA nephropathy is shown in Box 26.14. These features have prognostic significance independent of clinical features, both at the time of presentation and during follow-up.

Pathogenesis

The disease may be a result of several events, including:

- exaggerated bone marrow and tonsillar IgA1 immune response to viral or other antigens
- an abnormality in O-linked galactosylation in the hinge region of the IgA1 molecule
- functional abnormalities of two IgA receptors: CD89 expressed on blood myeloid cells and the transferring receptor (CD71) on mesangial cells
- circulating immune complexes composed of a glycan-specific IgG and a galactose-deficient IgA1 antibody.

IgA-rich immune complexes deposited in the glomerular mesangium induce mesangial cell activation and proliferation, as well as matrix synthesis and deposition. Removal of these complexes by bacterial proteases attenuates injury, and it is thought that glycan-specific autoantibodies, rather than IgA1 itself, play a key role in pathogenesis. Up to 50% of patients exhibit elevated serum IgA (polyclonal). Several diseases are associated with IgA deposits including Henoch–Schönlein purpura, chronic liver disease, malignancies (especially bronchial carcinoma), seronegative spondyloarthritis, coeliac disease, mycosis fungoides, and psoriasis.

Clinical features

IgA nephropathy usually presents with asymptomatic microscopic haematuria, or recurrent macroscopic haematuria with upper respiratory or gastrointestinal viral infection. Proteinuria also occurs and is nephrotic range in 5% of cases. Prognosis is usually good, especially in those with normal blood pressure, normal kidney function, and no proteinuria at presentation. Surprisingly, recurrent macroscopic haematuria is a good prognostic sign, although this may be

due to 'lead-time bias' as patients with overt haematuria come to medical attention at an earlier stage of their illness. The risk of eventual development of ESKD is about 25% in those with proteinuria of more than 1 g per day, elevated serum creatinine, hypertension, ACE gene polymorphism (DD isoform), and tubulointerstitial fibrosis on kidney biopsy.

Management

- All patients should receive an ACE inhibitor or an AII-RA, to reduce proteinuria and preserve kidney function.
- SGLT-2 inhibitors have shown benefit even in the absence of diabetes mellitus and are increasingly used with ACE or AII-RA as standard of care.
- Patients with proteinuria of >1–3 g/day, mild glomerular changes, and normal kidney function may be treated with steroids aiming to reduce proteinuria and stabilize kidney function. Oral targeted released budesonide (Nefecon) for 9 months has recently been shown to be effective in reducing proteinuria and decline in kidney function.
- In patients with higher-risk progressive disease with proteinuria of >750 mg/day and or eGFR < 60 mL/min/1.73 m^2, two separate randomized placebo-controlled studies have demonstrated benefit with low-dose oral methylprednisolone or mycophenolate.
- Tonsillectomy can reduce proteinuria and haematuria in those with recurrent tonsillitis.

Alport's syndrome

Alport's syndrome is a rare hereditary nephritis with haematuria, proteinuria (usually <1–2 g/day), progressive kidney disease, and high-frequency sensorineural deafness. Approximately 15% have ocular abnormalities including bilateral anterior lenticonus and macular or perimacular retinal flecks.

About 85% have an X-linked inherited mutation in the gene encoding the COL4α5 collagen chain, a critical component of the GBM. Penetrance is variable in females and depends on the type of mutation or degree of mosaicism following hybridization of the X chromosome. The remaining 15% of cases are due to autosomal recessive or dominant inherited mutations mainly affecting COL4α3 or 4 collagen chains. These mutations cause post-translational defects in α3, α4, and α5 chains, and result in incorrect assembly or folding of monomers, which are rapidly degraded. Over time, the basement membrane undergoes selective proteolysis, and glomerular membranes thicken unevenly, split, and ultimately deteriorate. Although the basement membrane is abnormal, podocyte function and the slit diaphragm are unaffected, and so proteinuria in Alport's syndrome is often mild and is the result of glomerular sclerosis, rather than primary loss of slit pores. In some patients with Alport's syndrome, a thin basement membrane, as seen in benign familial haematuria, is the only abnormality detected on histology. For this reason, the boundary between Alport's and benign familial haematuria has become increasingly vague. Thin basement disease (thinning of the glomerular capillary basement membrane on electron microscopy) which initially was thought to be a benign condition is now regarded as part of the inherited collagen 4 disease spectrum.

Management

Patients are treated with ACE inhibitors or AII-RAs to attenuate proteinuria. The disease is usually progressive and accounts for some 5% of cases of ESKD in childhood or adolescence. A recent study of bardoxolone methyl (anti-inflammatory and antioxidant inhibitor of proinflammatory cytokines) has demonstrated preservation of eGFR.

Anti-GBM glomerulonephritis

This condition is rare, with an incidence of 1 per 2 million. It accounts for 15% to 20% of RPGN although it represents less than 5% of all forms of GN. About two-thirds of these patients have ***Goodpasture's syndrome*** with associated lung haemorrhage.

Anti-GBM antibodies (detected by enzyme-linked immunosorbent assay [ELISA]) are present in serum and directed against the non-collagenous component of α3 (IV) and rarely α5 (IV) in the basement membrane. Anti-GBM glomerulonephritis is characterized by linear (Fig. 26.18A) capillary loop staining with IgG and C3, and extensive crescent formation. Anti-GBM disease never occurs in patients with Alport's syndrome; although it can develop after patients with Alport's receive a kidney transplant as anti-GBM alloantibodies evolve in response to the 'foreign' α3, α4, α5 (IV) collagen that was absent in the patient's own kidneys. Recently laminin-521, a novel target of autoantibodies has been identified in patients with lung haemorrhage.

The thymus expresses α3 (IV) peptides that eliminate autoreactive $CD4^+$ helper T cells and the few that escape deletion are kept in check by circulating regulatory (Treg) cells. Breakdown of this peripheral tolerance (the mechanism of which is unknown) results in

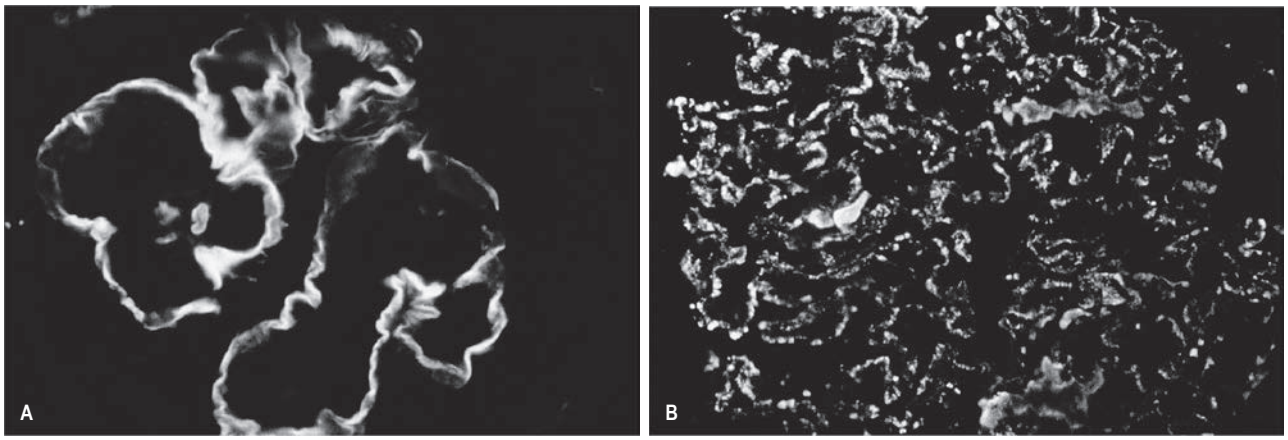


Fig. 26.18 Immunofluorescence. (A) Anti-glomerular basement membrane antibody (anti-GBM) deposition in a linear pattern typical of Goodpasture's syndrome. **(B)** Immune complex deposition in a diffuse granular pattern.

autoreactive $CD4^+$ cells producing anti-GBM antibodies. Since the $\alpha3$ (IV) non-collagenous epitope is hidden within the $\alpha3$, $\alpha4$, and $\alpha5$ (IV) promoter, it is presumed that an environmental factor, such as exposure to hydrocarbons or tobacco smoke, is required to reveal cryptic epitopes to the immune system.

The mechanism of kidney injury is complex. When anti-GBM antibody binds the basement membrane, it activates complement and proteases, causing disruption to the filtration barrier and Bowman's capsule, with proteinuria and the formation of crescents. Crescent formation is facilitated by IL-12 and interferon-gamma, which are produced by resident and infiltrating inflammatory cells.

Management involves:

- ***plasma exchange*** to remove circulating anti-GBM antibodies
- ***steroids*** to suppress inflammation from deposited antibody
- ***cyclophosphamide*** to suppress further antibody synthesis
- ***Imlidifase*** is an IgG-cleaving endopeptidase. In a small study of 15 patients with eGFR of less than 15 mL/min/1.73 m^2, the response rate was 67% compared to the 15% expected with current therapies in high-risk patients. Randomized trials are awaited.

Prognosis is determined by the extent of glomerular damage (measured by percentage of glomeruli containing crescents, serum creatinine, and need for dialysis) at treatment initiation. When oliguria occurs or serum creatinine is greater than 600 to 700 μmol/L, kidney failure is usually irreversible. Successfully treated disease does not recur. If left untreated, autoantibodies diminish spontaneously within 3 years and autoreactive T cells cannot be detected in convalescent patients. This suggests re-establishment of peripheral tolerance, which coincides with re-emergence of regulatory $FOXP3^+$-$CD25^+$ cells in the peripheral blood, which play a key role in inhibiting the autoimmune response. The emergence and persistence of these regulatory cells may underlie the 'single hit' nature of this condition.

ANCA-positive small-vessel vasculitis

Inflammation and necrosis of the blood vessel wall occur in many primary vasculitic disorders. The small-vessel vasculitides affecting the kidney include:

- ANCA-associated vasculitides:
  - granulomatosis with polyangiitis (GPA)
  - microscopic polyangiitis (MPA)
  - renal-limited vasculitis (without systemic features)
- eosinophilic granulomatosis with polyangiitis (frequently ANCA-negative).

GPA and MPA share a common pathology with focal necrotizing lesions affecting different vessels and organs:

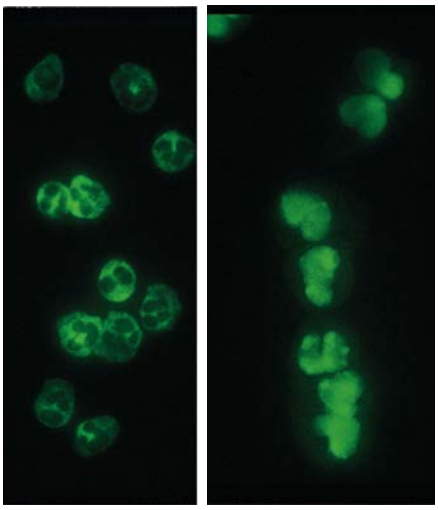
- Lungs: capillaritis causes lung haemorrhage
- Kidney: crescentic GN and/or focal necrotizing lesions causing AKI (Fig. 26.15)
- Skin (Fig. 26.19): purpuric rash or vasculitic ulceration.

Kidney histology is regarded as a 'gold standard' for diagnosis and prognosis in ANCA-associated GN. Classification is based on four categories of lesions:

- ***Focal:*** ≥50% normal glomeruli not affected by the disease process
- ***Crescentic:*** ≥50% of glomeruli with cellular crescents
- ***Mixed:*** a heterogeneous phenotype in which no glomerular feature predominates
- ***Sclerotic:*** ≥50% of glomeruli with global sclerosis

This system has been shown to have a prognostic value for 1- and 5-year kidney outcomes and may guide therapy.

Fig. 26.19 Vasculitic rash.



A B

Fig. 26.20 Antineutrophil cytoplasmic antibody (ANCA)-positive glomerulonephritis: immunofluorescence. **(A)** Cytoplasmic ANCA (cANCA). **(B)** Perinuclear ANCA (pANCA). *(From Schrier RW.* The Schrier Atlas of Diseases of the Kidney, Vol. IV: Systemic Diseases and the Kidney. *1999; Wiley–Blackwell, with permission.)*

There are two forms of ANCA: PR3-ANCA (cANCA) and MPO-ANCA (pANCA); Fig. 26.20. If ELISA and indirect immunofluorescence techniques are combined, diagnostic specificity is 99%. ANCA and anti-GBM antibodies can occur together; such patients tend to follow the natural history of anti-GBM disease.

PR3-ANCA positivity is found in the large majority (>90%) of patients with active GPA and in up to 50% of patients with MPA. ***MPO-ANCA positivity*** is present in most patients with renal-limited vasculitis and in a variable number of cases of MPA. Positivity for both types of ANCA antibodies occurs in up to 10% of patients, who have a variable clinical course and worse kidney outcome. Eosinophilic GPA may have either anti-MPO or anti-PR3-ANCA.

Drugs (e.g., propylthiouracil, hydralazine, minocycline, penicillamine) may induce vasculitis associated (most commonly) with MPO-ANCA, often in very high titres. Cases of drug-induced ANCA-associated vasculitis present with constitutional symptoms, arthralgia, and cutaneous vasculitis. Crescentic GN and lung haemorrhage also occur.

In patients with ANCA vasculitis, there is aberrant expression of PR3 and MPO in mature neutrophils, possibly due to disrupted

epigenetic silencing. Inappropriate expression of PR3 and MPO may alter the availability of these antigens by targeting these proteins to granules that are more readily exocytosed. Patients with anti-PR3 have autoantibodies to a peptide translated from the antisense DNA strand of PR3 (complementary PR3). This suggests that autoimmunity is initiated via a response against a complementary or antisense peptide, which then induces cross-reacting anti-idiotypic antibodies. Infection by fimbriated bacteria (e.g., *Escherichia coli*, *Klebsiella pneumoniae*, *Proteus mirabilis*) can, through molecular mimicry, trigger cross-reactive autoimmunity to lysosomal membrane protein 2 (LAMP-2), a membrane protein co-localized with PR3 and MPO in the intracellular vesicles of neutrophils.

Management

The sooner treatment is instituted, the greater chance there is of recovery of kidney function. Pulmonary haemorrhage and kidney failure at presentation are associated with worse prognosis.

High-dose oral corticosteroids and ***intravenous pulsed cyclophosphamide*** induce remission. Once remission has been achieved, cyclophosphamide can be substituted with azathioprine, which was more effective than mycophenolate in a head-to-head comparison trial.

Rituximab is effective in inducing remission in the short term (6–12 months). Use of rituximab every 3 months in fixed doses has also been shown to be superior to azathioprine in the maintenance of remission.

Colonization of the upper respiratory tract with *S. aureus* increases the risk of relapse, and treatment with ***sulfamethoxazole/trimethoprim*** reduces relapse.

Avacopan is an orally administered small-molecule C5a receptor antagonist that selectively blocks the effects of C5a including neutrophil chemoattraction and activation. It has recently been approved for treatment of vasculitis as a steroid-sparing agent following a positive randomized controlled in which it was found to be superior to placebo with steroid taper at 12 months. It is an exciting development for patients at risk of glucocorticoid toxicity.

Up to 25% of patients with PR3-ANCA harbour antibodies against human plasminogen and/or tissue plasminogen activator. Their presence has been associated with venous thromboembolism and fibrinoid necrotic glomerular lesions, suggesting functional interference with fibrinolysis. However, a formal role for anticoagulation in patients with ANCA-associated GN remains uncertain.

Relapse after complete withdrawal of immunosuppression occurs relatively frequently, and long-term, relatively low-dose immunosuppression may be necessary.

The PEXIVAS study of more than 700 patients with severe ANCA-positive RPGN, with or without lung haemorrhage, demonstrated the addition of plasma exchange neither saved lives nor avoided ESKD. This study also demonstrated that lower-dose oral glucocorticoids than conventionally used significantly reduced serious infection rates without less effective disease control.

Long-term survival depends on early diagnosis, severity of disease, frailty and co-morbidities, response to therapy, and the adverse effects of immunosuppressive drugs, particularly the occurrence of infections. Overall survival in ANCA-associated vasculitis is 90%, 85%, and 78% at 1, 2, and 5 years, respectively.

Mixed nephritic and nephrotic syndrome

Injury involves mesangial cells, endothelium, the basement membrane, and podocytes in this heterogeneous group of conditions.

Mesangiocapillary (membranoproliferative) glomerulonephritis

Mesangiocapillary (membranoproliferative) glomerulonephritis (MCGN) is an uncommon lesion. It has three subtypes with a similar clinical phenotype: nephrotic syndrome, haematuria, hypertension, and kidney impairment. They also produce similar microscopic findings, although pathogenesis may be different. Electron microscopy defines:

- ***Type 1 MCGN or immune complex C3 nephropathy:*** mesangial cell proliferation, with mainly subendothelial immune deposition and apparent splitting of the capillary basement membrane, giving a 'tramline' effect. It can be associated with persistently reduced plasma concentrations of C3 and normal concentrations of C4, as well as activation of the classical complement cascade. It is often idiopathic but can occur with chronic infection (abscesses, infective endocarditis, infected ventriculoperitoneal shunt) or cryoglobulinaemia secondary to hepatitis C infections (Fig. 26.21A).

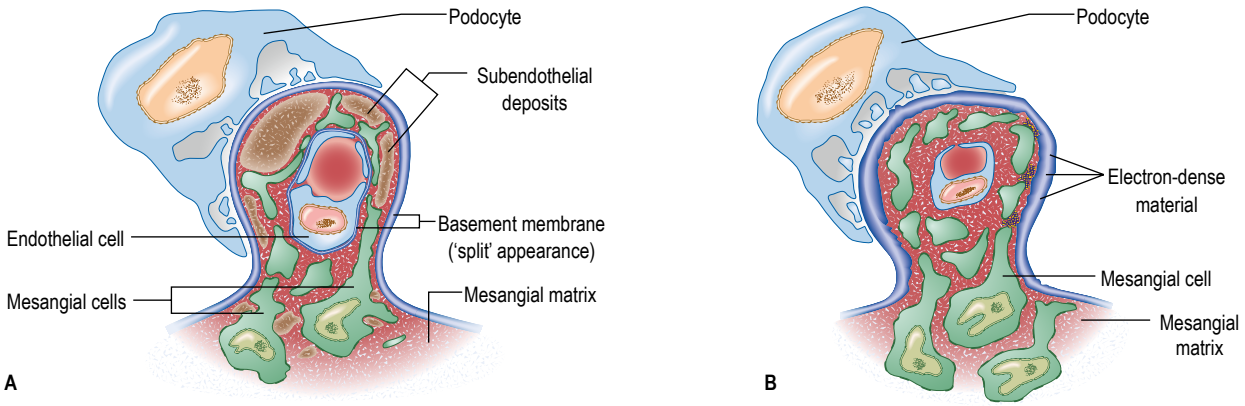


Fig. 26.21 Mesangiocapillary glomerulonephritis (MCGN). **(A)** Type 1 MCGN showing expanded mesangial matrix and mesangial cells, thickened capillary wall, large subendothelial deposits, and formation of a new layer of basement membrane (tramline effect). **(B)** Type 2 MCGN showing a variable glomerular appearance: electron-dense material has replaced Bowman's capsule, tubular basement membrane, and part of the capillary. There is proliferation of mesangial cells. *(After Marsh FP. Postgraduate Nephrology. Oxford: Butterworth–Heinemann; 1985.)*

- ***Type 2 MCGN, or C3 nephropathy:*** mesangial cell proliferation with electron-dense, linear (ribbon-like) intramembranous deposits that usually stain for C3 only (**Fig. 26.21B**). This ***dense deposit disease*** may be idiopathic or associated with factor H deficiency and partial lipodystrophy (loss of subcutaneous fat on the face and upper trunk). Alternatively, ***C3 nephropathy*** presents with low C3 concentrations (as in type 1 MCGN), due in this case to activation of the alternative complement cascade, where loss-of-function mutations in complement factor H or other factors allow uncontrolled alternative pathway activation. Autoantibodies to the C3 convertase enzyme (C3 nephritic factor) are present. Increasingly, monoclonal gammopathy is recognized with C3 nephropathy, which requires plasma cell clone depletion therapy. In patients of Cypriot origin, an autosomal dominant heterozygous duplication in the *CFHR5* gene (CFHR5 nephropathy) leads to a similar presentation.
- ***Type 3 MCGN*** has features of both type 1 and type 2 disease. Complement activation appears to be via the final common pathway of the cascade.

Most patients develop ESKD over several years. C3 nephropathy, particularly with an underlying genetic cause, recurs in a high proportion of kidney transplants, usually within the first year. This is less common (25%) in type 1 MCGN.

Management

In idiopathic MCGN with normal kidney function and non–nephrotic-range proteinuria, no specific therapy is required. Good blood pressure control, ideally with an ACE inhibitor or AII-RA, is of benefit.

In ***children*** with the nephritic syndrome and/or impaired kidney function, a trial of steroids may be warranted (alternate-day prednisolone 40 mg/m^2 for a period of 6–12 months). If no benefit is seen, this treatment is discontinued. Regular follow-up, with control of blood pressure, use of agents to reduce proteinuria, and correction of lipid abnormalities are necessary.

In ***adults*** with nephritic syndrome and/or kidney impairment, a trial of prednisolone with mycophenolate is recommended. Again, if no benefits are seen, treatment should be stopped. C3 nephropathy due to monoclonal gammopathy of renal significance requires treatment for underlying plasma cell dyscrasia. Treatment to slow the rate of progression of CKD is instituted (see 'Management of CKD' section). In C3 nephropathy, due to loss-of-function mutation in complement factor H, anti-C5 antibody (eculizumab) has been used successfully in several patients and in recurrence following transplantation. Several orally active complement B and C3 inhibitors are in trial after early promising results.

IgM nephropathy

This disorder is characterized by increased mesangial cellularity in most of the glomeruli, associated with granular immune deposits of IgM and complement. People present with episodic or persistent haematuria and the nephrotic syndrome. Unlike in minimal-change disease, the prognosis is not uniformly good, as steroid response is only 50%. Between 10% and 30% develop progressive kidney disease with evidence of secondary FSGS (see 'Secondary FSGS' section) on repeat biopsy.

C1q nephropathy

C1q nephropathy is very similar to IgM nephropathy in terms of presenting features and microscopic appearance, with the exception of C1q deposits in the mesangium. Sometimes, it is misdiagnosed as lupus nephritis, particularly in people with negative serology (so-called 'seronegative lupus'). The distinguishing features are intense C1q staining and absence of tubuloreticular inclusions (attributable to high circulating interferon-alfa) on electron microscopy.

Monoclonal gammopathy of renal significance

A monoclonal gammopathy of undetermined significance (MGUS) occurs in asymptomatic people with a paraprotein or M-band detected in blood (often incidentally). Kidney (usually glomerular) disease is increasingly recognized in association, related to deposition or precipitation of monotypic whole immunoglobulin or light chains in the kidney. Renal injury is driven by a clonal B-cell or plasma cell population.

The following are included under the term monoclonal gammopathy of renal significance (MGRS):

- ***deposition diseases:*** AL amyloid (see 'AL amyloidosis [immunoglobulin light chain associated]' section), immunotactoid and fibrillary GN, monoclonal immunoglobulin deposition diseases (MIDDs, including light chain and heavy chain DD), proliferative GN with monoclonal immunoglobulin deposition and light chain tubulopathy
- ***precipitation diseases:*** cryoglobulinaemic GN
- ***complement activation diseases:*** paraprotein-associated C3 nephropathy and atypical haemolytic uraemic syndrome (HUS, see 'Haemolytic uraemic syndrome' section).

Diagnosis is usually made in the context of a detectable paraprotein and characteristic kidney biopsy findings, and progression to ESKD is common unless treatment is given. Conventional immunosuppression is of no benefit, but clone-directed therapy with bortezomib, rituximab, or potentially daratumumab offers better outcomes.

Idiopathic fibrillary glomerulopathy

In this rare condition, microfibrillary structures are seen in the mesangium and glomerular capillary wall, visible on electron microscopy, and staining is positive for a specific marker, DNAJB9. Fibrils are larger than in amyloidosis (20–30 nm as opposed to 10 nm in diameter) and do not stain with Congo red. Patients present with proteinuria, mostly in the nephrotic range (60%), and microscopic haematuria (70%), hypertension, and CKD (50%) that may progress rapidly; 40% to 50% develop ESKD within 2 to 6 years. No treatment is known to be of benefit.

Immunotactoid glomerulopathy

In this disorder, microtubules that are much larger (30–40 nm in diameter) than the fibrils in fibrillary glomerulopathy are seen on electron microscopy. Most patients have circulating paraproteins, or monoclonal immunoglobulin deposition is seen in the glomeruli on immunofluorescence microscopy. A lymphoproliferative disease is the underlying cause in over 50% of cases. Clinical presentation and course are similar to fibrillary glomerulopathy. Complete or partial remission of the nephrotic syndrome can be achieved with chemotherapy in 80% of patients.

Fibronectin glomerulopathy

Glomerulonephritis due to fibrillar deposits is negative for Congo red staining. It is inherited as an autosomal dominant disorder and is associated with the massive deposition of fibronectin, a large dimeric glycoprotein consisting of two similar subunits (approximately 250 kDa in weight). The possible genetic abnormality in this disorder is a loss-of-function mutation in uteroglobin. It presents

with varying degrees of proteinuria seen first between the ages of 20 and 40, followed by hypertension, microscopic haematuria, and slow progression to ESKD in most patients.

Lysozyme nephropathy

A rare and under-recognized complication of chronic myelomonocytic leukaemia, other monocytic leukaemias, sarcoidosis, tuberculosis, and inflammatory bowel disease. Lysozyme, a small bactericidal cationic protein, is released into the circulation, filtered by the glomerulus, and reabsorbed by the proximal tubule, causing toxic tubular injury. Lysozyme concentrations in the blood are very high in all cases. The predominant abnormalities involve proximal tubular epithelial cells, which exhibit widespread degenerative changes including luminal ectasia, cytoplasmic simplification and vacuolization, irregular luminal profiles, loss of brush border, enlarged nuclei with prominent nucleoli, and focal apoptosis. Patient presents with AKI or CKD with superimposed AKI, and partial Fanconi syndrome. Early recognition and treatment of underlying cause result in satisfactory renal recovery.

Systemic lupus erythematosus (lupus nephritis)

Overt kidney disease occurs in at least one-third of patients with systemic lupus erythematosus (SLE) and 25% of those with lupus nephritis reach ESKD within 10 years. **Box 26.15** shows clinical and histological findings in class I to VI disease.

Serial biopsies show that in approximately 25% of patients, histological appearances change from one class to another. Immune deposits in the glomeruli and mesangium are characteristic of SLE (tubuloreticular structure in glomerular endothelial cells) and stain positive for IgG, IgM, IgA, and the complement components C3, C1q, and C4 on immunofluorescence.

Pathophysiology

SLE is a multifactorial, autoantigen-driven, T-cell dependent, B-cell mediated autoimmune disorder:

- Multiple circulating autoantibodies to cellular antigens (particularly anti-dsDNA, anti-Ro)
- Complement activation, which leads to reduced serum concentrations of C3, C4, and C1q. C1q is the first component of the classical complement cascade (see Chapter 6) involved in activation of complement and clearance of self-antigens generated during apoptosis.
- Nucleosomes (structures comprising DNA and histone, generated during apoptosis) are thought to be antigenic-generating nucleosome-specific T cells, antinucleosome antibodies, and nephritogenic immune complexes.
- Positively charged histone components of the nucleosome bind to the negatively charged heparan sulphate of the GBM, leading to an inflammatory reaction.
- T-helper 2 (Th2) cells, with basophils and B-cell survival factors (such as BLyS), enhance B-cell differentiation and survival, and stimulated nucleic acid–binding receptors promoting production of autoreactive antibodies.
- Type I and type II IFN-stimulated genes are up-regulated in the majority of adult SLE.
- Inflammation stimulates complement activation, mesangial cell proliferation, mesangial matrix expansion, and recruitment of inflammatory leucocytes.

Extraglomerular features of lupus nephritis include tubulointerstitial nephritis (in 75%), renal vein thrombosis, and renal artery stenosis. Thrombotic manifestations are associated with autoantibodies to phospholipids (lupus anticoagulant, anticardiolipin, beta-2 glycoprotein 1, see Chapter 40), and may present as infarction of glomerular segments, thrombotic microangiopathy, or vasculitis.

Management

Management depends on the clinical presentation. Hypertension should be treated. Disease activity, histology, and extrarenal manifestations guide therapy. ***Class I lupus nephritis*** requires no specific treatment. ***Class II*** usually runs a benign course, but some patients are treated with hydroxychloroquine and/or steroids alone. There have been several clinical trials with immunosuppressive agents in ***classes III, IV, and V***. Outcomes are affected by clinical characteristics, initial response to treatment, frequency of kidney flares, and irreversible damage on kidney biopsy. Steroids used

Box 26.15 Classification of lupus nephritis

Class	Histology	Clinical
I	***Minimal mesangial lupus nephritis:*** Immune deposits but normal light microscopy	Asymptomatic
II	***Mesangial proliferative lupus nephritis:*** Mesangial hypercellularity, matrix expansion	Mild CKD
III	***Focal lupus nephritis:*** Affects <50% of glomeruli Subepithelial deposits Subdivisions for active or chronic lesions	10%–20% of lupus nephritis Haematuria and proteinuria
IV	***Diffuse lupus nephritis:*** Affects >50% of glomeruli Segmental and global lesions Active and chronic lesions Subendothelial deposits	Most common form Severe kidney disease: progression to nephrotic syndrome, hypertension, and kidney failure
V	***Membranous lupus nephritis:*** Affects >50% of glomeruli Subepithelial deposits, C1q present Mesangial and small subendothelial deposits	10%–20% of lupus nephritis. Proteinuria and nephrotic syndrome. Can occur in combination with class III or IV
VI	***Advanced sclerosing lupus nephritis:*** ≥90% globally sclerosed glomeruli	Progression or healing following other stages Progressive CKD: no benefit to immunosuppression

(Modified from International Society of Nephrology/Renal Pathological Society 2004.)

for induction include intravenous methylprednisolone, followed by tapering oral prednisolone. Steroids are combined with one of the following:

- Mycophenolic acid analogues, e.g., MMF. Shown to be as effective as high-dose intravenous cyclophosphamide
- Voclosporin (plus MMF). A novel calcineurin inhibitor. Randomized control trial data demonstrated a fixed dose of 23.7 mg twice daily in combination with MMF and low-dose steroids led to a measurably better renal response compared to MMF and low-dose steroids alone, with a comparable safety profile. Unlike other calcineurin inhibitors, it does not require drug concentration monitoring.
- Low-dose cyclophosphamide.
- Belimumab (plus MMF or low-dose cyclophosphamide). A recombinant human IgG-1λ monoclonal antibody that inhibits B-cell activating factor (BAFF). Two randomized controlled trials have demonstrated belimumab combined with standard treatment is associated with an improved renal response. Recent immune profiling has shown that high baseline IgA2 anti-dsDNA antibodies are associated with a better treatment response to belimumab (after rituximab), with high baseline IL-6 concentrations conferring a poor response. Molecular prediction markers could guide precision-targeted therapies in the future.

Patients with patients with persistent disease activity or inadequate response to initial standard-of-care therapy are considered for B-cell depletion. Rituximab is a chimeric type 1 anti-CD20 antibody which causes B-depletion by complement-dependent cytotoxicity. Inconsistency in controlled trial data may be due to variable B-lymphocyte depletion. Obinutuzumab is a humanized type II anti-CD20 monoclonal antibody with direct B lymphocyte toxicity, which induces potent B-cell depletion and is well tolerated.

Chimeric antigen receptor T-cell (CAR-T cell) therapy is emerging as a potential treatment option in refractory disease. CAR-T cell therapy directed against CD19 (an early-stage B-lymphocyte surface marker) has been used successfully in an exploratory study of five patients with SLE patients refractory to other therapies, with drug-free remission after one year of follow-up, even after reappearance of B cells. CAR T-cell treatment was well tolerated with only mild cytokine-release syndrome.

Optimum duration of maintenance immunosuppression treatment is unclear. A randomized study of discontinuation of immunosuppression after 2 to 3 years of remission demonstrated a 25% relapse rate within a year in the absence of maintenance immunosuppression, compared to 12% when immunosuppression was continued. Although this study was unable to prove non-inferiority, 75% were able to discontinue treatment without relapse. Better molecular and histological markers are needed to identify patients with lupus nephritis who can safely discontinue immunosuppression.

Prognosis

Treatment leading to the normalization of proteinuria, hypertension, and kidney dysfunction indicates a good prognosis. The prognosis is better in patients with types I, II, and V disease. Glomerulosclerosis (type VI) usually predicts ESKD.

Cryoglobulinaemic kidney disease

Cryoglobulins (CGs) are individual or mixed immunoglobulins and complement components, which precipitate reversibly in the cold. Three types are recognized:

- ***Type I:*** monoclonal immunoglobulin, as is found in multiple myeloma and lymphoproliferative disorders.
- ***Type II:*** mixed monoclonal IgM and polyclonal IgG component. Most often, the monoclonal IgM is directed against the Fc portion of the polyclonal IgG ('rheumatoid factor'). Typically found in hepatitis C infections (80%–90% of cases) but also other infections and autoimmune disease.
- ***Type III:*** polyclonal IgM and IgG components in autoimmune disorders or infection, including hepatitis C.

Glomerular disease is more common in type II than in type III cryoglobulinaemia. In approximately 30% of these 'mixed' cryoglobulinaemias, no underlying or associated disease is found (essential cryoglobulinaemia). Glomerular pathological changes resemble those of MCGN (see Fig. 26.21).

Presentation is usually in the fourth or fifth decade of life. Systemic features include purpura, arthralgia, leg ulcers, Raynaud's phenomenon, evidence of systemic vasculitis, a polyneuropathy, and hepatic involvement. Glomerular disease typically presents as asymptomatic proteinuria, microscopic haematuria (or both), but presentation with an acute nephritic and nephrotic syndrome (most commonly) or features of CKD also occurs. Complement is consumed and cryoglobulins can be detected. Protein electrophoresis, rheumatoid factor, autoantibodies, and antiviral antibodies or mRNA of hepatitis C should be sought.

The underlying cause should be treated. As hepatitis C is the most common cause, antivirals are highly efficacious. Intensive plasma exchange or cryofiltration has been used in selected cases where vasculitis is limb-, organ- or life-threatening to achieve rapid removal of cryoglobulins. Uncontrolled studies of the anti-CD20 chimeric monoclonal antibody rituximab, which depletes B cells, appear promising.

Immunoglobulin A vasculitis (Henoch–Schönlein syndrome)

This clinical syndrome comprises a characteristic skin rash, abdominal colic, joint pain, and glomerulonephritis. Approximately 30% to 70% have clinical evidence of kidney disease with haematuria and/or proteinuria. The kidney disease is usually mild, but nephrotic syndrome and AKI can occur. The lesion is a focal segmental proliferative glomerulonephritis, sometimes with mesangial hypercellularity. In more severe cases, epithelial crescents may be present. Immunoglobulin deposition is mainly IgA in the glomerular mesangium distribution, similar to IgA nephropathy. There is no proven treatment; steroid therapy is ineffective for kidney involvement and management is usually supportive. In crescentic glomerulonephritis, aggressive immunosuppression has been tried with variable outcome.

Other glomerular disorders

Fabry's disease

This X-linked lysosomal storage disease results from a deficiency of the enzyme α-galactosidase. Glycosphingolipids (globotriaosylceramide) accumulate in many cells including podocytes in the kidney. Systemic features include angiokeratomas in the skin, cardiovascular disease, neuropathies, and progressive proteinuric CKD. Enzyme replacement with agalsidase (α-galactosidase) slows progression and injury to the heart, kidney, and skin. In certain cases, migalastat, an oral molecular chaperone, is effective.

Sickle cell nephropathy

Sickle cell disease is complicated relatively commonly by papillary sclerosis or necrosis, AVP-resistance (also referred to as 'nephrogenic diabetes insipidus'), and incomplete renal tubular acidosis. Glomerular lesions are rare and were historically traced to hepatitis B or C infection acquired through repeated blood transfusions. Occasionally, proteinuria or nephrotic syndrome with progressive

kidney disease is seen, particularly in *APOL1* G1/G2 gene variant. Rarely, membranous GN or mesangiocapillary glomerulonephritis with IgG deposits occurs. No form of effective therapy is known.

Glomerulopathy associated with pre-eclampsia

Presents with new hypertension and proteinuria after 20 weeks' gestation, often of rapid onset, and resolves after delivery. The glomerular lesion is characterized by marked endothelial swelling and obliteration of capillary lumina. In severe cases, there may be cortical necrosis or microangiopathic haemolysis. The disease is caused by abnormal placental development, which leads to an imbalance of pro- and anti-angiogenic factors, which are measurable in the maternal circulation prior to development of the clinical phenotype. An increase in soluble fms-like tyrosine kinase (sFlt1) antagonizes angiogenic factors including placental growth factor (PlGF) leading to systemic endothelial dysfunction. Concentrations of sFlt1 and PLGF can be used for prognosis and diagnosis. Delivery of the placenta is the only definitive treatment, the timing of which balances fetal maturity with maternal disease. Persistent proteinuria 6 months after delivery raises the possibility of undiagnosed CKD as a precipitant of pre-eclampsia, and warrants investigation.

Paraneoplastic glomerulonephritis

A rare complication of malignancy, paraneoplastic GN is usually misdiagnosed as idiopathic. Several cancers involve the glomerulus:

- ***Thymoma or Hodgkin lymphoma:*** polarization of the immune response towards a Th2 profile and possibly excessive production of IL-13 leads to the development of minimal change disease, mesangiocapillary GN, or membranous nephropathy.
- ***B-cell lymphoma and leukaemia:*** can induce injury via monoclonal immunoglobulin and cryoglobulin.
- ***Polycythaemia vera, essential thrombocythaemia, or primary myelofibrosis:*** severe thrombocytosis may induce FSGS, possibly due to elevated concentrations of platelet-derived growth factor.
- ***Myelodysplastic syndromes:*** autoimmunity can trigger a variety of glomerulonephritides.
- ***Epithelial carcinoma:*** glomerular inflammatory cells and subepithelial immune IgG1- and IgG2-containing complexes may aid the diagnosis of paraneoplastic membranous nephropathy.

Further reading

KDIGO 2024 Clinical practice guideline for the management of lupus nephritis. *Kidney Int* 2024; 105(Suppl 1S):S1–S69. https://kdigo.org/wp-content/uploads/2024/01/KDIGO_2024_Lupus_Nephritis_Guideline.pdf.
Kidney Disease: Improving Global Outcomes (KDIGO) Glomerular Diseases Work Group. KDIGO 2021 clinical practice guideline for the management of glomerular diseases. *Kidney Int* 2021; 100:S1.
Noura Alforaih and others. A review of lupus nephritis. *J Appl Lab Med* 2022; 7(6): 1450–1467.
Walters GD, Willis NS, Cooper TE, Craig JC. Interventions for renal vasculitis in adults. *Cochrane Database Syst Rev* 2020; 1(1):CD003232.
Watts RA Hatemi G, Burns JC et al. Global epidemiology of vasculitis. *Nat Rev Rheumatol* 2022; 18:22–34.

KIDNEY INVOLVEMENT IN OTHER DISEASES

Polyarteritis nodosa

Polyarteritis nodosa (PAN) is a multisystem disorder (see also Chapter 32, 'Polyarteritis nodosa' section). Aneurysmal dilation of medium-sized arteries presents with hypertension, polyneuropathy, and ischaemic infarction of a number of organs including skin, gut, heart, and brain. Aneurysms may be seen on arteriography. PAN may be associated with drug use and hepatitis B infection. Typically, patients are ANCA-negative. PAN is typically with slowly progressive CKD, often accompanied by severe hypertension. Rapidly progressive kidney failure is rare. Treatment with immunosuppression is less effective than for microscopic polyangiitis.

Systemic sclerosis (scleroderma)

Systemic sclerosis (SSc) (see Chapter 32, 'Systemic sclerosis [scleroderma]' section) is a chronic multisystem disease characterized by fibrosis and vasculopathy of the skin and visceral organs. Plasmacytoid dendritic cells appear to produce high plasma concentrations of CXCL4, which is associated with both skin and lung fibrosis, as well as with pulmonary arterial hypertension.

Approximately 10% of SSc patients develop scleroderma renal crisis, characterized by accelerated hypertension, rapidly progressive kidney failure, and proteinuria. 'Onion skinning' in arcuate and interlobular arteries is seen on biopsy resulting from vessel intimal proliferation, fibrin thrombi, and fibrinoid necrosis. The treatment of choice is ACE inhibitors, which have led to a remarkable improvement in outcomes, though up to 30% of patients progress to ESKD.

Haemolytic uraemic syndrome

Haemolytic uraemic syndrome (HUS) is characterized by intravascular haemolysis with red cell fragmentation (microangiopathic haemolysis), thrombocytopenia, and AKI due to thrombosis in small arteries and arterioles (**Fig. 26.22**). These features are also seen in disseminated intravascular coagulation, but coagulation tests are typically normal in HUS.

Diarrhoea-associated HUS

Diarrhoea-associated HUS (D+ HUS) often follows a febrile illness, particularly gastroenteritis associated with *E. coli*, notably strain O157. The latter produces verocytotoxin (or Shiga toxin), with a pathogenic A unit that inhibits protein synthesis and initiates endothelial damage. B units facilitate entry of the A unit into the endothelial cells by binding to a receptor (Gb3) on the endothelial cell. Toxins are transported to and into endothelial cells from the gut on neutrophils. Most patients with D+ HUS recover kidney function but commonly require supportive care, including fluid and electrolyte balance, blood pressure control, nutritional support, and dialysis. Plasmapheresis is not beneficial. About 30% develop CKD with persistent proteinuria, and 5% develop ESKD.

Atypical HUS

Atypical HUS (aHUS) is a complement-driven illness linked to abnormalities in complement regulatory proteins leading to unopposed complement activation and development of HUS. Factor

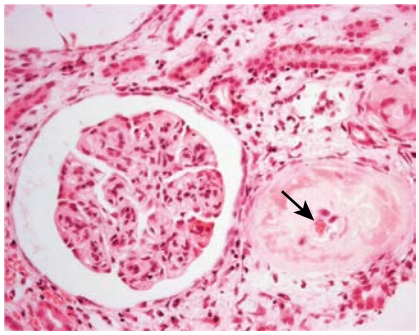


Fig. 26.22 Light microscopy of a typical haemolytic uraemic syndrome (HUS) kidney lesion. The *arrow* shows microthrombi.

H is a soluble protein produced by the liver, which regulates the activity of the alternative complement activation pathway; in particular, it protects host cell surfaces from complement-mediated damage. Factor I is a serine proteinase that plays a key role in inhibition of the alternative pathway amplification loop, which generates C3 convertase from C3b. CD46 (previously known as membrane co-factor protein, MCP) is highly expressed in the kidney and normally prevents glomerular C3 activation. Thrombomodulin has anti-inflammatory properties and plays a role in the inactivation of complement-derived anaphylatoxins. Deficiency of these factors can occur as an isolated phenomenon or due to autoantibody formation. Rarely, gain-of-function mutations can affect genes encoding the alternative pathway C3 convertase components, CFB and C3.

Severe hypertension and recurrence are frequent. The course of the disease is often indolent and progressive.

Management

Plasmapheresis or plasma infusion may be used at initiation of therapy, until thrombotic thrombocytopenic purpura (TTP) has been excluded and the diagnosis of aHUS is clear.

C5 activation is one of the critical steps in the activation of complement cascade. ***Eculizumab***, a monoclonal humanized anti-C5 antibody, has transformed outcomes, both at acute presentation and in prevention of recurrence in native kidneys and in transplants with 5-year ESKD survival improved from 40% in a control cohort to 86% in an eculizumab treated cohort. Biallelic pathogenic mutations in RNA processing genes, including EXOSC3, lead to eculizumab non-responsive aHUS, allowing targeted treatment to those who can benefit. Ravulizumab is an re-engineered form of eculizumab with an extended terminal elimination half-life allowing administration every 6 to 8 weeks (2–4 weeks with eculizumab). Liver transplantation is a potentially curative treatment in patients harbouring CFH and CFI mutations.

Sporadic cases of aHUS

These can be associated with pregnancy, SLE, scleroderma, malignant hypertension, metastatic cancer, HIV infection, and various drugs including oral contraceptives, ciclosporin, tacrolimus, chemotherapeutic agents (e.g., cisplatin, mitomycin C, bleomycin), and heparin. Management is supportive, with removal of the offending agent or specific treatment of the underlying cause. There is no evidence in favour of plasma infusion or plasmapheresis.

Streptococcus pneumoniae produces an enzyme (possibly neuraminidase) that can expose an antigen (Thomsen antigen) present on red blood cells, platelets, and glomeruli. Antibodies to the antigen result in an antigen–antibody reaction that can lead to HUS and anaemia. Awareness of this complication, judicious use of (washed) blood products, and avoidance of plasma infusion and plasmapheresis are warranted.

Cobalamin C disease is a hereditary disorder of vitamin B_{12} metabolism that may cause HUS and multiple organ damage in infants and, rarely, adults. It is caused by mutations in a gene encoding the methylmalonic aciduria and homocystinuria type C protein *(MMACHC)*. The resulting deficiency in methylcobalamin causes hyperhomocysteinaemia, decreased plasma methionine concentrations, and methylmalonic aciduria. Abnormal cobalamin C metabolism is associated with platelet activation, generation of reactive oxygen species, endothelial dysfunction, increased tissue factor expression, coagulation activation, and HUS. Parenteral hydroxocobalamin is the principal treatment for infants.

Thrombotic thrombocytopenic purpura

TTP (see Chapter 38, 'Thrombotic thrombocytopenic purpura' section) is characterized by microangiopathic haemolysis, neurological disturbance, and impaired kidney function. Distinction from aHUS requires measurement of serum activity of *ADAMTS13*. *ADAMTS13* activity is reduced in TTP, but normal in HUS and other conditions that cause thrombotic microangiopathy.

Antiphospholipid syndrome

Binding of antiphospholipid antibodies (aPL) to beta 2 glycoprotein I (β2GPI) induces endothelial cell–leucocyte adhesion and thrombus formation through the inhibition of endothelial nitric oxide synthase (eNOS). Antibodies may be primary, or secondary to infection (HIV, hepatitis C) or autoimmune disease (SLE).

A key clinical feature of antiphospholipid syndrome (APS) is recurrent thrombosis (both venous and arterial) in the presence of antiphospholipid antibodies (see Chapter 32, 'Antiphospholipid syndrome' section). Thrombotic microangiopathy is a rare but recognized. In some cases, a lupus nephritis-like lesion (usually mesangiocapillary GN) is seen. The only proven treatment for APS is systemic anticoagulation with warfarin as direct oral anticoagulants have inferior efficacy and increased bleeding risk. Use of steroids or plasmapheresis is reserved for patients with APS and life-threatening thrombotic microangiopathy.

Multiple myeloma

AKI is relatively common in myeloma, occurring in 20% to 30% of affected individuals at the time of diagnosis. It is often irreversible. Kidney involvement may present as:

- ***Light chain cast nephropathy:*** intratubular deposition of light chains, particularly kappa chains, which characteristically appear on histology as fractured casts with a giant cell reaction (Fig. 26.23)
- ***AL amyloidosis:*** deposition of amyloid fibrils of light chains (Congo-red positive)
- ***Light chain deposition disease:*** nodular glomerulosclerosis with granular deposits of usually lambda light chains (Congo red-negative)
- ***C3 nephropathy:*** a light chain variant of proliferative GN with monotypic immunoglobulin. Glomerular co-deposition of C3 is found in addition to monotypic light chain, implying complement activation via the alternative pathway
- ***Fanconi's syndrome:*** tubular toxicity due to light chains
- ***Hypercalcaemic nephropathy:*** bone resorption causing hypercalcaemia

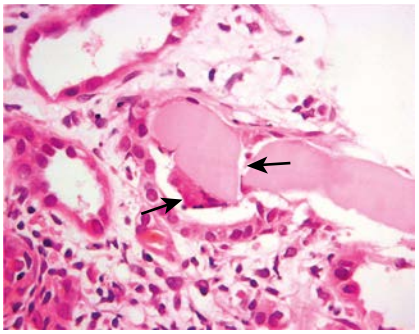


Fig. 26.23 Cast nephropathy in multiple myeloma. Light microscopy showing a characteristic fractured cast and giant cell reaction *(arrowed)*.

- ***Hyperuricaemic nephropathy:*** tumour lysis causing tubular crystallization of uric acid.

Treatment of the underlying myeloma is key to recovery (see Chapter 39, 'Myeloma' section). If a patient with cast nephropathy and severe AKI remains dialysis-dependent, prognosis is poor. Commencement of effective bortezomib-based chemotherapy decreases light chain production. Plasma exchange and high-cut-off haemodialysis are ineffective.

Further reading

Bu F, Zhang Y, Wang K et al. Genetic analysis of 400 patients refines understanding and implicates a new gene in atypical hemolytic uremic syndrome. *J Am Soc Nephrol* 2018; 29:2809–2819.
Saha M, McDaniel JK, Zheng XL. Thrombotic thrombocytopenic purpura: pathogenesis, diagnosis and potential novel therapeutics. *J Thromb Haemost* 2017; 15:1889–1900.
Syed YY. Ravulizumab: A review in Atypical Haemolytic Uraemic Syndrome. *Drugs* 2021; 81(5):587–594.

HYPERTENSION AND THE KIDNEY

Hypertension can be the cause or the result of kidney disease. It is often difficult to differentiate between the two on clinical grounds. Kidney function testing should be performed on all patients with hypertension, and kidney imaging may be indicated.

The ***mechanisms*** responsible for regulation of arterial blood pressure and the development of essential primary hypertension are unclear. Long-term regulation of arterial pressure is closely linked to the ability of the kidneys to excrete sufficient salt to maintain normal sodium balance, extracellular fluid volume, and normal blood volume at normotensive arterial pressures. Cross-transplantation experiments suggest that hypertension travels with the kidney, as hypertension develops in the normotensive recipient of a kidney genetically programmed for hypertension. Similarly, patients with ESKD due to hypertension become normotensive after receiving a kidney transplant from normotensive donors, provided the new kidney functions well.

One factor contributing to hypertension is the total number of nephrons per kidney ('nephron dose'). People with hypertension and normal kidney function have a significantly reduced number of nephrons in each kidney, and individual nephrons are enlarged because of glomerular hyperfiltration. Increased glomerular volume (a surrogate marker for reduced nephron number) is found in populations where hypertension is more common. Whether reduced nephron number is genetic or environmental in origin is unclear. Changes in the intrauterine environment may lead to reduced kidney volume, suggesting lower nephron number before birth, and hypertension in adult life.

Excess renal sympathetic activity may also contribute. Strategies to achieve bilateral renal sympathetic denervation were effective in patients with refractory hypertension in uncontrolled studies; to date, this benefit has not been borne out in controlled trial data.

Essential hypertension

Arterial pressure is a product of cardiac output and systemic vascular resistance (SVR). The compliance of systemic arteries is a critical component: 'stiff' (sclerotic) arteries cannot modulate pressure surges in systole, and systolic pressures rise. Over time, vessels (and the kidney) remodel:

- Small vessels and arterioles: intimal thickening with reduplication of the internal elastic lamina, hyalinization of the vessel wall
- Large vessels: concentric reduplication of the internal elastic lamina and endothelial proliferation produce an 'onion skin' appearance
- Kidney size: reduction in size of both kidneys. This may be asymmetrical if one major renal artery is more affected than the other
- Glomeruli: an increased proportion of sclerotic (scarred) glomeruli compared to age-matched controls.

Kidney function deteriorates with these changes though the risk of progression to advanced CKD varies with ethnicity. There is a fourfold higher incidence of ESKD due to hypertensive kidney disease in those of black African descent compared to white ethnicity. This difference may be due to overestimation of diagnosis, a higher incidence of hypertension (usually salt-sensitive), reduced nephron number, and/or a higher frequency of susceptibility alleles for ESKD. A genome-wide study found measurable associations between two sequence variants in the apolipoprotein L1 gene *(APOL1)* and non-diabetic nephropathy in African Americans with hypertension-attributed ESKD. These kidney disease risk variants are thought to have been positivity selected for evolutionary advantage in trypanosomal infection and protection from African sleeping sickness. Natural selection might therefore protect from one disease but allow another to develop.

In malignant hypertension arteriolar fibrinoid necrosis occurs, probably as a result of plasma entering the media of the vessel through splits in the intima. It is prominent in afferent glomerular arterioles. Fibrin deposition within small vessels is often associated with thrombocytopenia and red cell fragmentation on a peripheral blood film (microangiopathic haemolytic anaemia). Microscopic haematuria, proteinuria (usually modest, 1–3 g daily), and progressive CKD occur. If untreated, fewer than 10% of patients survive 2 years.

Management

The management of hypertension is described in Chapter 18. If treatment is effective before CKD has developed, the prognosis for kidney function is good, despite an initial fall in eGFR due to haemodynamic effects. Stabilization or improvement in kidney function, with healing of intrarenal arteriolar lesions and resolution of microangiopathic haemolysis, occurs with effective treatment of malignant hypertension.

Renal hypertension

Hypertension commonly complicates bilateral kidney disease such as chronic glomerulonephritis, reflux nephropathy, polycystic disease, and analgesic nephropathy. Two main mechanisms are responsible:

- activation of the renin–angiotensin–aldosterone system
- retention of salt and water as excretory function declines, leading to an increase in blood volume and blood pressure.

Hypertension occurs earlier, is more common, and tends to be more severe in patients with glomerular disease (GN) compared to tubulointerstitial diseases such as reflux or analgesic nephropathy.

Meticulous control of blood pressure may prevent ongoing or new vascular or parenchymal changes, and further deterioration of kidney function. There is good evidence that ACE inhibitors and angiotensin receptor antagonists offer additional renoprotective benefits for the same degree of blood pressure control as other agents, particularly when there is coexisting proteinuria.

RENOVASCULAR DISEASE

A fall in perfusion pressures in the afferent glomerular arterioles leads to increased production and release of renin from the juxtaglomerular apparatus. This causes an increase in angiotensin II, a potent vasoconstrictor, which also triggers aldosterone secretion, stimulates thirst, and leads to vascular smooth muscle hypertrophy and fibrosis in the kidney.

In renal artery stenosis, narrowing of the renal artery causes a fall in kidney perfusion pressure beyond the stenosis leading to angiotensin release, and increased salt and water. GFR is decreased. Urine from the ischaemic kidney is more concentrated, with a lower sodium concentration than urine from an unaffected kidney.

Renal artery stenosis is caused by one of two pathologies: ***atherosclerotic renovascular disease*** or ***fibromuscular dysplasia***.

Atherosclerotic renovascular disease

Atherosclerotic renovascular disease (ARVD) is a common cause of hypertension and CKD due to ischaemic nephropathy. Incidence increases with age:

- 5% in those under 60 years of age
- 16% in those over 60 years of age

Most commonly, the atherosclerotic lesion is ostial (within 1 cm of the origin of the renal artery) and usually associated with atherosclerotic vascular disease elsewhere. Patients with peripheral vascular disease (39%), coronary artery disease (10%–29%), congestive cardiac failure (34%), and aortic aneurysm (38%) have a high risk of significant renal artery stenosis. Many patients are asymptomatic and are discovered incidentally during investigation for other conditions. Aortography experience from the USA shows that 11% of asymptomatic patients have significant unilateral stenosis and 4% have bilateral disease. ARVD results in hypertension (present in 50%), sodium retention (ankle and flash pulmonary oedema), proteinuria (usually subnephrotic), and decreased GFR. Beyond the stenosis, affected kidneys demonstrate histological changes including vascular sclerosis, tubular atrophy, interstitial fibrosis with inflammatory cellular infiltrate, atubular glomeruli, cholesterol emboli, and secondary FSGS. These changes are often termed 'ischaemic nephropathy'. The affected kidney loses volume (as fibrosis replaces normal tissue) and becomes smaller on ultrasound measurement. These morphological changes are responsible for the fall in GFR whereas the arterial stenosis drives hypertension.

Renovascular disease should be considered in:

- patients with hypertension and/or CKD with abdominal or other audible bruits
- patients with kidney asymmetry, particularly where ultrasound suggests >1.5 cm difference in length
- recurrent flash pulmonary oedema without cardiopulmonary disease
- progressive CKD in patients with evidence of generalized atherosclerosis.

Screening

- ***Radionuclide studies:*** decreased kidney perfusion on the affected side (see 'Renal scintigraphy' section). In unilateral stenosis, a disproportionate fall in isotope uptake occurs on the affected side after administration of captopril. A normal result makes renovascular disease unlikely.
- ***Doppler ultrasound:*** higher renal artery velocity may mean a higher pressure differential across a stenosis. Operator dependency has led to a decline in use (in native kidneys) in clinical practice. Doppler also measures intrarenal vascular resistance, which can be valuable in predicting revascularization success. A resistive index of >80 is a predictor of poor response following intervention.
- ***MRA:*** used to visualize the renal arteries with good correlation between MRA findings and renal arteriography.
- ***CT:*** less expensive than MRA but less reliable. Also exposes patient to ionizing radiation and contrast injection.
- ***Renal arteriography:*** used to confirm the diagnosis and allow therapeutic intervention.

Management

The ***aims of treatment*** are to prevent decline in kidney function and to reverse salt and water overload. Most patients with ARVD will die a cardiovascular death and modifiable risk factors should be addressed. All patients with ARVD should be managed with a combination of aspirin, statin and optimal control of blood pressure, ideally with an ACE inhibitor or AII-RA. This is the paradox of ARVD: angiotensin drives kidney and cardiac events, so ACE inhibition is ideal medical therapy. However, because ACE inhibitors will cause a fall (often significant) in kidney perfusion beyond the stenosis, ACE inhibitors may cause AKI if there is bilateral, haemodynamically significant ARVD. Renal artery stenosis can progress to occlusion, particularly in patients with stenosis of more than 75%, as shown by serial angiography.

Revascularization offers definitive treatment of stenotic lesions in ARVD. Indications for revascularization are vessels with stenosis of >75% and recurrent flash pulmonary oedema, drug-resistant severe hypertension, ARVD affecting a solitary functioning kidney, patients with cardiac failure in whom ACE inhibition causes an unacceptable decline in kidney function, unexplained progressive CKD and dialysis-dependency. Options include transluminal angioplasty, stent insertion (sometimes the only endoscopic option when stenosis occurs close to the origin of the renal artery, making angioplasty technically difficult or impossible), reconstructive vascular surgery, and nephrectomy.

In several trials, renal artery stenting did not confer a significant benefit with respect to prevention of clinical events (cardiovascular events, progressive kidney dysfunction, or the need for RRT) when added to comprehensive medical therapy in people with atherosclerotic renal artery stenosis and hypertension or CKD. Up to 30% of renal artery interventions are complicated by atheroembolism, which may lead to further deterioration in kidney function.

Prognosis

Mortality is high from associated co-morbidities, and endothelial dysfunction. ARVD patients with ESKD have higher death rates than those with other causes of ESKD. Five-year survival is only 18% in patients with ESKD due to ARVD.

Fibromuscular dysplasia of the renal arteries

Fibromuscular dysplasia is often asymptomatic, typically occurring in younger females with hypertension. It is far less common than ARVD. FMD affects many arterial beds, including carotid and coronary arteries. Kidney function is often normal.

MRA (gadolinium-enhanced) reveals a characteristic string of beads appearance in the mid-renal artery (rather than the ostium in ARVD).

Angioplasty (occasionally with stent insertion) or, rarely surgery, was historically offered achieving cure (blood pressure <140/90 without treatment) rates of only 36% and 54%, respectively, in a study of over 2000 patients with blood pressure strongly related to

age. In addition, procedural complications were common: 12% after angioplasty and 17% after surgery. Current medical antihypertensive therapy is very effective and angioplasty is not usually performed.

Further reading

Chang AR, Lóser M, Malhotra R et al. Blood pressure goals in patients with chronic kidney disease: a review of evidence and guidelines. *Clin J Am Soc Nephrol* 2019; 14:161–169.
Hicks CW, Clark TWI, Cooper CJ et al. Atherosclerotic renovascular disease: A KDIGO (Kidney Disease: Improving Global Outcomes) Controversies Conference. *Am J Kidney Dis* 2022; 79:289–301.
Malhotra R, Craven T, Ambrosius WT et al. SPRINT Research Group. Effects of intensive blood pressure lowering on kidney tubule injury in CKD: a longitudinal subgroup analysis in SPRINT. *Am J Kidney Dis* 2019; 73:21–30.
Sol Carriazo and others. Blood pressure targets in CKD 2021: the never-ending guidelines debacle, *Clinical Kidney Journal* 2022:15;845–851.

Renal artery occlusion

This occurs from thrombosis *in situ*, usually in a damaged atherosclerotic vessel, or more commonly from embolization, for example, in atrial fibrillation. Leads to partial (wedge) or whole kidney infarction. Occlusion of a small branch artery may produce no symptoms but larger vessel occlusion can cause dull flank pain, macroscopic haematuria, and CKD. Intra-arterial thrombolytic therapy has been tried with mixed results.

Cholesterol embolization (atheroembolic kidney disease)

Showers of cholesterol-rich atheromatous material from ulcerated plaques reach the kidney from the aorta and/or renal arteries, particularly after catheterization of the abdominal aorta or attempts at renal artery angioplasty. Anticoagulants and thrombolytic agents also precipitate non-traumatic cholesterol embolization. Kidney failure from cholesterol emboli may be acute or slowly progressive. Clinical features include fever, eosinophilia, back and abdominal pain, and evidence of embolization elsewhere, such as to the retina or digits. Atheroembolic kidney disease occurs spontaneously in 25% of cases. It is more common in males, >70 years, and patients with cardiovascular disease. Over 80% have abnormal kidney function. Diagnosis can be confirmed by kidney biopsy (**Fig. 26.24**). The 2-year mortality is 30% and a similar percentage develop CKD. Acute/subacute presentations and co-morbidities including CKD, diabetes, heart failure, and gastrointestinal tract involvement are significant predictors of event occurrence. The risk of dialysis and death is 50% lower among those treated with statins.

Renal vein thrombosis

This is usually of insidious onset. It occurs in patients with an increased risk of venous thrombosis including nephrotic syndrome, renal cell carcinoma, and thrombophilia. Anticoagulation is indicated.

KIDNEY CALCULI AND NEPHROCALCINOSIS

Kidney and vesical calculi

Nephrolithiasis (kidney stones) usually occurs in the upper urinary tract. It is very common worldwide, with a lifetime risk of about 10%. Most stones are composed of calcium oxalate and phosphate (**Box 26.16**). Mixed infective stones, which account for about 15% of all calculi, are twice as common in females compared to males though the overall male to female ratio of stone disease is 2:1. Stone disease is frequently recurrent: more than 50% of patients will experience a second stone within 10 years of their first. The risk of recurrence increases if a metabolic or structural abnormality predisposes to stone formation and is not modified by treatment. Nephrolithiasis is not a benign condition; observational studies show an association with an increased risk of ESKD, bone diseases, hypertension, and myocardial infarction.

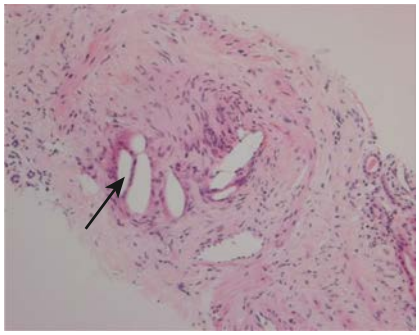


Fig. 26.24 Kidney cholesterol emboli showing characteristic cholesterol biconvex clefts *(arrowed)*.

Box 26.16 Type and frequency of kidney stones in the UK

Type of kidney stone	Percentage of stones
Calcium oxalate usually with calcium phosphate	65
Calcium phosphate alone	15
Magnesium ammonium phosphate (struvite)	10–15
Uric acid	3–5
Cystine	1–2

Aetiology

Stones form because solute concentrations exceed saturation, often in the context of a trigger or nidus to crystallize around. Inhibitors of crystal formation prevent stone formation in normal urine, despite the concentrations of stone-forming substances exceeding maximum solubility. Factors predisposing to stone formation are:

- ***Chemical composition of urine*** that favours stone crystallization.
- ***A concentrated urine*** resulting from dehydration (e.g., living/working in a hot climate or environment).
- ***Impairment of inhibitors that prevent crystallization*** in normal urine including inorganic magnesium, pyrophosphate, and citrate. Organic inhibitors include glycosaminoglycans and nephrocalcin (an acidic protein of tubular origin). Tamm–Horsfall protein may have a dual role in both inhibiting and promoting stone formation; it usually inhibits crystallization of urinary oxalate but in an undersialylated form promotes stone formation.

Recognized risk factors for stone formation are listed in **Box 26.17**.

Stones may be single or multiple, and vary enormously in size from minute, sand-like particles to staghorn calculi or large stone concretions in the bladder. They may be located within the kidney parenchyma or within the collecting system. Stones regularly cause obstruction, leading to hydronephrosis. A combination of obstruction and infection often causes lasting damage to the kidney.

Box 26.17 Risk factors for urinary tract stones

- Dehydration
- Hypercalcaemia
- Hypercalciuria
- Hyperoxaluria
- Hyperuricaemia and hyperuricosuria
- Infection
- Cystinuria
- Primary kidney disease (polycystic kidneys, medullary sponge kidneys, renal tubular acidosis)
- Drugs (see text)

Hypercalcaemia

If the GFR is normal, hypercalcaemia usually leads to hypercalciuria. Common causes of hypercalcaemia leading to stone formation are:

- primary hyperparathyroidism
- vitamin D ingestion
- sarcoidosis.

Hypercalciuria

This is the most common metabolic abnormality in calcium stone-formers. Approximately 8% of males excrete more than 7.5 mmol of calcium per day and calcium stone formation is more common in this group. In females, a 24-hour calcium excretion of more than 6.25 mmol makes stone formation more likely. About 90% of ionized calcium filtered by the kidney is reabsorbed, controlled largely by PTH. Causes of hypercalciuria include:

- hypercalcaemia
- high dietary calcium intake
- excessive resorption of calcium from the skeleton, e.g., prolonged immobilization or weightlessness
- idiopathic hypercalciuria: accelerated calcium transport, enhanced gut uptake, and increased urine excretion are thought to play a role.

Hyperoxaluria

Three autosomal recessive inborn errors of glyoxylate metabolism cause increased endogenous oxalate biosynthesis and calcium oxalate stone formation:

- ***Type 1:*** alanine–glyoxylate aminotransferase deficiency
- ***Type 2:*** glyoxylate reductase hydroxypyruvate reductase deficiency
- ***Type 3:*** 4-hydroxy-2-oxoglutarate aldolase 1 deficiency.

Widespread calcium oxalate crystal deposition in the kidneys leads to CKD in late teens or early twenties. Liver transplantation cures the metabolic defect.

More common causes of mild hyperoxaluria include:

- excess ingestion of oxalate-rich food, e.g., spinach, rhubarb, tea
- dietary calcium restriction with compensatory increased absorption of oxalate
- gastrointestinal disease (e.g., Crohn's), usually with intestinal resection, leads to increased colonic absorption of oxalate. Dehydration secondary to fluid loss from the gut may contribute.

Hyperuricaemia and hyperuricosuria

Uric acid stones account for 3% to 5% of all stones in the UK, but up to 40% in Israel. Uric acid is the end-point of purine metabolism. Hyperuricaemia increases uric acid excretion and stones can develop. Hyperuricaemia can occur as a primary defect, for example, idiopathic gout, or secondary to increased cell turnover, for example, myeloproliferative disease. Hyperuricosuria (>4 mmol/24 h on a low-purine diet) can occur without hyperuricaemia. Dehydration can also lead to uric acid stone formation. Patients with ileostomies are at particular risk due to a combination of dehydration in conjunction with gastrointestinal bicarbonate loss, which produces an acidic urine in which uric acid is less soluble. Some patients with calcium stones have hyperuricaemia and/or hyperuricosuria, with precipitation of calcium salts on an initial nidus of uric acid hypothesized.

Infective stones

Mixed infective stones are composed of magnesium ammonium phosphate (struvite) with variable amounts of calcium. Such stones are often large, forming a cast of the collecting system (staghorn calculus). Urinary tract infection with organisms that hydrolyze urea (e.g., *Proteus mirabilias*) leads to the formation of the strong base ammonium hydroxide. The availability of ammonium ions and the alkalinity of the urine favour stone formation. An increased production of mucoprotein from infection creates an organic matrix on which stone formation can occur.

Cystinuria

Cystinuria results in the formation of cystine stones (about 1%–2% of all stones).

Primary kidney disease

- ***Polycystic kidney disease*** (see 'Autosomal dominant polycystic kidney disease' section) has a high prevalence of stone disease.
- ***Medullary sponge kidney*** leads to dilation of collecting ducts with urinary stasis and calcification (Fig. 26.25). Approximately 20% have hypercalciuria and a similar proportion have a tubular acidification defect.
- ***Renal tubular acidosis***, either inherited and acquired, is associated with nephrocalcinosis and stone formation. Persistently alkaline urine and reduced urinary citrate excretion lead to stone formation, as there is less citrate to form a soluble complex with calcium in urine.

Drugs

Drugs can promote calcium stone formation (e.g., loop diuretics, antacids, glucocorticoids, theophylline, vitamins D and C, acetazolamide), uric acid stone formation (e.g., thiazides, salicylates), or precipitate into stones (e.g., indinavir, triamterene, sulfadiazine).

Other

Calcified papillae may mimic calculi, so causes of papillary necrosis such as analgesia misuse should be considered.

Clinical features

Most people with urinary tract calculi are asymptomatic. When symptoms exist (Box 26.18), pain is the most common symptom. It may be sharp or dull, constant, intermittent, or colicky. Classically, pain radiates from the flank to the iliac fossa, testis, or labia (in the distribution of the first lumbar nerve root). Ureteric colic occurs when a stone enters the ureter and either obstructs it or causes spasm during its passage down the ureter. Pallor, sweating and vomiting occur, and the patient is restless, trying to obtain relief from the pain. Fluids or diuretics (including alcohol) make the pain worse as peristaltic flow increases. Exertion may cause mobile calculi to move, precipitating pain and haematuria. Untreated, the pain of ureteric colic typically subsides after a few hours. When urinary tract

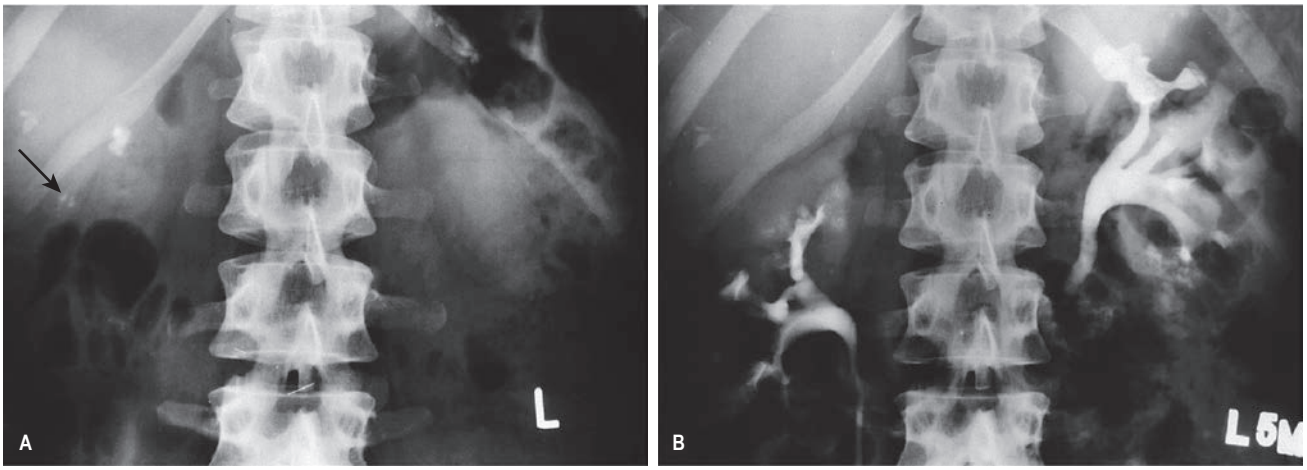


Fig. 26.25 Medullary sponge kidney. **(A)** Plain film showing 'spotty' calcification in the kidney *(arrowed)*. **(B)** Contrast reveals the calcification to be small calculi in the papillary zones.

Box 26.18 Clinical features of urinary tract stones

- Asymptomatic
- Pain, renal colic
- Haematuria
- Urinary tract infection
- Urinary tract obstruction

obstruction and infection are present, features of acute pyelonephritis or Gram-negative septicaemia may dominate the clinical picture.

Bladder stones

Bladder stones occur where stasis (e.g., bladder outflow obstruction), infection, and a nidus for stone formation (e.g., migratory stone, thrombus, foreign body) occur together. They are usually associated with bacteriuria and present with frequency, dysuria, and haematuria. Vaginal or perineal pain may occur if trigonitis is present. A calculus at the bladder neck or an obstruction in the urethra may cause bladder outflow obstruction, resulting in anuria and painful bladder distension.

Investigations

- ***Dipstick:*** test for red cells, protein, and glucose.
- ***Chemical analysis:*** of passed stones.
- ***Mid-stream specimen of urine:*** microscopy (crystals) and culture.
- ***Serum urea, electrolyte, creatinine (eGFR), and calcium concentrations.***
- ***Ultrasonography:*** kidney stones and kidney pelvis dilation (ureteric stones can be missed).
- ***Computed tomography of kidneys, ureters, and bladder (CT-KUB):*** sensitivity of >95% for detection of stones, see **Figs. 26.26 and 26.27**. CT has largely replaced plain radiographs and excretion urography.
- ***Excretion urography/intravenous pyelography:*** largely replaced with CT. Pure uric acid stones are radiolucent and show as filling defects. Mixed infective stones in which organic matrix predominates are barely radio-opaque.

Management

NSAIDs (e.g., diclofenac 50–75 mg i.m.) compare favourably with opiates (e.g., diamorphine, pethidine). Stones smaller than 0.5 cm diameter usually pass spontaneously, particularly if hydration is maintained (patients that can drink should aim for over 2.5 L intake/24 h). Alpha-blockers (e.g., tamsulosin) were thought to help expulsion of distal ureteral stones smaller than 6 mm (alpha receptors are predominantly present in the distal ureter and detrusor) but a randomized controlled trial has shown no benefit. Larger stones usually need urological or radiological intervention:

- Extracorporeal shock-wave lithotripsy (ESWL) fragments most stones, which then pass spontaneously
- Ureteroscopy with a YAG laser
- Percutaneous nephrolithotomy.

Investigation of the cause of stone formation

- ***Kidney imaging:*** for structural abnormalities of the urinary tract.
- ***Urine culture:*** for significant bacteriuria.
- ***Chemical analysis:*** of any passed stone (required for diagnosis of cystinuria or uric acid stone).
- ***Serum calcium:*** with investigation for cause if hypercalcaemia present (e.g., hyperparathyroidism, malignancy).
- ***Serum urate concentration:*** often elevated in uric acid stone-formers.
- ***Screening for cystinuria:*** a purple colour can be seen when sodium nitroprusside is added to an unacidified urine sample if cystinuria is present. Urine chromatography is required for definitive diagnosis.
- ***Urinary calcium, oxalate, and uric acid concentrations*** in two consecutive 24-hour urine samples. Aliquots are withdrawn for estimation of uric acid, after which the urine is acidified to prevent crystallization of calcium salts on the walls of the container, which would give falsely low results for urinary calcium and oxalate.
- ***Plasma bicarbonate:*** low in renal tubular acidosis. A urine pH that does not fall below 5.5 in metabolic acidosis is diagnostic (see Chapter 27, 'Renal tubular acidosis' section).

Prevention of recurrent stones

Idiopathic stone formers:

- High fluid intake: a daily urine volume of 2 to 2.5 L reduces saturation of solute.
- Reduction of animal protein intake to less than 50 g/day. Ingested protein is a source of purine, produces an acid load, and increases urinary calcium and oxalate excretion.

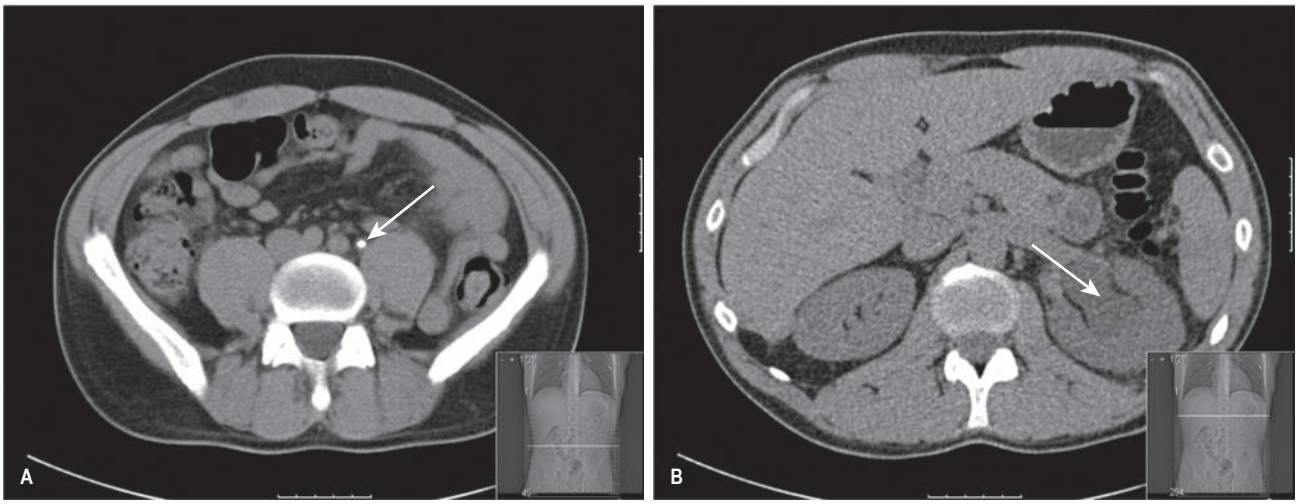


Fig. 26.26 Computed tomography of kidneys, ureters, and bladder (CT-KUB) in ureteric stone obstruction. **(A)** Left ureteric calculus *(arrowed)*. **(B)** A dilated renal pelvis *(arrowed)* proximal to the ureteric stone in A.

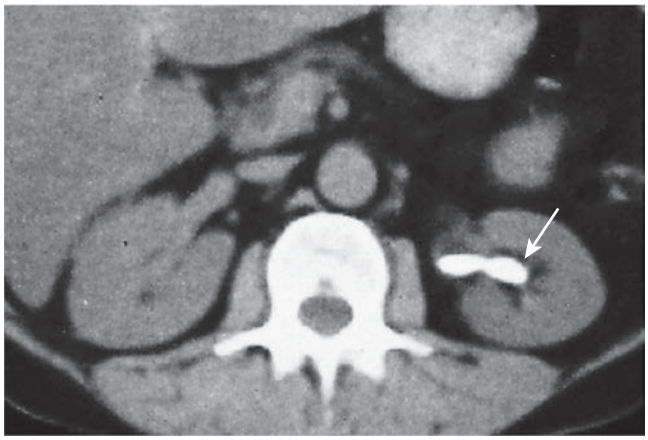


Fig. 26.27 CT scan showing a uric acid stone. The stone appears as a bright lesion in the left kidney *(arrowed)*.

Hypercalciuria:

- Dietary calcium restriction is inappropriate as restriction results in hyperabsorption of oxalate. Foods rich in oxalate should be avoided (nuts, spinach, chocolate, rhubarb). Patients in hard-water areas may benefit from drinking softened water.
- Limit salt intake to 50 mmol/day. Calcium excretion increases by 0.6 mmol/day for every 100 mmol Na^+ in the urine.
- Thiazide diuretic (e.g., bendroflumethiazide 2.5–5 mg) is hypothesized to reduce urinary calcium excretion due to mild volume contraction increasing proximal tubule calcium absorption. However, no benefit demonstrated in a randomized controlled trial.

Infective stones:

- Low-dose prophylactic antibiotics may prevent bacteriuria.

Uric acid stones:

- Allopurinol or febuxostat to maintain serum urate and urinary uric acid excretion in the normal range.
- Bicarbonate supplementation alkalinizes the urine making uric acid more soluble. However, alkalinization also facilitates precipitation of calcium oxalate and phosphate and so it is used with care.

Cystine stones:

- Can be prevented and will dissolve slowly with a high fluid intake, for example, 5 L of water in 24 hours. Patients also need to wake twice during the night to drink 500 mL or more of water. Understandably, many patients cannot tolerate this regimen. Alkalinization to a urine pH of 7 requires high doses of potassium citrate or bicarbonate. An alternative option is long-term use of penicillamine, which converts cystine to a more soluble penicillamine–cysteine complex. Side-effects include drug rashes, blood dyscrasias, and immune complex-mediated glomerulonephritis. Other cystine-binding drugs include captopril and tiopronin.

Hyperoxaluria:

- Lumasiran reduces concentrations of glycolate oxidase by targeting hydroxyacid oxidase 1 mRNA in hepatocytes through RNA interference. It is now available as first-line therapy for type 1 primary hyperoxaluria. Oral high-dose pyridoxine can also be used for type 1 disease (no therapeutic response in type 2). Unfortunately, there is currently no proven pharmacotherapy for 'idiopathic' hyperoxaluria present in up to 40% of stone-formers. Probiotic *Oxalobacter formigenes* has shown some promise.

Nephrocalcinosis

'Nephrocalcinosis' describes diffuse parenchymal calcification that is detectable radiologically (**Fig. 26.28**). The condition is typically painless. Hypertension and CKD commonly occur. The main causes of nephrocalcinosis are listed in **Box 26.19**.

Dystrophic calcification occurs following cortical necrosis. In ***hypercalcaemia and hyperoxaluria***, deposition of calcium oxalate results from the high concentrations of calcium and oxalate within the kidney. In ***renal tubular acidosis*** (see Chapter 27, 'Renal tubular acidosis' section), failure of urinary acidification favours calcium and oxalate precipitation in alkali urine, and a reduction in urinary citrate excretion reduces calcium chelation.

Management and prevention of nephrocalcinosis consist of treating the cause.

Further reading

Campschroer T, Zhu X, Vernooij RWM, Lock MTWT. Alpha-blockers as medical expulsive therapy for ureteral stones. *Cochrane Database Syst Rev* 2018; (4):CD008509. Accessed 27 August 2023.

Garrelfs SF, Frishberg Y, Hulton SA, et al. Lumasiran, an RNAi therapeutic for primary hyperoxaluria type 1. *N Engl J Med* 2021; 384 (13):1216–1226.

Zisman AL. Effectiveness of treatment modalities on kidney stone recurrence. *Clin J Am Soc Nephrol* 2017; 12:1699–1708.

Fig. 26.28 Radiograph of nephrocalcinosis.

Box 26.19 Causes of nephrocalcinosis

Mainly cortical (rare)
- Cortical necrosis (tramline calcification)

Mainly medullary
- Hypercalcaemia (primary hyperparathyroidism, hypervitaminosis D, sarcoidosis)
- Renal tubular acidosis (inherited and acquired)
- Primary hyperoxaluria
- Medullary sponge kidney
- Tuberculosis

URINARY TRACT OBSTRUCTION

Urinary tract obstruction can occur at any point between the kidney and the urethral meatus. Obstruction can be partial or complete, and leads to delayed transit of urine, rising urinary tract (and intrarenal) pressure, and—if obstructing both kidneys—eventual kidney impairment.

Aetiology

Obstructing lesions may lie within the lumen, in the wall of the urinary tract or outside the wall causing obstruction by external pressure. Obstruction causing dilation of the renal pelvis and/or ureter is termed hydronephrosis. Causes of obstruction are shown in **Box 26.20**. In children, obstruction usually results from congenital abnormalities, for example, urethral valves. Pelvic tumours, pregnancy, and stones cause obstruction in females. In males, stones are the dominant cause in young adults, but bladder outflow obstruction from prostatic disease dominates later in life.

Pathophysiology

Urine continues to be formed despite obstruction to flow. This leads to:
- progressive rise in intraluminal pressure
- dilation proximal to the site of obstruction
- compression and thinning of the renal parenchyma, eventually reducing it to a thin rim and a decrease in kidney size.

Box 26.20 Causes of urinary tract obstruction

Within the lumen
- Calculus
- Blood clot
- Sloughed papilla: diabetes, analgesia misuse, sickle cell disease or trait
- Tumour of renal pelvis, ureter, or bladder

Within the wall
- Pelviureteric neuromuscular dysfunction: congenital, 10% bilateral
- Ureteric stricture: tuberculosis (especially after treatment), calculus, after surgery
- Ureterovesical stricture: congenital, ureterocele, calculus, schistosomiasis
- Congenital megaureter
- Congenital bladder neck obstruction
- Neuropathic bladder
- Urethral stricture: calculus, gonococcal, after instrumentation
- Congenital urethral valve
- Pin-hole meatus

Pressure from outside
- Pelviureteric compression: bands, aberrant vessels
- Tumours: retroperitoneal tumour, colon, pelvis, ovarian, cervix, malignant lymphadenopathy
- Diverticulitis
- Aortic aneurysm
- Retroperitoneal fibrosis: peri-aortitis, e.g., IgG4 disease
- Accidental ligation of ureter
- Retrocaval ureter: right-sided obstruction
- Prostatic obstruction
- Phimosis

Acute obstruction leads to transient renal arterial vasodilation, followed by vasoconstriction probably mediated mainly by angiotensin II and thromboxane A_2. Ischaemic interstitial damage mediated by free oxygen radicals and inflammatory cytokines compounds the damage due to pressure.

Clinical features

Symptoms

Loin pain occurs, which can be dull or sharp, and constant or intermittent. Pain can be worse if urine flow and volume increase as the collecting system distends. A high fluid intake or diuretics, including alcohol and coffee, may provoke pain. Anuria is strongly suggestive of complete bilateral obstruction or obstruction of a single kidney. Polyuria may occur in partial obstruction, as rising pressures impair tubular concentrating capacity. Infection is a major complication and may cause malaise, fever, and signs and symptoms of sepsis.

Acute bladder retention causes significant discomfort and distress. It may present as confusion and agitation in the elderly. Chronic bladder outflow obstruction can occur with few symptoms, though if due to prostate disease, symptoms of prostatism (hesitancy, poor stream, terminal dribbling, and a sense of incomplete bladder emptying) may be present. Infection is common (and may precipitate acute retention) with symptoms including frequency, urgency, urge incontinence, dysuria, and offensive urine.

Signs

On abdominal palpation, loin tenderness may be present. Occasionally an enlarged hydronephrotic kidney is palpable. In bladder retention, the enlarged bladder can be felt or percussed. Examination of the genitalia, prostate, and rectum are warranted as prostatic obstruction and pelvic malignancy are common causes of urinary tract obstruction.

Investigations

- ***Urinalysis:*** for haematuria.
- ***Mid-stream sample of urine:*** for infection.
- ***Kidney function:*** serum potassium, urea, and creatinine to assess for kidney injury.
- ***Haemoglobin:*** to assess for anaemia of chronic disease.
- ***Prostate-specific antigen*** (see Chapter 47, 'Benign conditions of the prostate' section).
- ***Ultrasonography*** is the investigation of choice to examine for upper urinary tract dilation. Stones in the ureter can be missed.
- ***Plain abdominal radiography*** may detect radiolucent stones/calcification but can miss radiolucent stones and stones lying over bone.
- ***CT scanning*** has a high sensitivity, can visualize radiolucent stones as small as 1 mm, and delineates the anatomy of any obstruction, local mass lesions, or lymphadenopathy.
- ***Excretion urography*** is now seldom used. A characteristic delayed nephrogram is seen on the obstructed side (owing to a reduction in GFR). With time, the nephrogram on the affected side becomes denser than normal, owing to the prolonged nephron transit time, and the site of obstruction with proximal dilation is seen (**Fig. 26.29**).
- ***Radionuclide studies*** (see 'Renal scintigraphy' section) have a role in chronic obstruction to differentiate true obstructive nephropathy from retention of tracer in a baggy, low-pressure, unobstructed pelvicalyceal system.
- ***Antegrade pyelography and ureterography*** define site and cause of obstruction.
- ***Retrograde ureterography***. Relief of ureteric obstruction from below may be possible at the same time. In obstruction due to neuromuscular dysfunction at the pelviureteric junction or retroperitoneal fibrosis, the collecting system may fill normally from below.
- ***Cystoscopy and urethroscopy*** can visualize obstructing lesions within the bladder and urethra directly.
- ***Urethrography*** involves introducing contrast medium into the bladder by catheterization or suprapubic bladder puncture and taking radiographs during voiding. It may be of value in the diagnosis of urethral valves.

Management

Treatment aims to:

- relieve the obstruction
- treat the underlying cause
- prevent and treat infection
- preserve kidney function.

An obstructed and infected urinary tract is a medical emergency and delay in treatment can lead to septicaemia. It should be managed with:

- Bladder catheterization for rapid relief of bladder outflow obstruction. If a urinary catheter cannot be passed urethrally, a suprapubic catheter should be placed.
- Image-guided percutaneous nephrostomy for obstructed and hydronephrotic kidneys.
- Cystoscopic (retrograde) stenting provided this can be done without delay to treatment.

In contrast, in partial urinary tract obstruction, particularly if spontaneous relief is expected—such as by passage of a calculus—immediate treatment may not be indicated.

Surgical management depends on the cause of the obstruction. Dialysis may be indicated for obstruction causing kidney failure.

Fig. 26.29 Intravenous urography 24 hours after injection of contrast. A delayed nephrogram and pyelogram are shown on the left side, with urinary tract dilation proximal to an obstruction. Within the same time period, contrast medium has disappeared from the normal right side.

Post-obstructive diuresis

This occurs after relief of obstruction at any site in the urinary tract. Massive diuresis may occur following relief of bilateral obstruction due to a combination of prior sodium and water retention, the osmotic effect of retained solutes, and defective tubular reabsorptive capacity. This diuresis is associated with increased blood volume and high concentrations of atrial natriuretic peptide (ANP). The diuresis is usually self-limiting, but a minority of patients develop severe sodium, water, and potassium depletion needing intravenous replacement. In milder cases, oral salt and potassium supplements, together with increased oral water intake, are sufficient.

Calculi

See 'Kidney calculi and nephrocalcinosis' section.

Pelviureteric junction obstruction

A functional disturbance in peristalsis of the collecting system, without mechanical obstruction, causing pelvic hydronephrosis (**Fig. 26.30**). In patients with recurrent loin pain or progressive kidney damage, surgical correction by open or percutaneous pyeloplasty is indicated.

Obstructive megaureter

This childhood condition (more common in males) may become evident only in adult life. It results from a region of defective peristalsis at the lower end of the ureter adjacent to the ureterovesical junction.

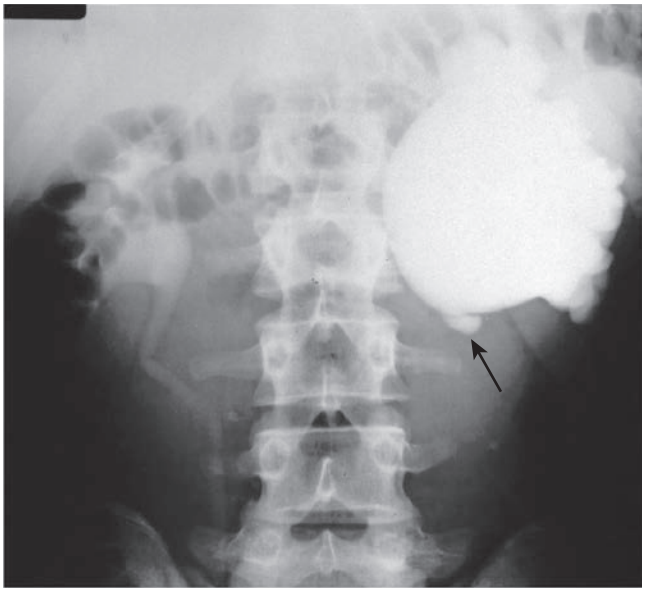


Fig. 26.30 Radiograph showing left pelviureteric junction obstruction *(arrowed)*.

It presents with urinary tract infection, flank pain, or haematuria. The diagnosis is made on imaging with ultrasound, CT, or, if necessary, ascending ureterography.

Excision of the abnormal portion of ureter with re-implantation into the bladder indicated in progressive deterioration in kidney function, bacteriuria that cannot be controlled by medical means, and recurrent stone formation.

Retroperitoneal fibrosis (chronic peri-aortitis)

Retroperitoneal fibrosis (RPF) is a descriptive term used when inflammatory fibrotic tissue encases the aorta and ureters. RPF is three times more common in males than in females and is an IgG4-related disease.

Extraluminal ureteric obstruction leads to unilateral or bilateral obstruction. The condition may extend from the level of the second lumbar vertebra to the pelvic brim. In up to 15% of patients, the fibrosis exists outside the retroperitoneum, consistent with it being a systemic condition. Mediastinal fibrosis, Riedel fibrosing thyroiditis, sclerosing cholangitis, fibrotic orbital pseudotumour, fibrotic arthropathy, and pleural, pericardial, and lung fibrosis have been reported in IgG4-related disease.

RPF is thought to be either an autoallergic inflammatory reaction to leaked material, probably ceroid material from atheromatous plaques, or a systemic autoimmune disease. There is an association with HLA-DRB1*03, an allele linked to various autoimmune diseases. RPF is possibly initiated as a vasa vasorum vasculitis in the aortic wall, which is often seen in chronic peri-aortitis. This inflammatory process causes medial wall thinning, promotes atherosclerosis, and extends into the surrounding retroperitoneum with a fibro-inflammatory reaction typical of chronic peri-aortitis. The autoimmune reaction to plaque antigens could be an epiphenomenon of this immune-mediated process. Activating antibodies against fibroblasts (detectable in one-third of patients) have also been implicated in the pathogenesis, as has the presence of IgG4-bearing plasma cells; the latter is a common finding in autoimmune chronic pancreatitis, a disorder sometimes associated with idiopathic retroperitoneal fibrosis. In addition, several infiltrating B cells

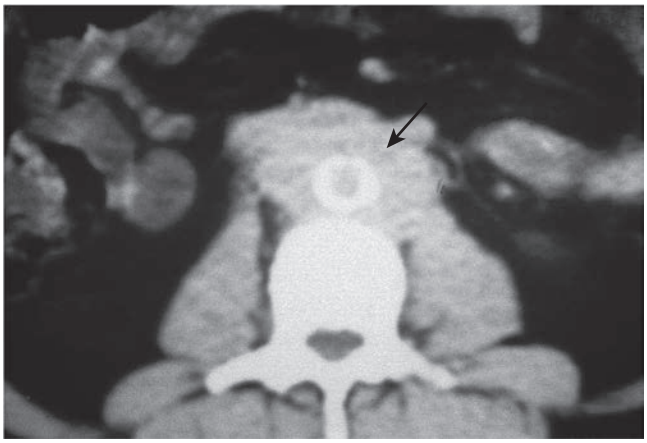


Fig. 26.31 Retroperitoneal fibrosis (peri-aortitis). Note the large mass *(arrowed)* surrounding the abdominal aorta on CT.

show clonal or oligoclonal immunoglobulin heavy chain rearrangement. These findings raise the possibility of RPF being a primary B-cell disorder.

Aetiology

Causes are many:

- Idiopathic in 60% to 70%.
- Secondary causes include:
  - Drugs: methysergide, lysergic acid, ergot-derived dopamine receptor agonists (cabergoline, bromocriptine, pergolide), ergotamine, methyldopa, hydralazine, beta-blockers.
  - Malignancy: colon, prostate, breast and stomach cancers, carcinoid, Hodgkin and non-Hodgkin lymphomas, sarcoma.
  - Infection: tuberculosis, syphilis, histoplasmosis, actinomycosis, fungal infection.
  - Surgery/radiotherapy: lymph node resection, colectomy, hysterectomy, aortic aneurysm repair.

Recognized associations include untreated abdominal aortic aneurysm, smoking, and asbestosis.

Clinical features and investigation

Presenting symptoms include malaise, low back pain, weight loss, testicular pain, claudication, and haematuria. Laboratory tests show normochromic anaemia, CKD, raised ESR, and increased CRP and serum IgG4 concentrations. Imaging with ultrasound shows a poorly circumscribed peri-aortic mass. Contrast-enhanced CT will also show the mass, as well as any tumour and lymphadenopathy (**Fig. 26.31**). MRI will demonstrate similar findings but does not require contrast. FDG-PET assesses the metabolic activity of the retroperitoneal mass; it also allows whole-body imaging and can detect occult malignant or infectious foci, particularly in secondary RPF.

Management

A biopsy is performed to exclude an underlying infection, lymphoma, or cancer. Initial management may involve decompressing obstructed ureters by stenting (anterograde or retrograde). Corticosteroids may reverse the obstructing inflammatory tissue, if only as a holding measure before definitive treatment of the underlying cause. Rituximab is effective in cases of proven IgG4-related disease. Response to treatment is assessed by serial measurements of ESR and repeat FDG-PET.

Obstruction is relieved surgically by ureterolysis. Long-term ureteric stent or stents can be used with chronic corticosteroid therapy,

but regular (usually 6-monthly) changes of the stent(s) are required if the peri-aortic mass does not regress.

Relapse after withdrawal of steroid therapy may occur and treatment may need to be continued for years. Mycophenolate or tamoxifen are also effective. Long-term follow-up is mandatory.

Benign prostatic hypertrophy

Benign prostatic hypertrophy is a common cause of urinary tract obstruction (see Chapter 47, 'Male health' section).

Prognosis of urinary tract obstruction

Prognosis is determined by aetiology and the stage at which obstruction is relieved. Four factors influence the extent of kidney damage and the degree and rapidity of recovery of kidney function after relief of obstruction:

- whether obstruction is partial or complete
- the duration of obstruction
- the presence or absence of infection
- the site of obstruction.

Complete obstruction for several weeks will lead to irreversible or only partially reversible kidney damage. If complete obstruction lasts several months, irreversible kidney damage occurs. Partial obstruction carries a better prognosis, depending on severity. Bacterial infection with obstruction increases kidney damage. Obstruction at or below the bladder neck leads to hypertrophy and trabeculation of the bladder without a rise in pressure within the upper urinary tract, in which case the kidneys are offered some protection from the effects of back-pressure.

Further reading

Perugino, C.A., Stone, J.H. IgG4-related disease: an update on pathophysiology and implications for clinical care. *Nat Rev Rheumatol* 2020; 16, 702–714.
Yaqoob MM, Bennett-Richards K, Junaid I. Retroperitoneal fibrosis. In: **Turner N, Lameire N, Goldsmith GJ et al. (eds).** *Oxford Textbook of Clinical Nephrology*, 4th edn. Oxford: Oxford University Press; 2016; pp. 2861–2864.
Yaqoob MM, Bennett-Richards K, Junaid I. The patient with urinary tract obstruction. In: **Turner N, Lameire N, Goldsmith GJ et al. (eds).** *Oxford Textbook of Clinical Nephrology*, 4th edn. Oxford: Oxford University Press; 2016; pp. 2854–2860.

DRUGS AND THE KIDNEY

The kidney eliminates many drugs, as well as the (often active) products of hepatic drug metabolism. Drug dosing often requires adjustment in CKD. Drugs can also impair kidney function either over time or acutely, for example, when an ACE inhibitor prevents an appropriate response to hypovolaemia.

Drug-induced impairment of the kidney can be caused by reduced kidney perfusion or direct kidney damage.

Causes of **reduced kidney perfusion** include:

- Hypovolaemia, e.g., loop diuretics such as furosemide.
- Salt and water loss, e.g., hypercalcaemia (adversely affects tubular salt and water conservation) induced by vitamin D therapy.
- Decrease in cardiac output which impairs kidney perfusion, e.g., beta-blocker.
- Decrease in kidney blood flow, e.g., ACE inhibitors, particularly in the presence of renovascular disease.

Direct kidney damage can be caused by:

- ***Acute tubular necrosis*** from prolonged or high-dose treatment with aminoglycosides (e.g., gentamicin, amikacin), amphotericin B, tenofovir, cisplatin, or calcineurin inhibitors. The combination of aminoglycosides with loop diuretics is particularly nephrotoxic.
- ***Crystal nephropathies*** caused by antivirals such as aciclovir and indinavir.
- ***Acute tubulointerstitial nephritis*** (see **Box 26.25**), a cell-mediated hypersensitivity nephritis occurring with many drugs including penicillins, cephalosporins, proton pump inhibitors, diuretics, sulphonamides, and NSAIDs (which have many other effects on the kidney. See **Box 26.21**).
- ***Chronic tubulointerstitial nephritis*** due to drugs (see **Box 26.26**).
- Drug-induced ***membranous glomerulonephritis*** e.g., penicillamine, gold, anti-TNF (see 'Membranous glomerulopathy' section).
- Retroperitoneal fibrosis with urinary tract obstruction secondary to medication (see 'Retroperitoneal fibrosis [chronic peri-aortitis]' section).

Box 26.21 Non-steroidal anti-inflammatory drugs and the kidney

Problem	Cause
Sodium and water retention	Reduced prostaglandin production
Acute tubulointerstitial nephritis	Hypersensitivity reaction
Nephrotic syndrome	Membranous glomerulopathy
Analgesic nephropathy	Papillary necrosis after chronic use
Acute kidney injury	Acute tubular necrosis
Hyperkalaemia	Decreased excretion of K^+

Using drugs in patients with impaired kidney function

See **Box 26.22**.

Absorption

Absorption can be unpredictable in uraemia, as gastric emptying may be delayed, and nausea and vomiting are frequent.

Metabolism

Drug metabolism by the kidney can be altered in CKD due to:

1. ***Reduced drug catabolism.*** Insulin catabolism is reduced in kidney disease meaning that insulin requirements in diabetes fall as kidney function declines.
2. ***Reduced conversion of a precursor to a more active metabolite.*** The 1α-hydroxylase enzyme responsible for conversion of 25-hydroxycholecalciferol to active 1,25-$(OH)_2D_3$ is in the kidney. In kidney disease, production of the enzyme declines, and deficiency of 1,25-$(OH)_2D_3$ results.

Protein binding

Reduced protein binding of a drug can increase its activity and potential for toxic side-effects. This occurs in hypoalbuminaemia (nephrotic syndrome) and when uraemic toxins bind to albumin and occupy drug-binding sites. Measurement of total plasma drug concentration (albumin-bound plus free drug) can give misleading results. For example, more free phenytoin than albumin-bound phenytoin is present in CKD for the same total plasma concentration, therefore lower total concentrations will control seizures.

Box 26.22 Safe prescribing in kidney disease

Adverse drug reactions are more than twice as common in patients with CKD compared to patients with normal kidney function. Safe prescribing in CKD demands knowledge of the clinical pharmacology of the drug and its metabolites in the context of reduced kidney function. Consider:

1. Is treatment indicated? Withhold if not.
2. Can the drug reach its site of action? For example, bacteriostatic concentrations of nitrofurantoin will not be attained in the urine in CKD with reduced glomerular filtration.
3. Is the drug's metabolism altered by CKD? If a drug or its metabolites undergo significant excretion by the kidneys, toxic accumulation may occur. Loading dose depends on the size of the patient and is unrelated to kidney function. After loading, avoidance of toxic concentrations requires administration of standard doses at longer time intervals, or smaller doses at the usual time intervals.
4. Is the drug toxic to the kidney? Toxicity is likely to be more significant if underlying CKD.
5. Is there a narrow therapeutic window? Should (can) blood concentrations of the drug be measured?
6. Is the drug a sodium or potassium salt? These are potentially hazardous when CKD is already complicated by salt (and water) overload or hyperkalaemia.
7. Can the dose be titrated? For example, observation of clinical effect, early detection of toxicity effects, measurement of drug concentrations, e.g., gentamicin.

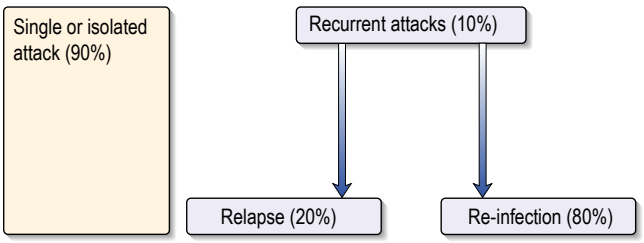


Fig. 26.32 **The natural history of urinary tract infection.**

Box 26.23 Organisms causing urinary tract infection

Organism	Approximate frequency (%)
Escherichia coli and other 'coliforms'	70
Proteus spp.	12
Staphylococcus saprophyticus or epidermidis[a]	10
Enterococcus faecalis[b]	6
Pseudomonas spp.	5
Klebsiella spp.	4

[a]More common in young females (20%–30%).
[b]More common in hospital practice.

Volume of distribution

Salt and water overload or depletion may occur in kidney disease. This affects the concentration of drug obtained from a given dose.

End-organ sensitivity

The renal response to drug treatment may be reduced in kidney disease. For example, thiazide diuretics may have little diuretic effect in patients with advanced CKD.

Drug elimination

Drugs eliminated by the kidneys are not excreted normally in CKD. This is usually more significant for water-soluble drugs (e.g., gentamicin), which are poorly absorbed from the gut, typically given by injection, and not metabolized by the liver, compared to lipid-soluble drugs (e.g., propranolol), which are well absorbed from the gut and principally metabolized by the liver.

Drugs affecting protein anabolism and catabolism

Tetracyclines, with the exception of doxycycline, have a catabolic effect. Increased nitrogenous waste products are not well cleared in advanced CKD and uraemia may be more marked. Corticosteroids have a catabolic effect and may lead to a disproportionate increase in serum urea compared to creatinine when used in high doses in CKD.

Changing kidney function

Patients with rapidly changing kidney function (e.g., evolving AKI, recovering ATN) and the removal of drugs by dialysis and haemofiltration will affect the dosage required.

URINARY TRACT INFECTION

Urinary tract infection (UTI) is common. Between 1% and 2% of patients presenting in primary care will have a UTI. Most UTIs occur in isolation (Fig. 26.32). UTI is most common in females, usually as cystitis. Half of all females have a UTI in their lifetime, most often occurring in a normal urinary tract. UTI is uncommon in males and children, often occurring in an abnormal urinary tract. UTI is not always uncomplicated; recurrent infection causes morbidity, and infection can lead to life-threatening Gram-negative septicaemia and kidney failure.

Aetiology and pathogenesis

Infection is most often caused by bacteria from a patient's own bowel flora (Box 26.23) and usually ascends the urethra. In females, the short urethra makes ascending infection more likely. Rarely, infection may arise from the bloodstream or lymphatics, or by direct extension, for example, vesico-colic fistula.

Bacterial virulence

How well an organism adheres to urothelium determines its virulence. The presence of flagella (for motility), aerobactin (used to acquire iron), haemolysin (to form pores), and most importantly, the presence of fimbriae (adhesins that attach organisms to the perineum and urothelium) on the bacterial cell surface make *E. coli* a common pathogen in UTI.

Innate host defence

Innate host defence prevents UTI in the following ways:

- ***Neutrophils.*** Bacterial adhesins activate receptors, e.g., Toll receptor 4 on the mucosal surface, resulting in IL-8 production and expression of its receptor CXCR1 on neutrophil surfaces. Activation of neutrophils is essential for bacterial killing.
- ***Urine osmolality and pH.*** Urinary osmolality >800 mOsm/kg and low or high pH reduce bacterial survival.
- ***Complement.*** Complement activation with mucosal IgA production by uroepithelium (acquired immunity) plays a major role in defence against UTI.
- ***Commensal organisms.*** Eradication of commensal organisms such as lactobacilli, corynebacteria, streptococci, and

bacteroides by spermicidal jelly or antibiotics results in overgrowth of *E. coli.*

- ***Urine flow.*** Good urine flow and normal micturition wash out bacteria. Urine stasis promotes UTI.
- ***Uroepithelium.*** Mannosylated Tamm–Horsfall proteins, present in the mucus and glycocalyx covering uroepithelium, have antibacterial properties and interfere with bacterial binding to uroepithelium. Cranberry juice and blueberry juice contain a large-molecular-weight factor (pro-anthrocyanidins) that prevents binding of *E. coli* to the uroepithelium.
- ***Blood group antigens.*** Females who are non-secretors of ABH blood group antigens are three to four times more likely to have recurrent UTIs.

Risk factors

- Female sex.
- New sexual activity, particularly in females.
- Indwelling urinary catheter or instrumentation of the urinary tract.
- Urinary tract stones.
- Urinary tract stasis (incomplete bladder emptying).
- Diabetes mellitus or immunosuppression.
- Pregnancy.

Clinical features

The most typical symptoms of lower UTI are:

- frequency of micturition by day and night
- dysuria (painful voiding)
- suprapubic pain and tenderness
- haematuria
- offensive urine.

These are the symptoms of bladder and urethral inflammation, or 'cystitis'. Loin pain and tenderness, with fever, night sweats, and rigors, suggest extension of infection to the renal pelvis and kidney, termed pyelitis or pyelonephritis. Localization of the site of infection based on symptoms alone is unreliable. UTI can also present with few or no symptoms (particularly in the immunocompromised), or with abdominal pain, fever, or haematuria in the absence of frequency and dysuria. In the elderly, new confusion may be the only symptom of UTI. Small children may not describe dysuria and symptoms are often 'atypical'. The possibility of UTI must always be considered in a febrile, sick child.

Diagnosis

Uncomplicated UTIs in females age ≤65 years can be diagnosed in those without known urinary tract abnormalities, recent urinary tract instrumentation, or systemic illness if they exhibit at least two of three key symptoms: dysuria, urgency, or frequency—along with an absence of vaginal discharge. Neither urine dipstick testing for leucocyte esterase nor urine culture enhances diagnostic sensitivity. Telephone-based protocols have outcomes similar to those of surgery-based diagnosis and treatment, and such methods are often preferred by patients. Patients with histories of uncomplicated UTIs can be taught to self-diagnose and initiate therapy.

Otherwise, diagnosis is based on culture of a clean-catch midstream specimen of urine and the presence or absence of pyuria. The criteria for the diagnosis of UTI are shown in **Box 26.24**. Most Gram-negative organisms reduce nitrates to nitrites and produce a red colour in the reagent square. False-negative results are common. Dipsticks that detect significant pyuria depend on the release of esterases from leucocytes. Dipstick tests that are positive for both nitrite and leucocyte esterase are highly predictive of acute infection (sensitivity of 75% and specificity of 82%).

Box 26.24 Criteria for the diagnosis of UTI

- Strongly symptomatic females: single isolate $>10^5$ cfu/L ($>10^2$ cfu/mL) in voided urine
- Males: 10^6 cfu/L (10^3 cfu/mL) of a pure or predominant organism
- Any single organism $>10^7$ cfu/L ($>10^4$ cfu/mL)
- *Escherichia coli (E.coli)* or *Staphylococcus saprophyticus* $>10^6$ cfu/L ($>10^3$ cfu/mL)
- Greater than 10^8 cfu/L ($>10^5$ cfu/mL) mixed growth with one dominant organism

CFU, Colony forming unit.

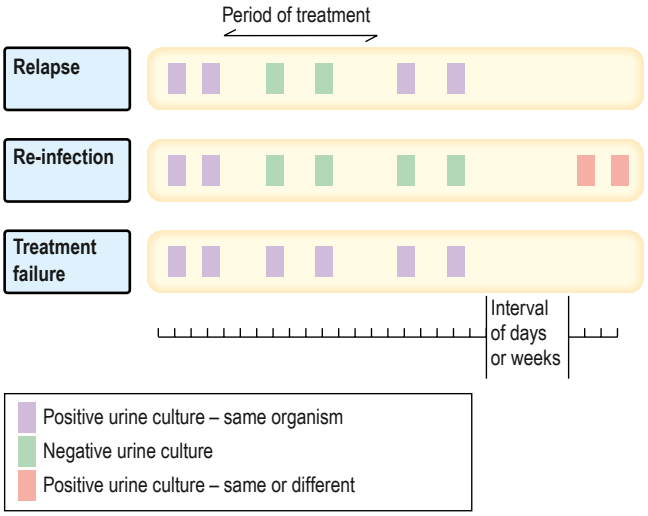


Fig. 26.33 A comparison of re-infection, relapse, and treatment failure in urinary tract infection.

Re-infection may be distinguished from relapsing infection (**Fig. 26.33**) in the following ways:

- ***Relapse*** is diagnosed by recurrence of bacteriuria with the same organism within 7 days of completion of antibacterial treatment. Treatment failure can occur with associated stones, scarred kidneys, polycystic kidney disease, or bacterial prostatitis.
- ***Re-infection*** is when bacteriuria is absent after treatment for at least 14 days, usually longer, followed by recurrence of infection with the same or different organisms. This is not due to failure to eradicate infection but is the result of re-invasion of a susceptible tract. Approximately 80% of recurrent infections are due to re-infection.

Where culture is negative and symptoms persist, consider:

- ***Abacteriuric frequency or dysuria ('urethral syndrome').*** This is caused by bladder trauma after intercourse, vaginitis, urethritis, and *Chlamydia* infection. Tuberculosis should be excluded in symptomatic 'sterile pyuria'.
- ***Interstitial cystitis.*** This is uncommon but distressing, affecting females over the age of 40. Patients present with frequency, dysuria, and often severe suprapubic pain but cultures are sterile. Cystoscopy shows inflammatory changes with ulceration of the bladder base. It is thought to be an autoimmune disorder. Treatments include oral prednisolone, bladder instillation of sodium cromoglicate or dimethyl sulphoxide, and bladder stretching

under anaesthesia. Unfortunately, relief of symptoms can be difficult to achieve.

- ***'Irritable bladder'.*** Predominant frequency and passage of small volumes of urine. It may occur after previous UTI.

Natural history

Serious morbidity is unusual. However, kidney scarring can result from recurrent UTI, and dissemination of infection as septicaemia can be fatal. In patients with a normal urinary tract (normal kidney imaging) outcomes are very good, and persistent or recurrent infection seldom results in serious kidney damage (uncomplicated UTI). In those with abnormal urinary tracts (stones or stasis), recurrence is more common, and outcomes are less good. The combination of infection and obstruction results in severe, sometimes rapid, kidney damage (obstructive pyonephrosis) and is a major cause of Gram-negative septicaemia from *Pseudomonas* and *Enterobacter* spp. Complicated and uncomplicated infection are compared in Fig. 26.34.

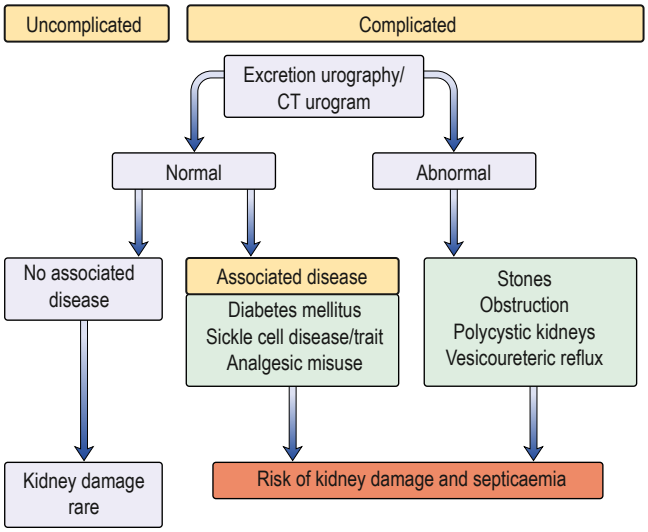


Fig. 26.34 Complicated versus uncomplicated urinary tract infection.

Acute pyelonephritis

Fever, loin pain with tenderness, and significant bacteriuria usually imply infection of the kidney (acute pyelonephritis). Small cortical abscesses and streaks of pus in the medulla can be present. Histologically, there is focal infiltration by polymorphonuclear leucocytes and many polymorphs in tubular lumina. Although, with antibiotics, significant permanent kidney damage in adults with normal urinary tracts is rare, CT scanning can show wedge-shaped areas of inflammation in the cortex (Fig. 26.35), where kidney function will be impaired.

Fig. 26.35 CT showing a wedge-shaped area of cortical loss *(arrowed)* following acute pyelonephritis.

Reflux nephropathy

This results from a combination of:

- vesico-ureteric reflux
- infection acquired in infancy or early childhood.

Normally, the vesico-ureteric junction acts as a one-way valve. Urine enters the bladder from above and the ureter is shut off during bladder contraction, thus preventing reflux of urine. In some infants and children—possibly even *in utero*—this valve mechanism is incompetent, and bladder voiding is associated with variable reflux of urine up the ureter. A secondary consequence is incomplete bladder emptying, as refluxed urine returns to the bladder after voiding. This latter event predisposes to infection, and the reflux of infected urine leads to kidney damage. Typically, there is papillary damage, tubulointerstitial nephritis, and cortical scarring in areas adjacent to 'clubbed calyces' (Fig. 26.36).

Diagnosis is based on CT scan of the kidneys, which shows irregular kidney outlines, clubbed calyces, and a variable reduction in kidney size. The condition may be unilateral or bilateral and may affect all or part of the kidney.

Reflux usually ceases around puberty with growth of the bladder base (a thickened bladder wall can prevent reflux with bladder contraction). Damage already done persists and progressive kidney fibrosis and further loss of function occur in severe cases, even in the absence of further infection.

Reflux nephropathy cannot occur in the absence of reflux, and it does not begin in adult life. Consequently, adult females with bacteriuria and a normal urogram can be reassured that kidney damage will not develop.

Chronic reflux nephropathy acquired in infancy predisposes to hypertension in later life and, if severe, is a relatively common cause of ESKD in childhood or adult life. Meticulous early detection and control of infection, with or without ureteral re-implantation to

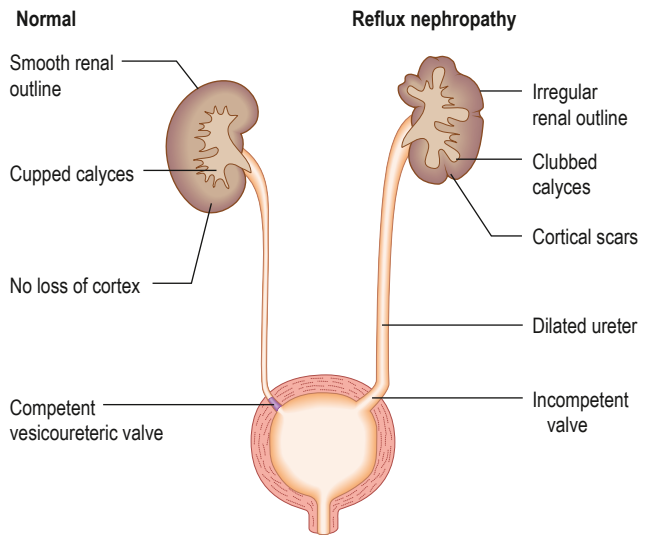


Fig. 26.36 Findings in patients with reflux nephropathy.

create a competent valve, can prevent further scarring and allow normal growth of the kidneys. No proof exists, however, that re-implantation surgery confers long-term benefit.

Continuous antibiotic prophylaxis compared with no treatment, significantly reduces the risk of symptomatic UTI in children with chronic reflux nephropathy but at the expense of increased risk for developing antimicrobial-resistant bacterial strains. However, it has not been shown to have a significant impact on the occurrence of new kidney scarring.

Investigations

Uncomplicated UTI usually does not require imaging. If infection is recurrent, unusually severe, affects males or children, or if the patient has diabetes or is immunosuppressed, UTI should be investigated.

- ***Ultrasound*** allows the detection of calculi, obstruction, abnormal urinary anatomy, and incomplete bladder emptying (post-micturition scan). It is helpful in the assessment of suspected pyelonephritis or an obstructed, infected kidney that may require drainage.
- ***CT*** is more sensitive than ultrasound. Contrast-enhanced CT allows different phases of excretion to be studied and can define extent of disease and identify significant complications or obstruction.
- ***MRI*** is particularly useful in those with iodinated contrast allergies, offering an ionizing radiation-free alternative for diagnosis.
- ***Nuclear medicine*** has a limited role in the evaluation of UTI in adults. Its main role is in the assessment of kidney function and detection of scars by DMSA scan, often prior to surgery.

Management

Management of a single isolated UTI

- ***Antibiotic*** choice should be guided by local susceptibilities. Typical antibiotics for females include trimethoprim (200 mg twice daily for 3 days) or nitrofurantoin (100 mg twice daily for 3 days). Fluoroquinolones offer no advantage in terms of cure rates. β-lactam antibiotics, e.g., amoxicillin–clavulanate, are less effective than the first-line recommendations. Patients who delay antibiotic treatment to encourage spontaneous resolution eventually receive antibiotics and have longer times to resolution. Males with uncomplicated UTIs should be treated as above but for 7 days.
- ***High fluid intake*** (>2 L daily) should be encouraged during treatment and for some subsequent weeks.
- ***Urine culture*** should be repeated 5 days after treatment.
- ***Acute pyelonephritis*** (fever, loin pain/tenderness) is usually managed with intravenous, for example aztreonam, cefuroxime, ciprofloxacin, or gentamicin, switching to oral therapy as symptoms improve. Intravenous fluids may be required. Kidney tract ultrasound examination is required to exclude an obstructed pyonephrosis requiring drainage by percutaneous nephrostomy.

Management of recurrent infection

Pre- and post-treatment urine cultures are necessary to confirm diagnosis and to establish whether recurrent infection is due to relapse or re-infection (Fig. 36.33).

- ***Relapse.*** A search should be made for a cause (e.g., stones or scarred kidneys). High-dose or prolonged treatment may be required. If this fails, long-term antibiotics may be necessary.
- ***Re-infection.*** This implies that the patient has a predisposition to peri-urethral colonization or poor bladder defence mechanisms. The use of a contraceptive diaphragm or spermicidal jelly may contribute, and alternative contraception should be offered. Atrophic vaginitis in postmenopausal females can be treated with topical oestrogen therapy.

  Preventative measures include:
- 2 L daily fluid intake
- voiding at 2- to 3-hour intervals, with double micturition if reflux is present
- voiding before bedtime and after intercourse
- avoidance of spermicidal jellies, bubble baths, and other chemicals when washing
- avoidance of constipation, which may impair bladder emptying.

Impaired bladder emptying on excretion urography/ultrasound requires urological assessment. If UTI continues to recur, treatment for 6 to 12 months with low-dose alternating monthly prophylaxis (trimethoprim 100 mg, co-trimoxazole 480 mg, cefalexin 125 mg, macrocrystalline nitrofurantoin) taken last thing at night when urine flow is low. Evidence for cranberry juice, D-mannose, and methenamine hippurate (Hipres) is limited.

Urinary infections in the presence of an indwelling catheter

Colonization of the bladder by a pathogen is common after a urinary catheter has been *in situ* for more than a few days, partly due to organisms forming biofilms. While the bladder catheter is in place, antibiotics are likely to be ineffective, with risk of antibiotic resistance developing. The patient should be treated only if there are symptoms or evidence of infection and the catheter should always be replaced. When changing catheters, a single dose of antibiotic (e.g., gentamicin) is recommended.

Infection by *Candida* is a complication of prolonged bladder catheterization. Treatment should be reserved for patients with evidence of invasive infection and those who are immunosuppressed. The catheter should be removed or replaced. In severe infections, continuous bladder irrigation with amphotericin may be useful.

Bacteriuria in pregnancy

Screening of asymptomatic pregnant females detects asymptomatic bacteriuria in 2% to 6%. Consider harmless outside of pregnancy, historical data associate asymptomatic bacteriuria in pregnancy with pyelonephritis and preterm labour. Ureteric dilation in response to gestational hormonal changes may allow ascending infection. Bacteriuria is therefore treated and eradication confirmed. Re-infection may require prophylactic therapy. Penicillins, nitrofurantoin (until 36 weeks'), and oral cephalosporins may be used safely in pregnancy. Tetracycline, trimethoprim (in the first trimester), and quinolones are avoided in pregnancy.

Bacterial prostatitis

Bacterial prostatitis is a relapsing infection that presents as perineal pain, recurrent epididymo-orchitis, and prostatic tenderness, with pus in expressed prostatic secretions. Treatment is for 4 to 6 weeks with drugs that penetrate the prostate, such as trimethoprim and ciprofloxacin. Long-term, low-dose treatment may be required. Prostadynia (prostatic pain in the absence of active infection) may persist after infection. Amitriptyline and carbamazepine may alleviate pain symptoms.

Kidney carbuncle

Kidney carbuncle is an abscess in the cortex caused by blood-borne staphylococcus, usually from a boil or carbuncle of the skin. It presents with a high, swinging fever, and loin pain, tenderness, and fullness. The urine shows no abnormality, as the abscess does not

communicate with the renal pelvis, more often extending into the perirenal tissue. Staphylococcal septicaemia is common. Diagnosis is by ultrasound or CT scanning. Treatment involves antibacterial therapy with flucloxacillin and surgical drainage.

Tuberculosis of the urinary tract

Tuberculosis of the urinary tract presents with frequency, dysuria, or haematuria. Cortical lesions result from haematogenous spread in the primary phase of infection. Most heal, but in some, infection persists and spreads to the papillae, with the formation of cavitating lesions and the discharge of mycobacteria into the urine. Infection of the ureters and bladder commonly follows, with the potential for development of ureteral strictures and a contracted bladder. Rarely, cold abscesses may form in the loin. In males, the disease may present with testicular or epididymal discomfort and thickening.

Diagnosis requires awareness, especially in patients with sterile pyuria. Imaging may show cavitating lesions in the papillary areas, commonly with calcification. There may also be evidence of ureteral obstruction with hydronephrosis. Culture of mycobacteria from early morning urine samples is not straightforward. Imaging may be normal in diffuse interstitial tuberculosis, when diagnosis is made instead by kidney biopsy, which demonstrates caseating granulomata with multinucleate giant cells and acid-fast bacilli on Ziehl–Neelsen staining (**Fig. 26.37**). Some patients present late with small, unobstructed kidneys, when the diagnosis is easy to miss.

Treatment is as for pulmonary tuberculosis (see Chapter 20, 'Tuberculosis/Management' section). Kidney ultrasonography and/or CT scanning should be carried out 2 to 3 months after treatment initiation, as ureteric strictures may develop in the healing phase.

Xanthogranulomatous pyelonephritis

This is an uncommon chronic interstitial infection of the kidney, most often due to *Proteus*, causing fever, weight loss, loin pain, and a palpable enlarged kidney. It is usually unilateral and associated with staghorn calculi and UTI. CT scanning shows radiolucent intrarenal abscesses. Nephrectomy is the treatment of choice as antibacterial treatment rarely eradicates infection.

Further reading

Johnson JR, Rosso TA. Acute pyelonephritis in adults. *N Engl J Med* 2020;378:48–59.
Mancuso G, Midiri A, Gerace E, Marra M, Zummo S, Biondo C. Urinary Tract Infections: The Current Scenario and Future Prospects. *Pathogens* 2023; 12:623.
Manu P Bilsen and others. Definitions of urinary tract infection in current research: A systematic review. *Open Forum Infect Dis* 2023; 10(7):ofad332.

TUBOLOINTERSTITIAL NEPHRITIS

Diseases of the kidney primarily affect the glomeruli, vasculature, or the substance of the kidney parenchyma, made up of tubules and the interstitium. Tubular and interstitial injury and recovery tend to go hand in hand.

Acute tubulointerstitial nephritis

In approximately 70% of cases, acute tubulointerstitial nephritis (TIN) is a hypersensitivity reaction to drugs (**Box 26.25**), most commonly penicillins, NSAIDs, or proton pump inhibitors.

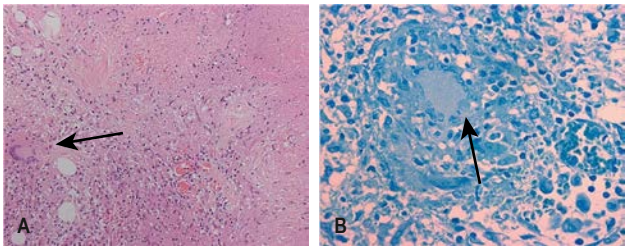


Fig. 26.37 Kidney tuberculosis. **(A)** Caseating granulomatous interstitial nephritis showing multinuclear giant cell *(arrowed)*. **(B)** Ziehl–Neelsen staining showing mycobacteria *(arrowed)*.

Box 26.25 Causes of acute tubulointerstitial nephritis

- Drugs (70%):
  - Antibiotics: cephalosporins, ciprofloxacin, erythromycin, penicillin, rifampicin, sulphonamides
  - Analgesics: NSAIDs
  - Diuretics: furosemide, thiazide diuretics
  - Miscellaneous: allopurinol, carbamazepine, cimetidine, phenytoin, proton pump inhibitors, sodium valproate
- Infection (15%):
  - Viruses, e.g., hantavirus
  - Bacteria, e.g., streptococci
- Idiopathic (8%)
- Tubulointerstitial nephritis with uveitis (TINU) (5%)
- Systemic inflammatory disorders, e.g.,:
  - Systemic lupus erythematosus (2%)
  - IgG4-related disorder

Drug-induced acute TIN

Patients may be asymptomatic (with impaired kidney function found incidentally), or present with fever, arthralgia, rash, and acute kidney injury. Many have eosinophilia and eosinophiluria. Kidney histology shows an intense interstitial cellular infiltrate, often including eosinophils, with variable tubular necrosis (**Fig. 26.38**). Urinary CXCL9 concentrations increase sevenfold and are an emerging diagnostic marker of drug-induced TIN. Rarely, NSAIDs cause a minimal-change lesion in addition to TIN and disease presents as nephrotic syndrome. Treatment involves stopping the offending drug. High-dose prednisolone (60 mg daily) may reduce risk of dialysis-requiring AKI and shorten time to recovery. Most patients make a good recovery in terms of kidney function, but some may be left with significant interstitial fibrosis and CKD.

Infection causing acute TIN

Acute pyelonephritis leads to tubular inflammation and a neutrophilic cellular infiltrate. TIN can complicate systemic infections with viruses (Hantavirus, Epstein–Barr virus, HIV, measles, adenovirus), bacteria (*Legionella*, *Leptospira*, streptococci, *Mycoplasma*, *Brucella*, *Chlamydia*) and others (*Leishmania*, *Toxoplasma*). Hantavirus causes haemorrhagic fever with TIN and can be fatal. Cytomegalovirus, polyoma (BK), and herpes simplex virus can cause acute TIN in a kidney transplant. Treatment involves eradication of infection with appropriate antibiotics or antiviral agents. In kidney transplantation, the immunosuppressive regimen is modified.

Acute TIN as part of multisystem inflammatory diseases

Several non-infectious inflammatory disorders (e.g., Sjögren's syndrome, SLE) can cause acute or chronic TIN rather than glomerulonephritis. Sjögren's syndrome can additionally present with

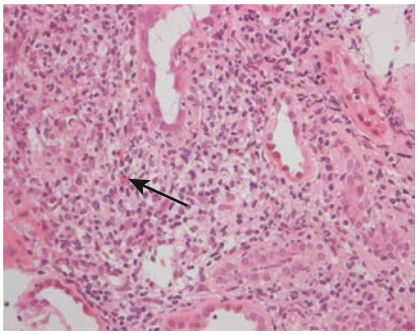


Fig. 26.38 **Tubulointerstitial nephritis showing diffuse interstitial infiltrate with red staining eosinophils *(arrowed)*.**

renal tubular acidosis. Sarcoidosis presents as granulomatous TIN in up to 20% of patients and associated hypercalcaemia causes AKI. These heterogeneous conditions with TIN generally respond to steroids.

TIN with uveitis syndrome

Patients with tubulointerstitial nephritis with uveitis (TINU) syndrome present with uveitis, an acute TIN, weight loss, anaemia, and a raised ESR. It is common in childhood, more frequently found in female, and can occur in adults. It may be associated with autoantibodies directed against modified CRP. A prolonged course of steroids leads to improvement in both kidney function and uveitis.

Chronic tubulointerstitial nephritis

The major causes of chronic TIN are given in **Box 26.26**. Patients present with any of polyuria, nocturia, proteinuria (usually <1 g daily), or uraemia. Chronic TIN may be linked with papillary necrosis (ischaemic damage to the papillae, which then slough into the calyces and urine) when associated with longstanding analgesic use, diabetes mellitus, and sickle cell disease or trait. A sloughed papilla may cause ureteric colic or produce acute ureteric obstruction. Visible and non-visible haematuria and sterile pyuria can occur. Histologically, a generalized chronic inflammatory cellular infiltration of the interstitium with tubular atrophy and generalized interstitial oedema or fibrosis is present. In many cases, no cause is found.

IgG4-related TIN is distinguished by very low concentrations of complement and increased serum IgG4. This may lead to immune complex formation.

Radiological appearances must be distinguished from reflux nephropathy (**Fig. 26.39**) and obstructive uropathy.

Tubular damage to the medullary area of the kidney leads to defects in urine concentration and sodium conservation with polyuria and salt wasting. Fibrosis progressing into the cortex leads to loss of excretory function and urea.

Analgesic nephropathy

The chronic consumption of large amounts of analgesics (classically, phenacetin but now NSAIDs) leads to chronic TIN and papillary necrosis. Analgesic nephropathy is more common in females and typically presents in middle age. Presentation may be with anaemia, CKD, UTI, haematuria, or urinary tract obstruction (owing to sloughing of a renal papilla). Salt- and water-wasting kidney disease may occur. Chronic analgesic use also predisposes to the development of uroepithelial tumours. Diagnosis is usually made on clinical grounds combined with the non-pathognomonic appearance of small, irregularly outlined kidneys on imaging.

? Box 26.26 Causes of chronic tubulointerstitial nephritis

Drugs/medication/toxins:
- All drug causes of acute tubulointerstitial nephritis (see **Box 26.25**)
- Ciclosporin
- 5-aminosalicylates
- Cadmium, lead, titanium
- Irradiation

Systemic disease:
- Diabetes mellitus
- Sickle cell disease or trait
- Systemic lupus erythematosus/vasculitis
- Sarcoidosis
- IgG4-related disease

Metabolic:
- Hyperuricaemia
- Nephrocalcinosis
- Hyperoxaluria

Infection:
- Human immunodeficiency virus
- Epstein–Barr virus

Miscellaneous:
- Hypertension
- Balkan nephropathy
- Herbal nephropathy
- Alport's syndrome

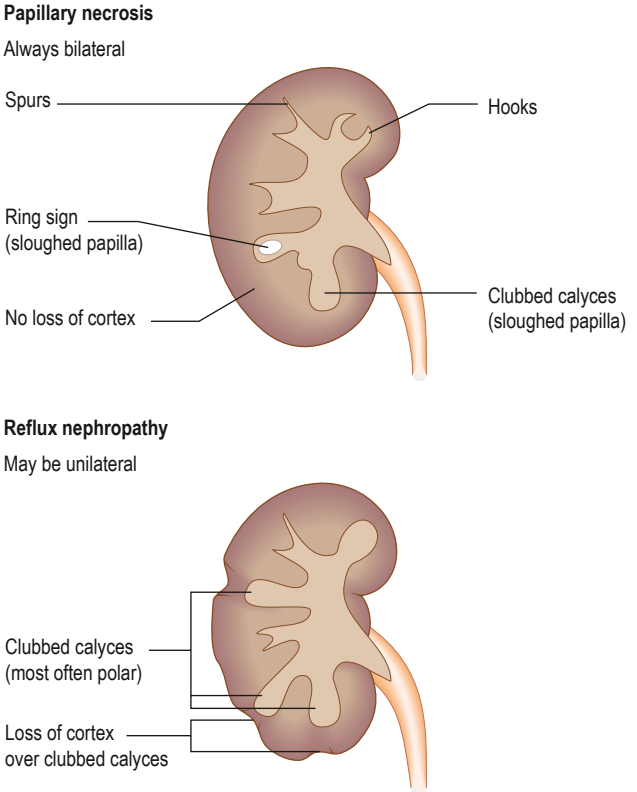


Fig. 26.39 **A comparison of the radiological appearances of papillary necrosis and reflux nephropathy.**

Patients should be encouraged to avoid NSAIDs. If necessary, dihydrocodeine or paracetamol may be offered as a reasonable alternative (this may stop further damage and even achieve an improvement). The development of flank pain or an unexpectedly

rapid deterioration in kidney function should prompt ultrasonography to screen for urinary tract obstruction due to a sloughed papilla.

Balkan nephropathy

Balkan nephropathy is a chronic TIN endemic in areas along the tributaries of the River Danube. Inhabitants of the low-lying plains, where the water supply comes from shallow wells, are affected, whereas the disease does not occur in hillside villages where surface water provides the supply. Chronic dietary poisoning by aristolochic acid is thought to be responsible for Balkan nephropathy and the urothelial cancers that accompany the diagnosis. The disease is insidious in onset, with mild proteinuria progressing to ESKD in 3 months to 10 years. There is no treatment.

Chinese herb nephropathy

Chinese herbal medicines may contain aristolochic acid, produced due to fungal contamination. Histology is similar to that of Balkan nephropathy, but the clinical course is aggressive, with relentless progression to ESKD and a high incidence of uroepithelial tumours.

Other forms of chronic TIN

These are rare (see **Box 26.26**). Diagnosis of all forms depends on a history of drug ingestion or industrial exposure to nephrotoxins. In patients with unexplained kidney impairment with normal-sized kidneys, kidney biopsy should be undertaken to exclude a treatable TIN, such as granulomatous TIN due to ***renal sarcoidosis*** (**Fig. 26.40**), which may be the first presentation of sarcoidosis. Renal sarcoidosis generally responds rapidly to steroids.

Hyperuricaemic nephropathy

Acute hyperuricaemic nephropathy causes AKI with very high serum urate concentrations as part of tumour lysis syndrome. It most commonly occurs following treatment of haematological malignancies but can occur prior. As tumour cells lyse in response to chemotherapy, large amounts of nucleoprotein are released with increased uric acid production. Kidney failure is due to intrarenal and extrarenal obstruction caused by deposition of uric acid crystals in the collecting ducts, pelvis, and ureters.

There may be flank pain or colic, and oliguria or anuria with increasing uraemia. Plasma urate concentrations (reference range 0.15–0.40 mmol/L) increase above 0.75 mmol/L and may be as high as 4.5 mmol/L.

Diagnosis is based on history and hyperuricaemia. Ultrasound may demonstrate extrarenal obstruction due to stones but may be normal.

Allopurinol 100 to 200 mg three times daily for 5 days is given prior to and throughout treatment with radiotherapy or cytotoxic drugs. A high rate of urine flow is maintained by oral or parenteral fluid, and the urine kept alkaline by the administration of sodium bicarbonate and acetazolamide, since uric acid is more soluble in alkaline than acidic mediums. ***Febuxostat***, a non-purine analogue inhibitor of xanthine oxidase, can be used if allopurinol cannot be tolerated and eGFR is more than 30 mL/min/1.73 m^2. ***Rasburicase***, a recombinant urate oxidase, and ***pegloticase***, a pegylated uricase, are successfully used to lower serum urate concentrations for both treatment and prevention. In severely oliguric or anuric patients, dialysis is required to lower plasma urate.

There is no convincing evidence for chronic hyperuricaemia nephropathy. However, a few observational studies have suggested that elevated concentrations of uric acid independently increase the risk for CKD, and that plasma urate reduction with allopurinol has a beneficial effect in slowing rate of CKD progression.

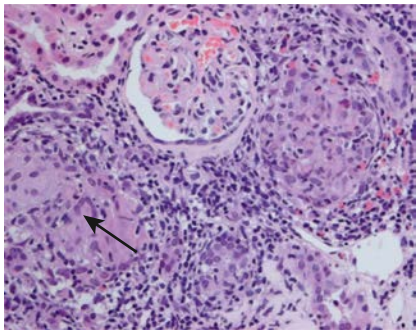


Fig. 26.40 Granulomatous tubulointerstitial nephritis. The granuloma consists of both giant cells *(arrowed)* and epithelioid cells. These findings are characteristic of renal sarcoid but can be seen in interstitial nephritis due to any cause (e.g., drug, infection)

Further reading

Erugino CA, Stone JH. IgG4-related disease: an update on pathophysiology and implications for clinical care. *Nat Rev Rheumatol* 2020; 16:702–714.

ACUTE KIDNEY INJURY

The assessment and management of acute kidney injury (AKI) are detailed in Chapter 11 ('Acute kidney injury' section).

Causes of AKI due to pathology affecting the parenchyma of the kidney ('renal'-AKI) include:

- acute tubular necrosis
- acute cortical necrosis
- rhabdomyolysis
- acute phosphate nephropathy
- tumour lysis syndrome
- hepatorenal syndrome.

The role of contrast in AKI is unclear. The historical term 'contrast-induced nephropathy' has been replaced with 'contrast-associated nephropathy'.

Acute tubular necrosis

Acute tubular necrosis (ATN) (see **Box 26.27**) can occur due to:

- sustained under-perfusion and reduced kidney blood flow, leading to tubular cell death
- nephrotoxins causing direct injury and tubular cell death.

Factors involved in the development of ATN include:

1. Intrarenal microvascular vasoconstriction with falling delivery of O_2 and increasing tubular hypoxia
   - Vasoconstriction in response to endothelin, adenosine, thromboxane A2, leukotrienes, and sympathetic nerve activity
   - Impaired vasodilation due to reduced sensitivity to nitric oxide, prostaglandins (PGE2), acetylcholine, and bradykinin.
2. Increased endothelial and vascular smooth muscle cell structural damage
   - Increased leucocyte–endothelial adhesion, vascular congestion and obstruction, leucocyte activation, and inflammation.
3. Ischaemic injury results in rapid depletion of intracellular ATP stores, resulting in cell death either by necrosis or apoptosis, with:
   - increased cytosolic cell calcium concentration
   - induction of nitric oxide synthases with increased production of nitric oxide, causing cell death

Box 26.27 Some causes of acute tubular necrosis

- Haemorrhage
- Burns
- Diarrhoea and vomiting
- Fluid loss from fistulae
- Pancreatitis
- Diuretics
- Myocardial infarction
- Congestive cardiac failure
- Endotoxic shock
- Snake bite
- Myoglobinaemia
- Haemoglobinaemia: due to haemolysis, e.g., *falciparum* malaria
- Hepatorenal syndrome
- Radiological contrast agents
- Drugs: aminoglycosides, non-steroidal anti-inflammatory drugs, angiotensin-converting enzyme inhibitors, platinum derivatives
- Placental abruption, postpartum haemorrhage
- Pre-eclampsia

  - increased production of intracellular proteases such as calpain, which cause proteolysis of cytoskeletal proteins and cell wall collapse
  - loss of cell polarity. Integrins on the basolateral side of the cell are translocated to the apical surface, forming casts when combined with other desquamated cells
  - tubular obstruction by desquamated viable or necrotic cells and back-leak of tubular fluid.

4. Tubular cellular recovery
   - Tubular cells can regenerate rapidly to reform the disrupted tubular basement membrane, which explains the reversibility of ATN.
   - Growth factors, e.g., insulin-like growth factor 1, epidermal growth factor, hepatocyte growth factor, and their receptors are upregulated.

Kidney blood flow is reduced in ATN, particularly blood flow to the cortex. Ischaemic tubular damage contributes to a reduction in glomerular filtration as glomeruli contract through afferent arteriolar spasm, filtrate leaks back towards the glomerulus, and tubular obstruction evolves (Fig. 26.41).

Non-oliguric AKI is usually the result of a less severe kidney insult, otherwise oliguria is common. Recovery of kidney function typically occurs after 7 to 21 days, although it may be delayed by continuing sepsis. There is no evidence that intravenous mannitol, furosemide, or 'renal-dose' dopamine are effective treatments.

Acute cortical necrosis

Kidney hypoperfusion results in diversion of blood flow from the cortex to the medulla, with a drop in GFR. Medullary ischaemic damage is largely reversible because tubular cells have the capacity for regeneration. In contrast, glomerular ischaemic injury does not heal with regeneration, but with scarring—glomerulosclerosis. Prolonged cortical ischaemia may lead to irreversible loss of kidney function termed 'cortical necrosis'. This may be patchy or complete. Any cause of ATN, if sufficiently severe or prolonged, may lead to cortical necrosis. Cortical necrosis occurs more commonly where the vasculature and endothelium have been damaged (with or without coagulopathy), for example, HUS, endothelial disease in pregnancy.

Rhabdomyolysis

Rhabdomyolysis, or 'crush syndrome', occurs when skeletal muscle injury leads to release of intracellular myoglobin, a direct tubular toxin leading to ATN. Causes include trauma, compartment syndrome, excessive exertion (marathon runners), status epilepticus, and muscle toxins (statins, malaria and antimalarials, snake/insect venom). Rapid rises in potassium, phosphate, and lactate concentrations accompany elevated muscle enzymes; characteristically, creatine phosphokinase (CK) is massively elevated. Early and vigorous fluid resuscitation is vital (injured muscle sequesters large amounts of fluid, causing hypovolaemia). Inflamed, injured muscle may become ischaemic in compartments, and a careful examination for typical 'woody-hard' muscle compartments may indicate the need for fasciotomies to release at-risk muscle. Sodium bicarbonate used to alkalinize the urine may limit myoglobin-induced injury. A similar syndrome can arise with massive haemolysis, where the tubular toxin is haemoglobin rather than myoglobin, and the biochemical derangements seen with muscle injury may be absent.

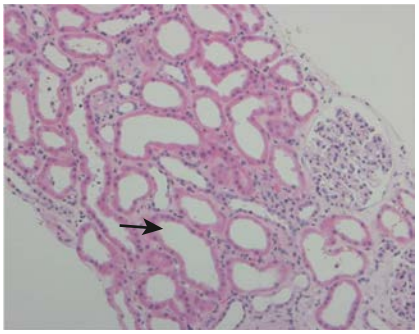


Fig. 26.41 Acute tubular necrosis. Effacement and loss of the proximal tubule brush border, patchy loss of tubular cells, and focal areas of proximal tubule dilation *(arrowed)*.

Acute phosphate nephropathy

Administration of oral sodium phosphate solution as bowel preparation for gastrointestinal investigations is a cause of AKI. Oral phosphate solution is contraindicated in patients with CKD, congestive heart failure, gastrointestinal obstruction, and pre-existing electrolyte disorders, for example, hypercalcaemia. Kidney biopsy shows abundant calcium phosphate deposits. Treatment is usually supportive, with dialysis if indicated, usually with good kidney recovery.

Tumour lysis syndrome

Tumour lysis syndrome complicates the initial treatment of cancer, most commonly lymphoproliferative tumours. Chemotherapy (including corticosteroids) causes rapid death of malignant cells with release of intracellular and membrane products including uric acid, potassium, and phosphate. A high filtered uric acid load leads to an obstructive crystal uropathy; serum urate may be four- or fivefold the upper limit of normal. The danger is hyperkalaemia. At-risk patients should be hydrated to maintain a good urine output (facilitating potassium excretion in the urine) and xanthine oxidase inhibitors, for example, allopurinol, may prevent uric acid formation. Rasburicase oxidizes uric acid to soluble allantoin and uric acid concentrations can be undetectable after administration.

Hepatorenal syndrome

Hepatorenal syndrome results from profound vasoconstriction with histologically normal kidneys. It only occurs once portal hypertension and ascites complicate liver disease. Patients are not hypovolaemic. Recovery is usually observed if successful liver transplantation is possible.

Contrast-associated acute kidney injury

In patients with impaired kidney function, iodinated radiological contrast media may cause kidney vasoconstriction and have a direct toxic effect on kidney tubules. However, a rise in serum creatinine concentration after contrast exposure is often non-specific, with other contributing factors including volume depletion, atheroembolization, and medication. Contrast-associated acute kidney injury may be a marker of an increased risk of serious adverse outcomes rather than a mediator of adverse outcomes and in many patients the effect is mild, transient, and fully reversible. Clinicians should therefore be cautious about withholding contrast in patients with kidney disease where it is indicated for diagnosis.

Risk factors for a creatinine rise in association with contrast use include:

- advanced CKD
- cardiac failure
- paraprotein kidney disease
- hypovolaemia or medications that reduce kidney perfusion e.g., ACE inhibitors, NSAIDs
- large amounts (>4 mL/kg, or repeated dosing) of contrast, e.g., angiography
- high-osmolality contrast agents.

Prevention is key: minimize the dose of contrast administered and use an iso-osmolar or low-osmolality contrast medium. Intravenous fluid (e.g., 1 L of 0.9% sodium chloride or 1.4% sodium bicarbonate 1–12 h before and 4–12 h after contrast) has shown benefit in observational studies, but randomized trial data are limited, and care is needed if coexisting hypertension or cardiac failure. N-acetylcysteine, dopamine, theophylline, and prophylactic haemodialysis (to remove contrast agent from the circulation) are of no benefit.

Renal replacement therapy in AKI

It is not clear when to start RRT in AKI. Early initiation has not been shown to improve outcomes and in up to 30% of cases there is kidney recovery without RRT. There are no data to favour haemodialysis over haemofiltration. However, in haemodynamically unstable patients, haemofiltration may be preferred as fluid volumes including nutrition, blood products, and antibiotics can be ultrafiltered with less swing in volume and more cardiovascular stability. More RRT is not necessarily better; two large trials have failed to show a survival benefit with augmented doses of RRT in critically ill patients. Peritoneal dialysis can be used in AKI, predominantly in low-income settings where haemodialysis may not be available. It has low efficiency (removing solutes slowly and less completely) and is not appropriate for patients with intra-abdominal causes or complications of AKI. The effect of increasing intra-abdominal pressure on lung function should also be considered.

CHRONIC KIDNEY DISEASE

CKD is a descriptive term and is used for a long-standing (>3 months) deterioration in kidney function of any underlying cause. A list of causes of CKD is given in **Box 26.28**. In many cases, there is no effective means to reverse the underlying disease. There are exceptions where reversal of CKD can occur:

- relief of urinary tract obstruction
- immunosuppressive therapy for glomerulonephritis or systemic vasculitis
- treatment of accelerated hypertension
- correction of critical narrowing of renal arteries.

Even if the underlying diagnosis cannot be reversed, the rate of deterioration in kidney function can be slowed.

? Box 26.28 Causes of chronic kidney disease

Congenital and inherited disease
- Polycystic kidney disease (adult and infantile forms)
- Medullary cystic disease
- Tuberous sclerosis
- Oxalosis
- Cystinosis
- Congenital obstructive uropathy

Glomerular disease
- Primary glomerulonephritides
- Secondary glomerular disease: for example, SLE, polyangiitis, granulomatosis with polyangiitis, amyloidosis, diabetic glomerulosclerosis, accelerated hypertension, haemolytic uraemic syndrome, thrombotic thrombocytopenic purpura, systemic sclerosis, sickle cell disease

Vascular disease
- Hypertensive nephrosclerosis
- Renovascular disease
- Small- and medium-sized-vessel vasculitis

Tubulointerstitial disease
- Tubulointerstitial nephritis
- Reflux nephropathy
- Tuberculosis
- Schistosomiasis
- Nephrocalcinosis
- Multiple myeloma: myeloma kidney
- Balkan nephropathy
- Papillary necrosis: for example, due to diabetes, sickle cell disease and trait, analgesic nephropathy
- Chinese herb nephropathy

Urinary tract obstruction
- Renal tract stone disease
- Prostatic disease
- Pelvic tumours
- Retroperitoneal fibrosis
- Schistosomiasis

Staging and prevalence

CKD is staged according to estimated GFR, calculated according to eGFR (see 'Estimated GFR [eGFR]' section). Staging is intended to reflect severity, prognosis, and outcomes (**Fig. 26.42**).

Staging is increasingly complex:

- Patients are staged both on eGFR (prefix G) and albuminuria (prefix A), reflecting the fact that both parameters correlate with progressive kidney impairment and cardiovascular risk, e.g., G3bA3 disease.
- Stage G3 CKD has been divided into two stages (G3a and G3b) in recognition of the increased cardiovascular complications seen in more advanced disease.
- Cystatin C is a more accurate marker of GFR and can be used to better stratify risk in stage G3 CKD.

Confusingly, patients with stage G2 CKD are not thought to have disease unless there is other evidence of kidney damage, for example, haematuria, proteinuria, structurally abnormal kidneys, inherited kidney disease, or biopsy changes consistent with kidney disease.

CKD affects 11% to 15% of the world's population, with the majority at stage 3 (**Box 26.29**). Diagnosis of CKD is becoming more likely as people live longer and a high prevalence of CKD

GFR and ACR categories and risk of adverse outcomes			Description	ACR (mg/mmol) / description		
				<3	3–30	>30
				A1	A2	A3
GFR (mL/min/1.73 m²)	≥90	G1	Normal or increased GFR with other evidence of kidney damage	No CKD markers of kidney damage		
	60–89	G2	Slight decrease in GFR with other evidence of kidney damage			
	45–59	G3a	Moderate decrease in GFR with or without other evidence of kidney damage			
	30–44	G3b	Moderate to severe increase in GFR			
	15–29	G4	Severe decrease in GFR with or without other evidence of kidney damage			
	<15	G5	End-stage renal disease			

Increasing risk

Increasing risk

Fig. 26.42 Classification of chronic kidney disease (CKD) using glomerular filtration rate (GFR) and urinary albumin:creatinine (ACR) categories. *(Adapted with permission from Kidney Disease: Improving Global Outcomes (KDIGO) CKD Work Group. KDIGO 2012 clinical practice guideline for the evaluation and management of chronic kidney disease.* Kidney Int *2013; Suppl. 3:1–150.)*

Box 26.29 Prevalence of CKD

CKD stage	Age in years (% affected)			
	<65	65–74	75–84	≥85
3	1.4	15.4	29.4	30.9
4	0.04	0.4	1.3	2.4
5	0.03	0.2	0.1	0.1
Total	1.5	15.9	30.9	33.4

is found in population-based surveys in elderly patients. There is wide geographical variation in incidence, prevalence, and causes of CKD across the globe. For instance, the most common cause of GN in sub-Saharan Africa is malaria. Schistosomiasis is a common cause of CKD due to urinary tract obstruction in parts of the Middle East. The incidence of CKD varies between ethnic groups, with hypertensive nephropathy and stage G5 CKD being three to four times more common in people of black African descent in the UK and USA than it is in white populations. The prevalence of diabetes mellitus, and hence of diabetic nephropathy, is higher in those of Asian ethnicities compared to white ethnicities. Over 70% of all cases of CKD are due to diabetes mellitus, hypertension, or atherosclerosis.

Progression

CKD tends to progress, although rate of progression may be slow. The speed of decline tends to depend on the underlying disease and blood pressure control. Patients with chronic glomerular diseases tend to deteriorate more quickly than those with chronic tubulointerstitial nephropathies. Regardless of cause, there may be common pathways to disease progression in CKD:

1. Each kidney has roughly a million nephrons. In CKD, when many nephrons have failed and scarred, the burden of filtration falls to fewer functioning nephrons.
2. Functioning ('remnant') nephrons experience increased flow per nephron (hyperfiltration), as kidney blood flow has not changed, and adapt with glomerular hypertrophy and reduced arteriolar resistance.
3. Increased flow, increased pressure, and increased shear stress set in motion a vicious circle of raised intraglomerular capillary pressure and strain, which accelerates remnant nephron failure.
4. Increased flow and strain may be detected as new or increasing proteinuria.

Angiotensin II produced locally modulates intraglomerular capillary pressure and GFR, causing vasoconstriction of postglomerular arterioles, and increasing the glomerular hydraulic pressure and filtration fraction (see Fig. 26.43). In addition, it indirectly upregulates transforming growth factor-beta (TGF-β), a potent fibrogenic cytokine, increasing collagen synthesis and epithelial cell transdifferentiation to myofibroblasts that contribute to excessive matrix formation. Angiotensin II also upregulates plasminogen activator inhibitor-1 (PAI-1), which inhibits matrix proteolysis by plasmin, with accumulation of excessive matrix and scarring in the glomeruli and interstitium.

Proteinuria itself may be harmful in the tubulointerstitium. In disease, albumin in the urinary space carries bound fatty acids, growth factors, and cytokines. When reabsorbed in the proximal tubule and degraded, these carried molecules may cause proximal tubular cell activation and interstitial scarring.

The prognosis of CKD correlates with:

- hypertension, particularly if poorly controlled
- proteinuria
- the degree of scarring in the interstitium (not glomeruli).

Inhibition of angiotensin II and reduction of proteinuria with ACE inhibitors or angiotensin-receptor antagonists slow the rate of progression of CKD. SGLT2 inhibitors have shown marked efficacy in slowing the rate of decline of kidney function in several seminal trials and are now used alongside angiotensin II inhibition. Finerenone, a non-steroidal mineralocorticoid receptor antagonist, is a new drug used in combination with both angiotensin II and SGLT2 inhibitors to slow progression of CKD.

Complications of CKD

Anaemia

Anaemia impairs quality of life and wellbeing. A normochromic, normocytic anaemia develops due to variable mechanisms:

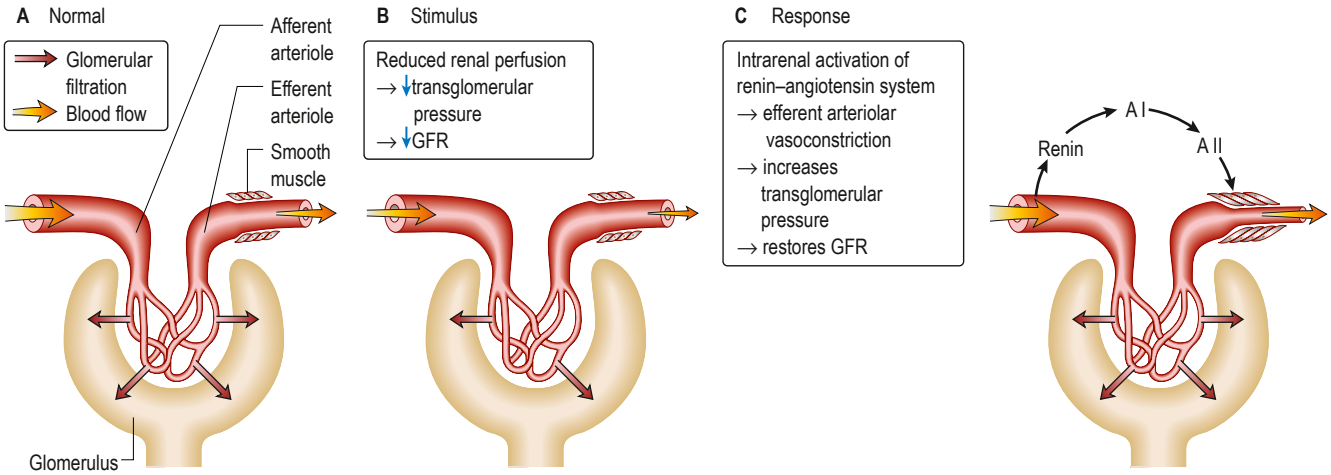


Fig. 26.43 Glomerular dynamics. Effect of the renin–angiotensin system. *AI,* Angiotensin I; *AII,* angiotensin II; *GFR,* glomerular filtration rate.

- ***Erythropoietin deficiency:*** the most significant mechanism in CKD.
- ***Increased blood loss:*** occult gastrointestinal bleeding, repeated blood sampling, blood loss during haemodialysis, or platelet dysfunction.
- ***Bone marrow dysfunction:*** toxins retained in CKD, or fibrosis secondary to hyperparathyroidism.
- ***Haematinic deficiency:*** decreased iron, vitamin B_{12}, or folate.
- ***Increased red-cell destruction:*** red cells have a shortened lifespan in uraemia, and haemodialysis may cause haemolysis.
- ***ACE inhibitors:*** interfering with the control of endogenous erythropoietin release.

Management

Synthetic (recombinant) human erythropoiesis-stimulating agents (ESAs) include epoetin-α or β, and long-acting darbepoetin-α. ESAs can be given subcutaneously or by intravenous injection. Supplemental intravenous iron may facilitate the response to erythropoietin and can be given prior to starting ESAs. Haemoglobin is monitored and the ESA dose adjusted to maintain a target haemoglobin of 100 to 120 g/L. Peripheral resistance increases in all patients, owing to loss of hypoxic vasodilation and to an increase in blood viscosity. Hypertension is a significant side-effect in 30% of patients new to ESAs requiring blood pressure monitoring and treatment if needed. Encephalopathy is rare.

Target haemoglobin is between 100 and 120 g/L. Studies in pre-dialysis CKD patients have not shown any benefit to a higher haemoglobin target (>120 g/L). Patients who ***fail to respond*** should be screened to exclude associated iron deficiency, bleeding, malignancy, infection, inflammation, or formation of anti-erythropoietin neutralizing antibodies. The increased demand for iron by the bone marrow is enormous when an ESA is started. Functional iron deficiency (poor mobilization of iron, despite adequate iron stores with ferritin >500 μg/L) is a common finding in patients with chronic inflammation. Synthesis of hepcidin, an acute phase reactant produced by the liver in response to cytokines (particularly IL-6), inhibits gastrointestinal iron absorption (see Chapter 38, 'Microcytic anaemia' section) and sequesters iron in the liver, preventing its release. Intravenous (rather than oral) iron supplements therefore optimize the response to ESAs in CKD. A trial has shown that high-dose intravenous iron and a target ferritin concentration of around 400 μg/L are safe and associated with a lower ESA requirement. Correcting anaemia with ESAs improves quality of life, exercise tolerance, and sexual and cognitive function in dialysis patients, and leads to regression of left ventricular hypertrophy. By avoiding blood transfusion, the risk of sensitization to HLA antigens, which may complicate future kidney transplantation, is reduced.

Reports of anti-erythropoietin antibody-mediated pure red-cell aplasia in patients receiving subcutaneous ESA therapy (particularly epoetin-α) have been described. Changes in the manufacturing process of the pre-filled syringes used to deliver ESAs have reduced the number of cases: the rubber stoppers interacted with the drug vehicle to act as an immunological adjuvant, stimulating anti-erythropoietin antibody production.

Oral agents that inhibit prolyl hydroxylase and prolong the life of hypoxia-inducible factor (HIF) 2α, a transcription factor for endogenous production of erythropoietin, include Roxadustat and Daprodustat. They are approved for use in predialysis CKD and patients on dialysis.

CKD mineral and bone disorder (CKD-MBD)

Once called 'renal osteodystrophy', the more appropriately termed mineral and bone disorder CKD-MBD encompasses:

- changes in calcium, phosphorus, PTH, FGF23, and vitamin D metabolism
- the various forms of bone disease that develop alone or in combination with CKD
- the vascular consequences, such as arterial stiffness and calcification.

Altered bone morphology in CKD (Fig. 26.44) includes:

- hyperparathyroid bone disease
- osteomalacia
- osteoporosis
- osteosclerosis
- adynamic bone disease.

Most patients with CKD are found on bone biopsy to have mixed bone disease. CKD-MBD is almost universal by stage G3b CKD.

Vascular calcification is common in CKD patients. Vascular smooth muscle cells, modulated by uraemia and/or phosphate, differentiate to an osteoblast-like phenotype and can synthesize and deposit matrix, which is then mineralized. Calcified vessel walls become increasingly stiff and non-compliant (leading to left ventricular hypertrophy over time), cardiac valves calcify, and soft-tissue calcium is deposited widely. There is an association between

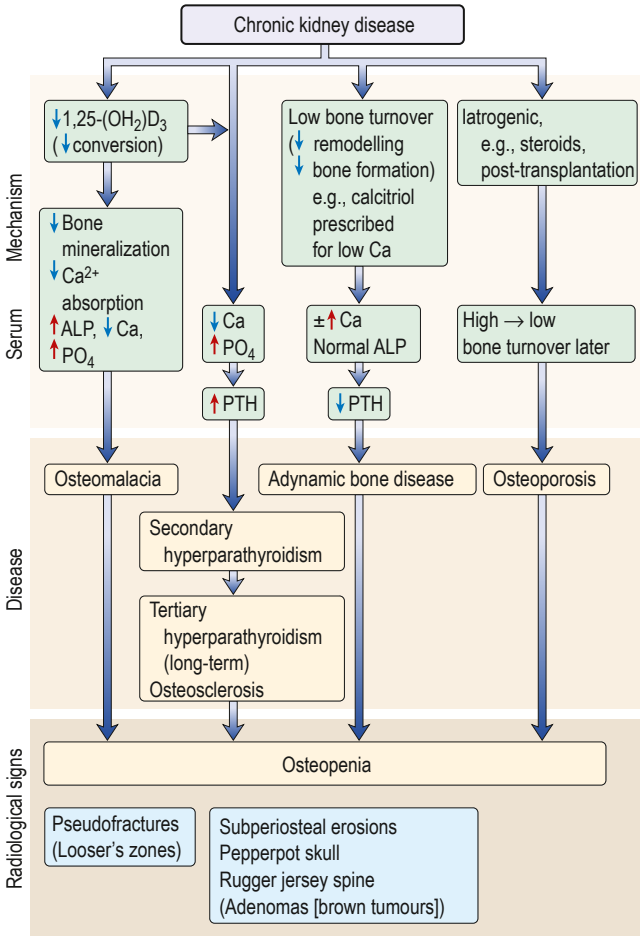


Fig. 26.44 Chronic kidney disease mineral and bone disorder (CKD-MB). Pathogenesis and radiological features of bone disease due to CKD. *ALP,* Alkaline phosphatase; *PTH,* parathyroid hormone.

vascular calcification and mortality in CKD, but a causal link has not yet been proven.

Pathogenesis

Phosphate excretion falls in the very early stages of CKD. Retained phosphate results in release of FGF23 and other phosphaturic agents by osteocytes as a compensatory mechanism. FGF23 causes phosphaturia, lowering plasma phosphate concentration to within the normal range. FGF23 also downregulates kidney 1α-hydroxylase, reducing the action of activated vitamin D in increasing intestinal absorption of phosphate. Despite consistently elevated concentrations of FGF23, phosphate concentrations in blood will increase as CKD progresses. Elevated FGF23 concentrations are the strongest independent predictor of mortality in patients with CKD. As CKD progresses, ***secondary hyperparathyroidism*** develops:

- Decreased renal production of 1α-hydroxylase results in reduced conversion of 25-$(OH)_2D_3$ to metabolically active 1,25-$(OH)_2D_3$ (1,25-dihydroxycholecalciferol).
- 1,25-$(OH)_2D_3$ deficiency results in reduced gut calcium absorption and a fall in serum calcium.
- Reduced activation of vitamin D receptors in the parathyroid glands by 1,25-$(OH)_2D_3$ increases the release of PTH.
- Calcium-sensing receptors in the parathyroid glands react rapidly to acute changes in extracellular calcium concentrations. Low calcium triggers PTH release.
- Retained phosphate also indirectly lowers ionized calcium (and probably directly via a putative but unrecognized phosphate receptor), resulting in increased PTH synthesis and release.
- PTH promotes reabsorption of calcium from bone and increases proximal tubular reabsorption of calcium. This mechanism attempts to reverse the hypocalcaemia caused by 1,25-$(OH)_2D_3$ deficiency and phosphate retention.

'Secondary' hyperparathyroidism leads to increased osteoclastic activity, cyst formation, and bone marrow fibrosis (osteitis fibrosa cystica). The typical biochemical findings are hypocalcaemia, hyperphosphataemia, an elevated PTH, and an increased serum alkaline phosphatase (marker of increased bone turnover). Radiologically, digital subperiosteal erosions and a 'pepperpot skull' are seen. Longstanding secondary hyperparathyroidism ultimately leads to hyperplasia of the glands with autonomous or ***tertiary hyperparathyroidism*** in which PTH release occurs independently of calcium or 1,25-$(OH)_2D_3$ control, and high bone turnover leads to hypercalcaemia.

Longstanding hyperparathyroidism causes increased bone density ***(osteosclerosis)***, seen particularly in the spine, where alternating bands of sclerotic and porotic bone give rise to a characteristic 'rugger jersey' radiograph appearance.

Deficiency of 1,25-$(OH)_2D_3$ and hypocalcaemia can also result in impaired mineralization of osteoid ***(osteomalacia)***.

'Adynamic bone disease' describes a state in which bone formation and resorption are depressed and the skeleton becomes inert. Bone turnover is reduced, usually where there is overtreatment with active vitamin D, low PTH (after surgical parathyroidectomy), accumulation of aluminium used as a phosphate binder, or diabetes. There may be hypercalcaemia, particularly if calcium intake is high (excess calcium cannot be laid down in bone). Serum alkaline phosphatase is normal and PTH is low. Radiographs and dual X-ray absorptiometry (DXA) scans show osteopenia. No treatment is of proven benefit.

Osteoporosis is commonly found in CKD, often after transplantation and the use of corticosteroids. Monitoring is with annual DXA scanning.

Management

The aim is to keep serum calcium and phosphate concentrations in the normal range as CKD progresses. Once dialysis is needed, PTH is kept within two- to threefold the upper limit of normal to ensure that bone turnover continues and adynamic bone does not develop. The less well-understood goal is to limit vascular calcification, by maintaining good bone health and avoiding calcium exposure.

Calcium, phosphate, and PTH should be measured 3-monthly. Imaging can define the extent of vascular calcification (e.g., spine or abdominal radiograph). Treatment and targets aim to reduce phosphate (and preferably limit calcium load), control PTH, and achieve a normal calcium.

Reduction of phosphate

- ***Dietary restriction.*** This is seldom effective alone because so many foods are phosphate-rich and phosphate-restricted diets lead to protein malnutrition.
- ***Gut phosphate binders.*** Sevelamer carbonate, lanthanum carbonate, calcium carbonate, calcium acetate, and ferric citrate all reduce phosphate absorption when taken with meals. In addition, sevelamer (unlike calcium-containing binders) attenuates vascular calcification and lowers cholesterol concentrations by 10% though it has not been shown to reduce mortality. Generally, adherence to phosphate binders is difficult.

- ***Tenapanor.*** A minimally absorbed, small-molecule inhibitor of the sodium/hydrogen exchanger isoform 3 (NHE3), inhibits both sodium and water transport locally in the gastrointestinal tract, as well as phosphate transport by indirectly inhibiting paracellular absorption of phosphate. It also downregulates intestinal phosphate transporters Na/Pi2b and reduces absorption directly. Clinically, utility has not yet been determined.

Control of PTH and achievement of normal calcium

- ***Calcitriol (1,25-dihydroxycholecalciferol) and vitamin D analogues.*** Include alfacalcidol, and novel vitamin D metabolites (22-oxacalcitriol, paricalcitol, doxercalciferol). They are given orally to suppress PTH once the concentration is three more times the upper reference limit. The newer vitamin D analogues, such as paricalcitol (19-nor-1,25 dihydroxyvitamin D_2), may be less likely to lead to hypercalcaemia, but utility over conventional and less expensive calcitriol or alfacalcidol remains to be established.
- ***Calcimimetic agents.*** Calcimimetics act by activating the calcium-sensing receptor, leading the parathyroid to respond as though serum calcium concentrations were high, reducing PTH synthesis and release. Agents include cinacalcet and etelcalcetide. These are used in secondary hyperparathyroidism to suppress PTH concentrations and lowering the calcium×phosphate product. The long-term safety and efficacy of these agents have been confirmed; however, survival benefit has not been shown.

Calciphylaxis

Also known as calcific uraemic arteriolopathy, this is a rare but serious life-threatening complication in CKD patients. It presents as painful skin patches, plaques, and ulcers, with non-healing eschars, panniculitis, and dermal necrosis. The characteristic feature on histology is vascular calcification and superimposed small-vessel thrombosis (Fig. 26.45). Risk factors include hyperparathyroidism, elevated serum phosphate, morbid obesity, and warfarin use (by inhibiting vitamin K-dependent carboxylation and activation of calcification inhibitor Matrix Gla Protein). Control of hyperparathyroidism with either surgical intervention or a calcimimetic agent is required. Sodium thiosulfate (an antioxidant) and bisphosphonates have been tried but have not been shown to be effective. Despite treatment, infection escalates risk, and as patients are often already frail, many die within a year of diagnosis.

Cardiovascular disease

CKD is a major risk factor for cardiovascular disease, and the greatly increased (16-fold) incidence of cardiovascular disease in CKD compared with the normal population has a significant impact on life expectancy. Sudden cardiac death, myocardial infarction, cardiac failure, stroke, and peripheral vascular disease all occur in excess as GFR declines (and the presence of proteinuria adds further to this risk). Kidney transplantation reverses the risk seen with stage G5 CKD.

Risk factors

- ***Hypertension:*** very common in CKD.
- ***Diabetes mellitus:*** the most common cause of CKD.
- ***Dyslipidaemia:*** universal in uraemia.
- ***Smoking:*** as common as the general population.

Ventricular hypertrophy is common in CKD, as is systolic and diastolic dysfunction. Left ventricular hypertrophy is a risk factor for early death in CKD, as in the general population. Systolic dysfunction is also a marker for early death. Myocardial fibrosis, abnormal myocyte function, calcium overload, hyperparathyroidism, and carnitine and selenium deficiencies all contribute to the systolic dysfunction seen in uraemic cardiomyopathy.

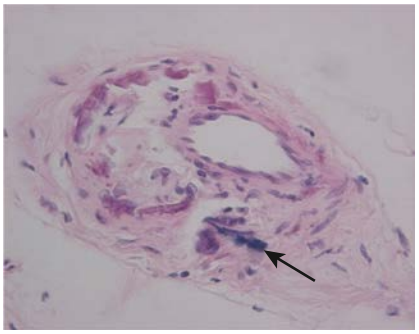


Fig. 26.45 Calciphylaxis. Characteristic arterial calcification of the skin microvasculature in the absence of vasculitic change *(arrowed)*.

Coronary artery and generalized vascular calcification

Traditional risk factors (e.g., smoking, diabetes) only partly explain the increased cardiovascular risk in patients with CKD. Coronary artery calcification is more common in patients with ESKD, and it is likely that this contributes to cardiovascular mortality. Peripheral vessel calcification increases vascular stiffness (reduced compliance), which manifests as increased pulse pressure, increased pulse wave velocity, and an increased afterload with advancing left ventricular hypertrophy. Risk factors related to calcification include:

- A raised (calcium × phosphate) product, which promotes calcification.
- Hyperparathyroidism, which increases intracellular calcium.
- Uraemia, which leads to loss of constitutive inhibitors of calcification, with vascular smooth muscle cells acquiring osteoblast-like characteristics.
- Inflammation, which inhibits fetuin: a glycoprotein synthesized by liver, and a potent inhibitor of vascular calcification.

Plain abdominal radiography (demonstrating 'pipe-stem' calcification of large arteries), electron beam or multislice CT of the coronary arteries, and vascular Doppler can all identify and quantify vascular calcification. Risk factor modification is comparable to that undertaken in patients without CKD, although clear evidence of benefit is less clear in CKD.

Other cardiovascular risk factors

Hyperhomocysteinaemia, *Chlamydophila pneumoniae* infection, malnutrition, inflammation, insulin resistance, oxidative stress, and elevated endogenous inhibitor of nitric oxide synthase and asymmetric dimethyl arginine (ADMA) all contribute to increased cardiovascular events. High ADMA concentrations in uraemia are caused in part by oxidative stress and offer an explanation for the 52% increase in risk of death and 34% increase in risk of cardiovascular events in patients with uraemia. Use of antioxidants, vitamin E, and acetylcysteine has been associated with reduced all-cause and cardiovascular mortality. However, trials to reduce concentrations of homocysteine with folic acid, B_6, and B_{12} supplementation have been unsuccessful.

The conclusion from a study of heart and kidney protection (SHARP) in over 9500 CKD patients was that around one-quarter of all heart attacks, strokes, and revascularizations could be avoided in CKD by using a combination of ezetimibe and simvastatin to

lower blood cholesterol. This combination did not, however, confer any survival advantage or prevent the development of ESKD.

The efficacy of coronary revascularization in stable coronary artery disease in patients with advanced CKD had historically been unclear as such patients were excluded from clinical trials. However, a recent randomized trial including over 750 patients with advanced kidney disease and moderate or severe ischemia on stress testing showed that an initial invasive strategy of coronary angiography and revascularization (if appropriate) added to medical therapy did not reduce death or non-fatal myocardial infarction compared to a conservative strategy of medical therapy alone and angiography reserved for only those in whom medical therapy failed. Clinical focus has therefore shifted to optimal medical therapy for ischaemic heart disease in patients with CKD, including those awaiting kidney transplantation.

Pericarditis

This usually occurs in two clinical settings:

- Uraemic pericarditis is a feature of severe, pre-terminal uraemia or of under-dialysis. Associated haemorrhagic pericardial effusion and atrial arrhythmias may occur. There is a risk of pericardial tamponade, and anticoagulants should be used with caution. Pericarditis may resolve with intensive dialysis.
- Dialysis pericarditis occurs because of an intercurrent illness or surgery in a patient receiving apparently adequate dialysis.

Skin disease

Pruritus (itching) is common and is caused by:

- accumulating nitrogenous waste products of protein catabolism (itching improves following dialysis)
- hypercalcaemia and hyperphosphataemia, or an elevated calcium × phosphate product
- hyperparathyroidism (even if calcium and phosphate concentrations are normal)
- iron deficiency.

Inadequate dialysis is a treatable cause of pruritus. However, for many, the cause of itching is unknown and no effective treatment exists.

Many patients with CKD suffer from dry skin, for which simple emollients are helpful. CKD may also cause porphyria cutanea tarda, a blistering, photosensitive skin rash. This results from a decrease in hepatic uroporphyrinogen decarboxylase combined with decreased clearance of porphyrins in the urine or by dialysis. Pseudoporphyria, a condition like porphyria cutanea tarda but without enzyme deficiency, is also seen with increased frequency in CKD.

Nephrogenic systemic fibrosis

Nephrogenic systemic fibrosis (NSF) is a systemic fibrosing skin disorder seen in patients with moderate to severe CKD (eGFR < 30 mL/min/1.73 m^2), particularly those on dialysis. Gadolinium-containing contrast agents, which are excreted exclusively by the kidney, have been implicated in over 95% of cases of NSF. Diagnosis is based on biopsy of an involved site, showing proliferation of dermal fibrocytes with excessive collagen deposition. Special testing may show gadolinium.

NSF is chronic and unremitting, with 30% having no improvement, 20% having modest improvement and 30% dying. No single or combination therapy offers benefit. Prevention is key and so high-risk (group I) gadolinium-based contrast agents should be avoided in patents with an eGFR below 30 mL/min/1.73 m^2. Lower-risk agents (from groups II and III) can be used if benefits outweigh risk.

Gastrointestinal complications

Hypergastrinaemia increases as GFR declines. Decreased gastric emptying, reflux oesophagitis, gastritis, and peptic ulceration are all common and the risk of gastrointestinal tract bleeding is increased. Constipation is particularly common in patients on continuous ambulatory peritoneal dialysis (CAPD).

Acute pancreatitis occurs more frequently, though elevations of serum amylase of up to three times the upper reference interval may be found in CKD without evidence of pancreatic disease due to retention of high-molecular-weight forms of amylase normally excreted in the urine.

Metabolic abnormalities

Gout

Urate is retained as GFR declines, and many drugs used to manage CKD increase the risk of gout. Nephrotoxic NSAIDs may be contraindicated due to CKD. Colchicine is helpful for an acute attack. Reduced-dose allopurinol or febuxostat are effective in prevention.

Insulin

Insulin is catabolized and to some extent excreted by the kidneys. Renal glucose production is also diminished with progressive CKD. For these reasons, insulin requirements in patients with diabetes decrease as CKD progresses. In contrast, end-organ resistance to insulin is a feature of advanced CKD, resulting in modestly impaired glucose tolerance. Insulin resistance may contribute to hypertension and lipid abnormalities.

Lipid metabolism

Abnormalities in lipid metabolism are common in CKD and include:

- impaired clearance of triglyceride-rich particles
- hypercholesterolaemia (particularly in advanced CKD).

The situation is further complicated in kidney failure when regular heparinization (in haemodialysis), excessive glucose absorption (in CAPD), and immunosuppressive drugs (in transplantation) all contribute to lipid abnormalities.

Endocrine abnormalities

These include:

- Hyperprolactinaemia, which may present with galactorrhoea.
- Decreased serum testosterone concentrations (seldom below normal reference limit), loss of libido, erectile dysfunction, and decreased spermatogenesis.
- Menstrual irregularities, oligomenorrhoea, or amenorrhoea.
- Complex abnormalities of growth hormone secretion and action, resulting in impaired growth in uraemic children (managed with recombinant growth hormone and insulin-like growth factor).
- Altered protein binding and increased thyroid-binding globulin loss with proteinuria make thyroid function tests difficult to interpret. Thyroid-stimulating hormone (TSH) is most commonly used to assess thyroid function. True hypothyroidism occurs with increased frequency in CKD.

Muscle dysfunction

Uraemia appears to interfere with muscle energy metabolism via an uncertain mechanism. Decreased physical fitness (cardiovascular deconditioning) is also a contributing factor.

Nervous system abnormalities

Uraemia affects on the central nervous system include depressed cerebral function, a decreased seizure threshold, asterixis, tremor, and myoclonus. Anxiety, depression, and impaired cognition also

occur. In people with very high blood urea prior to their first dialysis, sudden correction can cause ***dialysis disequilibrium:*** urea does not equilibrate rapidly across the blood–brain barrier, so blood urea falls with dialysis before urea in the central nervous system leading to rapid movement of water into the higher-urea environment, osmotic cerebral swelling and seizures. Dialysis disequilibrium can be avoided by correcting uraemia gradually using short, repeated haemodialysis treatments or by the use of peritoneal dialysis.

Increased circulating catecholamines lead to downregulation of α-receptors, impaired baroreceptor sensitivity, and impaired efferent vagal function in the autonomic nervous system. Overactivity of the sympathetic nervous system may contribute to hypertension in CKD. These abnormalities correct to some extent with regular dialysis and resolve after successful kidney transplantation.

Median nerve compression in the carpal tunnel is common, usually due to β_2-microglobulin-related amyloidosis. Advanced uraemia leads to a symmetrical polyneuropathy (which may recur in inadequately dialyzed patients).

Management of CKD

General principles

- Make a diagnosis and treat any modifiable underlying cause.
- Address cardiovascular risk factors: smoking cessation, obesity, and exercise.
- Avoid nephrotoxic drugs.
- Arrange follow-up, depending on the stage of CKD, to identify those most likely to progress early.

As stage G3 CKD progresses to stage G4 CKD, attention should be paid to correcting the complications associated with more advanced CKD and described earlier: anaemia, CKD-MBD, and metabolic abnormalities. Primary prevention of cardiovascular disease is a major part of CKD management at this stage.

Renoprotection

Preservation of kidney function and reduction in cardiovascular risk are key principles in the management of CKD (**Box 26.10**).

Correction of specific complications

Hyperkalaemia often responds to dietary restriction of potassium intake. Drugs that cause potassium retention should be stopped. Ion-exchange resins remove potassium in the gastrointestinal tract. Patiromer and sodium zirconium cyclosilicate are orally active, have been shown to control hyperkalaemia in phase 3 trials and are now available for clinical use. Emergency treatment of severe hyperkalaemia is described in Chapter 11 ('Hyperkalaemia/Management' section).

Correction of acidosis helps to address hyperkalaemia in CKD and may also decrease muscle catabolism. Sodium bicarbonate supplements (4.8 g of $NaHCO_3$ contains 57 mmol of both Na^+ and HCO_3) are effective, without significant risk of oedema or hypertension. Correction of metabolic acidosis with sodium bicarbonate at a mean dose of 1.8 g/day reduces rate of decline of GFR in advanced (stage G4 and G5) CKD. Calcium carbonate, also used as a phosphate binder, has a beneficial effect on acidosis.

Drug therapy should be reviewed and nephrotoxins avoided in patients with CKD. These include:

- tetracyclines (except for doxycycline)
- drugs excreted by the kidneys, such as gentamicin (prescribed only in the absence of any alternative, and drug concentrations monitored if feasible)
- NSAIDs

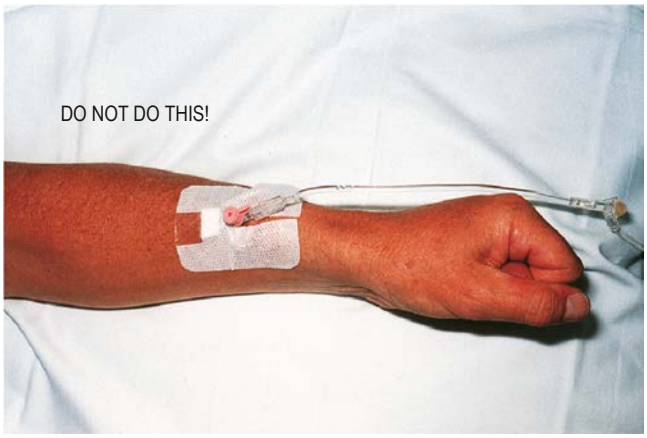


Fig. 26.46 Intravenous cannula in exactly the WRONG place. The cannula is shown in a right-handed patient with CKD, who will, in future, need a left (non-dominant arm) radiocephalic fistula.

- potassium-sparing agents, e.g., spironolactone, amiloride (and artificial salt substitutes which contain potassium).

Antibiotics, anticoagulants (particularly anti-thrombin oral anticoagulants, which currently lack a practical therapeutic monitoring), beta-blockers, oral hypoglycaemics, insulin, antidepressants, and analgesics (especially opioids, which can accumulate, leading to an opiate narcosis) often need dose adjustment in patients with CKD.

Early referral

Patients need time to adjust to the demands of CKD and its treatment, and to absorb information. Those who share in decisions and who are able to understand and choose treatments are better able to transition to dialysis or transplantation. As eGFR declines below 20 mL/min/1.73 m^2, a specialist team should counsel patients about diet, review medication, encourage risk factor modification, and support their choice about renal replacement (including the choice not to undertake dialysis). Where suitable, patients should be offered and prepared for ***pre-emptive living donor transplantation*** as the best option for RRT.

If patients opt for regular haemodialysis, an arteriovenous fistula should be fashioned in advance of the need for dialysis (with earliest planning in patients with diabetes and poor vasculature). Fistulae require several weeks to mature and become usable for vascular access. Veins required for future arteriovenous fistulae should not be rendered useless by cannulation (**Fig. 26.46**). Patients choosing peritoneal dialysis might have a buried catheter placed and left subcutaneously for some time, able to be externalized when dialysis is required.

Kidney replacement therapy

About 3 million people worldwide are either treated by haemodialysis or peritoneal dialysis or live with a functioning kidney transplant. In the UK, around 53 000 people receive treatment for kidney failure: 50% are transplanted, 42% are on haemodialysis, and 8% are on peritoneal dialysis.

Dialysis aims to mimic excretory function of normal kidneys by:

- eliminating nitrogenous and small molecular waste products
- maintaining normal electrolyte concentrations
- preventing systemic acidosis
- maintaining a normal extracellular volume.

Initiation of dialysis

It is not widely agreed *when* to start dialysis in patients with stage G5 CKD. Trials have suggested that early initiation is not associated

with improved survival or clinical outcomes. There is a general tendency to start dialysis at an eGFR close to 10 mL/min/1.73 m^2 but this decision is best shared with an individual patient, considering symptoms and life plans.

Some, particularly those who are frail or living with other comorbidities, may choose not to undergo dialysis as a shared, informed decision with their physician and families. For some, the *quantity* of extended life offered by dialysis is not matched by *quality* of life on dialysis. For others, events and complications of dialysis make the experience of treatment difficult and they choose to withdraw from dialysis. Respect for an individual's informed choice, support, and symptom control should be offered as part of end-of-life care.

Across the globe, RRT is expensive. Resources may be allocated to other pressing priorities, or available for a defined period only, or one form of RRT may available. Cost-effective, low-technology solutions to treat kidney failure are required.

Haemodialysis

Basic principles

Anticoagulated blood from a patient is pumped around an extracorporeal circuit and through a biocompatible, semipermeable membrane (the dialyzer, or 'artificial kidney') before being returned to the circulation. In the dialyzer, ultrapure dialysate flowing in the opposite direction is in close contact with blood. Small solutes (but not cells and larger-molecular-weight proteins) cross the membrane and move by diffusion down a concentration gradient (Fig. 26.47). A transmembrane pressure allows controlled fluid removal by ultrafiltration (and with fluid, more solute removal by convection). Over a standard 4-hour session, up to 80 L of blood (at around 300 mL/min) is circulated through the dialyzer. This provides an equivalent GFR of only 10 to 12 mL/min per 1.73 m^2. Box 26.30 describes a typical dialysate; dialysis is individually prescribed for each patient to obtain optimal results.

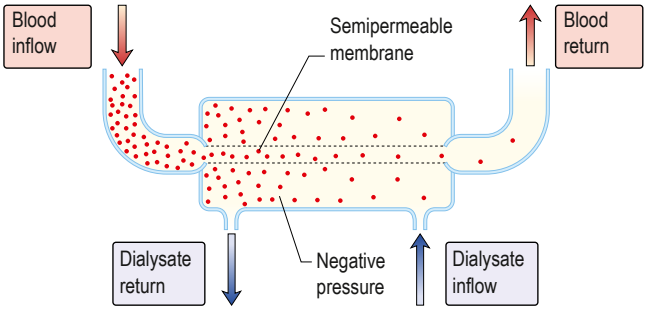


Fig. 26.47 Changes across a semipermeable dialysis membrane.

Box 26.30 Range of concentrations in standard dialysates used for haemodialysis

Dialysate	Range of concentration (mmol/L)
Sodium	130–145
Potassium	0.0–4.0
Calcium	1.0–1.6
Magnesium	0.25–0.85
Chloride	99–108
Bicarbonate	35–40
Glucose	0–10

Access

Effective dialysis needs blood flows of between 250 and 450 mL/min. To achieve this, a surgically fashioned ***arteriovenous fistula*** (AVF, Fig. 26.48) is formed using the radial or brachial artery and cephalic vein. Large-bore needles are inserted into the arterialized vein of the AVF to take blood to and from the dialysis machine. In patients with poor-quality veins or arterial disease (e.g., diabetes mellitus), synthetic arteriovenous grafts offer an alternative.

If AVF is not an immediate or viable option, a ***semi-permanent dual-lumen venous catheter*** can be inserted into the jugular or femoral vein via a subcutaneous tunnel. Although easy to place and suitable for immediate use, there is increased risk of bloodstream infection (with a foreign body directly accessing the circulation), catheter malfunction (thrombosis), and venous stenosis or occlusion compared to AVF. For urgent dialysis, a temporary (untunnelled) large-bore, double-lumen dialysis catheter may be inserted into a central vein—usually the jugular or femoral vein.

Aims

- ***Euvolaemia:*** an ideal or 'dry' weight is set for each individual patient. Each patient is weighed before each dialysis session and dialysis fluid losses are set with the aim of returning the patient to their 'dry weight' each time. Fluid gains between each dialysis session are the sum of fluid intake minus fluid losses over the interdialytic interval. For example, a patient who is anuric and drinks 1.5 L/day will gain 3 kg over 48 hours if no insensible or stool losses occur. This means they will be 3 kg above their dry weight before dialysis, and their dialysis prescription will be set to remove 3 L of fluid. In this way, dialysis delivers long-term volume and blood pressure control.
- ***Maintain electrolytes.*** A low dialysate potassium (1–3 mmol/L) means that dialysis confers a negative potassium balance (1–2 mmol/kg per session), allowing control of hyperkalaemia. Dialysate sodium is controlled (individualized between 137 and 141 mmol/L) to prevent large fluid shifts. Calcium concentration is set to prevent net calcium gain during dialysis.
- ***Prevent acidosis.*** Dialysate is buffered with bicarbonate (e.g., 36 mmol/L). This will diffuse into blood to correct acidosis during treatment.
- ***Balance dialysis frequency/duration with quality of life.*** On average, between 4 and 5 hours of dialysis, three times per week, will maintain volume and solute balance over time. However, recent evidence shows that more frequent dialysis (5–6 dialysis sessions per week) or longer treatment duration (6–8 h three times weekly) offers better quality of life, solute clearance, and cardiovascular health; though these benefits need to be balanced with the possibility of earlier loss of residual kidney function, vascular access loss, and quality-of-life impact. The provision of haemodialysis in the patient's home, rather than in a hospital setting, can help facilitate more intensive dialysis regimens. Shorter or less frequent dialysis is only sufficient in patients with ***residual kidney function*** (usually assessed by measuring 24 hours or 48 hours urine volume and clearance).
- ***Deliver enough dialysis*** (adequacy). The size, number, and nature of 'uraemic toxins' are not clear. The only true measures of dialysis adequacy are patient mortality and morbidity. Symptoms of under-dialysis are non-specific: insomnia, itch, fatigue (despite correction of anaemia), restless legs, and peripheral sensory neuropathy. Urea is used as a surrogate measure in adequacy calculations:

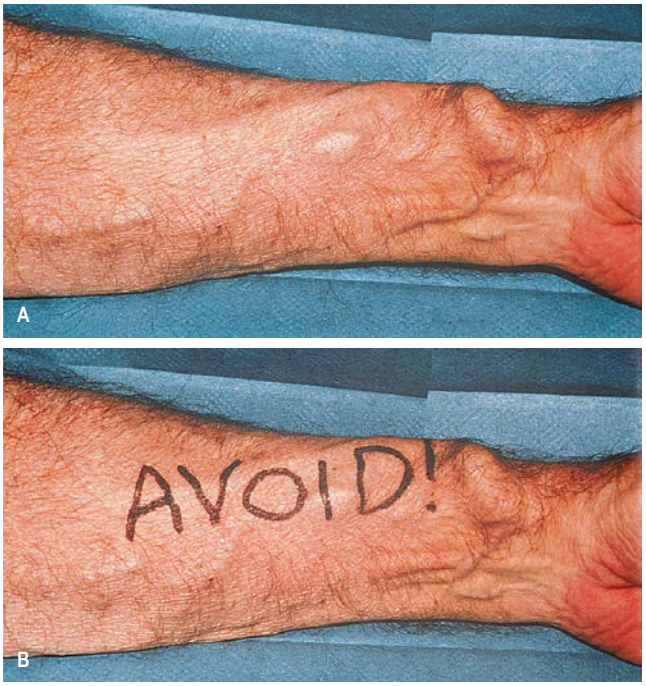


Fig. 26.48 Arteriovenous (radiocephalic) fistula. (A) Fistula in the left forearm. **(B)** A reminder for clinical teams about the presence of an existing arteriovenous fistula in a haemodialysis patient about to undergo surgery. To avoid thrombosis, the fistula should not be compressed during surgery.

- Urea reduction ratio (URR)—>65% reduction in serum urea concentration per dialysis session
- *eKt/V: K* = dialyzer clearance, *t* = duration of dialysis in minutes, *V* = urea distribution volume estimated as total body water. Minimum thresholds are required for well-nourished dialysis patients dialyzed three times per week. There are no trial data to show that a higher *eKt/V* than this minimum requirement associates with better outcomes.

Specific complications

- Access (AVF or catheter): malfunction, thrombosis, or bleeding.
- Bloodstream infections: may disseminate to soft tissue (septic arthritis), cardiac valves (endocarditis), or spinal column (vertebritis/discitis).
- Dialysis disequilibrium: dialysis is initiated with too rapid early urea removal. Subsequent movement of fluid towards the higher urea seen across the blood–brain barrier can cause cerebral oedema and seizures.
- Intradialytic hypotension: too rapid fluid removal exceeds refill of the circulation from the extravascular space.
- Dialysis-related amyloidosis: failure of clearance of β_2-microglobulin, a molecule of 11.8 kDa, leads to amyloid deposits causing median nerve compression in the carpal tunnel, dialysis arthropathy, bone cysts and fractures, pseudotumours, and gastrointestinal bleeding.

Haemofiltration

Haemofiltration differs from haemodialysis in that there is no dialysate. Plasma water (along with suspended solutes) is removed by convection across a high-flux semipermeable membrane and substitution fluid (with the desired biochemical composition) is infused to replace (large) fluid losses (Fig. 26.49). Haemofiltration may be preferred in the critical care setting, where haemodynamic instability is common. Modern dialysis machines have capacity to generate online ultra-pure water, which has minimized cost. It also means that filtration can be combined with dialysis as ***haemodiafiltration***, which increases middle molecule clearance (e.g., β_2-microglobulin) and prevents long-term dialysis complications in non-transplantable patients. Post-hoc analysis of several trials and recent randomized trial data suggest that haemodiafiltration confers a survival advantage, compared to haemodialysis.

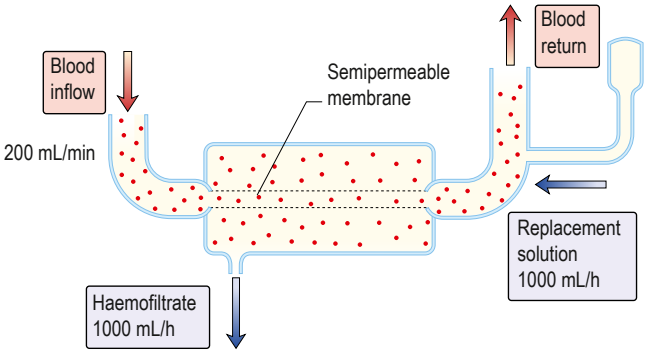


Fig. 26.49 Principles of haemofiltration.

Peritoneal dialysis

Peritoneal dialysis uses the peritoneal membrane as a semipermeable membrane, avoiding the need for extracorporeal circulation of blood. It is a simple, low-technology, effective treatment (Fig. 26.50).

- A soft catheter is placed through a skin tunnel into the peritoneal cavity through the midline of the anterior abdominal wall (Fig. 26.51).
- Dialysate is run into the peritoneal cavity, usually under gravity.
- Urea, creatinine, phosphate, and other uraemic toxins pass into the dialysate down their concentration gradients.
- Water (with solutes) moves into the peritoneal cavity by osmosis. Osmolarity, and therefore fluid removal, is controlled by the glucose or polymer (icodextrin) content of the dialysate (Box 26.31). More hypertonic solutions improve fluid removal.
- The dialysate fluid is exchanged regularly to repeat the process.

Continuous ambulatory peritoneal dialysis (CAPD): dialysate fluid is always present in peritoneal cavity and exchanged three to five times a day. Exchanges are performed by the patient (or trained carer) using a sterile no-touch technique to connect 1.5 to 3 L bags of dialysate to the peritoneal catheter. Each exchange takes 20 to 40 minutes.

Automated peritoneal dialysis (or nightly intermittent peritoneal dialysis): an exchange machine performs continuous low-volume exchanges at night while the patient is asleep. Dialysate can be left in the peritoneal cavity after the last exchange, adding extra dialysis time during the day.

Specific complications

- Bacterial peritonitis: fever, abdominal pain, and a cloudy peritoneal dialysate effluent (>100 cells/mm^3 on microscopy is suggestive) progressing to frank peritonitis. Rate of infection is about one episode every 2 patient-years. Peritoneal effluent is sent for culture and empirical antibiotic treatment is started while sensitivities are awaited. Common causative organisms are listed in Box 26.32.
- Catheter exit site infections: may progress to skin tunnel infections and peritonitis.

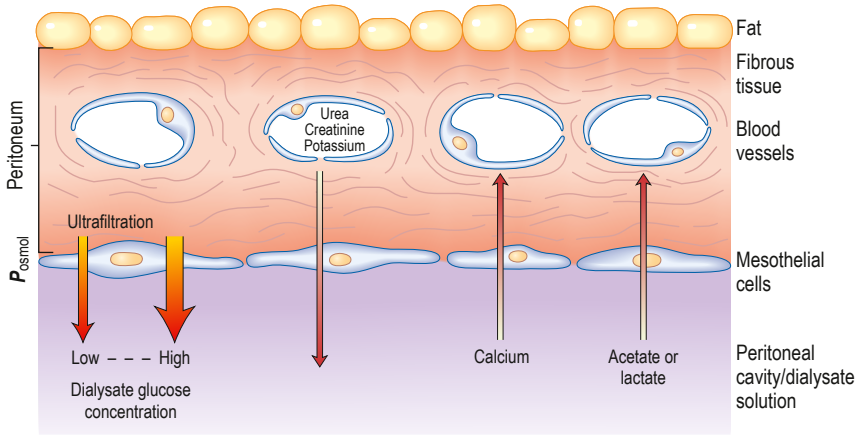


Fig. 26.50 Principles of peritoneal dialysis. Water moves into the peritoneal cavity, depending on the osmolarity of the dialysate.

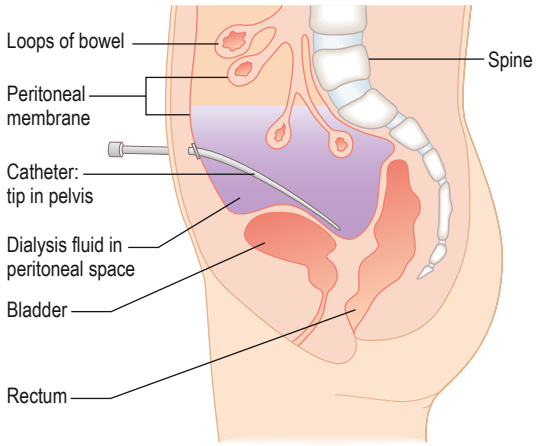


Fig. 26.51 The siting of a Tenckhoff peritoneal dialysis catheter.

- Constipation: may impair flow of dialysate in and out of the pelvis.
- Hernias: caused by raised intra-abdominal pressure. Dialysate can then 'leak' into the pleural cavity or scrotum (down a patent processus vaginalis).
- Sclerosing peritonitis: a potentially fatal complication in which long-term patients develop progressive thickening of the peritoneal membrane. This causes adhesions and strictures and repeated episodes of small bowel obstruction.

Adequacy

Most clinicians aim for a weekly *Kt/V* of 2.0 (see earlier), coupled with a creatinine clearance of 60 L per week. Once patients stop passing urine (and have no residual kidney function), peritoneal dialysis inadequacy becomes common, and a switch to haemodialysis may be required.

Dialysis in the frail

Dialysis can prolong life but individual patient benefit, especially for those who are frail, varies widely. Dialysis outcomes are poor in patients who are old and frail. Small studies suggest that mortality and quality-of-life outcomes are not improved by dialysis. In one study, over 50% died within the first year of initiating dialysis and around 30% had a decrease in functional status. Medical management of symptoms and end-of-life care should be provided for those who chose not to dialyze.

Box 26.31 Range of concentrations in standard continuous ambulatory peritoneal dialysates

Dialysate	Range of concentration (mmol/L)
Sodium	130–134
Potassium	0
Calcium	1.0–1.75
Magnesium	0.25–0.75
Chloride	95–104
Lactate	35–40
Glucose[a]	77–236
Total osmolality	356–511 mOsm/kg

[a]Glucose content is often expressed as g/dL of anhydrous glucose (e.g., 1.36% = 77 mmol/L). An even more hypertonic dialysate (6.36%) is available for acute (intermittent) peritoneal dialysis.

Box 26.32 Causes of continuous ambulatory peritoneal dialysis peritonitis

Cause	Approximate % of cases
Staphylococcus epidermidis	40–50
Gram-negative organisms: *Escherichia coli*, *Pseudomonas*, other	25
Staphylococcus aureus	15
Mycobacterium tuberculosis	2
Candida and other fungal species	2
No bacteria detected	20

Kidney transplantation

Successful kidney transplantation offers the potential for almost complete rehabilitation in ESKD.

- Survival is significantly better compared to dialysis patients on transplant waiting lists.
- Transplantation allows freedom from dietary and fluid restrictions.
- Anaemia and infertility are corrected.
- Need for parathyroidectomy is reduced.

It is the treatment of choice for all *appropriate* patients with ESKD. However, demand exceeds the supply of donor organs (in the UK, 44 per million population per year). Donor organs are a scarce and valuable resource that must be used optimally. The possibility of living donors, who can donate one kidney without impact on their long-term health, should be explored for all patients.

Kidney transplantation involves the anastomosis of an explanted human kidney on to the iliac vessels of the recipient (Fig. 26.52). The donor ureter is placed into the recipient's bladder. Unless the donor is genetically identical (i.e., an identical twin), immunosuppressive treatment is needed for as long as the transplant remains in place to prevent rejection. Patient and graft survival has steadily improved with:

- appropriate patient selection and assessment prior to transplantation
- better donor–recipient compatibility
- improvements in surgical techniques
- safer immunosuppressive regimens.

Some 80% of grafts survive for 5 to 10 years in the best centres, and 50% for 10 to 30 years. However, the half-life of kidney transplants is 13 to 16 years. The three most common causes of late graft loss are death of a patient with a functioning graft, recurrence of the original (or new) kidney disease, and chronic allograft nephropathy.

Considerations in successful kidney transplantation

ABO (blood group) compatibility between donor and recipient is preferred. ABO-incompatible kidney transplants (where donor and recipient have different blood groups) are successful with directed immunosuppression (including immunoadsorption to remove preformed antibodies, anti-CD20 antibodies to remove B lymphocytes, and intravenous pooled immunoglobulins for immunomodulation or anti-idiotypic antibodies).

Donor and recipient HLA mismatches should ideally be minimized. Nationwide matching schemes exist for kidneys retrieved from deceased donors. Complete compatibility at A, B, and DR loci offers the best long-term outcomes when compared to mismatched organs (i.e., antigens possessed by the donor and not possessed by the recipient). However, transplanting completely mismatched kidneys is performed routinely. With living donation, outcomes with a mismatched kidney are as good as, if not better than, matched deceased donor kidneys.

Pre-formed anti-HLA antibodies in recipients, resulting from sensitization from prior blood transfusions, kidney transplants, or pregnancy, predict less good outcomes. This is particularly true if existing antibodies are donor-specific (recipient anti-HLA antibodies against donor antigens). Transplantation can still be successful but pre-transplant strategies are necessary to minimize these antibodies. These include plasmapheresis, intravenous immunoglobulin, anti-B cell monoclonal antibody (rituximab), and IgG endopeptidase (which cleaves off IgG).

Most countries allow the removal of kidneys and other organs from patients who have suffered irretrievable brain damage ('brainstem death') while the heart is still beating (donation after brainstem death, DBD). Due to the shortage of solid organs and increasing numbers of patients waiting for them, several countries now allow the retrieval of organs after cardiac death (donation after cardiac death, DCD), with comparable results to 'heart-beating' donations. Increasing use is also made of expanded criteria donors (ECDs) that are aged over 60 years, or 55 and 59 years with co-morbidity including hypertension, diabetes, pre-retrieval AKI, and intracranial haemorrhage.

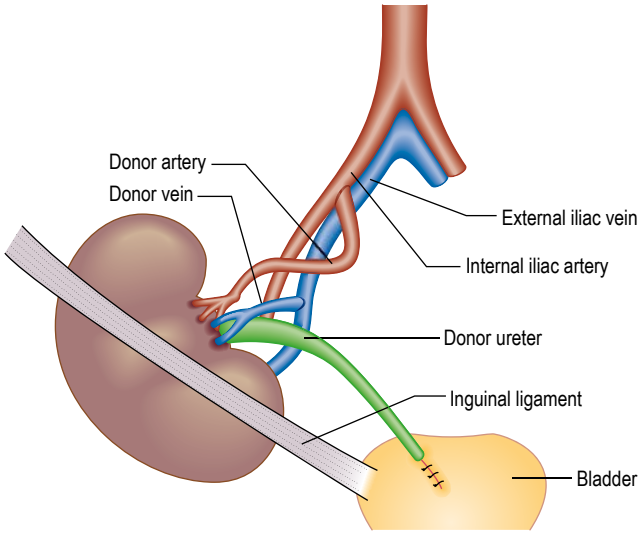


Fig. 26.52 Anatomy of a kidney transplant operation.

Living donation offers the best kidney outcomes. Potential living donors undergo comprehensive preoperative evaluation to ensure that they will come to no harm through donating a kidney to another person. Recent evidence suggests that live donors carry a statistically increased risk of CKD/ESKD post donation, but this risk is too low to be of clinical significance. All living donors should be consented carefully and monitored regularly.

Timing transplantation so that it takes place prior to dialysis initiation (pre-emptive transplantation) offers benefits to both recipient and graft.

Immunosuppression for transplantation

Long-term inhibition of the recipient immune system is needed to prevent 'non-self' immune-mediated injury to transplants. This is almost universally the case, except where living-related donation from an identical twin occurs. Some degree of immunological tolerance does develop, and the risk of rejection is highest in the first 3 months after transplantation. In the early months, rejection episodes occur in less than 20% of cadaver kidney recipients on current immunosuppression protocols, and most cases are reversible. A combination of immunosuppressive drugs is usually used (Box 26.33). Individualizing immunosuppression to a specific recipient remains an inexact science and preventing the complications of over-immunosuppression is as necessary as preventing rejection. Therapeutic drug monitoring is useful in delivering intense immunosuppression in the early post-transplant phase, with reduced drug doses/concentrations later.

Early complications

Early (technical) failure

Occlusion or stenosis of the arterial anastomosis, occlusion of the venous anastomosis and urinary leaks owing to damage to the lower ureter or to defects in the anastomosis between ureter and recipient bladder can occur.

Kidney transplant acute tubular necrosis

Delayed graft function resulting from ATN is the most common cause of deceased donor graft dysfunction (up to 40%–50%), particularly in kidneys from DCD or ECD donors (see earlier). Hypotension or loss of cardiac output prior to the death of the donor impacts the retrieved kidney. A prolonged 'cold ischaemia time' (the

Box 26.33 Immunosuppressive drugs used in kidney transplantation

Class	Examples	Mechanism of action	Clinical role	Side-effects
Calcineurin inhibitors	Ciclosporin Tacrolimus	Disrupt T-cell signalling	Used in most regimens	Nephrotoxicity (monitor concentrations) Hypertension Diabetes Hirsutism Virilization
Inhibitors of purine synthesis	Azathioprine Mycophenolate mofetil (MMF)	Inhibit purine synthesis and cell proliferation (especially lymphocytes)	Used in most regimens	Neutropenia Pancytopenia Deranged LFTs (azathioprine) Diarrhoea (MMF) Interaction with allopurinol (azathioprine)
Steroids	Prednisolone (oral) Methylprednisolone (i.v.)	Inhibit cytokine-regulated T-cell signalling	Used in most regimens. Dose tapers over weeks/months after transplantation. Methylprednisolone used on induction and to treat acute rejection	Osteoporosis Hypertension Diabetes Weight gain Poor wound healing Lipid abnormalities
Rapamycin inhibitor	Sirolimus	Inhibits cytokine-dependent cell proliferation	Alternative to calcineurin inhibitors	Delayed graft function Myelosuppression Impaired wound healing Thrombocytopenia Monitor concentrations
Anti-CD25 antibodies	Daclizumab Basiliximab	Monoclonal antibody, blocking IL-2	Given on induction	Well tolerated
Antibodies causing T-cell depletion	Anti-thymocyte globulin (ATG, polyclonal) OKT3 anti-CD3 antibody (monoclonal)	Destroy T cells	For steroid-resistant rejection May be used as induction agent for patients at high risk of rejection	Severe T-cell depletion (risk of sepsis) Late development of malignancy, e.g., lymphoma
Anti-CD52 antibody	Alemtuzumab	Depletes T and B cells	Used as induction agent	Risk of sepsis Malignancy
Anti-B7 antibody	Belatacept	Prevents engagement of B7 and CD28 Blocks co-stimulatory signal	Alternative to calcineurin inhibitors	Rejection Post-transplant lymphoma in EBV-negative patients
Anti-C5a antibody	Eculizumab	Inhibits complement activation by blocking activated C5	Post-transplant atypical HUS Antibody-mediated rejection	Infections, particularly meningococcal meningitis. Vaccinated against meningitis prior to use

EBV, Epstein–Barr virus; *HUS,* haemolytic uraemic syndrome; *IL-2,* interleukin 2; *LFTs,* liver function tests.

period during which the retrieved organ is cooled on ice in transit and awaits implantation) also leads to delayed graft function due to ATN. Finally, calcineurin inhibitors used to prevent rejection are themselves nephrotoxic and high peak and trough concentrations cause tubular injury. Kidneys transported on a perfusion machine, and controlled hypothermia in deceased donors prior to organ retrieval are associated with lower rates of delayed graft function.

Acute rejection

Acute rejection is seen in between 10% and 30% of transplant recipients and usually presents with declining kidney function within the first 3 months. Transplant biopsy (Fig. 26.53) confirms the diagnosis, severity, and type of rejection. It can be T cell or antibody-mediated and occur with or without endothelial injury ('vascular rejection'). Urinary measurement of RNA can be used as a marker of acute cellular rejection.

Cellular rejection may respond to high-dose intravenous corticosteroids, increased or switched calcineurin inhibition, or the use of T-cell–depleting agents such as anti-thymocyte globulin (ATG) or anti-CD52 antibody (Campath). Antibody-mediated rejection (diagnosed by the presence of circulating donor-specific anti-HLA antibodies and evidence of complement activation on kidney biopsy by C4d staining, Fig. 26.53B) is usually treated empirically by a combination of intravenous immunoglobulin (to neutralize and promote the clearance of anti-HLA antibodies), plasmapheresis (to remove antibodies), and anti-CD20 antibody administration (to deplete B lymphocytes), with variable success.

More than one episode of rejection within the first 3 months, vascular and/or antibody-mediated rejection, delayed rejection, requirement of dialysis within the first week after transplantation, and failure of serum creatinine to return to baseline after a rejection episode are associated with worse long-term outcomes.

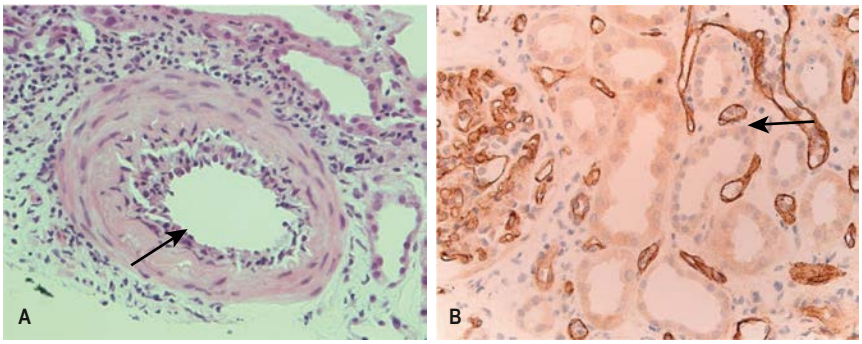


Fig. 26.53 Histology of kidney transplant in rejection. **(A)** Acute vascular rejection: the mononuclear inflammatory cell infiltrate is limited to the expanded intima and does not involve the entire vascular wall as in systemic vasculitis. The arrow indicates endothelialitis. **(B)** Antibody-mediated rejection: an acute tubular necrosis with capillary glomerulitis, and peritubular and glomerular capillary C4d positivity *(arrowed)*.

Infection

- Bacterial infections occur early (<1 month post transplant) and usually involve the urinary tract, wound, or chest.
- Cytomegalovirus infection develops weeks or months after transplantation. It occurs in 70% of cytomegalovirus-seronegative recipients from seropositive donors, and in patients receiving biological agents as induction or therapy for rejection, unless prophylaxis with valganciclovir is given.
- Opportunist infections including *Pneumocystis jirovecii* occur and prophylactic co-trimoxazole is given in the early months.
- Infections in the months after transplantation and are associated with heavier immunosuppressant burdens.
- Polyomavirus infection ***(BK nephropathy)*** can cause an aggressive TIN that may lead to graft loss. Immunosuppression is minimized in the hope of spontaneous viral clearance.

Late complications

- ***Post-transplantation lymphoproliferative disorders***, often associated with Epstein–Barr virus, are more frequent in patients who received profound immunosuppression. Immunization is minimized and standard chemotherapy offered, with mixed outcomes.
- ***Cancers*** which occur with increased frequency in patients receiving long-term immunosuppression include skin (basal and squamous cell), kidney, cervical, and vaginal malignancies.
- ***Cardiovascular disease*** causes 50% of patient deaths after transplantation. Hypertension, obesity, diabetes, insulin resistance, and lipid disorders all play a role, as does the history of CKD.
- ***Post-transplant osteoporosis*** may occur because of steroid use.
- ***Recurrent kidney disease*** is common. Primary FSGS often recurs and is a cause of early graft loss. Mesangiocapillary GN, diabetic nephropathy (unless simultaneous pancreas transplantation), and IgA nephropathy also commonly recur, with variable effects on long-term transplant survival.

Most grafts fail eventually through lasting immunological injury and/or immunosuppressant drug toxicity. A common histological finding is the combination of interstitial fibrosis and tubular atrophy (historically termed 'chronic allograft nephropathy'). There is a growing understanding of the role of subclinical chronic antibody-mediated rejection due to existing or *de novo* donor-specific antibodies, but non-immunological factors are also important. Progressive irreversible decline in graft function is associated with mild to modest proteinuria (<3 g/day), and interventions to change this trajectory currently lack evidence of benefit.

Further reading

Abramyan S, Hanlon M. Kidney transplantation. [Updated 2023 Jan 2]. In: *StatPearls* [Internet]. Treasure Island (FL): StatPearls Publishing; 2023. Available from: https://www.ncbi.nlm.nih.gov/books/NBK567755/.
Kovesdy CP. Epidemiology of chronic kidney disease: an update 2022. *Kidney Int Suppl (2011)* 2022; 12(1):7–11.
Sawinski D; Poggio ED. Introduction to kidney transplantation: Long-term management challenges. *CJASN* 2021; 16(8):1262–1263.
Thongprayoon C, Hansrivijit P, Leeaphorn N et al. Recent advances and clinical outcomes of kidney transplantation. *J Clin Med* 2020 Apr 22;9(4):1193.

CYSTIC KIDNEY DISEASE

Solitary or multiple kidney cysts are common, especially with advancing age. One or more kidney cysts are found in 50% of those aged over 50 years. Cysts are often asymptomatic and are found on ultrasound examination performed for some other reason. They have no clinical significance, except in the differential diagnosis of kidney tumours. Occasionally, they may cause pain and/or haematuria if large, or bleeding may occur into a cyst. Cystic degeneration (the formation of multiple cysts that enlarge with time) occurs regularly in the non-functioning kidneys of patients with ESKD treated by dialysis and/or transplantation. These acquired cysts have malignant potential.

Autosomal dominant polycystic kidney disease

Autosomal dominant polycystic kidney disease (ADPKD) is an inherited disorder that usually presents in adulthood. It is characterized by the development of multiple kidney cysts, variably associated with extrarenal (mainly hepatic and cardiovascular) abnormalities. ADPKD is the most common inherited nephropathy, with a prevalence rate ranging from 1:400 to 1:1000. It accounts for 3% to 10% of all patients commencing regular dialysis in high-income countries.

Mutations in *PKD1* (on chromosome 16) are responsible for 85% of cases, with mutations in *PKD2* (on chromosome 4) accounting for the remainder. These genetic abnormalities are distinct from autosomal recessive polycystic disease (due to mutations in the *PKHD1* gene on chromosome 6p21.1-p12), which is often lethal in early life.

PKD1 encodes polycystin 1, involved in cell–cell and/or cell–matrix interactions. *PKD2* encodes polycystin 2, which functions as a calcium ion channel. The polycystin complex occurs in cilia, responsible for sensing flow in the tubule. Disruption of the polycystin pathway results in reduced cytoplasmic calcium, which causes defective ciliary signalling and disoriented cell division in principal cells of the collecting duct leading to cyst formation. Extra-renal manifestations intracranial aneurysms, which can be associated with significant morbidity. Missense mutations in *GANAB*, encoding glucosidase II subunit α (GIIα) are associated with polycystic liver disease. The *GANAB* gene product GIIα is required for maturation and surface and ciliary localization of the ADPKD proteins (PKD1 and PKD2).

Progressive loss of kidney function is usually attributed to mechanical compression, apoptosis of the healthy tissue, and reactive fibrosis. Kidney function declines at a variable rate, depending on growth and size of cysts. Patients with rapidly growing cysts on MRI lose kidney function more rapidly.

Clinical features

Symptoms can occur from the second decade:

- loin pain and/or haematuria due to haemorrhage into a cyst, cyst infection, or stone formation
- loin or abdominal discomfort as the size of the kidneys increases
- subarachnoid haemorrhage associated with berry aneurysm rupture
- complications of hypertension
- complications of liver cysts (in around 50% of those with cysts)
- symptoms of anaemia and/or uraemia associated with CKD.

Complications and associations

- Pain from large cysts can be difficult to manage. Surgical decompression may offer some benefit in about two-thirds of patients.
- Cyst infection requires lipophilic antibiotics that penetrate infected cysts and are active against Gram-negative bacteria.
- Kidney calculi occur in 10% to 20%. They are often composed of uric acid and are radiolucent.
- Hypertension is an early and common feature. Intrarenal activation of the renin–angiotensin system is thought to be contributory, so ACE inhibitors and AII-RAs are logical first-line treatments. Early control of blood pressure is essential as left ventricular hypertrophy is common, and cardiovascular complications are a major cause of death.
- Progressive CKD is the most serious complication. At GFRs below 50 mL/min, the rate of decline in GFR averages 5 mL/min/1.73 m^2 each year, which is more rapid than in other primary kidney disorders. The probability of being alive without requiring dialysis or transplantation by the age of 70 years is in the order of 30%. Survival rates on haemodialysis and after kidney transplantation are better than those in patients with other primary kidney diseases.
- Approximately 30% of patients have hepatic cysts that can be complicated by massive enlargement with pain, and infection. More rarely, compression of the bile duct, portal vein, or hepatic venous outflow occurs.
- Asymptomatic intracranial aneurysm are found in about 10%, but prevalence is twice as high if there is a family history of aneurysms or intracranial haemorrhage. Screening for intracranial aneurysm in ADPKD is recommended for patients aged 18 to 40 years who have a positive family history.
- Mitral valve prolapse is found in 20%.

Fig. 26.54 Ultrasound image of a polycystic kidney. There is almost complete replacement of normal kidney tissue.

Box 26.34 Ultrasonographic diagnostic criteria for testing individuals at risk of autosomal dominant polycystic kidney disease (ADPKD)

Diagnostic criteria

- 15–39 years: ≥3 cysts (unilateral or bilateral)
- 40–59 years: ≥2 cysts in each kidney
- ≥60 years: ≥4 cysts in each kidney

ADPKD is excluded in patients meeting the following criteria:

- ≥40 years: <2 cysts
- 30–39 years: 0 cysts (excludes ADPKD in 98% of cases)

Negative ultrasound < 30 years:

- Unable to exclude ADPKD
- CT/MRI may have higher sensitivity for cysts
- Molecular genetic testing (if known mutation)

Diagnosis and screening

Physical examination commonly reveals large, irregular kidneys and possible hepatomegaly. Definitive diagnosis is established by ultrasound examination (Fig. 26.54 and Box 26.34). However, imaging techniques may be equivocal, especially in subjects under the age of 15 years. Several conditions can mimic the clinical and radiological appearance of ADPKD (Box 26.35).

Children and siblings of patients with ADPKD are offered screening. Affected individuals should have regular blood pressure checks and should be offered genetic counselling, including pre-implantation genetic diagnosis for females planning pregnancy. Gene linkage analysis can be utilized in many families.

Management

- ***Vasopressin receptor antagonists:*** inhibit cAMP in principal cells, reduce cyst growth, and slow progression of kidney disease. Approved for use if eGFR declines >3 mL/min/1.73 m^2 per year. Reversible liver toxicity requires close monitoring.
- ***Octreotide:*** a long-acting somatostatin analogue (which also inhibits cAMP) offers benefit in slowing growth of both liver and kidney cysts. However, lanreotide, an alternative somatostatin analogue, did not slow decline in kidney function in more than 300 patients with eGFR 30 to 60 mL/min/1.73 m^2 over 2.5 years of follow-up, despite a measurable effect on the rate of cyst growth.

Box 26.35 Conditions that clinically and radiologically mimic polycystic kidney disease

- Tuberous sclerosis: includes lymphangiomyomatosis and lymphangioleiomyomatosis, and can be associated with the *TSC2–PKD1* contiguous gene syndrome
- Von Hippel–Lindau syndrome
- Multicystic dysplastic kidney: a non-heritable unilateral syndrome or a systemic syndrome (e.g., prune belly syndrome)
- Juvenile nephronophthisis and medullary cystic kidney disease: can be autosomal recessive (presenting in the second decade of life) or autosomal dominant (presenting in the third to fourth decades)
- Glomerulocystic kidney disease: involves the expansion of Bowman's space, autosomal dominant
- Acquired cystic disease: associated end-stage kidney disease and hypokalaemia
- Renal-cell carcinoma: with cystic changes

- ***Blood pressure control:*** In patients eGFR > 60 mL/min/1.73 m^2, lower blood pressure (95–110/60–75 mmHg) is associated with slower increase in kidney volume, less albuminuria, and lower left ventricular mass compared to standard blood pressure (120–130/70–80 mmHg). However, the addition of an angiotensin receptor blocker to an ACE inhibitor (dual blockade) does not confer additional benefit.

Medullary cystic disease ('juvenile nephronophthisis')

Juvenile nephronophthisis is an autosomal recessive (or sporadic) condition which develops early in childhood. Mutations in the *NPHP1–4* genes, encoding nephrocystin and inversin (both co-localized in the cilia of the tubules) lead to multiple cyst formation. Polyuria, polydipsia, and growth retardation result from impaired tubular function. Medullary cystic disease is a similar condition that develops later in childhood, inherited as an autosomal dominant trait. Histological findings include interstitial inflammation and tubular atrophy, with later development of medullary cysts. Diagnosis is based on family history and kidney biopsy, and the cysts rarely being visible on imaging.

Medullary sponge kidney

Medullary sponge kidney is an uncommon cystic condition presenting with intermittent renal colic, passage of small stones, and haematuria. Although it is most often sporadic, affected families have been reported. Dilation of the collecting ducts in the papillae occurs, sometimes with cystic change. In severe cases, the medullary area has a sponge-like appearance. The condition may affect one or both kidneys, or only part of one kidney. Cyst formation is commonly associated with the development of small calculi within the cyst. In about 20% of patients, there is associated hypercalciuria or renal tubular acidosis. Hemi-hypertrophy of the skeleton has been described in this condition. CT or IVU (see **Fig. 26.25**) is diagnostic. Kidney function is usually well maintained. CKD is unusual, except where obstructive nephropathy develops owing to the presence of stones in the pelvis or ureters.

Further reading

Bae KT, Shi T, Tao C, et al. Expanded imaging classification of autosomal dominant polycystic kidney disease. *J Am Soc Nephrol* 2020; 31:1640–1651.
Meijer E, Visser FW, van Aerts RMM et al. Effect of lanreotide on kidney function in patients with autosomal dominant polycystic kidney disease: the DIPAK 1 Randomized Clinical Trial. *JAMA* 2018; 320:2010–2019.
Müller RU, Messchendorp AL, Birn H, et al. An update on the use of tolvaptan for ADPKD: consensus statement on behalf of the ERA Working Group on Inherited Kidney Disorders (WGIKD), the European Rare Kidney Disease Reference Network (ERKNet) and Polycystic Kidney Disease International (PKD-International). *Nephrol Dial Transplant* 2022; 37:825–839.

TUMOURS OF THE KIDNEY AND GENITOURINARY TRACT

Renal cell carcinoma

Renal cell carcinomas (RCCs) arise from proximal tubular epithelium. They are the most common kidney tumour in adults, accounting for 1% to 2% of all malignancies, and affect males more often than females (2:1). They usually present after 50 years of age and rarely before the age of 40. RCCs are highly vascular tumours. Microscopically, most are composed of large cells containing clear cytoplasm.

Clinical features

- Often asymptomatic and discovered incidentally
- Haematuria, loin pain, flank mass
- Malaise, anorexia, and weight loss (30%)
- Polycythaemia (5%)
- Hypertension (30%) due to renin secretion by the tumour
- Anaemia (30%) due to depression of erythropoietin
- Pyrexia (20%)
- Symptoms of metastases
- A left-sided varicocele due to renal vein invasion and testicular vein obstruction by a left-sided tumour (rare)

Diagnosis

Ultrasound may demonstrate a solid lesion. CT scanning is used to identify the lesion and involvement of the renal vein or inferior vena cava. MRI is used for tumour staging. Urine cytology for malignant cells is of no value. The ESR is usually raised. Liver biochemistry may be abnormal, returning to normal after surgery.

Management

A nephrectomy is performed unless bilateral tumours are present or the contralateral kidney functions poorly, in which case conservative surgery, such as partial nephrectomy, may be indicated. If metastases are present, nephrectomy may still be warranted since regression of metastases has been reported after removal of the main tumour mass. Medical management of RCC is discussed in Chapter 41 ('Renal cancer' section).

Von Hippel–Lindau (VHL) disease is an autosomal dominant disorder presenting with bilateral RCCs, retinal and cerebellar haemangioblastomas, phaeochromocytoma, pancreatic neuroendocrine tumours, and kidney cysts. VHL protein suppresses tumour activity by tagging hypoxia-inducible gene products (VEGF, TGF) for degradation. Inactivating mutations of the *VHL* gene lead to uncontrolled angiogenesis and neoplasia. Mutations of the same tumour suppressor gene are responsible for RCC and VHL.

Nephroblastoma (Wilms' tumour)

Usually seen within the first 3 years of life, Wilms' tumour presents as an (occasionally bilateral) abdominal mass, rarely with haematuria. Diagnosis is established by ultrasound, CT, or MRI. A combination of nephrectomy, radiotherapy, and chemotherapy has improved survival rates, even in children with metastatic disease. Overall, the 5-year survival rate is 90%.

Urothelial tumours

The calyces, pelvis, ureter, bladder, and urethra are lined by transitional cell epithelium. Transitional cell cancers account for about 3% of cancer mortality. They are uncommon below 40 years of age, and more common in males than females (4:1). Bladder tumours are

most prevalent, about 50 times as common as tumours of the ureter or renal pelvis. Predisposing factors include:

- cigarette smoking
- exposure to industrial carcinogens: β-naphthylamine and benzidine (petroleum, chemical, cable, and rubber industry workers), aristolochic acid (herbal weight-loss preparations)
- medication: phenacetin, cyclophosphamide
- chronic inflammation: schistosomiasis (usually squamous carcinoma).

Clinical features

Painless haematuria is the most common (80%) presenting symptom of bladder malignancy. Pain may occur from clot retention or local nerve involvement. Cystitis symptoms occur in the absence of significant bacteriuria. Flank pain may occur with urinary tract obstruction in ureteric lesions.

Investigations

- Urine cytology for malignant cells
- Urinary tumour markers
- Cystoscopy to assess tumour burden and for biopsy
- CT or MRI of the pelvis

Management

Medical management of urothelial and testicular tumours is outlined in Chapter 41 ('Urothelial [bladder and ureteric] cancer' and 'Testicular and ovarian germ cell tumours' sections).

DISEASES OF THE PROSTATE GLAND

Benign prostatic enlargement is discussed in Chapter 47 ('Benign conditions of the prostate' section). The pathogenesis, diagnosis, and prognosis of prostate cancer are described in Chapter 41 ('Prostate cancer' section).

THE URINARY TRACT IN THE ELDERLY

Progressive sclerosis of glomeruli occurs with ageing. Together with atheromatous renovascular disease, this accounts for a progressive reduction in GFR with age. A GFR of 50 to 60 mL/min/1.73 m^2 (about half the reference value for a young adult) may therefore be regarded as 'normal' in patients over 80 years of age. A reduction in muscle mass with ageing may mask this deterioration in kidney function as serum creatinine falls despite the decline in GFR. Accurate estimation of GFR in the elderly is necessary, particularly when prescribing drugs that are renally excreted.

Urinary tract infections

UTIs are common in advanced age, as impaired bladder emptying (e.g., due to prostatic disease or a neuropathic bladder) is more common. Symptoms may be atypical: confusion, incontinence, nocturia, offensive urine, or a vague change in wellbeing with little in the way of dysuria are common (see Chapter 49).

Further reading

FitzGerald TJ et al. A review of supportive care for older people with advanced chronic kidney disease. *Clin Kidney J* 2023; 16:2023;635–646.
Merchant AA, Ling E. An approach to treating older adults with chronic kidney disease. *CMAJ* 2023; 195(17):E612–E618.

Bibliography

Current Opinion in Nephrology and Hypertension. A monthly journal with review articles, each issue devoted to one or two topics.
Journal of the American Society of Nephrology. The highest-impact journal in nephrology with bi-monthly self-assessment programme (SAP) supplements.
Kidney International. The major journal associated with the International Society of Nephrology—monthly with original and review articles.
Nephrology, Dialysis, Transplantation. The major European journal devoted to the subject, with review articles, editorial comments and original papers.

Significant websites

http://www.kidney.org.uk *UK charity run by and for patients.*
http://www.tinkershop.net/nephro.htm *Nephrology calculator.*

Questions for this chapter can be found on Elsevier eBooks+.

27 Water balance, fluids, and electrolytes

M. Magdi Yaqoob and Kieran McCafferty

Clinical skills 854
Assessing a patient's fluid status— 854
Assessing acid–base disturbance 855
Background 855
Disorders of extracellular volume 855
Increased extracellular volume 859
Decreased extracellular volume 863
Disorders of sodium concentration 865
Hyponatraemia 866
Hypernatraemia 869
Disorders of potassium concentration 869
Hypokalaemia 870
Hyperkalaemia 871
Disorders of magnesium concentration 872
Hypomagnesaemia 873
Hypermagnesaemia 873
Disorders of phosphate concentration 874
Hypophosphataemia 874
Hyperphosphataemia 876
Acid–base disorders 877
Respiratory acidosis and alkalosis 879
Metabolic acidosis and alkalosis 879

CORE SKILLS AND KNOWLEDGE

Learning about fluid and electrolyte disorders and acid–base physiology can be a daunting experience. However, these disorders form some of the most common presentations to hospital and may complicate a wide range of other conditions, and so all doctors need to be confident when approaching them.

This chapter reviews the physiology that underpins normal fluid, electrolyte, and acid–base balance, and discusses disorders of blood chemistry caused by dietary factors, underlying medical conditions, and medical treatments. These imbalances may be acute or chronic, may occur with varying degrees of severity, and may or may not be partially compensated by the body's regulatory mechanisms.

Key skills to master for undergraduates include:

- Comprehensively assessing a patient's fluid status and gaining confidence in prescribing intravenous fluids and diuretics appropriately
- Developing an appropriate diagnostic approach to abnormalities of common electrolytes including sodium and potassium
- Understanding the implications of an abnormal blood pH and diagnosing the common causes of acidosis and alkalosis.

These skills can be best learned by experience—particularly in assessing patients presenting to the hospital emergency department or who have been admitted to critical care units. Take the opportunity to review patients' fluid balance charts on ward rounds and try (with supervision) to practise prescribing intravenous fluids appropriately. Review the results of blood gas analyses and work through any abnormalities shown.

CLINICAL SKILLS

Assessing a patient's fluid status is a key skill for all doctors, requiring key clinical skills in history-taking (Box 27.1), examination (Fig. 27.1), and data interpretation (Box 27.2). Many patients are hypovolaemic on admission to hospital as a result of acute illness. Others will exhibit signs of increased extracellular volume and require diuretic therapy (see 'Clinical use of diuretics' section).

Assessing a patient's fluid status

Box 27.1 Clues to volume status in the history

	Increased extracellular volume	Decreased extracellular volume
Symptoms	Peripheral oedema Breathlessness on exercise Orthopnoea Paroxysmal nocturnal dyspnoea	Thirst Dizziness, especially on standing
Medical history	Heart failure Cirrhosis Advanced chronic kidney disease Intravenous fluid administration	Recent diarrhoea or vomiting Symptoms suggesting sepsis Use of diuretics

Box 27.2 Clues to volume status on basic investigations

Evidence for increased extracellular volume

- 'Haemodilution': fall in blood haemoglobin concentration and packed cell volume (haematocrit)
- Fluid overload on chest X-ray
- Evidence of volume overload or cardiac dysfunction on transthoracic echocardiogram

Evidence for decreased extracellular volume

- 'Haemoconcentration': rise in blood haemoglobin and packed cell volume (haematocrit)
- Acute kidney injury (see Chapter 26, 'Acute kidney injury' section)
- Disproportionate rise in serum urea compared with serum creatinine
- Changes in serum sodium and osmolality
- Concentrated urine: specific gravity (measured on dipstick) >1.035

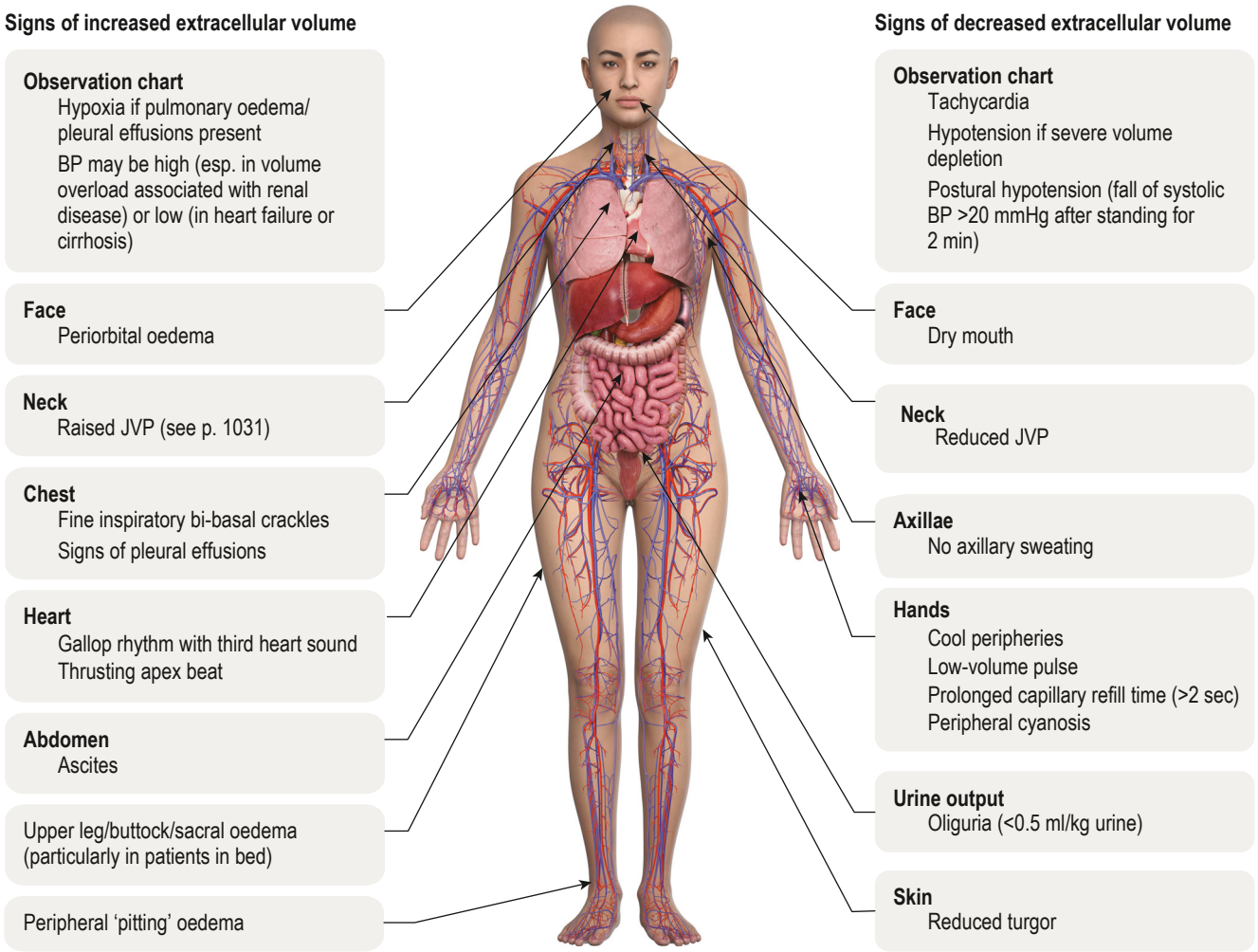


Fig. 27.1 Clues to volume status on clinical examination.

Fig. 27.2 The ABL800FLEX blood gas analyzer. *(From Al-Shaikh B, Stacey SG.* Essentials of Equipment in Anaesthesia, Critical Care and Peri-Operative Medicine, *5th edn. Edinburgh: Elsevier, 2019, Fig. 13.5, courtesy Radiometer Ltd.)*

Assessing acid–base disturbance

Arterial or venous blood gas analysis is readily carried out in most hospitals using table-top blood gas analyzers (Fig. 27.2), which measure:

- blood pH
- partial pressures of carbon dioxide (PCO_2) and oxygen (PO_2)
- calculated base excess (BE) and bicarbonate (HCO_3^-)
- blood electrolytes including sodium and potassium concentrations
- other molecules including glucose and lactate.

This information is vital in assessing respiratory failure, critical illness (failure to maintain blood pH in the normal range is often a poor prognostic marker), as well as chronic disorders affecting acid–base homeostasis.

A stepwise approach to interpreting blood gas analyses is shown in Fig. 27.3. Always go through a blood gas analysis report fully, describing all the abnormalities present, before pulling the different abnormalities together into a complete description.

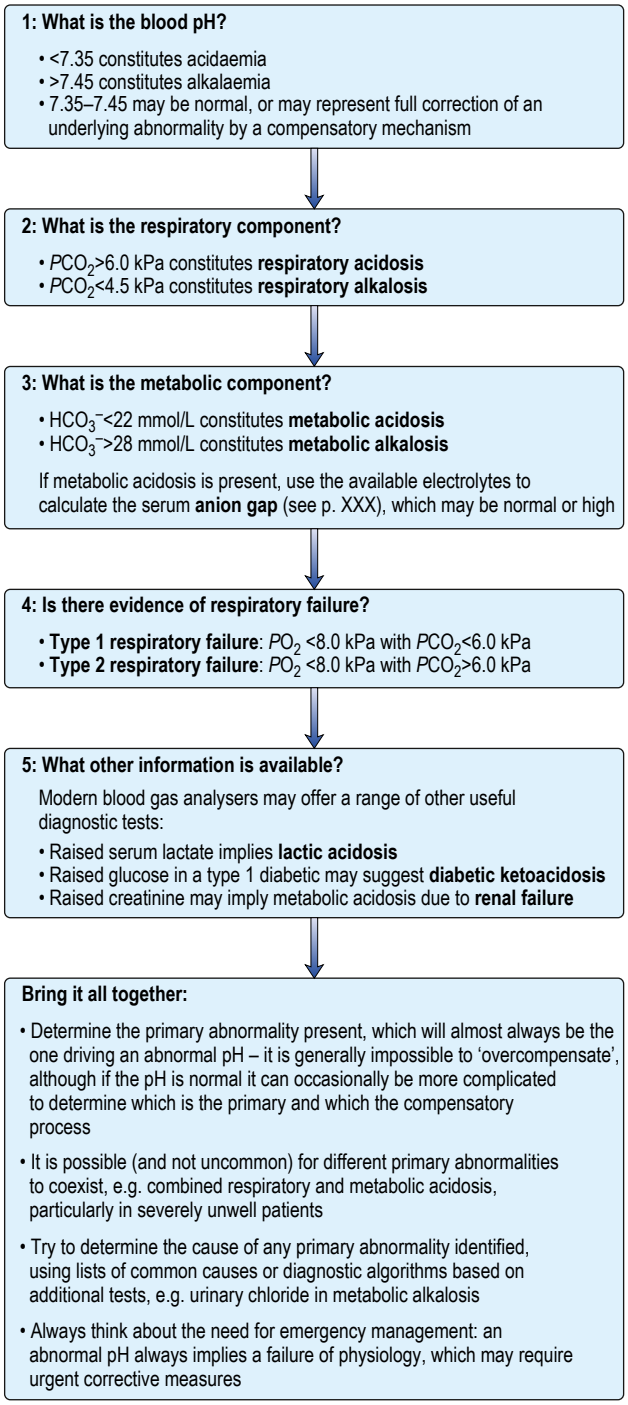


Fig. 27.3 Stepwise approach to interpreting blood gas analysis.

BACKGROUND

To understand the disorders described in this chapter, it is crucial to separate out:

- disorders of total extracellular volume (hyper- and hypovolaemia), which primarily relate to the increase or decrease of total body sodium
- disorders of blood osmolality (which, since sodium is the most abundant extracellular electrolyte, are generally encountered in clinical practice as hypo- and hypernatraemia); these conditions relate to inappropriate retention or loss of free water causing dilution or concentration of blood
- disorders in the serum concentrations of other major electrolytes including potassium, magnesium, and phosphate
- disorders of blood pH (acidaemia and alkalaemia), reflecting abnormal serum concentrations of hydrogen and bicarbonate ions and changes in the partial pressure of carbon dioxide.

Although this chapter deals with each of these groups in turn, it is important to remember how interconnected they are. For example, it would not be unusual to encounter a patient who is simultaneously hypervolaemic, hyponatraemic, hyperkalaemic, and acidaemic as a result of the same underlying disease process. Interventions must be individualized to address the particular physiological requirements of each patient. Once the underlying cause of abnormalities is diagnosed, this can be corrected while appropriate measures are put in place to support the body's physiological mechanisms to maintain homeostasis.

DISORDERS OF EXTRACELLULAR VOLUME

In normal healthy people, the total body water constitutes 50% to 60% of lean body weight in males and 45% to 50% in females. Thus in a healthy 70 kg male, total body water is approximately 42 L. This is contained in three major compartments:

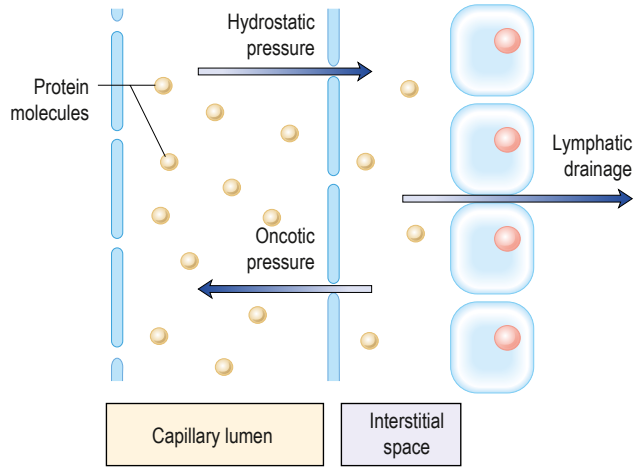


Fig. 27.4 Distribution of water between the vascular and extravascular (interstitial) spaces. This is determined by the equilibrium between hydrostatic pressure, which tends to force fluid out of the capillaries, and oncotic pressure, which acts to retain fluid within the vessel. The net flow of fluid outwards is balanced by 'suction' of fluid into the lymphatics, which returns it to the bloodstream. Similar principles govern the volume of the peritoneal and pleural spaces.

- intracellular fluid (28 L, about 35% of lean body weight) extracellular—the interstitial fluid that bathes the cells (9.4 L, about 12%)
- plasma (also extracellular) (4.6 L, about 4%–5%).

In addition, small amounts of water are contained in bone, dense connective tissue, and epithelial secretions, such as the digestive secretions and cerebrospinal fluid (CSF).

The intracellular and interstitial fluids are separated by the cell membrane; the interstitial fluid and plasma are separated by the capillary wall (Fig. 27.4). In the absence of solute, water molecules move randomly and in equal numbers in either direction across a semipermeable membrane. However, if solutes are added to one side of the membrane, intermolecular cohesive forces reduce the kinetic activity of the water molecules. As a result, water tends to stay in the solute-containing compartment because there is less free diffusion across the membrane. This ability to 'hold' water in the compartment can be measured as the osmotic pressure.

Osmotic pressure

Osmotic pressure is the primary determinant of the distribution of water among the three major compartments. The concentrations of the major solutes in the compartments differ, each having one solute that is primarily limited to that compartment and therefore determines its osmotic pressure:

- The ***intracellular fluid*** contains mainly potassium (K^+) (most intracellular Mg^{2+} is bound to proteins and therefore osmotically inactive).
- In the extracellular compartment, Na^+ salts predominate in the ***interstitial fluid***, along with proteins in the plasma.

Regulation of plasma volume is somewhat more complicated because of the tendency of the plasma proteins to hold water in the vascular space by an oncotic effect that is partly counterbalanced by the hydrostatic pressure in the capillaries that is generated by cardiac contraction (see Fig. 27.4). The composition of intracellular and extracellular fluids is shown in Box 27.3.

Osmotically active solutes cannot freely leave their compartment. The capillary wall, for example, is relatively impermeable to plasma proteins, and the cell membrane is 'impermeable' to Na^+ and K^+ because the Na^+/K^+-adenosine triphosphatase (ATPase) pump largely restricts Na^+ to the extracellular fluid and K^+ to the intracellular fluid. By contrast, Na^+ freely crosses the capillary wall and achieves similar concentrations in the interstitium and plasma; as a result, it does not contribute to fluid distribution between these compartments. Similarly, urea crosses both the capillary wall and the cell membrane, and is osmotically inactive. Thus, the retention of urea in kidney failure does not alter the distribution of the total body water.

Box 27.3 Electrolyte composition of intracellular and extracellular fluids (mmol/L)

	Plasma	Interstitial fluid	Intracellular fluid
Na^+	142	144	10
K^+	4	4	160
Ca^{2+}	2.5	2.5	1.5
Mg^{2+}	1.0	0.5	13
Cl^-	102	114	2
HCO_3^-	26	30	8
PO_4^{2-}	1.0	1.0	57
SO_4^{2-}	0.5	0.5	10
Organic acid	3	4	3
Protein	16	0	55

Body Na^+ stores are the primary determinant of extracellular fluid volume. Specialist imaging techniques (sodium magnetic resonance imaging, Na-MRI) have shown that remarkable amounts of Na are stored in the muscle, bones, and skin of hypertensive patients with hyperaldosteronism who are in effect chronic salt retainers.

Thus, the extracellular volume—and therefore tissue perfusion—are maintained by appropriate alterations in Na^+ excretion. For example, if Na^+ intake is increased, extra Na^+ will initially be added to the extracellular fluid. The associated increase in extracellular osmolality will cause water to move out of the cells, leading to extracellular volume expansion. Balance is restored by excretion of the excess Na^+ in the urine.

Regulation of extracellular volume

The extracellular volume is determined by the total body sodium content, which is closely controlled by the kidneys (Fig. 27.5). Plasma fluid (4.6 L in a 70 kg male) constitutes one-third of extracellular volume; of this, 85% (3.9 L) lies in the venous circulation, and only 15% (0.7 L) resides in the arteries.

The fullness of the arterial vascular compartment (***effective arterial blood volume***, EABV) is the primary determinant of renal sodium and water excretion: that is, the ***effective circulatory volume*** for the purposes of body fluid homeostasis.

The fullness of the arterial compartment depends on a relationship between cardiac output and peripheral arterial resistance. Thus, diminished EABV may occur when there is a fall in cardiac output or a fall in peripheral arterial resistance (an increase in the holding capacity of the arterial vascular tree). When the EABV is expanded, urinary Na^+ excretion is increased and can exceed 100 mmol/L. By contrast, the urine can be rendered virtually free of Na^+ in the presence of EABV depletion and normal kidney function.

These changes in Na^+ excretion can result from alterations both in the filtered load, determined primarily by the glomerular filtration rate (GFR), and in tubular reabsorption, which is affected by multiple factors. In general, changes in tubular reabsorption constitute the

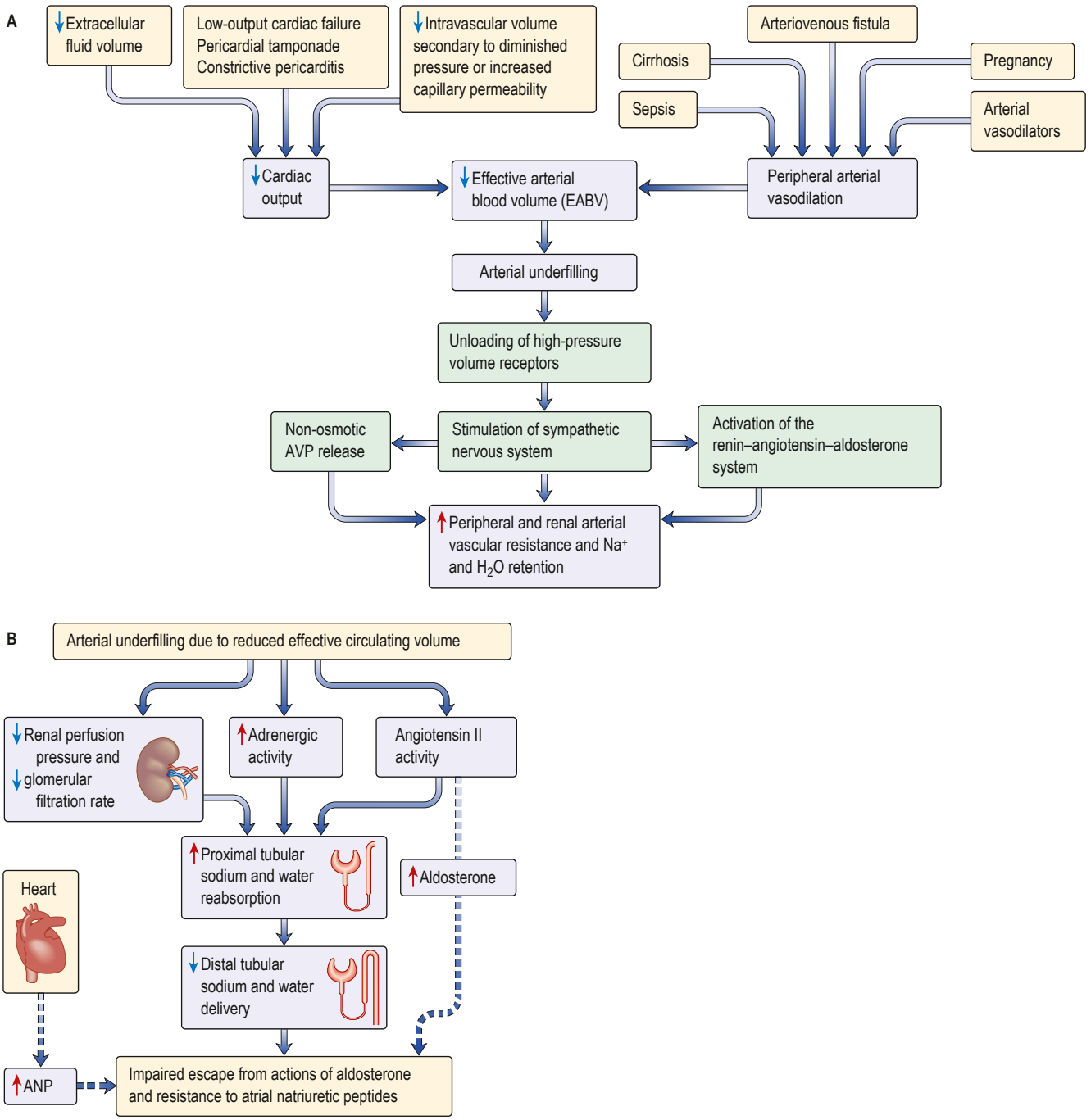


Fig. 27.5 Regulation of extracellular volume. **(A)** Sequence of events in which a decrease in cardiac output or peripheral arterial dilation initiates renal sodium and water retention. **(B)** Mechanism of impaired escape from the actions of aldosterone and resistance to atrial natriuretic peptides (ANP). *AVP,* Arginine vasopressin. *(Modified with permission from Schrier RW.* Renal and Electrolyte Disorders, *7th edn. Philadelphia: Lippincott Williams & Wilkins, 2010.)*

main adaptive response to fluctuations in the effective circulating volume. How this occurs can be appreciated from **Box 27.4** and **Fig. 27.6**, which describe the sites and determinants of segmental Na^+ reabsorption. Although the loop of Henle and distal tubules make a major overall contribution to net Na^+ handling, transport in these segments primarily varies with the amount of Na^+ delivered; that is, reabsorption is flow-dependent. In comparison, the neurohumoral regulation of Na^+ reabsorption according to body needs occurs primarily in the proximal tubules and collecting ducts. Diseases associated with malfunction of each section of the tubule are shown in **Box 27.5**.

Neurohumoral regulation of extracellular volume

This is mediated by volume receptors that sense changes in the EABV rather than alterations in the sodium concentration. These receptors are distributed in both kidney and cardiovascular tissues.

- ***Intrarenal receptors.*** Receptors in the walls of the afferent glomerular arterioles respond, via the juxtaglomerular apparatus, to changes in kidney perfusion, and control the activity of the renin–angiotensin–aldosterone system (see Chapter 26). In addition, sodium concentration in the distal tubule and sympathetic nerve activity alter renin release from the juxtaglomerular cells.

Box 27.4 Mechanisms of sodium transport in the various nephron segments

	Filtered Na reabsorbed (%)	Major mechanisms of luminal Na^+ entry	Major factors regulating transport
Proximal tubule	60–70	Na^+–H^+ exchange and co-transport of Na^+ with glucose, phosphate and other organic solutes	Angiotensin II Noradrenaline (norepinephrine)
Loop of Henle	20–25	Na^+–K^+–$2Cl^-$ co-transport	Flow Pressure natriuresis mediated by nitric oxide
Distal tubule	5	Na^+–Cl^- co-transport	Flow
Collecting ducts	4	Na^+ channels	Aldosterone Atrial natriuretic peptide

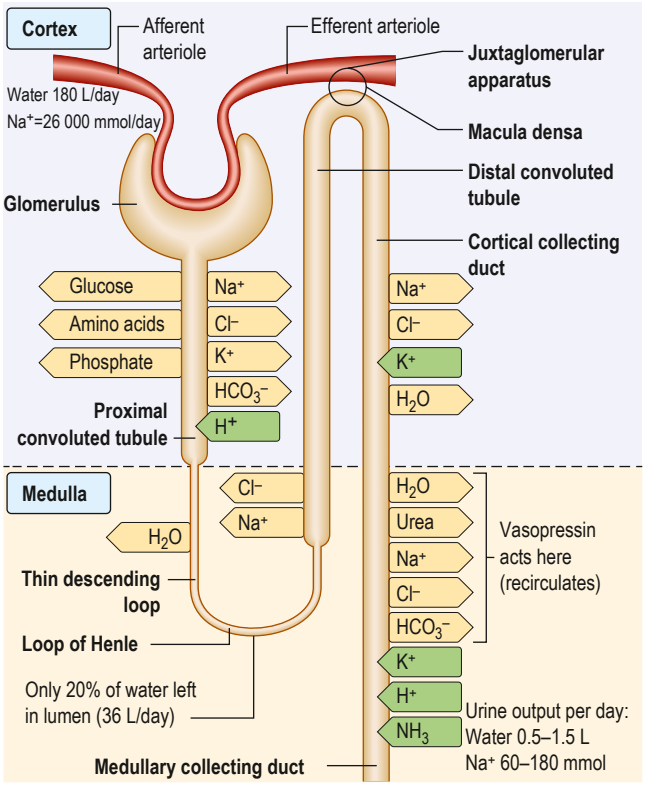


Fig. 27.6 The nephron—electrolyte and water exchange. Some 180 L of water and 26 000 mmol of sodium/day enter the nephrons via the afferent arterioles to the kidneys. Removal or addition of electrolytes results in the excretion of approximately 1 L of water and 60 to 180 mmol of sodium/day.

Prostaglandins I_2 and E_2 are generated within the kidney in response to angiotensin II, acting to maintain GFR and sodium and water excretion, and modulating the sodium-retaining effect of this hormone.

- ***Extrarenal receptors.*** These are located in the vascular tree in the left atrium and major thoracic veins, and in the carotid sinus body and aortic arch. These volume receptors respond to a slight reduction in effective circulating volume, which results in increased sympathetic nerve activity and a rise in catecholamines. In addition, volume receptors in the cardiac atria control the release of a powerful natriuretic hormone—atrial natriuretic peptide (ANP)—from granules located in the atrial walls (see Chapter 17).

High-pressure arterial receptors (carotid, aortic arch, juxtaglomerular apparatus) predominate over low-pressure volume receptors in volume control in mammals.

Aldosterone and ANP are responsible for day-to-day variations in Na^+ excretion, through their respective ability to augment and diminish Na^+ reabsorption in the collecting ducts.

- A ***salt load***, for example, leads to an increase in the effective circulatory and extracellular volume, raising both kidney perfusion pressure, and atrial and arterial filling pressure. The increase in the kidney perfusion pressure reduces the secretion of renin, and subsequently that of angiotensin II and aldosterone, whereas the rise in atrial and arterial filling pressure increases the release of ANP. ANP secretion also directly inhibits renin release. These factors combine to reduce Na^+ reabsorption in the collecting duct, thereby promoting excretion of excess Na^+.

Box 27.5 Major functions of each section of the tubule and associated diseases

Site	Function	Diseases associated with this section of nephron
Proximal convoluted tubule (PCT)	Early PCT: Solute reabsorption: glucose, amino acids, bicarbonate, phosphate, low-molecular-weight proteins Late PCT: Urate secretion/absorption Straight PCT: Secretion of drugs (loop and thiazide diuretics) and drug metabolites	Fanconi's syndrome: failure to reabsorb solutes and low-molecular-weight proteins. Leads to glycosuria, phosphaturia, aminoaciduria, and low-molecular-weight proteinuria. May be caused by drugs, myeloma, Wilson's disease and rare genetic diseases Type 2 renal tubular acidosis
Loop of Henle	Site of counter-current multiplier to hyper-concentrate the medullary interstitium, enabling control of final urine concentration	Bartter's syndrome
Distal convoluted tubule	Involvement in sodium and chloride transport, and calcium and magnesium reabsorption	Gitelman's syndrome Gordon's syndrome
Collecting ducts	Na^+ and water reabsorption Secretion of potassium Secretion of bicarbonate or H^+	Liddle's syndrome Pseudohypoaldosteronism type 1 AVP-resistance: resistance to actions of vasopressin. Leads to large volume of dilute urine with polyuria/polydipsia Syndrome of inappropriate antidiuretic hormone secretion (SIADH) Type 1 renal tubular acidosis

- By contrast, in patients with a ***low Na+*** intake or in those who become volume-depleted as a result of vomiting or diarrhoea, the ensuing decrease in effective volume enhances the activity of the renin–angiotensin–aldosterone system and reduces the secretion of ANP. The net effect is enhanced Na^+ reabsorption in the collecting ducts, leading to a fall in Na^+ excretion. This increases the extracellular volume towards normal.

With ***more marked hypovolaemia***, a decrease in GFR leads to an increase in proximal and thin ascending limb Na^+ reabsorption, which contributes to Na^+ retention. This is brought about by enhanced sympathetic activity acting directly on the kidneys and indirectly by stimulating the secretion of renin/angiotensin II (see **Fig. 27.5B**) and non-osmotic release of arginine vasopressin (AVP), also called vasopressin.

The regulation of salt excretion in response to rises or falls in vascular pressure acts to defend the body against changes in the effective circulating volume. Marked persistent hypovolaemia leads to systemic hypotension and increased salt and water absorption in the proximal tubules and ascending limb of Henle. This process is partly mediated by changes in interstitial hydrostatic pressure and local prostaglandin and nitric oxide production. These mechanisms are not always effective, and recurrent episodes of hypovolaemia may, over time, lead to chronic kidney disease (CKD), such as in the recently described syndrome Mesoamerican nephropathy, where recurrent episodes of heat-stress in agricultural workers along the Pacific coast in Central America is thought to be a cause.

Increased extracellular volume

Sodium and water are retained despite increased extracellular volume in oedematous conditions such as cardiac failure, hepatic cirrhosis, and hypoalbuminaemia.

Aetiology

In any condition where patients become oedematous, a key principle in explaining salt and water retention is the concept of 'arterial underfilling'. This may be due either to reduced cardiac output or diminished peripheral arterial resistance. Arterial underfilling in these settings leads to reduction of pressure or stretch (i.e., 'unloading' of arterial volume receptors), which results in activation of the sympathetic nervous system and the renin–angiotensin–aldosterone system, and non-osmotic release of AVP. These neurohumoral mediators promote salt and water retention in the face of increased extracellular volume (see **Fig. 27.5A**), a condition often described as 'secondary hyperaldosteronism'.

The mechanism for 'arterial underfilling' and subsequent sodium retention varies between different disease states:

- In ***heart failure***, reduction in cardiac output and the consequent fall in effective circulatory volume and arterial filling lead to sustained activation of the renin–angiotensin–aldosterone system, non-osmotic release of AVP, and increased activity of the renal sympathetic nerves via volume receptors and baroreceptors (see **Fig. 27.5B**).
- In ***hepatic cirrhosis***, the mechanism is complex, but involves peripheral vasodilation, due to increased nitric oxide generation from nitrogenous dietary substances, which are not effectively metabolized by the failing liver, resulting in reduced EABV and arterial filling. This leads to an activation of a chain of events that are common also to other conditions with marked peripheral vasodilation and heart failure (see **Fig. 27.5**).
- In ***nephrotic syndrome***, relative arterial underfilling was traditionally explained by loss of plasma protein in urine, causing loss of intravascular fluid into the interstitium. However, various animal and human studies have shown that the transcapillary oncotic gradient remains unchanged in nephrotic syndrome and movement of fluid from the vascular into the interstitial space is not explained by alterations in the transcapillary hydrostatic pressure gradient. Instead, the mechanism for oedema seen in nephrotic syndrome is increased capillary wall permeability (or 'hydraulic conductivity') caused by incompletely understood mechanisms affecting intercellular junctional complexes.

As well as relative arterial underfilling causing secondary hyperaldosteronism, nephrotic syndrome is also characterized by functional abnormalities in the kidney itself: in the cortical collecting duct, Na^+/K^+-ATPase expression and activity is increased threefold along the basolateral surface of cells (see **Fig. 27.6**), and activity of the amiloride-sensitive epithelial sodium channel (ENaC) activity is also increased.

- In ***protein-losing enteropathy***, uncompensated loss of plasma proteins in the intestine causes non-selective hypoproteinaemia (with reduced albumin and immunoglobulin levels) and loss of plasma oncotic pressure. Patients present with oedema, nutritional deficiencies, infections, and gastrointestinal symptoms including vomiting, diarrhoea, steatorrhea, and abdominal pain. The condition may be confirmed by the presence of elevated α1-antitrypsin levels in the stool. Underlying causes of intestinal protein loss including conditions which damage the intestinal mucosa (inflammatory, infective, malignancy) or block gastrointestinal lymphatic flow (through obstruction and lymphangiectasia).

The mechanism for patients becoming oedematous in each of these conditions relates to sodium delivery to the distal tubule:

- A combination of alpha-adrenergic stimulation, elevated angiotensin II, reduced kidney perfusion, and reduced GFR results in avid sodium absorption from the proximal tubules.
- Sodium delivery to the distal portion of the nephron is correspondingly reduced, so that under the influence of aldosterone, there is significant sodium reabsorption in the collecting duct.
- Patients continue to retain sodium in the face of expanded extracellular volume and become progressively oedematous (see **Fig. 27.5B**). They often display a substantial natriuresis when given spironolactone, which blocks aldosterone receptors.

Primary hyperaldosteronism and 'escape' from hypervolaemia

The physiology of secondary hyperaldosteronism described above differs somewhat from primary hyperaldosteronism, which is caused by intrinsic oversecretion of aldosterone, typically by an adrenal adenoma (see Chapter 34, 'Primary hyperaldosteronism' section). In such individuals, high doses of mineralocorticoids initially cause renal sodium retention so that the extracellular volume is increased by 1.5 to 2 L. However, renal sodium retention then ceases, sodium balance is re-established, and there is no detectable oedema. This 'escape' from mineralocorticoid-mediated sodium retention explains why oedema is not a characteristic feature of primary hyperaldosteronism. The escape is dependent on an increased delivery of sodium to the site of action of aldosterone in the collecting ducts, where aldosterone-dependent sodium reabsorption is overwhelmed and sodium excretion continues and sodium is not progressively retained. The increased distal sodium delivery is achieved by high extracellular volume-mediated arterial overfilling. This suppresses sympathetic activity and angiotensin II generation, and increases cardiac release of ANP with resultant increase in kidney perfusion pressure and GFR.

Other causes of sodium retention

Numerous drugs and various other physiological states can cause renal sodium retention, particularly in patients whose kidney function is already impaired. Typically the degree of oedema is more modest than in heart failure, cirrhosis, and nephrotic syndrome.

- ***Oestrogens*** cause mild sodium retention, due to a weak aldosterone-like effect. This is the reason for weight gain in females in the premenstrual phase.
- ***Mineralocorticoids and liquorice*** (the latter potentiating the sodium-retaining action of cortisol) have aldosterone-like actions.
- ***Non-steroidal anti-inflammatory drugs (NSAIDs)*** cause sodium retention in the presence of activation of the renin–angiotensin–aldosterone system.
- ***Thiazolidinediones*** have been associated with salt and water retention and are contraindicated in patients with heart failure. (By contrast, newer antidiabetic agents e.g., sodium–glucose co-transporter-2 inhibitors [Chapter 35, 'Sodium-glucose transporter 2 inhibitors ('flozins')' section] have established beneficial effects in patients with diabetes and heart failure.)
- Initiation of ***insulin*** treatment for type 1 diabetes, and ***re-feeding after malnutrition*** are both associated with the development of transient oedema. The mechanism is complex but involves up-regulation of ENaC in the principal cell of the collecting duct.
- ***A decreased GFR*** decreases the kidney's capacity to excrete sodium. This may be acute, as in the acute nephritic syndrome (see Chapter 26), or may occur as part of the presentation of CKD. In ***end-stage kidney failure***, extracellular volume is controlled by the balance between salt intake and its removal by dialysis.
- ***Idiopathic oedema*** sometimes occurs where there is no evidence of heart, liver, kidney, or endocrine disease. Females are more affected than males, and the condition may be caused by alterations in effective circulating volume related to posture, with sodium excretion reduced during the day and increased at night.
- Oedema is also caused by raised interstitial oncotic pressure as a result of increased capillary permeability to proteins. This can occur as part of a rare complement-deficiency syndrome; with therapeutic use of interleukin 2 in cancer chemotherapy; or in ovarian hyperstimulation syndrome.

Localized oedema

This does not reflect disturbances of extracellular volume control *per se* or relates to impaired local venous or lymphatic drainage. Examples include:

- ***lymphoedema*** affecting any part of the body where lymphatic drainage is impaired, for example, in the upper limb after surgical dissection of axillary lymph nodes during treatment for breast cancer
- ***post-thrombotic syndrome***, where venous drainage is chronically impaired following deep vein thrombosis (see Chapter 40, 'Post-thrombotic syndrome' section)
- facial oedema due to ***superior vena caval obstruction***.

Clinical features

Peripheral oedema is caused by expansion of the extracellular volume by at least 2 L (15%). The ankles are normally the first part of the body to be affected. Oedema may be noted in the face, particularly in the morning, or if the patient is in bed, oedema may accumulate in the sacral area. Expansion of the interstitial volume also causes pulmonary oedema, pleural effusions, pericardial effusion, and ascites. Substantial amounts of sodium and water may accumulate in the body without clinically obvious oedema or evidence of raised venous pressure. In particular, several litres may accumulate in the pleural space or as ascites; these spaces are then referred to as 'third spaces'.

Expansion of the blood volume causes a raised jugular venous pressure, cardiomegaly, added heart sounds, and basal crackles. Blood pressure is often elevated, but may be normal or low depending on the cause of hypervolaemia: in heart failure and cirrhosis, there may be hypotension in the face of volume overload, caused by low cardiac output or reduced peripheral vascular resistance, respectively.

Management

The underlying cause should be treated where possible. Heart failure, for example, should be treated and offending drugs such as NSAIDs withdrawn.

Sodium restriction has only a limited role but is useful in patients who are resistant to diuretics. Sodium intake can be reduced to approximately 100 mmol (2 g) daily; reductions below this are often difficult to achieve without affecting the palatability of food.

The mainstay of treatment is the use of ***diuretic agents***, which increase sodium, chloride, and water excretion in the kidney (**Box 27.6**). These agents act by interfering with membrane ion pumps that are present on numerous cell types; they mostly achieve specificity for the kidney by being secreted into the proximal tubule, resulting in much higher concentrations in the tubular fluid than in other parts of the body.

Clinical use of diuretics

Loop diuretics

These potent diuretics are useful in the treatment of any cause of systemic extracellular volume overload. They stimulate excretion of both sodium chloride and water by blocking the sodium–potassium-2-chloride (NKCC2) channel in the thick ascending limb of Henle (**Fig. 27.7**) and are useful in stimulating water excretion in states of relative water overload. They also act by causing increased venous capacitance, resulting in rapid clinical improvement in patients with left ventricular failure, preceding the diuresis. Although generally well-tolerated, unwanted effects include:

- urate retention, causing gout
- hypokalaemia
- hypercalciuria leading to increased risk of calcium-based kidney stones
- hypomagnesaemia
- decreased glucose tolerance
- allergic tubulointerstitial nephritis and other allergic reactions
- myalgia—especially with high-dose bumetanide
- ototoxicity (due to an action on sodium pump activity in the inner ear)—particularly with high-dose furosemide
- interference with excretion of lithium, resulting in toxicity.

There is little to choose between the drugs in this class. Bumetanide has a better oral bioavailability than furosemide, particularly in patients with severe peripheral oedema, and has more beneficial effects than furosemide on venous capacitance in left ventricular failure.

Thiazide diuretics

Thiazide diuretics are less potent than loop diuretics. They act by blocking a sodium chloride channel in the ***distal convoluted tubule*** (producing an induced form of Gitelman's syndrome, which is associated with reduced function of this channel; see Chapter 26) (**Fig. 27.8**). They cause relatively more urate retention, glucose intolerance and hypokalaemia than loop diuretics. They interfere

Box 27.6 Types and clinical uses of diuretics

Class	Major action	Examples	Clinical uses	Potency
Loop diuretics	↓ Na^+–Cl^-–K^+ co-transport in thick ascending limb of loop of Henle	Furosemide Bumetanide Torasemide	Volume overload (heart failure, nephrotic syndrome, CKD)	+++
Thiazide and related diuretics	↓ Na^+–Cl co-transport in early distal convoluted tubule	Bendroflumethiazide Chlortalidone Metolazone Indapamide	Hypertension Volume overload ('sequential nephron blockade') Hypercalciuria	++
Potassium-sparing diuretics	↓ Na^+ reabsorption (in exchange for K^+) in collecting duct (principal cells)	Aldosterone antagonists, e.g., spironolactone, eplerenone, finerenone Others: amiloride, triamterene	Hyperaldosteronism (primary and secondary) Bartter's syndrome Heart failure Cirrhosis with fluid overload Prevention of K^+ deficiency in combination with loop or thiazide	+
Carbonic anhydrase inhibitors	↓ Na^+ HCO_3^- reabsorption in proximal collecting duct ↓ Aqueous humour and CSF formation	Acetazolamide	Seldom used for diuretic effects Glaucoma Idiopathic intracranial hypertension (see Chapter 28, 'Idiopathic intracranial hypertension' section)	±
Vasopressin/AVP receptor blockers (aquaretics)	Block V2 receptor in collecting ducts producing free water diuresis	Lixivaptan Tolvaptan Satavaptan i.v. conivaptan (also blocks V_{1A})	SIADH Polycystic kidney disease Heart failure	+

CKD, Chronic kidney disease; *SIADH,* syndrome of inappropriate antidiuretic hormone secretion.

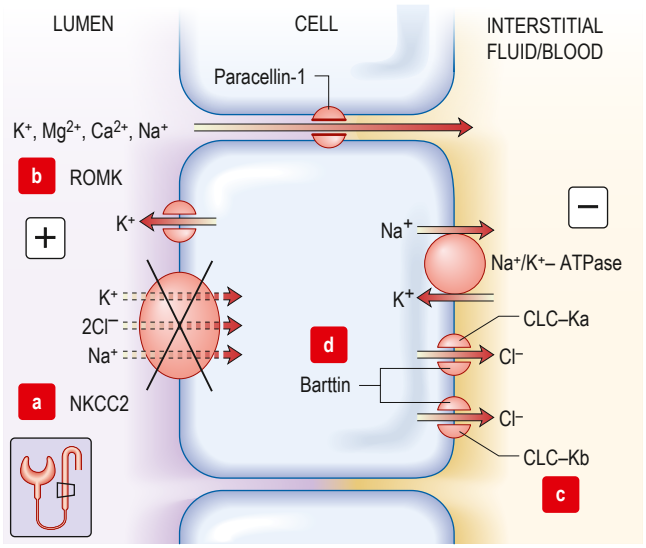


Fig. 27.7 Transport mechanisms in the thick ascending limb of the loop of Henle. Sodium chloride is reabsorbed in the thick ascending limb by the bumetanide-sensitive sodium–potassium-2–chloride co-transporter (NKCC2), which is the target of loop diuretics such as furosemide. The electroneutral transporter is driven by the low intracellular sodium and chloride concentrations generated by the Na^+/K^+-ATPase and the kidney-specific basolateral chloride channel (CLC-Kb). The availability of luminal potassium is rate-limiting for NKCC2, and recycling of potassium through the ATP-regulated potassium channel (renal outer medulla K^+ channel, ROMK) ensures the efficient functioning of the NKCC2 and generates a lumen-positive transepithelial potential. Calcium and magnesium, along with potassium and sodium, are also absorbed paracellularly in the thick ascending loop. Bartter's syndrome: loss of function mutations for the different types of syndrome. **(a)** NKCC2 (type I). **(b)** ROMK (type II). **(c)** CLC-Kb (type III). **(d)** Barttin (type IV). In addition, type V may be caused by a gain-of-function mutation of the calcium-sensing receptor (not shown).

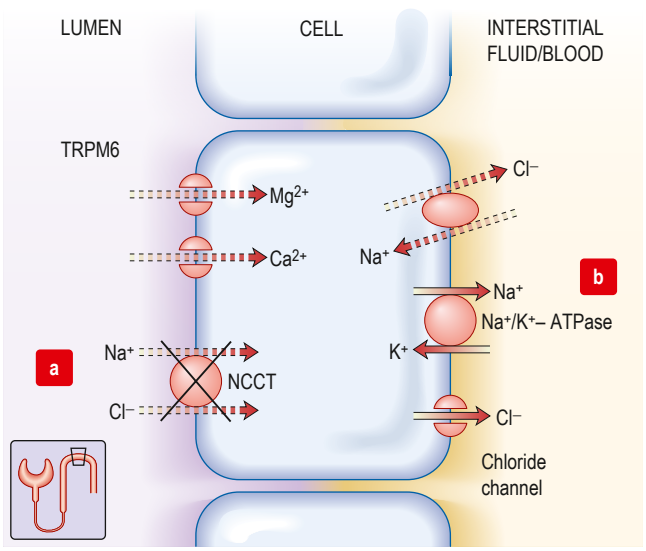


Fig. 27.8 Transport mechanisms in the distal convoluted tubule. **(a)** Sodium chloride is usually reabsorbed by the apical thiazide-sensitive sodium–chloride co-transporter (NCCT) in the distal convoluted tubule. **(b)** The electroneutral transporter is driven by the low intracellular sodium and chloride concentrations generated by the Na^+/K^+-ATPase, and an undefined basolateral chloride channel. In this nephron segment, there is an apical calcium channel and a basolateral sodium-coupled exchanger. The mechanisms for the transport of magnesium are similar to those for calcium, involving TRPM6, a member of the transient receptor potential family. Mutation of NCCT leads to decreased reabsorption of sodium chloride and increased absorption of calcium (Gitelman's syndrome). Overactivity of NCCT leads to Gordon's syndrome.

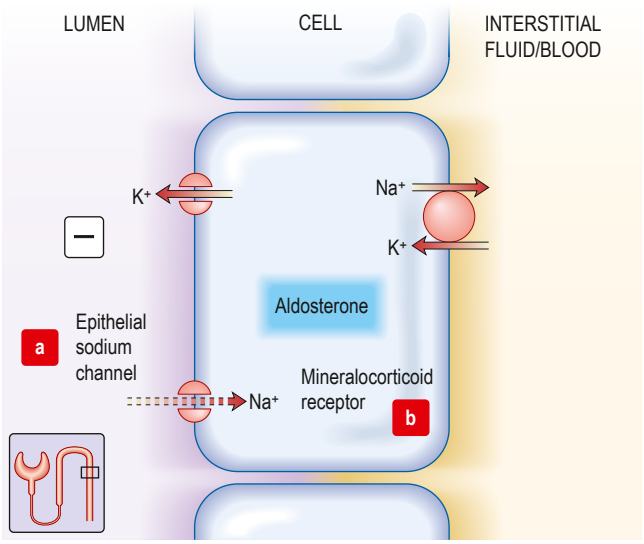


Fig. 27.9 Aldosterone-regulated transport in the cortical collecting ducts. Under normal conditions, the ENaC is the rate-limiting barrier for the normal entry of sodium from the lumen into the cell. The resulting lumen-negative transepithelial voltage (indicated by the minus sign) drives potassium secretion from the principal cells and proton secretion from the α-intercalated cells (see Fig. 27.15). **(a)** Site of action of amiloride and triamterene: mutation of the epithelial sodium channel gene (Liddle's syndrome) produces reabsorption of sodium and increased secretion of potassium (not shown). **(b)** Site of action of aldosterone antagonists: mutation of the gene encoding mineralocorticoid receptor leads to salt wasting and decreased secretion of potassium.

with water excretion and may cause hyponatraemia, particularly if combined with amiloride or triamterene. This effect is clinically useful in diabetes insipidus. Thiazides reduce peripheral vascular resistance by mechanisms that are not completely understood but do not appear to depend on their diuretic action, and are widely used in the treatment of hypertension. Thiazides reduce calcium excretion. This effect is useful in patients with idiopathic hypercalciuria but may cause hypercalcaemia. Numerous agents are available, with varying half-lives but little else to choose between them. Metolazone is not dependent for its action on glomerular filtration, and therefore retains its potency in kidney impairment.

Potassium-sparing diuretics

Potassium-sparing diuretics (Fig. 27.9) are unique among the class of diuretics in that (as their name suggests) they lead to hyperkalaemia rather than hypokalaemia. There are two types:

- ***Mineralocorticoid receptor antagonists (MRA).*** Mineralocorticoid receptors act as intracellular receptors and nuclear transcription factors. Following binding of ligands such as aldosterone the receptors interact with multiple different co-factors affecting gene transcription. They are present in the heart, kidney, bone marrow, and endothelium, and following activation have multiple effects including sodium reabsorption, potassium excretion, inflammation, and fibrosis. MRAs block the binding of aldosterone to the mineralocorticoid receptor. Two classes exist with different spectrums of biological activity:
  - Steroidal MRAs such as spironolactone and eplerenone have proven efficacy in reducing mortality in heart failure (in part by antagonizing the fibrotic effect of aldosterone on the heart). Eplerenone may be preferred because it is devoid of antiandrogenic or antiprogesterone properties.
  - Non-steroidal MRAs such as finerenone have recently been shown to slow the progression of kidney failure, lower proteinuria, and improve cardiovascular outcomes in patients with diabetes and CKD.
- ***Amiloride and triamterene***, which inhibit sodium uptake by blocking ENaC in the collecting duct and reduce renal potassium excretion by reducing lumen-negative transepithelial voltage. They may be used as potassium-sparing agents with thiazide or loop diuretics.

Carbonic anhydrase inhibitors

These are relatively weak diuretics and are seldom used, except in the treatment of glaucoma. They cause metabolic acidosis and hypokalaemia. Recent evidence suggests a role for acetazolamide in acute heart failure.

Sodium–glucose co-transporter 2 (SGLT2) inhibitors

These drugs, originally developed to provide glycaemic control in diabetes, have diuretic effects by blocking glucose and sodium reabsorption in the proximal tubule causing an osmotic diuresis and natriuresis, and by increasing sodium delivery to the juxtaglomerular apparatus, restoring normal tubule-glomerular feedback. Clinical trials have demonstrated significant cardiac and kidney protection and reduction in mortality in people with diabetes, heart failure, and CKD.

Resistance to diuretics

Resistance may occur as a result of:

- poor bioavailability
- reduced GFR, which may be due to decreased circulating volume despite oedema (e.g., nephrotic syndrome, cirrhosis with ascites) or intrinsic renal disease
- activation of sodium-retaining mechanisms, particularly aldosterone.

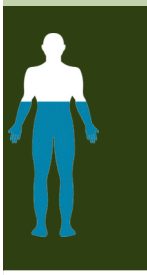
Intravenous administration of diuretics may be required to establish a diuresis, and high doses of loop diuretics are needed to achieve adequate concentrations in the tubule if GFR is depressed. However, the daily dose of furosemide must be limited to a maximum of 1.5 g for an adult because of ototoxicity. Intravenous albumin solutions are occasionally used to temporarily restore plasma oncotic pressure in the nephrotic syndrome and allow mobilization of oedema, but do not increase the natriuretic effect of loop diuretics.

Combinations of various classes of diuretics are extremely helpful in patients with resistant oedema, an approach referred to as ***'sequential nephron blockade'***. A loop diuretic plus a thiazide inhibit two major sites of sodium reabsorption; this effect may be further potentiated by addition of a potassium-sparing agent. Metolazone in combination with a loop diuretic is particularly useful in refractory congestive cardiac failure because its action is less dependent on glomerular filtration. However, this potent combination can cause severe electrolyte imbalance.

Effects on kidney function

All diuretics may increase plasma urea concentrations by increasing urea reabsorption in the medulla. In certain situations, diuretics also decrease GFR:

- Excessive diuresis causes volume depletion and pre-renal failure.
- Diuretics can cause allergic tubulointerstitial nephritis.
- Thiazides may directly cause a drop in GFR; the mechanism is complex and not fully understood.



Further reading

Ellison DH, Felker GM. Diuretic treatment in heart failure. *N Engl J Med* 2017; 16:1964–1975.
Kermond R, Mallett A, McCarthy H. A clinical approach to tubulopathies in children and young adults. *Pediatr Nephrol* 2023; 38:651–662.
Knepper MA, Kwon T-H, Nielson S. Molecular physiology of water balance. *N Engl J Med* 2015:1349–1358.
Warren AM, Grossmann M, Christ-Crain M, Russell N. Syndrome of inappropriate antidiuresis: From pathophysiology to management. *Endocr Rev* 2023; 30 (2):175–87.

Decreased extracellular volume

Deficiency of sodium and water causes shrinkage both of the interstitial space and of the blood volume, and may have profound effects on organ function.

Clinical features

Symptoms

Thirst, muscle cramps, nausea and vomiting, and postural dizziness occur. Severe depletion of circulating volume causes hypotension and impairs cerebral perfusion, causing confusion and eventual coma.

Signs

Signs can be divided into those due to loss of interstitial fluid and those due to loss of circulating volume.

- ***Loss of interstitial fluid*** leads to loss of skin elasticity ('turgor')—the rapidity with which the skin recoils to normal after being pinched. Skin turgor decreases with age, particularly at the peripheries making this test less useful in detecting fluid status. The turgor over the anterior triangle of the neck or on the forehead is less dependent on age and so may be more useful.
- ***Loss of circulating volume*** leads to decreased pressure in the venous and (if severe) arterial compartments.

Loss of up to 1 L of extracellular fluid in an adult may be compensated for by venoconstriction and may cause no physical signs. Loss of more than this amount causes the following:

- ***Postural hypotension.*** Normally, the blood pressure rises if a subject stands up, as a result of increased venous return due to venoconstriction (this maintains cerebral perfusion). Loss of extracellular fluid (underfill) prevents this and causes a fall in blood pressure. This is one of the earliest and most reliable signs of volume depletion, as long as the other causes of postural hypotension are excluded (**Box 27.7**).
- ***Low jugular venous pressure.*** In hypovolaemic patients, the jugular venous pulsation can be seen only with the patient lying completely flat, or even head down, because the right atrial pressure is lower than 5 cmH_2O.
- ***Peripheral venoconstriction.*** This causes cold skin with empty peripheral veins (which are difficult to cannulate just when the patient needs intravenous therapy the most.) This sign is often absent in sepsis, where peripheral vasodilation contributes to effective hypovolaemia.
- ***Tachycardia.*** This is not always a reliable sign. Beta-blockers and other antiarrhythmics may prevent tachycardia, and hypovolaemia may activate vagal mechanisms and actually cause bradycardia.

Aetiology

Salt and water may be lost from the kidneys, the gastrointestinal tract, or the skin (**Box 27.8**).

In addition, there are a number of situations where signs of volume depletion occur despite a normal or increased body content of sodium and water:

- Septicaemia causes vasodilation of both arterioles and veins, resulting in greatly increased capacitance of the vascular space and fall in peripheral venous resistance. In addition, increased capillary permeability to plasma proteins leads to loss of fluid from the vascular space to the interstitium.
- Diuretic treatment of heart failure or nephrotic syndrome may lead to rapid reduction in plasma volume. Mobilization of oedema may take much longer.
- There may be inappropriate diuretic treatment of oedema (e.g., when the cause is local rather than systemic), causing systemic hypovolaemia.

Investigations

Blood tests are, in general, not helpful in the assessment of extracellular volume. Plasma urea may be raised owing to increased urea reabsorption and, later, to pre-renal failure (when the creatinine rises as well), but this is very non-specific. Urinary sodium is low if the kidneys are functioning normally, but is misleading if the cause of the volume depletion involves the kidneys (e.g., diuretics, intrinsic kidney disease).

Assessment of volume status is shown in **Fig. 27.1** and **Boxes 27.1 and 27.2**.

Management

Distribution of different types of replacement fluid

Fig. 27.10 shows the relative effects on the compartments of the addition of identical volumes of water, saline, and colloid solutions. One litre of water given intravenously as 5% glucose (which is rapidly metabolized to generate energy, water, and carbon dioxide) is distributed equally into all compartments, whereas the same amount of 0.9% saline remains in the extracellular compartment. The latter

? Box 27.7 Postural hypotension: some causes of a fall in blood pressure from lying to standing

Decreased circulating volume (hypovolaemia)

Autonomic failure
- Diabetes mellitus
- Systemic amyloidosis
- Shy–Drager syndrome
- Parkinson's disease
- Ageing

Interference with autonomic function by drugs
- Tricyclic antidepressants

Interference with peripheral vasoconstriction by drugs
- Nitrates
- Calcium-channel blockers
- α-Adrenoceptor-blocking drugs

Prolonged bed rest (cardiovascular deconditioning)

? Box 27.8 Causes of extracellular volume depletion (hypovolaemia)

Haemorrhage
- External
- Concealed, for example, leaking aortic aneurysm

Burns

Gastrointestinal losses
- Vomiting
- Diarrhoea
- Ileostomy losses
- Ileus

Renal losses
- Diuretic use
- Impaired tubular sodium conservation
- Reflux nephropathy
- Papillary necrosis
  - Analgesic nephropathy
  - Diabetes mellitus
  - Sickle cell disease

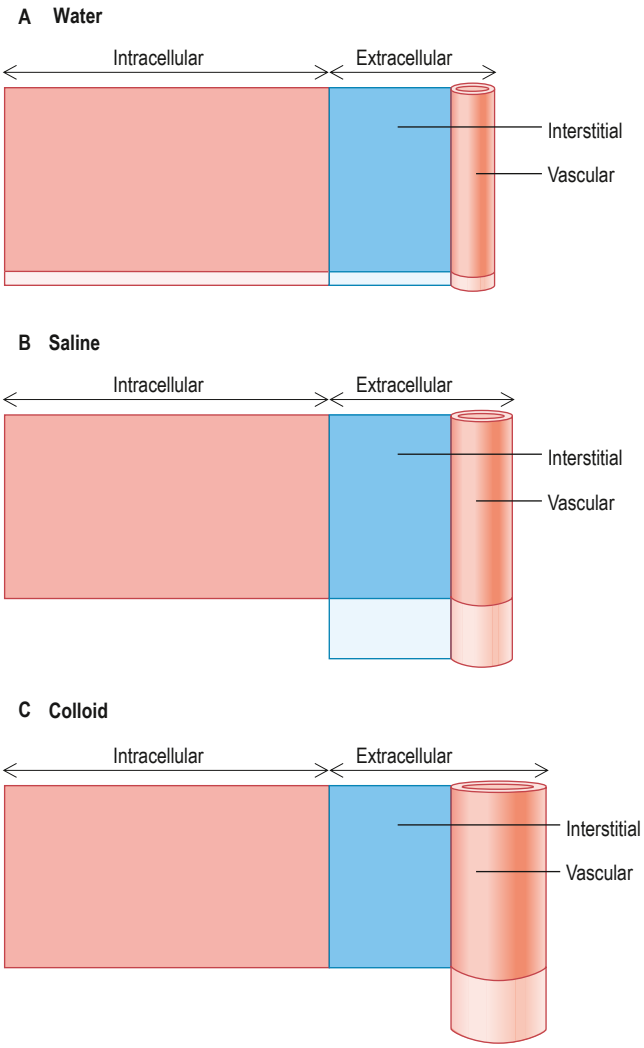


Fig. 27.10 Relative effects of the addition of 1 L of (A) water, (B) saline 0.9%, and (C) a colloid solution. See text for a fuller explanation.

is thus the correct treatment for extracellular water depletion, with sodium keeping the water in this compartment. The addition of 1 L of a colloid solution (such as colloidal starch or gelatine) with its high oncotic pressure stays in the vascular compartment, and colloids were formerly commonly used in the treatment for hypovolaemia (although they are much less used now after studies did not demonstrate any clinical outcome benefits compared to crystalloid solutions such as saline).

Goals of fluid therapy

When prescribing intravenous fluid, the initial consideration should be the goal of therapy, which will be to guide the type, volume rate, and duration of fluid replacement, as well as clarify how response to therapy is assessed:

- ***Resuscitation fluids*** are required in the context of hypotension and critical illness. Typically boluses of crystalloid fluid are used and the blood pressure response assessed to determine whether further fluid is required (see Chapter 13).
- ***Replacement fluids*** are required when a disease state has led to losses of particular components of fluid, electrolytes, and other components of blood. The principles of fluid replacement are to estimate the likely total volume of losses, and to select a replacement fluid to replace what is missing (the composition of different types of replacement fluid is given in **Box 27.9**):
  - *Loss of whole blood* (haemorrhage) is treated by infusion of a combination of red cells and a plasma substitute, or (if unavailable) whole blood.
  - *Loss of plasma*, as occurs in burns or severe peritonitis, should be treated with human plasma or a plasma substitute.
  - *Loss of salt and water* (e.g., through vomiting, diarrhoea, or excessive diuresis) may be treated by oral salt and water administration (if mild) or intravenous crystalloid (if more severe). ***Normal saline*** contains large amounts of chloride, which can cause a normal anion-gap metabolic acidosis as well as renal vasoconstriction when administered in large quantities. Thus, the use of ***balanced crystalloids*** (Hartmann's solution or Plasma-Lyte) is generally preferred, especially in critically ill patients of those requiring large volumes of fluid; these solutions have been shown to result in a lower rate of the composite outcome of death from any cause, new renal-replacement therapy or persistent renal dysfunction than the use of saline.
  - *Loss of water alone* can occur in diabetes insipidus or in a patient who is unable to drink (e.g., after surgery). This causes extracellular volume depletion only in severe cases because the loss is spread evenly among all the compartments of body water. The correct treatment is to give large volumes of oral or nasogastric water. If intravenous treatment is required, water is given as 5% glucose with K^+ because pure water would lead to osmotic lysis of blood cells, although in severe hypernatraemia (see below), it is necessary to begin replacement with 0.9% saline to avoid correcting the osmolality too rapidly.
  - *Loss of other electrolytes* (e.g., potassium, magnesium, calcium) can be treated by oral supplementation, by addition of the relevant salts into bags of replacement intravenous fluids, or (if severe) by administration of concentrated electrolyte solutions (e.g., 10% calcium gluconate). Because of the risk of dysrhythmia, this is often done in a critical care unit with cardiac monitoring.
  - *Loss of H^+ or bicarbonate* resulting in alkalosis or acidosis, respectively, often resolves with treatment of the underlying condition. In advanced CKD, oral bicarbonate supplementation is often required to treat acidosis; occasionally intravenous bicarbonate if used in patients where profound gastrointestinal bicarbonate loss has caused a normal anion gap acidosis (see later in the chapter).
- Maintenance fluids are required in patients who are unable to drink, for instance those with upper gastrointestinal pathology

? Box 27.9 Assessment of volume status

Best achieved by simple clinical observations, which you should carry out yourself (see Fig. 27.1). Check:

- Jugular venous pressure
- Central venous pressure: both basal and after intravenous fluid challenge (see Chapter 13, 'Central venous pressure' section)
- Postural changes in blood pressure
- Chest X-ray
- Serial weights of the patient
- Urine output measurements at regular intervals

or who have been made 'nil by mouth' ahead of surgery. If all that is required is avoidance of fluid depletion, 1 to 2 L or a balanced crystalloid can be given over 24 hours, remembering that surgery is a stimulus to sodium and water retention and that over-replacement may be as dangerous as under-replacement.

Repeated clinical assessments are vital when large volumes of intravenous fluids are being administered. Severe hypovolaemia induces venoconstriction, which maintains venous return; over-rapid correction does not give time for this to reverse, resulting in signs of circulatory overload (e.g., pulmonary oedema), even if a total body extracellular fluid deficit remains. In less severe extracellular fluid depletion, fluid should be replaced at a rate of 1000 mL every 4 to 6 hours, with repeated clinical assessment.

Further reading

Clinical Guideline CG174: Intravenous fluid therapy in adults in hospital. *National Institute of Healthcare and Clinical Excellence (NICE)* 2013. https://www.nice.org.uk/guidance/cg174.

Curran Jeffrey D, Major Paityn, Tang Kent et al. Comparison of balanced crystalloid solutions: A systematic review and meta-analysis of randomized controlled trials. *Crit Care Explor* 2021; 3(5):e0398.

El Gkotmi N, Kosmeri C, Filippatos TD et al. Use of intravenous fluids/solutions: a narrative review. *Curr Med Res Opin* 2017; 33:459–471.

Florez ID, Sierra J, Pérez-Gaxiola G. Balanced crystalloid solutions versus 0.9% saline for treating acute diarrhoea and severe dehydration in children. *Cochrane Database Syst Rev* 2023;(5):CD013640.

DISORDERS OF SODIUM CONCENTRATION

These are generally best thought of as disorders of body water content. As discussed above, while total body sodium content is regulated by volume receptors; water content is adjusted to maintain a normal osmolality and (in the absence of abnormal osmotically active solutes) a normal sodium concentration. Disturbances of sodium concentration are caused by disturbances of water balance.

Regulation of water excretion

The chief mechanisms for regulating osmolality are thirst and the ability of the kidney to concentrate and dilute urine. Both mechanisms are principally controlled by intracellular osmoreceptors in the hypothalamus (and to some extent by volume receptors in capacitance vessels close to the heart, and via the renin–angiotensin system). Changes in the plasma Na^+ concentration and osmolality sensed by osmoreceptors influence both thirst sensation and the release of AVP (vasopressin) from the supraoptic and paraventricular nuclei of the anterior hypothalamus.

AVP plays a central role in urinary concentration by increasing the water permeability of the cortical and medullary collecting ducts. There are three major G-protein coupled receptors for vasopressin (AVP) which underlie its different physiological effects:

- ***V_{1A}*** found in vascular smooth muscle cells: activation induces vasoconstriction.
- ***V_{1B}*** found in the anterior pituitary and throughout the brain: this mediates the effect of AVP on the pituitary, leading to adrenocorticotrophic hormone (ACTH) release.
- ***V_2*** receptors in the principal cells of the kidney distal convoluting tubule and collecting ducts: these mediate the AVP response on urinary concentration.

The ability of AVP to increase the urine osmolality is first dependent on active transport of sodium in the ascending limb of the loop of Henle (which occurs constitutively and is not dependent on AVP). In the loop of Henle, active reabsorption of NaCl without water along with the counter-current multiplier has the effect of making the medullary interstitium highly concentrated, while also diluting the tubular fluid. The newly formed urine subsequently passes through this area of high osmolality as the distal collecting ducts flow through the medullary pyramids into the renal pelvis, so the large osmolar gradient between lumen and interstitium can be used to reabsorb water and concentrate the urine (Fig. 27.6).

Effect of AVP on the kidney

The cortical collecting duct has two cell types with very different functions:

- ***Principal cells*** (about 65%) have sodium and potassium channels in the apical membrane and, as in all sodium-reabsorbing cells, Na^+/K^+-ATPase pumps in the basolateral membrane. They are the primary targets of AVP (regulating osmolality), and also aldosterone (regulating extracellular volume).
- ***Intercalated cells*** play a role in hydrogen and bicarbonate handling and in potassium reabsorption and are not directly involved in sodium or water excretion.

AVP acts on V_2 (vasopressin) receptors located on the basolateral surface of principal cells, resulting in pre-formed cytoplasmic vesicles that contain unique water channels (called aquaporins) moving to, and then being inserted into, the luminal membrane. The water channels span the luminal membrane and permit water movement out of urine into the cells down a favourable osmotic gradient (Fig. 27.11). This water is then rapidly returned to the systemic circulation across the basolateral membrane. When the AVP effect has worn off, the water channels are removed from the luminal membrane by endocytosis and returned to the cytoplasm.

The effect is that in the absence of AVP, the collecting ducts are relatively impermeable to water and little is reabsorbed, resulting in dilute urine being excreted in the context of low plasma osmolality.

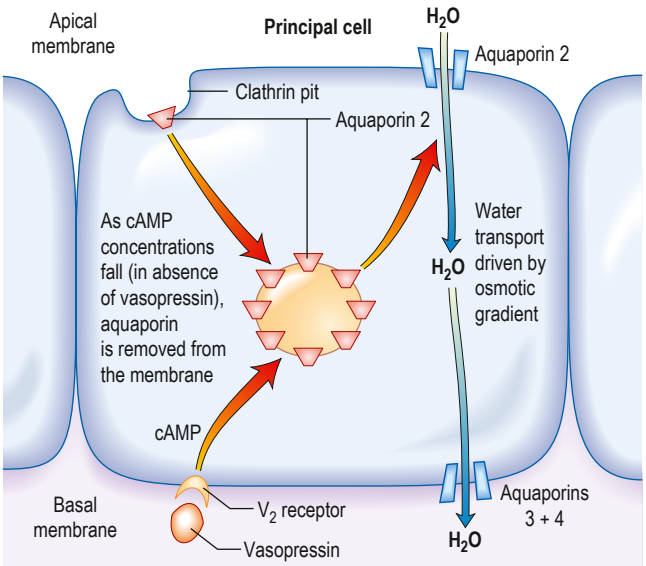


Fig. 27.11 Aquaporin-mediated water transport in the collecting duct. Stimulation of the vasopressin 2 receptor causes cyclic adenosine monophosphate (cAMP)-mediated insertion of the aquaporin into the apical membrane, allowing water transport down the osmotic gradient. *(Adapted from Connolly DL, Shanahan CM, Weissberg PL. Water channels in health and disease.* Lancet *1996; 347:211, with permission from Elsevier.)*

By contrast, the presence of AVP significantly increases the permeability of the collecting duct tubular cells when plasma osmolality is high, promoting water reabsorption down the favourable osmotic gradient. As a result, water is retained in the body, and a small volume of highly concentrated urine is passed.

Plasma osmolality and AVP release

The rate of AVP secretion is directly affected by plasma osmolality. At a plasma osmolality of <275 mosmol/kg (which usually represents a plasma Na^+ concentration of <135–137 mmol/L), there is essentially no circulating AVP. As the plasma osmolality rises above this threshold, the secretion of AVP increases progressively.

Two simple examples will illustrate the basic mechanisms of osmoregulation, which is so efficient that the plasma Na^+ concentration is normally maintained within 1% to 2% of its baseline value.

- ***Ingestion of a water load*** leads to an initial reduction in the plasma osmolality, thereby diminishing the release of AVP. The ensuing reduction in water reabsorption in the collecting ducts allows the excess water to be excreted in a dilute urine. This process is normally so efficient that volume regulation is not affected and there is no change in ANP release or in the activity of the renin–angiotensin–aldosterone system. Thus, a dilute urine is excreted and there is little alteration in the excretion of Na^+.
- ***Administration of both water and sodium*** (e.g., by giving intravenous 0.9% saline), by contrast, causes an increase in volume but no change in plasma osmolality. In this setting, ANP secretion is increased, aldosterone secretion is reduced, and AVP secretion does not change. The net effect is the appropriate excretion of the excess Na^+ in a relatively isoosmotic urine.

By contrast, ***water loss*** resulting from sweating is followed by a rise in both plasma osmolality and AVP secretion, causing water reabsorption in the collecting duct and the (appropriate) excretion of a small volume of concentrated urine. This renal effect of AVP minimizes further water loss but does not replace the existing water deficit; this is accomplished by the stimulation of thirst. The importance of thirst is illustrated by studies in patients with central diabetes insipidus, who are deficient in AVP. These patients often complain of marked polyuria, which is caused by an inability to appropriately reabsorb water in the collecting ducts. However, they do not typically become hypernatraemic because urinary water loss is offset by the thirst mechanism.

Clinical disorders of osmoregulation

Diabetes insipidus (see also Chapter 34, 'Diabetes insipidus' section) describes a failure in the mechanism of urinary concentration relating to AVP deficiency or resistance. Patients experience excessive polyuria and thirst, and are prone to becoming profoundly dehydrated and hyperosmolar if their access to water is restricted. There are two forms:

- ***AVP-deficiency (also referred to as 'cranial diabetes insipidus')*** occurs when AVP secretion is impaired. It usually occurs in the context of injury to the hypothalamus or pituitary, for example, after skull base neurosurgery. AVP-deficiency can be treated using desmopressin, a synthetic analogue of AVP (see Chapter 34, 'The thirst axis' section).
- ***AVP-resistance (also referred to as 'nephrogenic diabetes insipidus')*** occurs where the cortical tubular cells are resistant to the effects of AVP. The commonest cause is long-term lithium treatment for bipolar affective disorder.

? Box 27.10 Causes of hyponatraemia with decreased extracellular volume (hypovolaemia)

Extrarenal (urinary sodium <20 mmol/L)
- Vomiting
- Diarrhoea
- Haemorrhage
- Burns
- Pancreatitis

Kidney (urinary sodium >20 mmol/L)
- Diuretics
- Osmotic diuresis (e.g., hyperglycaemia, severe uraemia)
- Adrenocortical insufficiency
- Tubulointerstitial renal disease
- Recovery phase of acute tubular necrosis

By contrast, in the ***syndrome of inappropriate diuresis*** (***SIAD***, see also Chapter 34, 'Syndrome of inappropriate antidiuresis [SIAD]' section), there is impaired water excretion and dilutional hyponatraemia caused by the persistent presence of AVP. Correction of the hyponatraemia in this setting initially requires restriction of water intake.

AVP is also secreted by non-osmotic stimuli such as stress (e.g., surgery, trauma), markedly reduced effective circulatory volume (e.g., cardiac failure, hepatic cirrhosis), psychiatric disturbance, and nausea, irrespective of plasma osmolality. This is mediated by the effects of sympathetic overactivity on supraoptic and paraventricular nuclei. In addition to water retention, AVP release in these conditions promotes vasoconstriction owing to the activation of V_{1A} (vasopressin) receptors distributed in the vascular smooth muscle cells.

Hyponatraemia

Hyponatraemia (Na <135 mmol/L) is the most common biochemical abnormality in hospitalized patients, with up to 35% of inpatients developing hyponatraemia during their stay. There is often an associated disorder of extracellular volume, so that a clinical approach to hyponatraemia begins with establishing whether a patient has:

- hyponatraemia with hypovolaemia
- hyponatraemia with euvolaemia
- hyponatraemia with hypervolaemia.

Rarely, hyponatraemia may be a 'pseudo-hyponatraemia'. This occurs in hyperlipidaemia (either high cholesterol or high triglyceride) or hyperproteinaemia where there is a spuriously low measured sodium concentration, the sodium being confined to the aqueous phase but having its concentration expressed in terms of the total volume of plasma. In this situation, plasma osmolality is normal and therefore treatment of 'hyponatraemia' is unnecessary. Note that artefactual 'hyponatraemia', caused by taking blood from the limb into which fluid of low sodium concentration is being infused, should be excluded.

Hyponatraemia with hypovolaemia

This is due to salt loss in excess of water loss; the causes are listed in **Box 27.10**. In this situation, AVP secretion is initially suppressed (via the hypothalamic osmoreceptors) but, as fluid volume is lost, volume receptors override the osmoreceptors and stimulate both thirst and the release of AVP. This is an attempt by the body to defend circulating volume at the expense of osmolality.

With extrarenal losses and normal kidneys, the urinary excretion of sodium falls in response to the volume depletion, as does water

Box 27.11 Average concentrations and potential daily losses of water and electrolytes from the gut

	Na^+ (mmol/L)	K^+ (mmol/L)	Cl^- (mmol/L)	Volume (mL in 24 h)
Stomach	50	10	110	2500
Small intestine recent ileostomy	120	5	110	1500
Small intestine adapted ileostomy	50	4	25	500
Bile	140	5	105	500
Pancreatic juice	140	5	60	2000
Diarrhoea	130	10–30	95	1000–2000+

Box 27.12 Causes of hyponatraemia with normal extracellular volume (euvolaemia)

Abnormal arginine vasopressin (AVP) release
- Vagal neuropathy (failure of inhibition of AVP release)
- Deficiency of adrenocorticotrophic hormone (ACTH) or glucocorticoids (Addison's disease)
- Hypothyroidism
- Severe potassium depletion

Syndrome of inappropriate antidiuretic hormone secretion (SIADH)
- See **Box 34.52**

Psychiatric illness
- 'Psychogenic polydipsia'
- Antidepressant therapy

Increased sensitivity to AVP
- Chlorpropamide
- Tolbutamide

AVP-like substances
- Oxytocin
- Desmopressin

Unmeasured osmotically active substances stimulating osmotic AVP release
- Glucose
- Chronic alcohol misuse
- Mannitol
- Sick-cell syndrome (leakage of intracellular ions)

excretion, leading to concentrated urine containing <10 mmol/L of sodium. However, in salt-wasting kidney disease, compensation cannot occur and the only physiological protection is increased water intake in response to thirst.

Clinical features

With sodium depletion, the clinical picture is usually dominated by features of volume depletion (see 'Decreased extracellular volume' section). The diagnosis is usually obvious where there is a history of gut losses, diabetes mellitus with significant hyperglycaemia causing osmotic diuresis, or diuretic abuse. Examination of the patient is often more helpful than the biochemical investigations, which include plasma and urine electrolytes and osmolality.

Box 27.11 shows the potential daily losses of water and electrolytes from the gut. Losses due to kidney or adrenocortical disease may be less easily identified but a urinary sodium concentration of >20 mmol/L, in the presence of clinically evident volume depletion, suggests a renal loss.

Management

This is directed at the primary cause whenever possible.

In a healthy patient:
- Give oral electrolyte–glucose mixtures (see Box 42.36).
- Increase salt intake with slow sodium 60 to 80 mmol/day.

In a patient with vomiting or severe volume depletion:
- Give intravenous fluid with potassium supplements.
- Correction of acid–base abnormalities is usually not required.

Hyponatraemia with euvolaemia

Hyponatraemia with euvolaemia (**Box 27.12**) results from an intake of water in excess of the kidneys' ability to excrete it (dilutional hyponatraemia); there is no change in body sodium content but the plasma osmolality is low.

- With normal kidney function, dilutional hyponatraemia is uncommon even if a patient drinks approximately 1 L per hour, although the term 'primary' or 'psychogenic polydipsia' is used to describe patients where hyponatraemia is driven by excessive water consumption. This can be demonstrated by evidence of large volume of highly dilute urine as the kidney responds appropriately by attempting to excrete all of the free water.
- Postoperative hyponatraemia is a common clinical problem (seen in almost 1% of patients), with *symptomatic* hyponatraemia occurring in 20% of these patients. Sometimes, this is exacerbated by inappropriate intravenous fluid administration, for example, over-generous administration of 5% glucose. Premenopausal females are at most risk for developing hyponatraemic encephalopathy postoperatively, with postoperative AVP values in young females being 40 times higher than in young males. To prevent iatrogenic hyponatraemia, hypotonic intravenous fluids should be avoided in favour of and administered 0.9% saline or balanced replacement fluids (e.g., Hartmann's solution) unless clinically contraindicated. The serum sodium should be measured daily in any patient receiving continuous parenteral fluid.
- Marathon runners who drink excess water can become hyponatraemic. Even so-called 'isotonic sports drinks' can lead to hyponatraemia, as they contain little sodium, their osmolality being made up with carbohydrates, which are metabolized into energy and water.
- Some degree of hyponatraemia is usual in acute oliguric kidney injury, while in CKD it is most often due to ill-given advice to 'push' fluids.

Syndrome of inappropriate AVP secretion

This is described in Chapter 34. There is inappropriate secretion of AVP, causing water retention and hyponatraemia.

Clinical features

Symptoms of hyponatraemia are common when hyponatraemia develops *acutely* (<48 h, often postoperatively). Overt neurological symptoms rarely occur until the serum sodium is less than 125 mmol/L, and are more usually associated with values around 115 mmol/L or lower, particularly when *chronic*. In elderly patients, however, even mild hyponatraemia with a serum sodium of 130 to 135 mmol/L has been associated with increased falls risk and subtle changes in cognition. They are due to the movement of water into brain cells in response to the fall in extracellular osmolality.

Hyponatraemic encephalopathy

Symptoms and signs include headache, confusion, and restlessness leading to drowsiness, myoclonic jerks, generalized convulsions, and eventually coma. Magnetic resonance imaging (MRI) of the brain may reveal cerebral oedema.

Risk factors for developing hyponatraemic encephalopathy

The brain's adaptation to hyponatraemia initially involves extrusion of blood and CSF, as well as sodium, potassium, and organic osmolytes, in order to decrease brain osmolality. Various factors can interfere with successful adaptation. These factors, rather than the absolute change in serum sodium, predict whether a patient will suffer hyponatraemic encephalopathy.

- ***Premenopausal females*** are more likely to develop encephalopathy than postmenopausal females and males because of inhibitory effects of sex hormones and the effects of vasopressin on cerebral circulation, resulting in vasoconstriction and hypoperfusion of the brain.
- ***Hypoxaemia*** is a major risk factor for hyponatraemic encephalopathy. Patients with hyponatraemia who develop hypoxia, from any cause, have a high risk of mortality. Hypoxia is the strongest predictor of mortality in patients with symptomatic hyponatraemia.

Investigations

The cause of hyponatraemia with apparently normal extracellular volume requires investigation:

- ***Plasma and urine electrolytes and osmolalities.*** The plasma concentrations of sodium, chloride, and urea are low, giving a low osmolality. The urine sodium concentration is usually high and the urine osmolality is typically higher than the plasma osmolality.
- ***Further investigations.*** These are needed to exclude Addison's disease and hypothyroidism.
- ***Potassium and magnesium depletion*** potentiate AVP release and are causes of diuretic-associated hyponatraemia.

SIADH is often over-diagnosed. Some causes are associated with a lower set-point for AVP release, rather than completely autonomous AVP release; an example is chronic alcohol use.

Management

The underlying cause should be corrected where possible (**Box 27.13**).

- ***Most patients.*** Most cases are simply managed by restriction of water intake (to 1000 or even 500 mL/day) with review of diuretic therapy. Magnesium and potassium deficiency must be corrected. In mild sodium deficiency (hypovolaemic hyponatraemia), 0.9% saline given slowly (1 L over 12 h) is sufficient.
- ***Acute onset with symptoms.*** This is an acute medical emergency and should be treated aggressively and immediately.

In patients who demonstrate severe neurological signs, such as fits, coma, or cerebral oedema, European guidelines suggest immediate infusion of 150 mL of ***hypertonic saline*** (3%, 513 mmol/L) over 20 minutes. Repeat measurement of sodium concentration allows repeat dosing, aiming for an increase in plasma osmolality by 5 mmol/L in the first hour, which is often enough to resolve symptoms. After the first hour (and achievement of the 5 mmol/L increase), sodium concentration should be repeated at 6, 12, and 24 hours; further hypertonic saline is not required unless symptoms persist.

Further management should be based around the cause of hyponatraemia, commonly comprising water restriction. The sodium concentration should not be allowed to rise by more than 10 mmol/L in the first 24 hours and 8 mmol/h during following 24 periods (especially in chronic hyponatraemia, where patients are especially at risk of osmotic demyelination if concentrations rise too fast). No further treatment is required once the sodium concentration reaches 130 mmol/L.

Box 27.13 Key points in the management of hyponatraemia

- Does the patient have symptomatic hyponatraemia (seizures/coma)?
  - Yes: involve critical care input early; 3% saline 150 mL with close monitoring and early repeat measurement. Aim for early increase of 5 mmol/L, avoiding excessive over-correction
- Always treat the underlying cause
- What is the volume status of the patient?
  - If hypovolaemic: rehydrate with 0.9% saline
  - If euvolaemic (SIADH): restrict fluid to 500–700 mL/day
  - If hypervolaemic (chronic kidney disease/heart failure): restrict fluid and salt, consider loop diuretic
- Is the hyponatraemia acute or chronic?
  - If acute: more likely to develop symptomatic hyponatraemia
  - If chronic: increased risk of osmotic demyelination due to rapid correction
- Check sodium levels frequently: may need to check Na every 1–2 h
- Correction must not exceed 10 mmol in first 24 h and 18 mmol in 48 h

Osmotic demyelination syndrome

A rapid rise in extracellular osmolality, particularly if there is an 'overshoot' to high serum sodium and osmolality, will result in the osmotic demyelination syndrome (ODS), formerly known as central pontine demyelination. This is a devastating neurological complication characterized by spastic quadriparesis, pseudobulbar palsy, and other features (e.g., parkinsonism) if there is also extrapontine myelinolysis. ODS is diagnosed by the appearance of characteristic hypointense lesions on T_1-weighted images and hyperintense lesions on T_2-weighted images on MRI; these can take 2 weeks or more to appear.

The pathophysiology of ODS is not fully understood. The most plausible explanation is that the brain loses organic osmolytes very quickly in order to adapt to hyponatraemia so that osmolarity is similar between the intracellular and extracellular compartments. However, neurons take up organic osmolytes slowly in the phase of rapid correction of hyponatraemia, resulting in a hypo-osmolar intracellular compartment and shrinkage of cerebral vascular endothelial cells. Consequently, the blood–brain barrier is functionally impaired, allowing lymphocytes, complement and cytokines to enter the brain, damage oligodendrocytes, activate microglial cells, and cause demyelination.

Prevention of ODS is based on avoiding rapid correction of hyponatraemia, especially where it is chronic. Factors predisposing to the development of ODS include alcohol excess, cirrhosis, malnutrition, hypokalaemia, hypoxaemia, and central nervous system radiation, and even greater care should be taken in correcting sodium in patients where these features are present. Even a modest water diuresis in the elderly with reduced muscle mass is large enough to cause a rapid rise in plasma sodium. If plasma sodium rises too quickly due to anticipated water diuresis, administration of desmopressin to stop the water diuresis is beneficial. If plasma sodium rises regardless, then it should be actively lowered again to the maximum limit of correction (<8 mmol/L per day) with the administration of 5% glucose solution.

Sodium correction using vasopressin antagonists

Vasopressin V_2 receptor antagonists, including tolvaptan and conivaptan, produce a selective water diuresis without affecting sodium and potassium excretion. They raise the plasma sodium concentration in patients with hyponatraemia caused by SIADH, heart failure, and cirrhosis, and are licensed in many countries for correction of euvolaemic hyponatraemia. They should be used with caution because of the risk of over-rapid Na correction, but may be particularly useful in patients with chronic hyponatraemia caused by SIADH (see Chapter 34).

Hyponatraemia with hypervolaemia

The common causes of hyponatraemia due to water excess are shown in **Box 27.14**. In all these conditions, there is usually an element of reduced GFR with avid reabsorption of sodium and chloride

Box 27.14 Causes of hyponatraemia with increased extracellular volume (hypervolaemia)

- Heart failure
- Liver failure
- Oliguric kidney injury
- Hypoalbuminaemia

Box 27.15 Causes of hypernatraemia

Arginine vasopressin (AVP) deficiency
- Pituitary diabetes insipidus

Iatrogenic
- Administration of hypertonic sodium solutions
- Administration of drugs with a high sodium content (e.g., piperacillin)
- Use of 8.4% sodium bicarbonate after cardiac arrest

Insensitivity to AVP (AVP-resistance)
- Lithium
- Tetracyclines
- Amphotericin B
- Acute tubular necrosis
- Osmotic diuresis
- Total parenteral nutrition
- Hyperosmolar hyperglycaemic state (see Chapter 11, 'Hyperosmolar hyperglycaemic state' section)

Plus
- Deficient water intake: impaired thirst or consciousness
- Excessive water loss through skin or lungs

in the proximal tubule. This leads to reduced delivery of chloride to the 'diluting' ascending limb of the loop of Henle and a reduced ability to generate 'free water', with a consequent inability to excrete dilute urine. This is commonly compounded by the administration of diuretics that block chloride reabsorption and interfere with the dilution of filtrate either in the loop of Henle (loop diuretics) or distally (thiazides). Oral fluid restriction is often required alongside diuretic therapy.

Hypernatraemia

This is much rarer than hyponatraemia and nearly always indicates a water deficit. Causes are listed in **Box 27.15**.

Hypernatraemia is always associated with increased plasma osmolality, which is a potent stimulus to thirst. None of the factors listed in **Box 27.15** causes hypernatraemia unless thirst sensation is abnormal or access to water is limited. For instance, a patient with diabetes insipidus will maintain a normal serum sodium concentration by maintaining a high water intake until an intercurrent illness prevents this. Thirst is frequently deficient in elderly people, making them more prone to water depletion. Hypernatraemia may occur in the presence of normal, reduced, or expanded extracellular volume, and does not necessarily imply that total body sodium is increased.

Clinical features

Symptoms of hypernatraemia are non-specific. Nausea, vomiting, fever, and confusion may occur. A history of longstanding polyuria, polydipsia, and thirst suggests diabetes insipidus. Assessment of extracellular volume status guides resuscitation. Mental state should be assessed. Convulsions occur in severe hypernatraemia.

Investigations

Simultaneous urine and plasma osmolality and sodium should be measured. Plasma osmolality is high in hypernatraemia. Passage of urine with an osmolality lower than that of plasma in this situation is clearly abnormal and indicates diabetes insipidus. In pituitary diabetes insipidus, urine osmolality will increase after administration of desmopressin; the drug (a vasopressin analogue) has no effect in AVP-resistance (see Chapter 34). If urine osmolality is high, this suggests either an osmotic diuresis due to an unmeasured solute (e.g., urea-induced diuresis caused by protein metabolism during parenteral feeding) or excessive extrarenal loss of water (e.g., heat stroke).

Management

Treatment is that of the underlying cause. For example:
- In AVP-deficiency (cranial diabetes insipidus), replace AVP in the form of desmopressin, a stable non-pressor analogue of AVP.
- Withdraw nephrotoxic drugs and replace water either orally or, if necessary, intravenously.

In ***severe*** (>170 mmol/L) hypernatraemia, 0.9% saline (150 mmol/L) should be used initially. Avoid too rapid a drop in serum sodium concentration; the aim is correction over 48 hours, as overrapid correction may lead to cerebral oedema.

In ***less severe*** (e.g., >150 mmol/L) hypernatraemia, the treatment is 5% glucose or 0.45% saline; the latter is obviously preferable in hyperosmolar diabetic coma. Very large volumes—5 L/day or more—may need to be given in diabetes insipidus.

If there is clinical evidence of volume depletion (see 'Decreased extracellular volume' section), this implies that there is a sodium deficit as well as a water deficit.

Further reading

Adrogué HJ, Tucker BM, Madias NE. Diagnosis and management of hyponatremia: *JAMA* 2022; 19;328(3):280–291.
Arzhan S, Roumelioti ME, Litvinovich I, Bologa CG, Myers OB, Unruh ML. Hypernatremia in hospitalized patients: A large population-based study. *Kidney360* 2022; 3 (7):1144–1157.
Cuesta M, Garrahy A, Thompson CJ. SIAD: practical recommendations for diagnosis and management. *J Endocrinol Invest* 2016; 39:991–1001.
Hureaux M, Vargas-Poussou R. Genetic basis of nephrogenic diabetes insipidus. *Mol Cell Endocrinol* 2023; 560:111825.
Mustajok S. Severe hyponatraemia (P-Na<116 mmol/l) in the emergency department: a series of 394 cases. *Intern Emerg Med* 2023; 18(3): 781–789.
Sulyok E, Farkas B, Nagy B, Várnagy Á, Kovács K, Bódis J. Tissue sodium accumulation: Pathophysiology and clinical implications. *Antioxidants (Basel)* 2022; 11(4):750.

DISORDERS OF POTASSIUM CONCENTRATION

Regulation of serum potassium concentration

The World Health Organization (WHO) recommends a dietary intake of 90 mmol/day to reduce blood pressure; most people have a dietary intake of between 80 and 150 mmol daily, depending on fruit and vegetable intake. Most of the body's potassium (3500 mmol in an adult) is intracellular. Serum potassium levels are controlled by:
- uptake of K^+ into cells
- urinary excretion
- extrarenal losses (e.g., gastrointestinal).

Uptake of potassium into cells is governed by the activity of the Na^+/K^+-ATPase in the cell membrane and by H^+ concentration.

Uptake is ***stimulated*** by:
- insulin
- β-adrenergic stimulation
- alkalosis—K^+ exchanged for H^+ across cell membranes.

Uptake is ***decreased*** by:

- α-adrenergic stimulation
- acidosis
- cell damage or cell death—resulting in massive K^+ release.

The ***kidney*** plays a pivotal role in the maintenance of potassium balance by varying its secretion with changes in dietary intake. Over 90% of the filtered potassium is reabsorbed in the proximal tubule and the loop of Henle and <10% of the filtered load is delivered to the early distal tubule. Potassium absorption in the proximal tubule is entirely passive and follows that of sodium and water, while its reabsorption in the thick ascending limb of the loop of Henle is mediated by the sodium–potassium-2–chloride co-transporter. However, potassium is secreted by the principal cells in the cortical and outer medullary collecting tubule. Secretion in these segments is very tightly regulated in health and can be varied according to individual needs; it is responsible for most of urinary potassium excretion.

Renal excretion of potassium is increased by aldosterone, which stimulates K^+ and H^+ secretion in exchange for Na^+ in the principal cells of the collecting duct (see **Fig. 27.9**). Because H^+ and K^+ are interchangeable in the exchange mechanism, acidosis decreases, and alkalosis increases the secretion of K^+. Aldosterone secretion is stimulated by hyperkalaemia and increased angiotensin II levels, as well as by some drugs, and this acts to protect the body against hyperkalaemia and against extracellular volume depletion. The body adapts to dietary deficiency of potassium by reducing aldosterone secretion. However, because aldosterone is also influenced by volume status, conservation of potassium is relatively inefficient, and significant potassium depletion may therefore result from prolonged dietary deficiency.

A number of drugs affect K^+ homeostasis by affecting aldosterone release (e.g., heparin, NSAIDs) or by directly affecting how the kidney handles potassium (e.g., diuretics).

Other endogenous proteins and metabolites also affect potassium homeostasis. Klotho, an anti-ageing protein expressed in the distal tubule (and other organs), increases potassium excretion. CD63, a tetra-spanning protein, inhibits its excretion. Moreover, protein kinase A- and C-mediated phosphorylation inhibits conductance of K^+ channels in the principal cells of the collecting duct but the cytochrome P450-epoxygenase-mediated metabolite of arachidonic acid (11–12-epoxyeicosatrienoic acid) activates these channels and plays a role in overall potassium homeostasis.

Normally, only about 10% of daily potassium intake is excreted in the gastrointestinal tract. Vomit contains around 5 to 10 mmol/L of K^+, but prolonged vomiting causes hypokalaemia by inducing sodium depletion, stimulating aldosterone, which increases renal potassium excretion. Potassium is secreted by the colon, and diarrhoeal fluid contains 10 to 30 mmol/L of K^+; profuse diarrhoea can therefore induce marked hypokalaemia. Colorectal villous adenomas may rarely produce profuse diarrhoea and marked K^+ loss.

Hypokalaemia

Aetiology

Common causes

The most common causes of chronic hypokalaemia are diuretic treatment (particularly thiazides) and hyperaldosteronism. In hospitalized patients, acute hypokalaemia is most often caused by diuretic (thiazide or loop) use and intravenous fluids without potassium, and redistribution into cells, particularly in the case of diabetic ketoacidosis (see Chapter 35), where the use of fluids without potassium combined with insulin treatment can cause a rapid fall in the serum potassium. The common causes are shown in **Box 27.16**.

? Box 27.16 Causes of hypokalaemia

Increased renal excretion
(Urinary K^+ >20 mmol/day)
- Diuretics:
  - Thiazides
  - Loop diuretics

Increased aldosterone secretion
- Liver failure
- Heart failure
- Nephrotic syndrome
- Cushing's syndrome
- Conn's syndrome
- Adrenocorticotrophic hormone (ACTH)–producing tumours

Exogenous mineralocorticoid
- Corticosteroids
- Liquorice (potentiates actions of cortisol in the kidney)

Kidney disease
- Renal tubular acidosis types 1 and 2
- Renal tubular damage (diuretic phase)
- Nephrotoxicity:
  - Amphotericin
  - Aminoglycosides
  - Cytotoxic drugs
- Release of urinary tract obstruction
- Bartter's syndrome
- Liddle's syndrome
- Gitelman's syndrome

Reduced intake of K^+
- Intravenous fluids without K^+
- Dietary deficiency

Redistribution into cells
- β-Adrenergic stimulation
- Acute myocardial infarction
- Beta-agonists, e.g., salbutamol
- Insulin treatment, e.g., treatment of diabetic ketoacidosis
- Correction of megaloblastic anaemia, e.g., B_{12} deficiency
- Alkalosis
- Hypokalaemic periodic paralysis

Gastrointestinal losses
(Urinary K^+ <20 mmol/day)
- Vomiting
- Severe diarrhoea
- Purgative abuse
- Villous adenoma
- Ileostomy or ureterosigmoidostomy
- Fistulae
- Ileus/intestinal obstruction

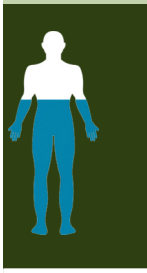
Rare causes

These rare causes are discussed in detail because they show the mechanisms of how diuretics can affect the kidney.

Bartter's syndrome

Bartter's syndrome (clinically similar to the effects of treatment with loop diuretics) consists of metabolic alkalosis, hypokalaemia, hypercalciuria, occasionally hypomagnesaemia (see 'Hypomagnesaemia' section), normal or low blood pressure, and an elevated plasma renin and aldosterone. The primary defect in this disorder is an impairment in sodium and chloride reabsorption in the thick ascending limb of the loop of Henle (see **Fig. 27.7**). Mutation in the genes encoding either the sodium–potassium-2–chloride co-transporter *(NKCC2)*, the ATP-regulated renal outer medullary potassium channel *(ROMK)* or kidney-specific basolateral chloride channels *(CLC-Kb)*—Bartter's syndrome types I, II, and III, respectively—causes loss of function of these channels, with consequent impairment of sodium and chloride reabsorption. Loss-of-function mutations of the protein Barttin cause type IV Bartter's syndrome, which is associated with sensorineural deafness and kidney failure.

In summary, these defects in sodium chloride transport initiate the following sequence, which is almost identical to that seen with chronic ingestion of a loop diuretic: the initial salt loss leads to mild volume depletion, resulting in activation of the renin–angiotensin–aldosterone system; the combination of hyperaldosteronism and increased distal flow (owing to the reabsorptive defect) enhances potassium and hydrogen secretion at the secretory sites in the collecting tubules, leading to hypokalaemia and metabolic alkalosis.



Diagnostic pointers include high urinary potassium and chloride despite low serum values, as well as increased plasma renin. Hyperplasia of the juxtaglomerular apparatus is seen on renal biopsy. Hypercalciuria is a common feature but magnesium wasting, though rare, also occurs. Differential diagnoses include diuretic abuse (where the urinary potassium and chloride concentrations are low) and primary hyperaldosteronism (where renin levels are low).

Treatment is with combinations of potassium supplements, amiloride, and rarely indometacin to block prostaglandin, which then restores vasoresponsiveness to angiotensin II and potassium.

Gitelman's syndrome

Gitelman's syndrome is characterized by hypokalaemia, metabolic alkalosis, hypocalciuria, hypomagnesaemia, normal blood pressure, and elevated plasma renin and aldosterone. There are striking similarities between Gitelman's syndrome and the biochemical abnormalities induced by chronic thiazide diuretic administration. Thiazides act in the distal convoluted tubule to inhibit the function of the apical sodium–chloride co-transporter (NCCT) (see **Fig. 27.8**). Analysis of the gene encoding the NCCT has identified loss-of-function mutations in Gitelman's syndrome.

Defective NCCT function leads to increased solute delivery to the collecting duct (as in Bartter's syndrome), with resultant solute wasting, volume contraction, and an aldosterone-mediated increase in potassium and hydrogen secretion. Impaired function of NCCT causes hypocalciuria. Impaired sodium reabsorption across the apical membrane, coupled with continued intracellular chloride efflux across the basolateral membrane, causes the cell to become hyperpolarized. This in turn stimulates calcium reabsorption via apical, voltage-activated calcium channels. Decreased intracellular sodium also facilitates calcium efflux via the basolateral sodium–calcium exchanger.

Treatment consists of potassium and magnesium supplementation ($MgCl_2$) and a potassium-sparing diuretic. Volume resuscitation is usually not necessary, because patients are not dehydrated.

Liddle's syndrome

This is characterized by potassium wasting, hypokalaemia, and alkalosis, but is associated with low renin and aldosterone production, and high blood pressure. There is a mutation in the gene encoding for the amiloride-sensitive ENaC in the distal tubule/collecting duct. This constitutive activation of the ENaC results in excessive sodium reabsorption with coupled potassium and hydrogen secretion. Unregulated sodium reabsorption across the collecting tubule causes volume expansion, inhibition of renin and aldosterone secretion, and development of low-renin hypertension (see **Fig. 27.9**).

Treatment consists of sodium restriction, along with amiloride or triamterene administration. Both are potassium-sparing diuretics that directly close the sodium channels. The mineralocorticoid antagonist spironolactone is ineffective, since the increase in sodium-channel activity is not mediated by aldosterone.

Clinical features

Hypokalaemia is usually asymptomatic but severe hypokalaemia (<2.5 mmol) causes muscle weakness. Potassium depletion may also cause symptomatic hyponatraemia.

Hypokalaemia is associated with an increased frequency of atrial and ventricular ectopic beats. This association may not always be causal because adrenergic activation (for instance, after myocardial infarction) results in both hypokalaemia and increased cardiac irritability. Hypokalaemia in patients without cardiac disease is unlikely to lead to serious arrhythmias.

Hypokalaemia seriously increases the risk of digoxin toxicity by increasing binding of digoxin to cardiac cells, potentiating its action, and decreasing its clearance.

Chronic hypokalaemia is associated with interstitial kidney disease but the pathogenesis is not completely understood.

Box 27.17 Treatment of hypokalaemia

Cause	Treatment
Dietary deficiency	Increase intake of fresh fruit/vegetables or oral potassium supplements (20–40 mmol daily)
Hyperaldosteronism, e.g., cirrhosis, thiazide therapy	Spironolactone/eplerenone Co-prescription of a potassium-sparing diuretic with a similar onset and duration of action
Intravenous fluid replacement	Add 20 mmol of K^+/L of fluid with monitoring Ensure potassium is checked at least daily until 48 h after potassium supplementation stopped

Management

The underlying cause should be identified and treated where possible. **Box 27.17** shows some examples.

Acute hypokalaemia may correct spontaneously. In most cases, withdrawal of oral diuretics or purgatives, accompanied by the oral administration of potassium supplements in the form of slow-release potassium or effervescent potassium, is all that is required. Intravenous potassium replacement is needed only in conditions such as cardiac arrhythmias, muscle weakness, or severe diabetic ketoacidosis. Intravenous replacement should be used cautiously and with regular monitoring in patients with impaired renal function. Pre-mixed intravenous fluids with potassium supplementation (e.g., 0.9% saline with 20 mmol potassium chloride) have replaced the addition of potassium ampoules to standard intravenous fluid to reduce the risk of medication errors.

The treatment of adrenal disorders is described in Chapter 34.

Failure to correct hypokalaemia may be due to concurrent hypomagnesaemia. Serum magnesium should be measured and any deficiency corrected.

Hyperkalaemia

Aetiology

Common causes

Acute self-limiting hyperkalaemia occurs normally after vigorous exercise and is of no pathological significance. Hyperkalaemia in all other situations is due either to increased release from cells or to failure of excretion (**Box 27.18**). The most common causes are renal impairment and drug interference with potassium excretion. The combination of angiotensin-converting enzyme (ACE) inhibitors with potassium-sparing diuretics or NSAIDs is particularly dangerous.

Rare causes

Hyporeninaemic hypoaldosteronism

This is also known as type 4 renal tubular acidosis (see 'Type 4 renal tubular acidosis' section). Hyperkalaemia occurs because of acidosis and hypoaldosteronism.

Box 27.18 Causes of hyperkalaemia

Decreased excretion
- Acute kidney injury[a]
- Drugs[a]:
  - Amiloride
  - Triamterene
  - Mineralocorticoid receptor antagonists
  - ACE inhibitors/angiotensin receptor blockers
  - NSAIDs
  - Ciclosporin
  - Heparin
- Aldosterone deficiency
- Hyporeninaemic hypoaldosteronism (RTA type 4)
- Addison's disease
- Acidosis[a]
- Gordon's syndrome

Increased release from cells
(Decreased Na^+/K^+-ATPase activity)
- Acidosis
- Diabetic ketoacidosis
- Rhabdomyolysis/tissue damage
- Tumour lysis
- Succinylcholine (amplified by muscle denervation)
- Digoxin poisoning
- Vigorous exercise (α-adrenergic; transient)

Increased extraneous load
- Potassium chloride
- Salt substitutes
- Transfusion of stored blood[a]

Spurious
- Increased *in vitro* release from abnormal cells
- Leukaemia
- Infectious mononucleosis
- Thrombocytosis
- Familial pseudohyperkalaemia, e.g., haemolysis in syringe
- Increased release from muscles
- Vigorous fist clenching during phlebotomy[a]

[a]Common causes.

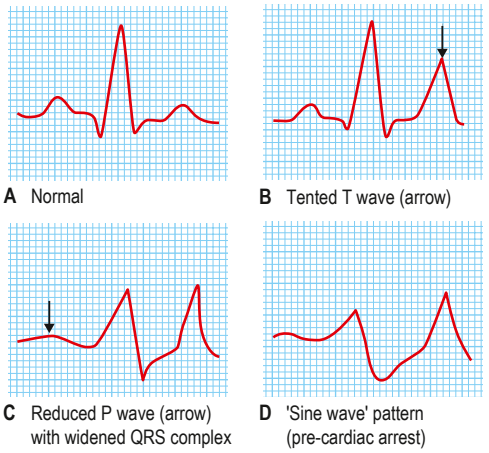


Fig. 27.12 Progressive ECG changes with increasing hyperkalaemia.

Hyperkalaemic periodic paralysis

Hyperkalaemic periodic paralysis (see Chapter 28) is precipitated by exercise, and is caused by an autosomal dominant mutation of the skeletal muscle sodium-channel gene.

Gordon's syndrome (familial hyperkalaemic hypertension, pseudohypoaldosteronism type 2)

This appears to be a mirror image of Gitelman's syndrome (see 'Gitelman's syndrome' section), in which primary renal retention of sodium causes hypertension, volume expansion, low renin/aldosterone, hyperkalaemia, and metabolic acidosis. There is also an increased sensitivity of sodium reabsorption to thiazide diuretics, suggesting that the thiazide-sensitive NCCT is involved (see **Fig. 27.8**). Genetic analyses, however, have excluded abnormalities in NCCT.

Suxamethonium and other depolarizing muscle relaxants

These cause release of potassium from cells. Induction of muscle paralysis during general anaesthesia may result in a rise of plasma potassium of up to 1 mmol/L. This is not usually a problem unless there is pre-existing hyperkalaemia.

Clinical features

Serum potassium of >7.0 mmol/L is a medical emergency and is associated with ECG changes (**Fig. 27.12**). Severe hyperkalaemia may be asymptomatic and may predispose to sudden death from asystolic cardiac arrest. Muscle weakness is often the only symptom, unless (as is commonly the case) the hyperkalaemia is associated with metabolic acidosis, causing Kussmaul respiration. Hyperkalaemia causes depolarization of cell membranes, leading to decreased cardiac excitability, hypotension, bradycardia, and eventual asystole.

Management

The acute management of hyperkalaemia is addressed in Chapter 11.

Further reading

[Guideline] Konrad M, Nijenhuis T, Ariceta G et al. Diagnosis and management of Bartter syndrome: executive summary of the consensus and recommendations from the European Rare Kidney Disease Reference Network Working Group for Tubular Disorders. *Kidney Int* 2021; 99(2):324–335.
Seyberth HW, Weber S, Kömhoff M. Bartter's and Gitelman's syndrome. *Curr Opin Pediatr* 2017; 29:179–186.

DISORDERS OF MAGNESIUM CONCENTRATION

Magnesium (Mg^{2+}) plays a pivotal role in many biological processes such as enzymatic reactions, gene transcription, bone remodelling, and neuromuscular stability. Approximately 99% of the Mg^{2+} in the body is in the intracellular compartment, mainly in bone (approximately 85%) and in muscle and soft tissues (approximately 14%). The other 1% is in the extracellular fluid.

Plasma magnesium levels are normally maintained within the range 0.7 to 1.1 mmol/L (1.4–2.2 mEq/L). The average daily magnesium intake is 15 mmol, which is absorbed mainly in the small intestine and, to a lesser extent, in the colon. In the healthy adult, there is no net gain or loss of magnesium from bone, so that balance is achieved by the urinary excretion of the net magnesium absorbed. The kidney reabsorbs between approximately 95% and 98% of the filtered Mg^{2+} and plays a major role in maintaining plasma Mg^{2+} concentrations within the normal range.

Control and renal handling of magnesium

Cortical thick ascending limb of Henle

Approximately 30% of Mg^{2+} is bound to plasma proteins but the remaining fraction is freely filterable. The major site of magnesium transport is the cortical thick ascending limb of the loop of Henle, where 65% to 70% of the filtered load is reabsorbed, with only 10% to 20% being reabsorbed in the proximal tubule (see **Fig. 27.7**). This transport is passive, paracellular and carried out by tight junction proteins (paracellin-1 and claudins). This process is driven by the lumen-positive electrochemical gradient, characteristic of this segment. This voltage gradient is created by the apical disproportionate net transport of two Cl^- to one Na^+ (by the bumetanide-sensitive sodium–potassium-2–chloride transporter) and the secretion of K^+ (via the ROMK) (see **Fig. 27.7**).

Loss-of-function mutations in these key reabsorptive processes lead to hypomagnesaemia as part of the distinctive clinical syndromes described later.

Bartter's syndrome

Hypomagnesaemia is rare in Bartter's syndrome (see 'Bartter's syndrome' section). This is because the transepithelial voltage, which is responsible for magnesium reabsorption, is preserved and any additional filtered magnesium will be offset by a compensatory increase in absorption in the distal convoluted tubule.

Familial hypomagnesaemia, hypercalciuria, and nephrocalcinosis

Familial hypomagnesaemia, hypercalciuria, and nephrocalcinosis (FHHNC) is characterized by excessive renal magnesium and calcium wasting; the main defect lies in the cortical thick ascending limb of Henle. Ten different mutations have been identified in a novel gene that encodes for paracellin-1 and claudins 16/19 complex, in the tight junction proteins.

Individuals develop bilateral nephrocalcinosis and progressive CKD. Patients also have elevated parathyroid hormone levels, which precedes any reduction in GFR. A substantial proportion of patients show incomplete distal renal tubular acidosis, hypocitraturia, and hyperuricaemia. Extrarenal involvement, such as myopia, nystagmus, and chorioretinitis, has been reported.

Distal convoluted tubule

The reabsorption rate in the distal convoluted tubule (DCT)—10%—is much lower than in the cTAL, but it defines the final urinary excretion, as there is no significant reabsorption in the collecting duct; 3% to 5% of filtered magnesium is finally excreted in the urine. Magnesium reabsorption in the DCT is transcellular and active (see Fig. 27.8).

Hypomagnesaemia

In addition to the familial causes described earlier, hypomagnesaemia most often develops as a result of deficient intake, defective gut absorption (associated with the use of proton pump inhibitors), or excessive gut or urinary loss (Box 27.19). It can also occur with acute pancreatitis, possibly owing to the formation of magnesium soaps in the areas of fat necrosis, and during drug treatment with aminoglycosides and cisplatinum compounds. Due to the severe effects of hypomagnesaemia, routine measurements of serum Mg^{2+} should be conducted in the critically ill, as well as in patients who are exposed to drugs and other conditions associated with Mg^{2+} deficiency.

Clinical features

Symptoms and signs (indicating a deficit of 0.5–1 mmol/kg) include irritability, tremor, ataxia, carpopedal spasm, hyper-reflexia, confusion, and seizures. An ECG may show a prolonged QT interval, broad flattened T waves, and occasional shortening of the ST segment.

Hypocalcaemia is often also present, and urinary calcium can help distinguish between causes of hypomagnesaemia: it is increased in cases associated with Bartter's syndrome, loop diuretics, and FHHNC and reduced in cases associated with Gitelman's syndrome and thiazide diuretics.

Box 27.19 Causes of hypomagnesaemia

Decreased magnesium absorption
- Malabsorption (severe)
- Malnutrition
- Alcohol excess
- Proton pump inhibitors

Increased renal excretion
- Drugs:
  - Loop diuretics
  - Thiazide diuretics
  - Digoxin
- Diabetic ketoacidosis
- Hyperaldosteronism
- SIADH
- Alcohol excess
- Hypercalciuria
- 1,25-dihydroxyvitamin D_3 deficiency
- Drug toxicity:
  - Amphotericin
  - Aminoglycosides
  - Cisplatin
  - Ciclosporin

Gut losses
- Prolonged nasogastric suction
- Excessive purgation
- Gastrointestinal/biliary fistulae
- Severe diarrhoea

Inherited tubular wasting
- Bartter's syndrome
- Familial hypomagnesaemia, hypercalciuria, and nephrocalcinosis
- Isolated dominant hypomagnesaemia
- Gitelman's syndrome
- Isolated recessive hypomagnesaemia
- Hypomagnesaemia with secondary hypocalcaemia
- EAST syndrome (Epilepsy, Ataxia, Sensorineural deafness and Tubulopathy)

Miscellaneous
- Acute pancreatitis

Box 27.20 Causes of hypermagnesaemia

- Impaired renal excretion:
  - Chronic kidney disease
  - Acute kidney injury
- Increased magnesium intake:
  - Purgatives, e.g., magnesium sulphate
  - Antacids, e.g., magnesium trisilicate
- Haemodialysis with high [Mg^{2+}] dialysate

Management

This involves the withdrawal of precipitating agents such as diuretics or purgatives. Oral magnesium replacement (e.g., magnesium glycerophosphate) is often ineffective due to poor absorption and diarrhoea, and if symptomatic (or with hypocalcaemia), intravenous replacement is generally preferred. This should be repeated daily and continued for 2 days after normal plasma levels have been achieved.

Relationship between hypomagnesaemia and plasma potassium

Magnesium depletion can lead to refractory hypokalaemia. The normal concentration of intracellular magnesium usually blocks secretory K^+ currents through ROMK channels. Magnesium depletion promotes K^+ loss by releasing ROMK from magnesium-mediated inhibition. Close monitoring, with potassium supplements if necessary, is required in patients presenting with primary symptomatic low plasma magnesium levels.

Hypermagnesaemia

This primarily occurs in patients with kidney disease who are given magnesium-containing laxatives or antacids. It can also be induced by magnesium-containing enemas. Causes are given in Box 27.20.

Clinical features

Symptoms and signs relate to neurological and cardiovascular depression, and include weakness with hyporeflexia proceeding to narcosis, respiratory paralysis, and cardiac conduction defects. Symptoms usually develop when the plasma magnesium level exceeds 2 mmol/L (4 mEq/L).

Management

Treatment requires withdrawal of any magnesium therapy. An intravenous injection of 10 mL of calcium gluconate 10% (2.25 mmol calcium) is given to antagonize the effects of hypermagnesaemia, along with glucose and insulin (as for hyperkalaemia) to lower the plasma magnesium level. Dialysis may be required in patients with severe kidney disease.

Further reading

Ayuk J, Gittoes NJ. Contemporary view of the clinical relevance of magnesium homeostasis. *Ann Clin Biochem* 2014; 51(Pt 2):179–188.
Wolf MT. Inherited and acquired disorders of magnesium homeostasis. *Curr Opin Pediatr* 2017; 29:187–198.

DISORDERS OF PHOSPHATE CONCENTRATION

Phosphate forms an essential part of most biochemical systems. The regulation of plasma phosphate level is both directly and closely linked to that of calcium.

About 85% of all body phosphorus is within bone, plasma phosphate normally ranging from 0.80 to 1.15 mmol/L and accounting for only 1% of the total body phosphate. Phosphate reabsorption from the glomerular filtrate occurs entirely and actively in the proximal tubule and is hormonally regulated. It is decreased by parathyroid hormone (PTH) via a cyclic adenosine monophosphate (cAMP)-dependent mechanism; thus, primary hyperparathyroidism is associated with low plasma levels of serum phosphate. Other factors that are known to control phosphate reabsorption in the proximal tubule are 1,25-dihydroxyvitamin D_3, sodium delivery to the proximal tubule, serum concentrations of calcium, bicarbonate, carbon dioxide tension, glucose, alanine, serotonin, dopamine and sympathetic activity.

Osteoblast-secreted phosphaturic factors (phosphatonins), such as fibroblast growth factor 23 (FGF23), matrix extracellular phosphoglycoprotein (MEPG), and frizzled-related protein 4 (FRP-4), play a role in phosphate homeostasis. FGF23 is the most extensively investigated phosphatonin. It binds to its receptor, FGFR1, in the kidney and causes phosphaturia; it also regulates vitamin D by inactivation of 1α-hydroxylase (CYP27B1) and upregulation of 24 hydroxylase (CYP24A1) enzymes, with the net result of low 1,25-vitamin D synthesis. Serum FGF23 is also the earliest marker of abnormal phosphate handling in CKD patients, occurring before changes in serum phosphate, calcium, or PTH (**Fig. 27.13**). FGF23 requires the enzyme Klotho for its activity. Loss-of-function mutation in either FGF23 or Klotho results in a similar phenotype of shortened lifespan, premature ageing, and hyperphosphataemia, and, as expected from the mode of action, increases 1,25 vitamin D levels.

Phosphate absorption is an active process carried out by a family of sodium–phosphate co-transporters (NPT) in the gut and kidneys. NPT2a and NPT2c are expressed in the brush border of the proximal tubule while NPT2b is expressed in lungs and intestine.

Under normal circumstances, plasma phosphate levels are kept constant; for example, after a phosphate-rich meal, the bone releases FGF23, which inhibits NPT2a and causes phosphate excretion. Moreover, phosphate in the plasma causes the release of PTH, which also inhibits NPT2a, NPT2c, and NHERF1, resulting in phosphaturia (see **Fig. 27.13**). These two principal mechanisms keep plasma phosphate levels within normal limits on a daily basis.

Hypophosphataemia

Significant hypophosphataemia (<0.4 mmol/L or <1.25 mg/dL) occurs in a number of clinical situations, owing to redistribution into cells, renal losses, or decreased intake (**Box 27.21**).

Clinical features include:

- muscle weakness, for example, diaphragmatic weakness, decreased cardiac contractility, skeletal muscle rhabdomyolysis. Diaphragmatic weakness may be severe in acute hypophosphataemia, and may impede the weaning of a patient from a ventilator
- a left shift in the oxyhaemoglobin dissociation curve (reduced 2,3-bisphosphoglycerate, 2,3-BPG) and rarely haemolysis
- confusion, seizures, and coma.

Aetiology

Hypophosphataemia related to parathyroid hormone

Primary hyperparathyroidism is a common cause of hypophosphataemia. Very rarely, gain-of-function mutations of the PTH1 receptor cause hypophosphataemia and Jansen's metaphyseal chondrodysplasia due to constitutive activation of PTH signalling, even in the presence of low or absent circulating PTH levels.

Hypophosphataemia related to vitamin D

Hypophosphataemia can be part of osteomalacia and rickets due to vitamin D deficiency, either dietary (globally, the most common cause) or genetic, and is usually accompanied by hypocalcaemia (calcipenic) and secondary hyperparathyroidism.

Vitamin D–dependent rickets type I

Also known as ***pseudo-vitamin D-deficient rickets***, this is caused by 1α-hydroxylase deficiency due to inactivating mutations in its gene. This condition manifests clinically in the first year of life with severe hypocalcaemia, often complicated by tetany, moderate hypophosphataemia, and enamel hypoplasia. The characteristic biochemical findings are normal serum levels of 25-hydroxyvitamin D, low values of 1,25-dihydroxyvitamin D and, usually, relatively high PTH levels. The treatment of choice is replacement therapy with calcitriol.

Vitamin D–dependent rickets type II

This is a form of vitamin D resistance and is known as ***hereditary vitamin D–resistant rickets***. It is an autosomal recessive disorder and is usually caused by loss-of-function mutations in the gene encoding the vitamin D receptor. The clinical manifestations vary widely, depending on the type of mutation within the vitamin D receptor and the amount of residual vitamin D receptor activity. Affected children usually develop rickets within the first 2 years of life, with alopecia in two-thirds of cases, which is due to lack of vitamin D receptor action within keratinocytes. The treatment involves calcitriol and calcium supplementation.

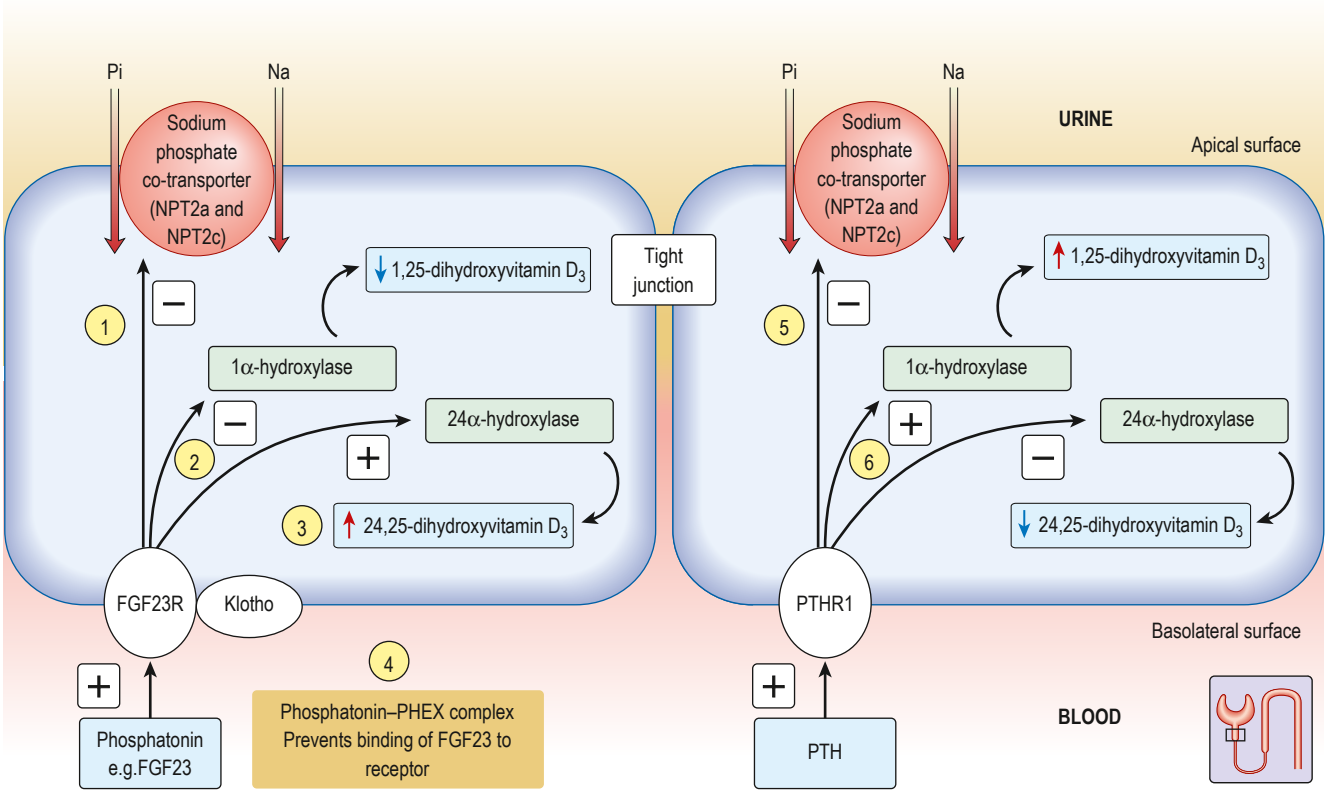


Fig. 27.13 PHEX and PTH regulation of phosphate transport and vitamin D metabolism in the proximal tubule. **(1)** Fibroblast growth factor 23 (FGF23) inhibits the sodium–phosphate co-transporter (NPT2a, c), causing phosphaturia. **(2)** Binding of FGF23 to the FGFR1–Klotho complex leads to lower levels of the active form of vitamin D (1,25-vitamin D) by inhibiting 1α-hydroxylase. **(3)** It also increases levels of inactive vitamin D (24,25-vitamin D) by stimulating 24α-hydroxylase. **(4)** PHEX (phosphate-regulating gene with homologies to endopeptidase on the X chromosome) acts as a competitive inhibitor, binding FGF23 and blocking its action. **(5)** Parathyroid hormone (PTH) also inhibits NPT2a, which leads to phosphaturia. **(6)** PTH has opposite effects on vitamin D metabolism compared with FGF23 by stimulating 1α-hydroxylase and inhibiting 24α-hydroxylase.

Box 27.21 Causes of hypophosphataemia

Redistribution
- Respiratory alkalosis
- Treatment of diabetic ketoacidosis (insulin drives phosphate into cells)
- Re-feeding (particularly with carbohydrate) after fasting or starvation
- After parathyroidectomy (hungry bone disease)

Decreased intake/absorption
- Diet
- Malabsorption
- Vomiting
- Gut phosphate binders, e.g., sevelamer
- Vitamin D deficiency
- Alcohol withdrawal

Renal losses
- Hyperparathyroidism
- Renal tubular defects
- Diuretics
- Tenofovir treatment

Hypophosphataemic rickets
- Vitamin D–dependent rickets types I and II
- Tumour-induced osteomalacia (TIO)
- Autosomal dominant and recessive
- X-linked dominant rickets (XLR)
- Dent's disease

Hypophosphataemia related to decreased renal reabsorption of phosphate

Excessive phosphatonins (FGF23)

This condition occurs in patients with ***tumour-induced osteomalacia*** (TIO), ***X-linked dominant hypophosphataemic rickets*** (XLR), and ***autosomal dominant hypophosphataemic rickets*** (ADHR). These syndromes have similar biochemical and osseous phenotypes. Patients have osteomalacia or rickets, reduced tubular phosphate reabsorption, hypophosphataemia, normal or low serum calcium, normal PTH and PTH-related protein concentrations, and normal or low 1,25-dihydroxyvitamin D_3.

In ***TIO***, there is excessive production of phosphaturic agents: for example, FGF23, MEPG, and FRP-4. The result is excess inhibition of the sodium–phosphate co-transporter in the proximal tubule and urinary phosphate wasting.

In ***ADHR***, FGF23 is mutated so that it is resistant to PHEX (phosphate-regulating gene with homologies to endopeptidases on the X chromosome) proteolysis.

In **XLR**, mutations in PHEX prevent binding to FGF23 and FRP-4, resulting in a net relative excess of phosphatonins.

Normal adaptive increases in 1,25-dihydroxyvitamin D_3 synthesis in response to low phosphate levels do not occur in TIO, ADHR, and XLR, aggravating phosphaturia (see Fig. 27.13).

Autosomal recessive hypophosphataemia with high FGF23 levels has been described in patients with mutations in the gene encoding dentin matrix protein 1 *(DMP1)*, which is a transcription factor produced by odontoblasts, osteoblasts, and osteocytes.

Reduced NPT2a/NHERF1 activity

Mutations in NPT2a (sodium–phosphate co-transporter 2a) and NHERF1 (sodium hydrogen regulator factor 1) have been identified in patients with low phosphate levels and a reduced ratio of the maximum tubular reabsorption of phosphate (TmP) normalized for GFR (TmP/GFR ratio). Heterozygous loss-of-function mutation in NPT2a and NHERF1 in patients is characterized by hypophosphataemia, low TmP/GFR, and normal PTH. Carriers of these mutations do not have hypercalcaemia, which means these mutations do not affect the action of PTH on bones.

Dent's disease

Dent's disease is the generally accepted name for a group of hereditary tubular disorders, including X-linked recessive nephrolithiasis with CKD, X-linked recessive hypophosphataemic rickets, and idiopathic low-molecular-weight proteinuria. It is characterized by low-molecular-weight proteinuria, hypercalciuria, hyperphosphaturia, nephrocalcinosis, kidney stones, and eventual kidney failure, with some patients developing rickets or osteomalacia.

Dent's disease is caused, in 60% of cases, by loss-of-function mutation of a proximal tubular endosomal chloride channel, CLC5. Mutations in OCRL1 occur in 15% of cases. This chloride channel, along with the proton pump, is essential for acidification of proximal tubular endosomes, required for normal endocytosis, degradation, and recycling of absorbed proteins, vitamins, and hormones. Defective endosomal acidification results in impaired endosomal degradation and recycling of endocytosed hormones such as PTH, causing the clinical phenotype of hyperparathyroidism. In addition, the receptors for reabsorption of low-molecular-weight proteins and albumin in the proximal tubules (megalin and cubilin) are decreased in Dent's disease, causing low-molecular-weight proteinuria and urinary leak of cytokines, hormones, and chemokines. The result is progressive kidney fibrosis and decline in kidney function.

Hypophosphataemia related to re-feeding syndrome

Hypophosphataemia may also occur as part of the 're-feeding syndrome' (see Chapter 25), which occurs in significantly malnourished patients who rapidly recommence feeding via oral, enteral, or parenteral routes.

Following starvation, the body's stores of phosphate are low due to reduced intake. After a carbohydrate load, insulin is released, which causes rapid cellular uptake of phosphate, along with potassium and magnesium, leading to a fall in the serum concentrations of these electrolytes.

Furthermore, insulin stimulates production of adenosine triphosphate (ATP) thus increasing the cellular requirement for phosphate. The inability to phosphorylate ADP to ATP due to inadequate phosphate concentrations leads to cardiac dysfunction and arrhythmias, respiratory failure due to weakness of the diaphragm, rhabdomyolysis, and seizures.

Re-feeding syndrome can be prevented by correcting electrolyte disturbances before re-feeding and avoiding rapid increases of caloric intake above the resting energy requirements.

Management is essentially supportive; patients may require hospitalization for electrolyte replacement and monitoring.

Diagnosis

Patients with hypophosphataemia should have their urinary fractional excretion of phosphate calculated from (urinary phosphate × serum creatinine)/(serum phosphate × urine creatinine). A value of <0.7 mmol/L indicates renal phosphate wasting. If PTH levels are high, then the patient is very likely to have hyperparathyroidism, either primary or secondary to vitamin D deficiency (acquired or genetic) or Dent's disease. If PTH levels are low, then the only possibility is gain-of-function mutation in PTH1R. If PTH levels are normal, then FGF23 concentrations can be measured: a high FGF23 concentration may indicate tumour-induced osteomalacia, or mutations in FGF23/Klotho, PHEX, or DMP1. Normal FGF23 and PTH levels point to mutations in NPT2a, NPT2c, and NHERF1.

Management

Oral phosphate supplementation and calcitriol (1,25-dihydroxyvitamin D) administration is required if there is vitamin D deficiency.

Treatment of ***acute hypophosphataemia*** is with intravenous phosphate at a maximum rate of 9 mmol every 12 hours. Repeated measurements of calcium and phosphate are required, as over-rapid administration of phosphate may lead to severe hypocalcaemia, particularly in the presence of alkalosis. ***Chronic hypophosphataemia*** can be corrected with oral effervescent sodium phosphate. The anti-FGF23 neutralizing antibody (burosumab) may be used to reverse the biochemical and clinical phenotypic manifestations in patients with X-linked osteomalacia or tumour-induced osteomalacia.

Hyperphosphataemia

Hyperphosphataemia is common in patients with CKD (see Chapter 26; Box 27.22), and is usually asymptomatic, but may be a cause of significant itching. Prolonged hyperphosphataemia causes hyperparathyroidism and periarticular and vascular calcification.

Familial tumoural calcinosis is characterized by calcifications of muscles, skin, eyelids, and vessels, as well as hyperostosis. In this condition, absence of glycosylation of FGF23 makes it unstable and more sensitive to proteolysis. This results in its deficiency and hyperphosphataemia due to increased renal phosphate reabsorption through increased NPT2a activity.

Usually, no treatment is required for acute hyperphosphataemia, as the causes are self-limiting. Treatment of chronic hyperphosphataemia in the context of CKD involves:

- reduction in phosphate intake
- use of gut-active phosphate binders such as sevelamer carbonate or calcium carbonate
- increased removal of phosphate through dialysis.

? Box 27.22 Causes of hyperphosphataemia

- Acute kidney injury
- Chronic kidney disease
- Phosphate-containing enemas
- Hyperparathyroidism/ pseudohyperparathyroidism
- Tumour lysis
- Rhabdomyolysis
- Familial tumoural calcinosis

Box 27.23 Relationship between [H⁺] and pH

pH	[H⁺] (nmol/L)
6.9	126
7.0	100
7.1	79
7.2	63
7.3	50
7.4	40
7.5	32
7.6	25

Further reading

Miller A, Mathew S, Patel S, Fordjour L, Chin V.L. Genetic disorders of calcium and phosphorus metabolism. *Endocrines* 2022; 3:150–167.
Walker EYX, et al. Burosumab in management of X-linked hypophosphataemia: a retrospective cohort study of growth and serum phosphate levels. *Arch Dis Child* 2023; 108:379–384.

ACID–BASE DISORDERS

The concentration of hydrogen ions in both extracellular and intracellular compartments is extremely tightly controlled, and very small changes lead to major cell dysfunction. Blood pH is tightly regulated and is normally maintained at between 7.38 and 7.42. Any deviation from this range indicates a substantial change in the hydrogen ion concentration [H⁺] because blood pH is the negative logarithm of [H⁺] (Box 27.23). The [H⁺] at a physiological blood pH of 7.40 is 40 nmol/L. An increase in the [H⁺] (a fall in pH) is termed acidaemia. A decrease in [H⁺] (a rise in the blood pH) is termed alkalaemia. The disorders that cause these changes in the blood pH are acidosis and alkalosis, respectively.

Normal acid–base physiology

The normal adult diet contains 70 to 100 mmol of acid (H⁺). Throughout the body, there are buffers that minimize any changes in blood pH that these ingested hydrogen ions might cause. Such buffers include intracellular proteins (e.g., haemoglobin) and tissue components (e.g., the calcium carbonate and calcium phosphate in bone), as well as the bicarbonate–carbonic acid buffer pair generated by the hydration of carbon dioxide. This buffer pair is clinically most relevant, in part because its contribution can be measured and because alterations in this buffer pair reveal changes in all other buffer systems. Bicarbonate ions [HCO_3^-] and carbonic acid (H_2CO_3) exist in equilibrium; and in the presence of carbonic anhydrase, carbonic acid dissociates to carbon dioxide and water, as expressed in the following equation:

$$H^+HCO_3^- \rightleftarrows H_2CO_3 \xleftrightarrow{\text{carbonic anhydrase}} CO_2 + H_2O.$$

The addition of hydrogen ions drives the reaction to the right, decreasing the plasma bicarbonate concentration [HCO_3^-] and increasing the arterial carbon dioxide pressure (P_aCO_2). As shown in the following ***Henderson–Hasselbalch equation***, a fall in the plasma [HCO_3^-] increases [H⁺] and thus lowers blood pH:

$$[H^+] = \frac{181 \times P_aCO_2}{[HCO_3^-]}$$

where [H⁺] is expressed in nmol/L, P_aCO_2 in kilopascals, [HCO_3^-] in mmol/L and 181 is the dissociation coefficient of carbonic acid. Alternatively, the equation can be expressed as:

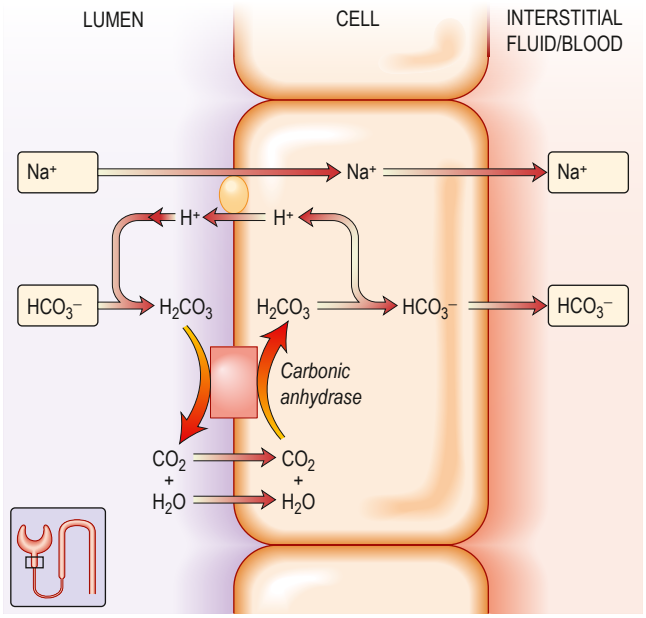


Fig. 27.14 Resorption of sodium bicarbonate (mainly proximal tubule). Bicarbonate is reclaimed by the secretion of H⁺ in exchange for Na⁺ into the tubule. This results in the formation of H_2CO_3, which is then broken down to CO_2. This is reabsorbed and converted back to H_2CO_3, which now dissociates into H⁺ and HCO_3^-. The net result is reabsorption of Na⁺ and HCO_3^-. This process is dependent on carbonic anhydrase within the cells and on the luminal surface of the tubular cell.

$$pH = pK + \frac{\log [HCO_3^-]}{[H_2CO_3]}$$

where pK = 6.1.

Although the acidaemia stimulates an increase in ventilation, which blunts this change in pH, increased ventilation does not regenerate the bicarbonate used in the buffering process. Thus, the bicarbonate used in the buffering process must be regenerated to maintain normal acid–base balance.

Consequently, the kidney must excrete hydrogen ions to return the plasma [HCO_3^-] to normal. Maintenance of a normal plasma [HCO_3^-] under physiological conditions depends not only on daily regeneration of bicarbonate but also on reabsorption of all bicarbonate filtered across the glomerular capillaries.

Renal reabsorption of bicarbonate

The plasma [HCO_3^-] is normally maintained at approximately 25 mmol/L. In individuals with a normal GFR (120 mL/min), about 4500 mmol of bicarbonate is filtered each day. If this filtered bicarbonate were not reabsorbed, the plasma [HCO_3^-] would fall, along with blood pH. Thus, maintenance of normal plasma [HCO_3^-] requires that essentially all of the bicarbonate in the glomerular filtrate be reabsorbed (Fig. 27.14).

The proximal convoluted tubule reclaims 85% to 90% of filtered bicarbonate; by contrast, the distal nephron reclaims very little. This difference is caused by the greater quantity of luminal (brush border) carbonic anhydrase in the proximal tubule than in the distal nephron. As a result of these quantitative differences, bicarbonate that escapes reabsorption in the proximal tubule is excreted in the urine.

Proximal tubular bicarbonate reabsorption is catalyzed by the Na^+/K^+-ATPase pump located in the basolateral cell membrane. By exchanging peritubular potassium ions for intracellular sodium

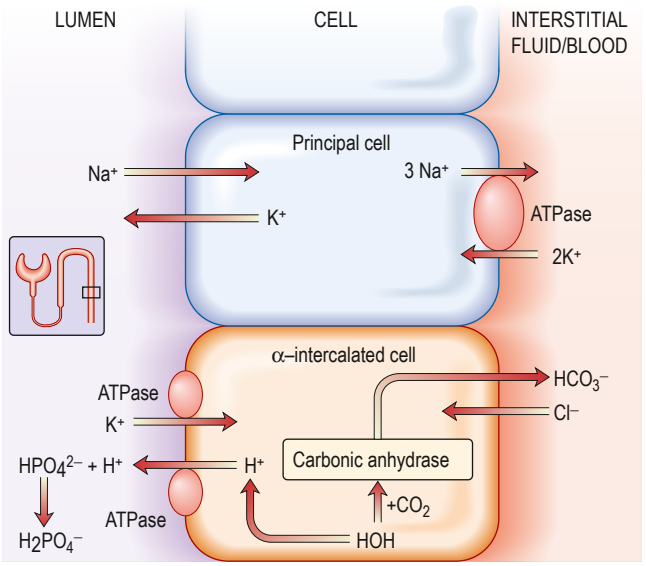


Fig. 27.15 Renal excretion of H^+. H^+ excretion is indirectly controlled by aldosterone in the cortical collecting ducts. Aldosterone enhances sodium absorption and potassium excretion by opening sodium channels and enhancing the activity of the Na^+/K^+-ATPase in the principal cell. These actions set up an electrochemical gradient favouring H^+ excretion. Aldosterone also facilitates H^+ excretion by stimulating H^+-ATPase in the intercalated cell.

ions, the pump keeps the intracellular sodium concentration low, allowing sodium ions to enter the cell by moving down the sodium concentration gradient from the tubule lumen to the cell interior. Hydrogen ions are transported in the opposite direction (at the Na^+–H^+ antiporter), thereby maintaining electroneutrality. Before bicarbonate enters the proximal tubule, it combines with secreted hydrogen ions, forming carbonic acid. In the presence of luminal carbonic anhydrase (CA-IV), carbonic acid rapidly dissociates into carbon dioxide and water, which can then rapidly enter the proximal tubular cell. In the cell, carbon dioxide is hydrated by cytosolic carbonic anhydrase (CA-II), ultimately forming bicarbonate, which is then transported down an electrical gradient from the cell interior, across the membrane into the peritubular fluid, and into the blood. In this process, each hydrogen ion secreted into the proximal tubule lumen is reabsorbed and can be resecreted; there is no net loss of hydrogen ions or net gain of bicarbonate ions.

Renal excretion of $[H^+]$

More acid is secreted into the proximal tubule (up to 4500 nmol of hydrogen ions each day) than in any other nephron segment (Fig. 27.15). However, the hydrogen ions secreted into the proximal tubule are almost completely reabsorbed with bicarbonate; consequently, proximal tubular hydrogen ion secretion does not contribute significantly to hydrogen ion elimination from the body. The excretion of the daily acid load requires hydrogen ion secretion in more distal nephron segments.

Most dietary hydrogen ions come from sulphur-containing amino acids that are metabolized to sulphuric acid (H_2SO_4), which then reacts with sodium bicarbonate as follows:

$$H_2SO_4 + 2NaHCO_3 \rightarrow Na_2SO_4 + 2CO_2 + 2H_2O.$$

Excess sulphate is excreted in the urine, whereas excess hydrogen ions are buffered by bicarbonate and lower the plasma $[HCO_3^-]$. This fall in plasma $[HCO_3^-]$ leads to a slight decrease in the blood pH, although a smaller decrease than would have occurred if buffer were unavailable. The subsequent excretion of hydrogen ions takes place primarily in the collecting duct and results in the regeneration of 1 mmol of bicarbonate for every mmol of hydrogen ions excreted in the urine.

The ***collecting duct*** has three types of cell:

- The ***principal cells*** with an aldosterone-sensitive Na^+ absorption site. These cells reabsorb Na^+ and H_2O and secrete K^+ under the influence of aldosterone.
- The ***α-intercalated cells***, which possess the proton pump for the active secretion of hydrogen ions in exchange for reabsorption of K^+ ions. Aldosterone increases H^+ ion secretion.
- The ***β-intercalated cells*** are mirror images of α-intercalated cells. Here the H^+-ATPase pump is located in the basolateral rather than the apical membrane, whereby H^+ ions are secreted into the peritubular capillary. The HCO_3^- ions, on the other hand, are secreted into the tubular lumen by an anion exchanger in the apical membrane.

Secretion of hydrogen ions from the cortical collecting ducts (CCD) is indirectly linked to sodium reabsorption.

- Aldosterone has several facilitating effects on hydrogen ion secretion. Aldosterone opens sodium channels in the luminal membrane of the ***principal cell*** and increases Na^+/K^+-ATPase activity. The subsequent movement of cationic sodium into the principal cell creates a negative charge within the tubule lumen. Potassium ions from the principal cells and hydrogen ions from the α-intercalated cells move out from the cells down the electrochemical gradient and into the lumen.
- Aldosterone also stimulates directly the H^+-ATPase in the α-intercalated cell, further enhancing hydrogen ion secretion. The H^+ to be secreted arises from the reassociation of H_2O and CO_2 in the presence of carbonic anhydrase; thus, a bicarbonate molecule is regenerated each time an H^+ is eliminated in the urine.

When hydrogen ions are secreted into the lumen of the collecting tubule, a tiny, but physiologically critical, fraction of these excess hydrogen ions remains in solution. Here, they increase the urinary $[H^+]$ and lower urinary pH below 4.0. Nevertheless, below this urine pH, inhibition of proton-secreting pumps, such as H^+-ATPase, severely restricts kidney secretion of more hydrogen ions. Consequently, secretion of hydrogen ions depends on the presence of buffers in the urine that maintain the urine pH at a level higher than 4.0.

In the presence of alkali excess, the homeostatic needs are reversed. Although the kidney can excrete excess alkaline load by reducing reabsorption of filtered bicarbonate in the proximal and distal tubule, the collecting ducts also contribute by secreting bicarbonate brought about by switching to β-intercalated cells. This switch enables kidneys to secrete bicarbonate and conserve H^+ ions.

Buffer systems in acid excretion

Two buffer systems are involved in acid excretion: the titratable acids, such as phosphate, and the ammonia system. Each system is responsible for excreting about half of the daily acid load of 50 to 100 mmol under physiological conditions (see Fig. 27.15).

Titratable acid

A titratable acid is a filtered buffer substance having a conjugate anion that can be titrated within the pH range occurring physiologically in the urine. Phosphoric acid (pK_a 6.8) is the usual titratable urinary buffer. Hydrogen ions bind to the conjugate anions of the titratable acids and are excreted in the urine. For each hydrogen ion excreted in this form, a bicarbonate ion is regenerated within the cell and returned to the blood (see Fig. 27.15).

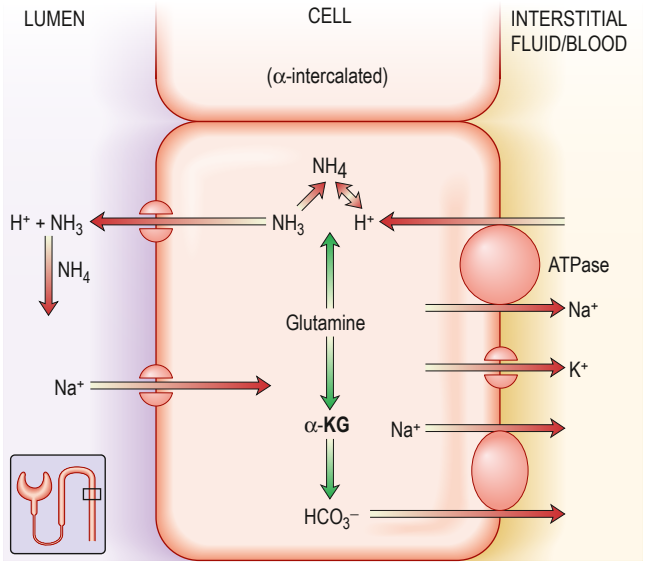


Fig. 27.16 The ammonia buffering system in the kidney. All ammonia used to buffer H^+ in the collecting duct is synthesized in the proximal convoluted tubule, and glutamine is the main source of this ammonia. As glutamine is metabolized, α-ketoglutarate (α-KG) is formed, which ultimately breaks down to bicarbonate that is then secreted into the peritubular fluid at an Na^+–HCO_3^- co-transporter.

Ammonium (NH_4^+)

In the setting of metabolic acidosis, titratable acids cannot increase significantly because the availability of titratable acid is fixed by the plasma concentration of the buffer and by the GFR. The ammonia buffer system, by contrast, can increase several hundred-fold when necessary. Consequently, impaired renal excretion of hydrogen ions is always associated with a defect in ammonium excretion (Fig. 27.16).

All ammonia used to buffer urinary hydrogen ions in the collecting tubule is synthesized in the proximal convoluted tubule. Glutamine is the primary source of ammonia. It undergoes deamination catalyzed by glutaminase, resulting in α-ketoglutaric acid (see Fig. 27.16) and ammonia. In the collecting duct, NH_3 reacts with the hydrogen ions secreted by the tubular cells to form ammonium. As ammonium (NH_4) is not lipid soluble, it is trapped in the lumen and excreted in the urine as ammonium chloride. Two conditions predominantly promote ammonia synthesis by the proximal tubular cell: systemic acidosis and hypokalaemia.

Aetiology of acid–base disturbance

Acid–base disturbance may be caused by:

- abnormal CO_2 removal in the lungs ('respiratory' acidosis and alkalosis)
- abnormalities in the regulation of bicarbonate and other buffers in the blood ('metabolic' acidosis and alkalosis).

Both may, and usually do, coexist. For instance, metabolic acidosis causes hyperventilation (via medullary), leading to increased removal of CO_2 in the lungs and partial compensation for the acidosis. Conversely, respiratory acidosis is accompanied by renal bicarbonate retention and a compensatory metabolic alkalosis. The situation is even more complex if a patient has both respiratory disease and a primary metabolic disturbance.

Box 27.24 Changes in arterial blood gases

	pH	P_aCO_2	HCO_3^-
Respiratory acidosis	N or ↓	↑↑	↑ (compensated)
Respiratory alkalosis	N or ↑	↓↓	↓ (slight)
Metabolic acidosis	N or ↓	↓	↓↓
Metabolic alkalosis	N or ↑	↑ (slight)	↑↑

Diagnosis

Clinical history and examination usually point to the correct diagnosis. Box 27.24 shows the typical blood changes, but in complicated patients the acid–base nomogram (Fig. 27.17) may be used to interpret the pH, P_aCO_2, and [HCO_3^-] measured in blood. If the values from a patient lie in one of the bands in the diagram, it is likely that only one abnormality is present. For example, if the P_aCO_2 is high (indicated by the curved lines running diagonally across the nomogram), a primary respiratory acidosis is present due to a failure to excrete CO_2. In this context, a normal [HCO_3^-] indicates the lack of metabolic compensation; consequently the pH remains abnormally low (typically seen in an acute respiratory acidosis before time has allowed metabolic compensation to occur). Rising serum [HCO_3^-] caused by metabolic compensation results in a move upwards and rightwards along the curved line; the pH normalizes but both P_aCO_2 and [HCO_3^-] remain abnormal (a chronic, or compensated respiratory acidosis).

Respiratory acidosis and alkalosis

Respiratory acidosis

This is caused by retention of CO_2, commonly seen in chronic obstructive pulmonary disease (COPD). The P_aCO_2 and [H^+] rise. Renal retention of bicarbonate may partly compensate, returning the [H^+] towards normal).

Respiratory alkalosis

Increased removal of CO_2 is caused by hyperventilation, so there is a fall in P_aCO_2 and [H^+].

Metabolic acidosis and alkalosis

Metabolic acidosis

This is due to the accumulation of any acid other than carbonic acid, and there is a primary decrease in the plasma [HCO_3^-]. Several disorders can lead to metabolic acidosis: acid administration, acid generation (e.g., lactic acidosis during shock or cardiac arrest), impaired acid excretion by the kidneys, or bicarbonate losses from the gastrointestinal tract or kidneys. Calculation of the plasma anion gap is extremely useful in narrowing this differential diagnosis.

Anion gap

The first step is to identify whether the acidosis is due to primary losses of bicarbonate, or addition of another acid. (Because HCO_3^- and Cl^- are the primary anions in blood, retention of $NaHCO_3^-$ is the physiological equivalent of retention of H^+Cl^-; Cl^- accumulates as HCO_3^- is lost, in order to maintain electrical neutrality.) Identification of whether the primary metabolic disturbance relates to Cl^-/HCO_3^- or another acid is achieved by calculation of the anion gap.

- The normal cations present in plasma are Na^+, K^+, Ca^{2+}, and Mg^{2+}.

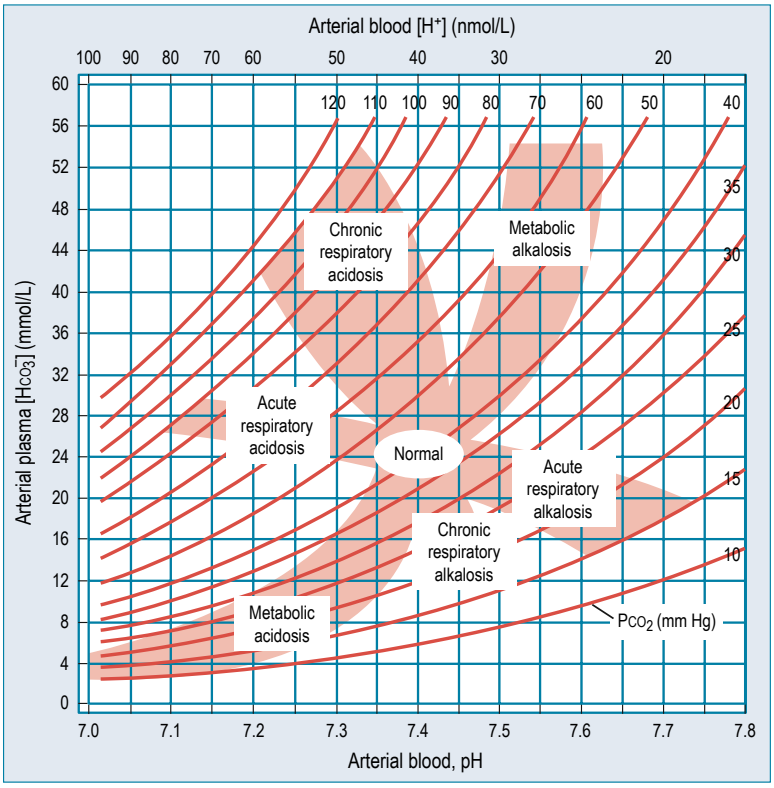


Fig. 27.17 The acid–base nomogram. This was derived from a large number of observations in patients with 'pure' respiratory or metabolic disturbances. The bands show the 95% confidence limits representing the individual varieties of acid–base disturbance. The curved lines represent the pCO_2. The central white box shows the approximate limits of arterial pH and PCO_2 in normal individuals.

- The normal anions present in plasma are Cl^-, HCO_3^- and negative charges present on albumin, phosphate, sulphate, lactate, and other organic acids.
- The sums of the positive and negative charges are equal.
- Measurement of plasma [Na^+], [K^+], [Cl^-], and [HCO_3^-] is usually easily available.

As there are more unmeasured anions than cations, the ***normal anion gap*** is 12 to 16 mmol/L, although calculations with more sensitive methods place this at 3 to 9 mmol/L. Albumin normally makes up the largest portion of these unmeasured anions. As a result, a fall in the plasma albumin concentration from the normal value of about 40 to 20 g/L may reduce the anion gap by as much as 6 mmol/L because each 1 g/L of albumin has a negative charge of 0.2 to 0.28 mmol/L.

Metabolic acidosis with a normal anion gap

If the anion gap is normal in the presence of acidosis, this suggests that H^+Cl^- is being retained or that $Na^+HCO_3^-$ is being lost. Causes of a normal-anion-gap acidosis are given in **Box 27.25**. In these conditions, plasma bicarbonate decreases and is replaced by chloride to maintain electroneutrality. Consequently, these disorders are sometimes referred to collectively as hyperchloraemic acidosis.

Renal tubular acidosis

The term 'renal tubular acidosis' (RTA) refers to systemic acidosis caused by impairment of the ability of the renal tubules to maintain acid–base balance. In patients with RTA, urinary pH is >5.5, even in the face of low blood pH, representing an inability to excrete adequate amounts of acid to maintain a normal blood pH.

? Box 27.25 Causes of metabolic acidosis with a normal anion gap

Increased gastrointestinal bicarbonate loss
- Diarrhoea
- Ileostomy
- Ureterosigmoidostomy

Increased renal bicarbonate loss
- Acetazolamide therapy
- Proximal (type 2) renal tubular acidosis
- Hyperparathyroidism
- Tubular damage, e.g., drugs, heavy metals, paraproteins

Decreased renal hydrogen ion excretion
- Distal (type 1) renal tubular acidosis
- Type 4 renal tubular acidosis (aldosterone deficiency)

Increased HCl production
- Ammonium chloride ingestion
- Increased catabolism of lysine, arginine

This group of disorders is uncommon and only rarely a cause of significant clinical disease.

Type 4 renal tubular acidosis

This results from a deficiency or unresponsiveness to aldosterone and, as such, type 4 RTA is also called 'hyporeninaemic hypoaldosteronism'. This is the most common form of RTA. The cardinal features are hyperkalaemia and acidosis occurring in a patient with mild CKD, usually caused by tubulointerstitial disease (e.g., reflux nephropathy) or diabetes. Hyperkalaemia maintains the acidosis through impairment of ammonia production in the

Box 27.26 Features of type 4 renal tubular acidosis (hyporeninaemic hypoaldosteronism)

- Hyperkalaemia (in the absence of drugs known to cause hyperkalaemia)
- Low plasma bicarbonate and hyperchloraemia
- Normal adrenocorticotrophic hormone (ACTH) stimulation test
- Low basal 24 h urinary aldosterone
- Subnormal response of plasma renin and plasma aldosterone to stimulation: samples taken over 2 h supine and again after 40 mg furosemide (80 mg if creatinine >120 μmol/L) and 4 h upright posture
- Correction of hyperkalaemia by fludrocortisone 0.1 mg daily

proximal tubule leading to a reduction in net acid excretion. Gordon's syndrome (see 'Gordon's syndrome [familial hyperkalaemic hypertension, pseudohypoaldosteronism type 2]' section) shares biochemical abnormalities but differs in having normal GFR and hypertension. Plasma renin and aldosterone are found to be low, even after measures that would normally stimulate their secretion. Features of type 4 RTA are shown in **Box 27.26**. An identical syndrome is caused by chronic ingestion of NSAIDs, which impair renin and aldosterone secretion. Treatment is with fludrocortisone, sodium bicarbonate, diuretics or gut-acting potassium binders, or a combination of these. Resolution of hyperkalaemia may correct the metabolic acidosis through increased ammonium excretion. Dietary potassium restriction alone is ineffective.

Type 3 renal tubular acidosis

This condition is vanishingly rare, and represents a combination of type 1 and type 2 RTA. Inherited type 3 RTA is caused by mutations resulting in carbonic anhydrase type II deficiency, which is characterized by osteopetrosis, RTA of mixed type, cerebral calcification, and learning difficulty.

Type 2 ('proximal') renal tubular acidosis

This is very rare in adult practice. It is caused by failure of sodium bicarbonate reabsorption in the proximal tubule. The cardinal features are acidosis, hypokalaemia, the inability to lower the urine pH below 5.5 despite systemic acidosis, and the appearance of bicarbonate in the urine despite a subnormal plasma bicarbonate. This disorder normally occurs as part of a generalized tubular defect ***(Fanconi's syndrome)***, together with other features of proximal tubular reabsorption such as glycosuria and amino-aciduria. Inherited forms of isolated type 2 RTA are described as both autosomal dominant and recessive patterns of inheritance, where putative mutations are in the Na^+–H^+ antiporter in the apical membrane and Na^+–HCO_3^- co-transporter in the basolateral membrane of proximal tubular cells, respectively. Treatment is with sodium bicarbonate; massive doses may be required to overcome the renal 'leak'.

Type 1 ('distal') renal tubular acidosis

This is due to a failure of H^+ excretion in the distal tubule (**Box 27.27**). It consists of:

- acidosis
- hypokalaemia (few exceptions)
- inability to lower the urine pH below 5.3 despite systemic acidosis
- low urinary ammonium production.

These features may be present only in the face of increased acid production, hence the need for an acid load test in diagnosis (**Box 27.28**). Other features include:

- low urinary citrate (owing to increased citrate absorption in the proximal tubule, where it can be converted to bicarbonate)

Box 27.27 Causes of type 1 distal renal tubular acidosis

Primary
- Idiopathic
- Genetic
- Marfan's syndrome
- Ehlers–Danlos syndrome
- Sickle cell anaemia

Nephrocalcinosis
- Chronic hypercalcaemia
- Medullary sponge kidney

Hypergammaglobulinaemic states
- Amyloidosis[a]
- Cryoglobulinaemia
- Chronic liver disease

Drugs and toxins
- Amphotericin B
- Lithium carbonate
- NSAIDs

[a]May also cause proximal renal tubular acidosis.

Box 27.28 Diagnosis of renal tubular acidosis

Plasma HCO_3^- < 21 mmol/L, urine pH > 5.3 = renal tubular acidosis.

Distal RTA: defect in dietary acid excretion. Urine pH always >5.5

Proximal RTA: defect in bicarbonate reabsorption. Urine pH variable

Bicarbonate infusion test (rarely performed)

Differentiates between proximal and distal RTA:
- Use a bicarbonate infusion (0.5–1 mmol/kg per hour) to raise serum bicarbonate to 18–20 mmol/L.
- If urine pH rises to >7.5 and the fractional excretion of bicarbonate (Fe HCO_3) rises to >15%, the defect is bicarbonate reabsorption and the diagnosis is proximal RTA.
- If there is little change in urine pH and fractional excretion of bicarbonate remains low, the diagnosis is distal RTA.

Acid load test (rarely performed)

Required if plasma HCO_3^- is >21 mmol/L but there is suspicion of partial RTA (e.g., nephrocalcinosis-associated diseases):
- Give 100 mg/kg ammonium chloride by mouth.
- Check urine pH hourly and plasma HCO_3^- at 3 h.
- Plasma HCO_3^- should drop below 21 mmol/L unless the patient vomits (in which case the test should be repeated with an antiemetic).
- If urine pH remains >5.3 despite a plasma HCO_3^- of 21 mmol/L, the diagnosis is confirmed.

- hypercalciuria.

These abnormalities result in osteomalacia, renal stone formation, and recurrent urinary infections:

- ***Osteomalacia*** is caused by buffering of H^+ by Ca^{2+} in bone, resulting in depletion of calcium from bone.
- ***Renal stone formation*** is caused by hypercalciuria, hypocitraturia (citrate inhibits calcium phosphate precipitation), and alkaline urine (which favours precipitation of calcium phosphate).
- ***Recurrent urinary infections*** are caused by kidney stones.

Both autosomal dominant and recessive inheritance patterns have been reported in primary distal RTA. In autosomal recessive distal RTA, a substantial proportion of patients have sensorineural deafness, and this is associated with a loss-of-function mutation in the H^+-ATPase at the apical surface of intercalated cells.

Treatment is with sodium bicarbonate, potassium supplements, and citrate. Thiazide diuretics are useful because they cause volume contraction and increased proximal sodium bicarbonate reabsorption.

Urinary anion gap

Another useful tool in the evaluation of metabolic acidosis with a normal anion gap is the urinary anion gap:

$$\text{URINARY ANION GAP} = \{\text{urinary } [Na^+] + \text{urinary } [K^+]\} - \text{urinary } [Cl^-].$$

Box 27.29 Causes of metabolic acidosis with an increased anion gap: think GOLD MARK

- Glycols (ethylene and propylene): ingestions
- Oxoproline (pyroglutamic acid): chronic paracetamol ingestion
- L-lactate: type A and type B
- D-lactate: small bowel bacterial overgrowth
- Methanol: acute ingestion (see Chapter 15, 'Methanol' section)
- Aspirin: salicylate overdose (see Chapter 15, 'Salicylates' section)
- Renal failure: accumulation of organic acids
- Ketoacidosis: diabetic, starvation, alcohol

(From Mehta AN, Emmett JB, Emmett M. GOLD MARK: an anion gap mnemonic for the 21st century. Lancet *2008; 372(9642):892).*

Here, the largest unmeasured cation is ammonium, and so raised urinary ammonium will result in a reduced urinary anion gap. This calculation can be used to distinguish between

- the normal-anion-gap acidosis caused by diarrhoea (or other gastrointestinal alkali loss), where the kidneys function appropriately, secreting lots of ammonium (also driven by hypokalaemia), and the urinary anion gap is reduced
- the normal-anion-gap acidosis caused by distal RTA where impaired acid secretion in the distal tubule prevents generation of ammonium and the anion gap is normal or increased.

In both disorders, the plasma [K^+] is characteristically low.

Metabolic acidosis with a high anion gap

If the anion gap is increased, there is an unmeasured anion present in increased quantities. This is either one of the acids normally present in small but unmeasured quantities, such as lactate, or an exogenous acid. See later and **Box 27.29** for causes of a high-anion-gap acidosis.

Chronic kidney disease

Chronic acidosis is most often caused by CKD, in which there is a failure to excrete fixed acid. Up to 40 mmol of hydrogen ions may accumulate daily. These are buffered by bone, in exchange for calcium. Chronic acidosis is therefore a major risk factor for renal osteodystrophy and hypercalciuria.

Chronic acidosis has also been shown to be a risk factor for muscle wasting in kidney failure, and may also contribute to the inexorable progression of some types of kidney disease.

Kidney disease causes acidosis in several ways:

- Reduction in the number of functioning nephrons from any cause decreases the capacity to excrete ammonia and H^+ in the urine.
- Tubular disease in particular may cause bicarbonate wasting, so acidosis is a particular feature of those types of CKD in which the distal tubules are particularly affected such as reflux nephropathy and chronic obstructive uropathy.

Uraemic acidosis should be corrected because of the effects on growth, muscle turnover, and bones. Oral sodium bicarbonate 2 to 3 mmol/kg daily is usually enough to maintain serum bicarbonate above 20 mmol/L but may contribute to sodium overload. Calcium carbonate improves acidosis and also acts as a phosphate binder and calcium supplement, and is commonly used. Acidosis in end-stage kidney disease is usually fully corrected by adequate dialysis.

Lactic acidosis

Increased lactic acid production occurs when cellular respiration is abnormal, because of either a lack of oxygen in the tissues ***('type A')*** or a metabolic abnormality, such as that induced by drugs such as metformin ***('type B')***. The most common cause in clinical practice is type A lactic acidosis, occurring in septic or cardiogenic shock. Significant acidosis can occur despite a normal blood pressure and P_aCO_2, owing to splanchnic and peripheral vasoconstriction. Acidosis worsens cardiac function and vasoconstriction further, contributing to a downward spiral and fulminant production of lactic acid.

Raised serum D-lactate occurs with fermentation of glucose by abnormal bowel flora in the short gut syndrome.

Ketoacidosis

See also Chapter 35. There is a high-anion-gap acidosis due to the accumulation of acetoacetic and hydroxybutyric acids, owing to increased production and some reduced peripheral utilization in diabetic ketoacidosis. Ketoacidosis without hyperglycaemia occurs in starvation and after alcohol overdose.

Mixed metabolic acidosis

Both types of acidosis may coexist. For instance, cholera would be expected to cause a normal-anion-gap acidosis owing to massive gastrointestinal losses of bicarbonate, but the anion gap is often increased owing to kidney failure and lactic acidosis as a result of hypovolaemia.

To help ascertain if there is an additional normal-anion-gap acidosis as well as, for example, a ketoacidosis, the delta anion gap can be calculated. This is the difference between the upper reference value of the anion gap and the calculated anion gap. This value should equal the fall in the bicarbonate concentration from the lower limit of normal. If the bicarbonate is significantly ***lower*** than predicted (>5 mmol/L), then an additional normal-anion-gap ***acidosis*** must be present. Similarly, if the bicarbonate is >5 mmol/L ***higher*** than expected, an additional normal-anion-gap ***alkalosis*** must be present.

Clinical features of acidosis

Clinically, the most obvious effect is stimulation of respiration, leading to the clinical sign of 'air hunger', or Kussmaul respiration.

Acidosis increases delivery of oxygen to the tissues by shifting the oxyhaemoglobin dissociation curve to the right, but it also leads to inhibition of 2,3-BPG production, which returns the curve towards normal. Cardiovascular dysfunction is common in acidotic patients, although it is often difficult to dissociate the numerous possible causes of this. Acidosis is negatively inotropic. Severe acidosis also causes venoconstriction, resulting in redistribution of blood from the peripheries to the central circulation, and increased systemic venous pressure, which may worsen pulmonary oedema caused by myocardial depression. Arteriolar vasodilatation also occurs, further contributing to hypotension.

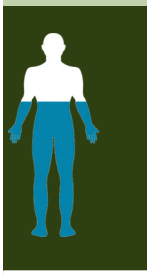
Cerebral dysfunction is variable. Severe acidosis is often associated with confusion and fits, but numerous other possible causes are usually present.

As mentioned earlier, acidosis stimulates potassium loss from cells, which may lead to potassium deficiency if kidney function is normal, or to hyperkalaemia if renal potassium excretion is impaired.

General treatment of acidosis

Treatment should be aimed at correcting the primary cause. In lactic acidosis caused by poor tissue perfusion ('type A'), treatment should be aimed at maximizing oxygen delivery to the tissues by protecting the airway, and improving breathing and circulation. This may require inotropic agents, mechanical ventilation, and invasive monitoring. In other forms of acidosis, treatment is directed at the underlying disorder—for example:

- insulin in diabetic ketoacidosis
- treatment of methanol and ethylene glycol poisoning with ethanol



- removal of salicylate by dialysis.

The question of whether severe acidosis should be treated with bicarbonate is controversial:

- Rapid correction of acidosis may result in tetany and fits owing to a rapid decrease in ionized calcium.
- Administration of sodium bicarbonate (8.4%) provides 1 mmol/mL of sodium, which may lead to extracellular volume expansion, exacerbating pulmonary oedema.
- Bicarbonate therapy increases CO_2 production and will therefore correct acidosis only if ventilation can be increased to remove the added CO_2 load.
- The increased amounts of CO_2 generated may diffuse more readily into cells than bicarbonate, worsening intracellular acidosis.

Administration of sodium bicarbonate (50 mmol, as 50 mL of 8.4% sodium bicarbonate i.v.) is still occasionally given during cardiac arrest and is often necessary before arrhythmias can be corrected. Correction of hyperkalaemia associated with acidosis is also of undoubted benefit. In other situations, there is no clinical evidence to show that correction of acidosis improves outcome, but sodium bicarbonate is sometimes administered in severe acidosis using intravenous 1.26% (150 mmol/L) bicarbonate infused over 2 to 3 hours with electrolyte and pH monitoring.

Metabolic alkalosis

Metabolic alkalosis is common, comprising half of all the acid–base disorders in hospitalized patients, with most cases associated with vomiting, hypovolaemia, the use of diuretics, and nasogastric suction. The mortality associated with metabolic alkalosis is substantial; the mortality rate is 45% in patients with an arterial pH of 7.55 and 80% when the pH is over 7.65. Although this relationship is not necessarily causal, severe alkalosis should be viewed with concern.

Classification and definitions

Metabolic alkalosis has been classified on the basis of underlying pathophysiology (**Box 27.30**):

- ***chloride depletion***, the most common cause, which can be corrected without potassium repletion
- ***potassium depletion***, usually with mineralocorticoid excess
- metabolic alkalosis due to ***both potassium and chloride depletion***.

? Box 27.30 Causes of metabolic alkalosis

Chloride depletion
- Gastric losses: vomiting, mechanical drainage, bulimia
- Chloruretic diuretics, e.g., bumetanide, furosemide, chlorothiazide, metolazone
- Diarrhoeal states: villous adenoma, congenital chloridorrhoea
- Cystic fibrosis (high sweat chloride)

Potassium depletion/ mineralocorticoid excess
- Primary aldosteronism
- Secondary aldosteronism
- Apparent mineralocorticoid excess: primary deoxycorticosterone excess: 11α- and 17α-hydroxylase deficiencies
- Drugs: liquorice (glycyrrhizic acid) as a confection or flavouring, carbenoxolone
- Liddle's syndrome
- Bartter's and Gitelman's syndromes and their variants
- Laxative abuse, clay ingestion

Hypercalcaemic states
- Hypercalcaemia of malignancy
- Acute or chronic milk–alkali syndrome

Others
- Amoxicillin or penicillin therapy
- Bicarbonate ingestion: massive or smaller ingestion with kidney disease
- Recovery from starvation
- Hypoalbuminaemia

Chloride may be lost from the gut, kidney, or skin. The loss of gastric fluid rich in acid results in alkalosis because bicarbonate generated during the production of gastric acid returns to the circulation. In Zollinger–Ellison syndrome (see Chapter 22) or gastric outflow obstruction, these losses can be massive. Although sodium and potassium loss in the gastric juice is variable, the obligate urinary loss of these cations is intensified by bicarbonaturia, which occurs during disequilibrium.

Chloruretic agents including most diuretics directly produce loss of chloride, sodium, and fluid in the urine. These losses in turn promote metabolic alkalosis by several mechanisms:

- Diuretic-induced increases in sodium delivery to the distal nephron enhance potassium and hydrogen ion secretion.
- Extracellular volume contraction stimulates renin and aldosterone secretion, which blunts sodium losses but accelerates potassium and hydrogen ion secretion.
- Potassium depletion augments bicarbonate reabsorption in the proximal tubule and stimulates ammonia production, which in turn will increase urinary net acid excretion.

Urinary losses of chloride exceed those for sodium and are associated with alkalosis, even when potassium depletion is prevented. The cessation of events that generate alkalosis is not necessarily accompanied by resolution of the alkalosis. A widely accepted hypothesis for the maintenance of alkalosis is chloride depletion rather than volume depletion. Although normal functioning of the proximal tubule is essential for bicarbonate absorption, the collecting duct appears to be the major segment of the nephron for altered electrolyte and proton transport in both maintenance and recovery from metabolic alkalosis. During maintenance, the α-intercalated cells in the CCD do not secrete bicarbonate because insufficient chloride is available for bicarbonate exchange. When chloride is administered and luminal or cellular chloride concentration increases, bicarbonate is promptly excreted and alkalosis is corrected.

Metabolic alkalosis in ***hypokalaemia*** is generated primarily by an increased intracellular shift of hydrogen ion, causing intracellular acidosis. Potassium depletion is also associated with enhanced ammonia production and increased obligate net acid excretion. This is the corollary to the acidosis seen with hyperkalaemia in type 4 RTA. Infusion of potassium alone can correct systemic alkalosis and intracellular acidosis as cells exchange extracellular potassium for intracellular hydrogen, which can then buffer extracellular bicarbonate.

Milk–alkali syndrome, in which both bicarbonate and calcium are ingested, produces alkalosis by vomiting, calcium-induced bicarbonate absorption, and reduced GFR.

Clinical features

The symptoms of metabolic alkalosis *per se* are difficult to separate from those of chloride, volume, or potassium depletion. Tetany, apathy, confusion, drowsiness, cardiac arrhythmias, and neuromuscular irritability are common when alkalosis is severe. The oxyhaemoglobin dissociation curve is shifted to the left. Respiration may be depressed.

Management

Chloride-responsive metabolic alkalosis

Although replacement of the chloride deficit is essential in chloride depletion states, selection of the accompanying cation—sodium or potassium—is dependent on the assessment of extracellular fluid volume status (see Fig. 27.1), the presence or absence of associated potassium depletion, and the degree and reversibility of any

depression of GFR. Since chloride and extracellular volume depletion coexist, isotonic (0.9%) saline solution is generally appropriate therapy. If kidney function is normal, bicarbonate will be excreted, and metabolic alkalosis will be rapidly corrected as soon as chloride is made available.

Chloride-resistant metabolic alkalosis

Metabolic alkalosis due to potassium depletion is managed by correction of the underlying cause (see 'Hypokalaemia' section). Mild to moderate alkalosis requires oral potassium chloride administration. However, the presence of a cardiac arrhythmia or generalized weakness requires intravenous potassium chloride, typically alongside chloride administration in the form of 0.9% saline with added potassium chloride (KCl).

Further reading

Berend K. Diagnostic use of base excess in acid-base disorders. *N Engl J Med* 2018; 378:1419–1428.
Berend K. Review of the diagnostic evaluation of normal anion gap metabolic acidosis. *Kidney Dis (Basel)* 2017; 3:149–159.
Kraut JA, Madias NE. Lactic acidosis. *N Engl J Med* 2015; 372:1078–1079.
Seifter JL, Chang HY. Disorders of acid-base balance: new perspectives. *Kidney Dis (Basel)* 2017; 2:170–186.
Weiner ID, Verlander JW. Ammonia transporters and their role in acid-base balance. *Physiol Rev* 2017; 97:465–494.

Questions for this chapter can be found on Elsevier eBooks+.

28 Neurology

Apeksha M. Shah

Clinical skills 886
History 886
Examination 886
Putting it all together 886
Background 888
Functional neuroanatomy 888
Patterns of neurological disease 889
Cortical disease 889
Cortical disorders of communication 889
Cortical disorders of movement 891
Cortical disorders of sensation and perception 891
Cortical disorders of memory 891
Cortical disorders of executive function 891
Cranial nerve lesions 891
I: Olfactory nerve 891
II: Optic nerve and visual system 892
III, IV, VI: Oculomotor, trochlear, and abducens nerves 896
V: Trigeminal nerve 897
VII: Facial nerve 898
VIII: Vestibulocochlear nerve 900
Lower cranial nerves IX, X, XI, XII 901
Disorders of motor control systems 902
Corticospinal (pyramidal) system 903
Extrapyramidal system 904
Cerebellum 905
Tremor 906
Dysarthria 906
Lower motor neuron lesions 906
Spinal reflex arc 907
Disorders of sensory pathways and pain 908
Lesions of the sensory pathways 908
Pain 910
Bladder control and sexual dysfunction 911
Investigation of neurological disease 913
Neurophysiological investigations 913
Neurological conditions 915
Headache, migraine, and facial pain 915
Primary headache disorders 916
Secondary headache disorders 919
Facial pain 920
Epilepsy and loss of consciousness 921
Epilepsy 921
Other causes of blackouts 927
Sleep disorders 928
Movement disorders 929
Parkinsonian disorders 929
Other akinetic–rigid syndromes 932
Hyperkinetic movement disorders 932
Neuroinflammatory disorders 934
Multiple sclerosis 934
Other neuroinflammatory disorders 939
Nervous system infection 941
Meningitis 941
Encephalitis 941
HIV and neurology 942
Other infections and post-infectious inflammatory conditions 942
Brain and spinal abscesses 944
Brain tumours 944
Hydrocephalus 946
Brain injury 947
Traumatic brain injury 947
Hypoxic–ischaemic brain injury 948
Disorders of consciousness 948
Spinal cord disease 950
Spinal cord compression 950
Other spinal cord disorders 951
Dementia 952
Dementia syndromes 954
Management of dementia 958
Motor neuron disease 959
Congenital disorders 960
Neurogenetic disorders 961
Neurocutaneous syndromes 961
Hereditary ataxias 961
Paraneoplastic syndromes 962
Peripheral nerve disease 962
Mononeuropathies 964
Mononeuritis multiplex 964
Polyneuropathies (peripheral neuropathy) 964
Plexus and nerve root lesions 969
Muscle diseases 970
Inflammatory myopathies 971
Metabolic and endocrine myopathies 971
Neuromuscular junction disorders 973
Muscular dystrophies 974
Myotonic syndromes 975
Skeletal muscle channelopathies 975
Stiff person syndrome 975
Mitochondrial diseases 975

CORE SKILLS AND KNOWLEDGE

Neurology is a large subject covering many conditions that have serious effects on the daily lives of patients and their families. We are now in an exciting era of rapid therapeutic expansion in neurology, much of which is based on insights from basic neuroscience, but multidisciplinary long-term coordinated care remains essential to improve the lives of patients. Neurology spans conditions as diverse as cognitive disorders involving higher-level mental functioning, to disorders of peripheral nerve and skeletal muscle. Many neurological disorders affect multiple domains both within and outside the nervous system, and cause complex, interacting problems that a patient and their care team must grapple with. It remains a firmly clinical specialty, and a good neurologist has mastery of clinical skills but more importantly needs to forge a close partnership with the patient.

Key skills in neurology include:

- Distilling information gathered from careful history-taking and physical examination into a well-defined clinical syndrome, suggesting likely diagnoses and a rational approach to investigation
- Managing neurological emergencies including seizures, coma, central nervous system infection, spinal cord compression, and acute neuromuscular weakness
- Managing disability and complex, chronic, or progressive disease, working with other health professionals and in partnership with patients to help them achieve good quality of life and effective symptom control.

Opportunities for learning neurology include attending ward rounds on a stroke unit or with acute neurology teams, practising neurological examination with consenting patients, attending a range of general and specialist neurology clinics, and observing specialist neurological investigations such as electrophysiology or lumbar puncture.

CLINICAL SKILLS

Pattern recognition in neurology—interpretation of history, symptoms, and examination—is very reliable. Practical experience is vital. There are three critical questions in formulating a clinical diagnosis:

- What is/are the site(s) of the lesion(s)?
- What is the likely pathology?
- Does a recognizable disease fit this pattern?

History

Neurological differential diagnosis relies heavily on a carefully taken history. Patterns emerge from the history that will suggest the likely causes, site and extent of disease. Most of the information will emerge from listening to the patient's narrative account. A collateral history is often vital in cases where the patient has cognitive impairment or has had altered awareness. An approach to history-taking in neurology is shown in **Box 28.1**.

Examination

If the neurological history focuses primarily on the likely *cause* of the neurological disease, the examination focuses on the likely *anatomical location* of the lesion. The diagram above illustrates a basic approach to focused examination of the nervous system with key signs to elicit or exclude.

An overview of the neurological examination is provided in **Boxes 28.2 and 28.3**; more details are provided in sections covering individual components of the nervous system, below. A focused examination should be guided by the history, and much useful information can be gained simply from observing the patient. It is necessary to practise the neurological examination in order to become deft and appreciate the range of normal.

Further reading

Clarke C. The language of neurology—symptoms, signs and basic investigations. In: **Clarke C, Howard R, Rossor M et al. (eds).** *Neurology: A Queen Square Textbook*. Oxford: Blackwell; 2009.

Putting it all together

Information from the history and examination should be used together in order to answer the questions in **Box 28.4**, which will lead to a diagnostic formulation, encompassing both the anatomical location and the possible underlying aetiology. Investigations should be directed at refining the clinical impression, which becomes the basis for transfer of information and management.

Box 28.1 Taking a history in neurology

Presenting symptoms

- Pattern—constant, episodic, progressive
- Time course—hyperacute, acute, subacute, chronic
- Onset—sequence of events when symptoms first began, or each time they recur
- Duration of symptoms, and any previous episodes
- Triggers—warm baths may worsen symptoms caused by demyelination; cold wind and chewing can trigger trigeminal neuralgia; dietary stimuli, tiredness, or stress can trigger migraines

Screening questions for other relevant features

- Systemic symptoms—fever (may suggest infection), weight loss, and anorexia (may suggest malignancy with paraneoplastic neurological phenomena)
- Raised intracranial pressure symptoms—headaches worse on waking, lying flat, or Valsalva manoeuvre; vomiting; diplopia
- Visual symptoms—blurring, diplopia, flashing lights
- Bulbar symptoms—changes in speech and difficulty swallowing
- Motor symptoms—weakness, stiffness, abnormal gait, tremor
- Sensory symptoms—loss of sensation, neuropathic ('tingling' or 'electric') pain
- Loss of coordination—dizziness, staggering gait, loss of fine motor control, altered speech
- Autonomic symptoms—postural dizziness, faecal incontinence, urinary retention
- Loss of consciousness—features suggesting a neurological cause include aura, muscle jerking, tongue biting, incontinence, and a drowsy post-ictal phase.
- Disturbance of higher functions—personality change, cognitive decline, loss of executive control, disinhibition, psychiatric symptoms

Past medical history

- Atherosclerotic risk factors—hypertension, diabetes, other cardiovascular disease
- Risk factors for embolic disease—atrial fibrillation, patent foramen ovale
- Diabetes—time since diagnosis, degree of glycaemic control, presence of other complications
- Malignancies, systemic inflammatory disorders, immunosuppressive conditions, e.g., HIV
- History of anxiety, depression, or other psychiatric disorders

Box 28.1 Taking a history in neurology—cont'd

Medication history
- Use of neuroactive medications, for instance antipsychotics

Family history
- History of inherited neurological disorders such as Huntington's chorea, myotonic dystrophy, and Charcot–Marie–Tooth disease in close relatives

Social history
- Occupation
- Family circumstances, caring responsibilities

Disability and effect on lifestyle
- Mobility
- Ability to perform activities of daily living
- Current provision of carers

Lifestyle factors
- Smoking, alcohol consumption, use of illegal drugs

Safety
- Do they drive?
- What is their employment? Might their illness pose a risk to themselves or others?
- Do they have caring responsibilities? Is support needed for any dependents or children?
- Do they take baths or go swimming alone? Any other potentially risky leisure activities?

General features
- Level of consciousness
- Affect
- Obvious deformity or disability
- Signs of systemic illness e.g., sepsis

Olfaction (CN I)

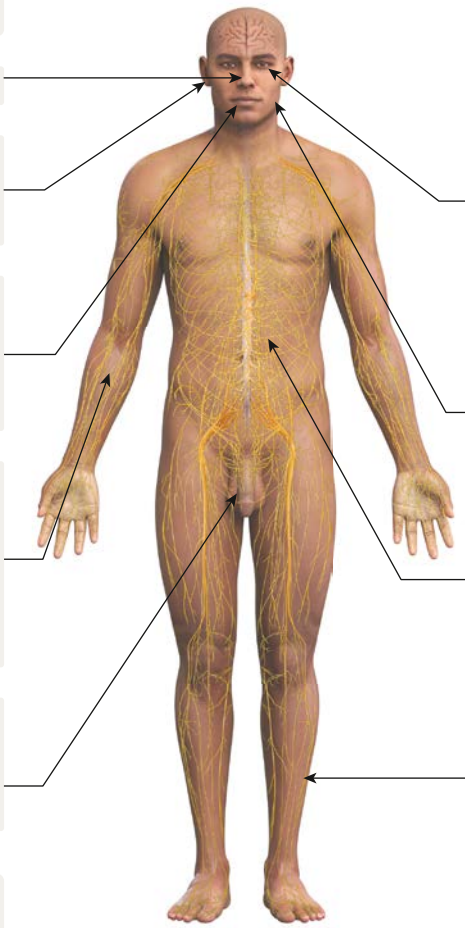
Hearing (CN VIII)
- Rinne's test
- Weber's test

Lower cranial nerves
- Speech
- Swallow
- Gag reflex
- Tongue

Upper limbs
- Posture (extensor, flexor)
- Fasciculation, wasting
- Tone, power
- Sensation
- Coordination
- Reflexes

Sphincter control (cord lesion, cauda equina)
- Urinary retention
- Faecal incontinence

Gait
- Hemiplegic (e.g., stroke)
- Spastic (e.g., cord lesions)
- High stepping (peripheral neuropathy)
- Waddling (myopathic)
- Shuffling (Parkinsonism)
- Ataxic (cerebellar)



Speech
- Language problems (aphasia)
  - Expressive (Broca's)
  - Receptive (Wernicke's)
- Voice problems (dysarthria, dysphonia)
  - Bulbar palsy (LMN)
  - Pseudobulbar palsy (UMN)

Eyes
- Visual fields
- Acuity, colour (CN II)
- Movements (CN III, IV, VI)
- Light reflex, accommodation (CN III)
- Ptosis (CN III)

Face
- Sensation (CN V)
- Movements (CN VII)
  - Whole face (LMN)
  - Upper face spared (UMN)

Trunk
- Dermatomal rash (shingles)
- Sensory level (spinal cord lesion)

Lower limbs
- Posture (extensor, flexor)
- Fasciculation, wasting
- Tone, power
- Sensation (dermatomes)
- Coordination
- Reflexes
- Planters

Movements
- Slow, hesitant (Parkinsonism)
- Dance-like, writhing (chorea)
- Repetitive, absent (epileptiform)

Higher cortical functions
- Aphasia (impaired naming)
- Agnosia (impaired recognition)
- Apraxia (impaired skillful movements)
- Loss of executive function
- Personality change

BACKGROUND

Neurology is a broad clinical specialty that requires good clinical skills and a rational diagnostic approach to link underlying pathophysiological processes with a diverse range of clinical syndromes. The skills required cannot be replaced with investigations or imaging alone. Many neurological diagnoses are life-changing chronic or progressive conditions, and some are life-limiting. Neurologists are privileged to look after patients facing these challenges, to hear their stories, to treat where possible, and to help people through the journey where treatments are limited. The value of careful listening and empathy cannot be over-emphasized, both when gathering information to reach a diagnosis, and when working to achieve the best possible therapeutic outcome.

We are entering an era of burgeoning therapeutics, and many conditions previously regarded as devastating now have effective treatments. Many other treatments will emerge in the coming decades. It is an exciting time to work in the field, and remaining up-to-date is vital to ensure the best possible outcome for patients. Disorders causing significant cognitive or physical disability require close cooperation with allied health professionals, primary care, psychiatrists, and community services.

Neurology covers the whole spectrum of healthcare, from primary care where many common conditions (such as headache) are initially managed, to acute inpatient settings, outpatient clinics, day-care units, and more highly specialized tertiary and quaternary care units. Some 17% of general practitioner consultations and 10% of emergency department visits are for neurological symptoms; 19% of all hospital admissions are for neurological disorders and 25% of chronic disability in adults below the age of 64 is due to neurological disease (**Box 28.5**).

FUNCTIONAL NEUROANATOMY

Neuron and synapse

The neuron is the functional unit of the entire nervous system (**Fig. 28.1**). Its cell body and axon terminate in a synapse. The size and type of each group of neurons vary. A thoracic spinal cord α-motor neuron has an axonal length of more than 1 metre and innervates between several hundred and 2000 muscle fibres in one leg—a motor unit. By contrast, some spinal or intracerebral interneurons have axons shorter than 100 μm, terminating on one neuronal cell body.

Neurotransmitters

Neurotransmitters are excitatory (acetylcholine, noradrenaline [norepinephrine], adrenaline [epinephrine], 5-hydroxytryptamine [5-HT, serotonin], dopamine, glutamate, and aspartate) or inhibitory (γ-aminobutyric acid [GABA], histamine and glycine). Neuropeptides, such as vasopressin, adrenocorticotrophic hormone (ACTH), substance P, and opioid peptides, as well as the purines adenosine triphosphate (ATP) and adenosine monophosphate (AMP) are both excitatory and inhibitory.

Synaptic transmission is mediated by neurotransmitters released by action potentials passing down an axon. Neurotransmitters activate postsynaptic receptors and are removed by transporter proteins. The neurotransmitter–receptor reaction increases ionic permeability and propagates a further action potential. Axonal electrical activity and synaptic chemical release are the basis of neurological function.

Box 28.2 Ten-part neurological examination

1. ***State of consciousness, arousal, appearance***
2. ***Mental state, attitude, insight***
3. ***Cognitive function***—orientation, recall, level of intellect, language, other cortical problems, e.g., apraxia
4. ***Gait and balance tests, including tandem walking and Romberg's test***
5. ***Neck***—stiffness, palpation, and auscultation of carotids
6. ***Cranial nerves*** (see **Box 28.6**)
   - I. Ask about sense of smell
   - II. Check visual acuity, fields, pupillary light reflexes, and accommodation
   - III, IV, VI. Eye movements (cross/H)—smooth pursuit and saccades
   - V. Check facial sensation in 3 divisions; assess power of muscles of mastication
   - VII. Observe facial expression and test facial muscle strength
   - VIII. Check hearing; Rinne's and Weber's tests
   - IX–X. Say 'ah' and observe palatal movement; gag reflex only if necessary
   - XI. Check sternocleidomastoid (turn head against resistance) and trapezius (shrug) power
   - XII. Protrude tongue and move from side to side
7. ***Motor system***

   Upper limbs:
   - Wasting and fasciculation
   - Posture of arms: drift, rebound, tremor
   - Tone: spasticity, clonus, extrapyramidal rigidity
   - Power: 0–5 scale (see **Box 28.3**)
   - Tendon reflexes: + or ++ normal; +++ pathological; 0 = absent with reinforcement

   Thorax and abdomen:
   - Respiration
   - Thoracic and abdominal muscles

   Lower limbs:
   - Wasting and fasciculation
   - Tone, power, and tendon reflexes
   - Plantar responses
8. ***Coordination and fine movements***

   Finger-nose and heel-shin testing; repetitive movements
9. ***Sensory system***
   - Chart area of sensory loss; start in area with abnormal sensation and move stimulus towards normal area to demarcate boundaries

   Posterior columns:
   - Vibration (128 Hz tuning fork)
   - Joint position—small movements of distal interphalangeal joints in toes and fingers
   - Light touch—use cotton wool

   Spinothalamic tracts:
   - Pain: use a split orange-stick or a single-use pin (never a hypodermic needle)

   Temperature: warm and cold objects

   Cortical sensory loss—dysgraphaesthesia and astereognosis
10. ***Specialized tests*** as required e.g., Hallpike for vertigo, Phalen's test, and so on

Box 28.3 Muscle power: the Medical Research Council (MRC) scale

Grade	Definition
5	Normal power
4	Active movement against gravity and resistance
3	Active movement against gravity only
2	Active movement with gravity eliminated
1	Flicker of contraction
0	No contraction

Box 28.4 Formulating a diagnosis in neurology

Is the lesion global or focal?
- **Global** lesions affect all (or most) brain functions, and typically affect cognition and consciousness
- **Focal** lesions affect one region of the body and resultant signs may be asymmetric

What is the timescale?
- **Immediate onset** symptoms are usually vascular (ischaemia or haemorrhage), or electrical (epilepsy) in aetiology
- **Rapidly progressive** symptoms (over days) may represent infective (e.g., meningitis) or inflammatory (e.g., demyelinating) disease
- **Slowly progressive** symptoms tend to be caused by degenerative diseases

Do motor lesions affect the upper or the lower motor neuron?
- **Upper motor neuron (UMN) signs** include increased tone (spasticity), weakness, increased reflexes, and upgoing plantars—a 'pyramidal' pattern of weakness where stronger muscles (upper limb flexors, lower limb extensors) overwhelm weaker ones
- **Lower motor neuron (LMN) signs** include flaccid weakness, decreased or absent reflexes, wasting, and fasciculations

Where is the lesion?
- **Cerebral cortex/internal capsule:** contralateral UMN signs
- **Cerebellum:** nystagmus, impaired balance/coordination
- **Brainstem:** impaired consciousness, global signs, cranial nerve abnormalities
- **Spinal cord:** UMN paraplegia/quadriplegia with sensory level
- **Nerve root:** LMN myotomal signs; dermatomal sensory loss
- **Single peripheral nerve:** LMN signs and sensory loss according to distribution of nerve
- **Peripheral nerves, diffuse:** 'length dependent' LMN signs (worst in feet), 'glove and stocking' sensory loss
- **Neuromuscular junction/muscle:** only motor signs present. Fatiguability common, wasting and fasciculations

Is there a functional component?
- Are there elements that don't fit within standard neuroanatomical patterns of disease?
- Are there positive features of a functional component? Hoover's sign (involuntary extension of the 'paralyzed' leg when raising the other leg against resistance) or entrainment of a tremor
- Beware of searching for 'stressors' or underlying triggers for functional disease; these are not universally present and can obscure the assessment
- Functional elements commonly exist alongside organic disease

Box 28.5 UK incidence of common neurological conditions

Conditions	Events per 100 000/year
Migraine	400
Stroke	240
Dementia over age of 85	68
Epilepsy	50
Parkinson's disease	19
Severe brain injury and subdural haematoma	13
All central nervous system tumours	9
Trigeminal neuralgia	8
Meningitis	7
Multiple sclerosis	7
Myasthenia, all muscle and motor neuron disease	5

PATTERNS OF NEUROLOGICAL DISEASE

Neurological symptoms and signs suggest the area of the nervous system that is malfunctioning (e.g., aphasia—dominant frontal lobe, hemiparesis—the internal capsule, or a Bell's palsy—VIIth cranial (facial) nerve).

Focal lesions of the cortex, and lesions throughout the nervous system, cause symptoms and signs by two processes:
- ***Suppression or destruction of neurons*** and surrounding structures (Fig. 28.2). This is the most common process: part of the system simply fails to work.
- ***Synchronous discharge of neurons*** by *irritative* lesions (Fig. 28.3), for example cortical lesions, causes epilepsy, either focal or generalized.

Essential elements of neuroanatomy

For clinical purposes, the complexity of neuroanatomy must be reduced to its core elements:
- cortical regions controlling higher behaviour
- cranial nerves
- three systems of motor control:
  - corticospinal or pyramidal system
  - extrapyramidal system
  - cerebellum
- the lower motor neuron (LMN)
- sensory pathways and pain
- control of the bladder and sexual function.

CORTICAL DISEASE

The cerebral cortex plays a key role in basic movement and sensation, as well as controlling higher functions such as language, memory, learning, consciousness, and emotion. Discrete cortical lesions most commonly arise from strokes but may also be caused by a variety of other pathologies including malignant disease, localized infection, or demyelinating lesions affecting cortical connections. While many higher functions are localized within the cortex so that anatomically defined lesions cause specific pattern of disease, work on neuronal networks, functional imaging, and plasticity questions the traditional views of highly specific localization of cortical function.

Cortical disorders of communication

The concept of cerebral dominance arose from a simple observation: right-handed stroke patients with acquired language disorders had destructive lesions within the left hemisphere. Right-handed (and 70% of left-handed) people have language function on the left. More specifically, destructive lesions within the dominant frontotemporo-parietal region cause disorders of communication:
- spoken language—***aphasia***, also called ***dysphasia***
- writing—***agraphia***
- reading—acquired ***alexia***.

Aphasia

Aphasia is loss of, or defective language from damage to, the speech centres within the dominant hemisphere. Numerous varieties have been described:

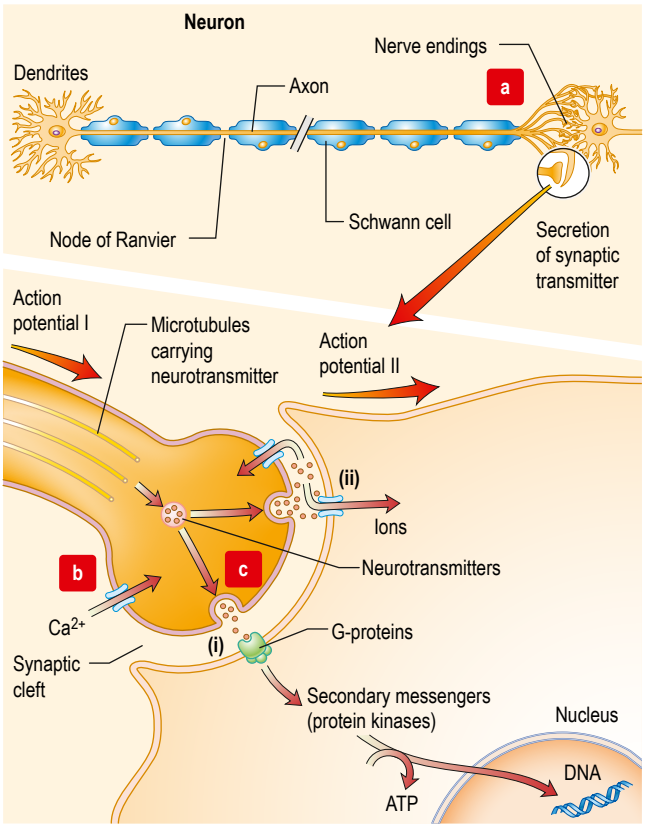


Fig. 28.1 The functional unit: neuron and neurotransmitters. (a) ***The action potential***—that is, the nerve impulse—travels down the axon. Microtubules carry neurotransmitters to nerve endings. **(b)** ***Action potential I*** depolarizes the synaptic membrane, opening voltage-gated calcium channels. **(c)** Influx of calcium ions causes vesicles to fuse with the membrane, allowing neurotransmitter binding to receptors **(i)** and activation of secondary messengers that modulate gene transcription and also open ligand-gated channels **(ii)**. This allows ions to enter, depolarize the membrane, and ***initiate action potential II***.

Site of lesion	Disorder	R	L
Frontal, either	Intellectual impairment Personality change Urinary incontinence Monoparesis or hemiparesis		
Frontal, left	Broca's aphasia		
Temporo-parietal, left	Acalculia Alexia Agraphia Wernicke's aphasia Right-left disorientation Homonymous field defect		
Temporal, right	Failure to recognize faces Homonymous field defect		
Parietal, either	Contralateral sensory loss or neglect Agraphaesthesia Homonymous field defect		
Parietal, right	Confusional states Dressing apraxia Failure to recognize faces		
Parietal, left	Limb apraxia		
Occipital/ occipitoparietal	Visual field defects Visuospatial defects Disturbances of visual recognition		

Fig. 28.2 Principal features of destructive cortical lesions in a right-handed individual.

Broca's (expressive, anterior) aphasia

Damage in the dominant frontal lobe causes reduced speech fluency with relatively preserved comprehension. The patient makes great efforts to initiate language, which becomes reduced to a few disjointed words with failure to construct sentences (sometimes described as ***telegrammatic***). Patients who recover say they knew what they wanted to say but could not get the words out.

Wernicke's (receptive, posterior) aphasia

Dominant hemisphere temporo-parietal damage leaves fluency of language unaffected but words are muddled. This varies from insertion of a few incorrect or unnecessary words into speech to a profuse outpouring of jargon (non-existent words). Severe jargon aphasia is bizarre and often mistaken for psychotic behaviour.

Patients who recover from Wernicke's aphasia say that they found speech, both their own and others', like an unintelligible foreign language—that is, incomprehensible—but they could neither stop speaking nor understand speech.

Nominal (anomic) aphasia

This means difficulty in naming objects. Naming difficulty is an early feature in all types of aphasia. This is often tested in practice by asking patient to name parts of a watch; a more sensitive test utilizes pictures to test low-frequency words, such as rhinoceros, violin, tricycle, or ladybird. Cultural and linguistic differences need to be taken into account.

Global (central) aphasia

This means the combination of the expressive problems of Broca's aphasia and the loss of comprehension of Wernicke's, with loss of both language production and understanding. It is due to widespread damage to speech areas and is the most common aphasia after a severe left hemisphere infarct. Writing and reading are also affected.

Conduction aphasia

Damage to the arcuate fasciculus, that connects Wernicke's with Broca's area, can lead to difficulty with repetition.

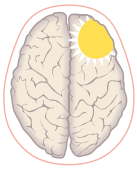
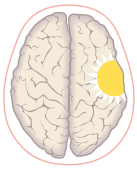
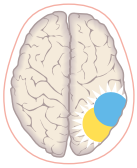
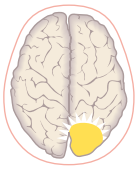
Site of lesion	Effects	R	L
Frontal	Focal motor seizures of contralateral limbs Conjugate deviation of head and eyes away from the lesion		
Temporal	Formed visual hallucinations Complex partial seizures Memory disturbances (e.g. *déjà vu*)		
Parietal Parieto-occipital	Focal sensory seizures of contralateral limbs Crude visual hallucinations (e.g. shapes in one part of the field)		
Occipital	Visual disturbances (e.g. flashes)		

Fig. 28.3 Effects of irritative cortical lesions.

Cortical disorders of movement

Hemiparesis

Hemiparesis describes weakness or inability to move one side of the body, and may arise from destructive lesions affecting the primary motor cortex or its main axonal connections, which run through the internal capsule, brainstem, and spinal cord.

Apraxia

Apraxia is a disorder of motor planning, affecting learned purposeful movement, and not explained by weakness or sensory loss. Dominant hemisphere lesions, particularly involving the parietal lobe, cause various types of (bilateral) ***limb apraxia***, including ***ideomotor apraxia*** (inability to imitate meaningless gestures, or to pantomime an action, such as 'pretend to brush your teeth').

More unusual forms of apraxia arise from lesions of the non-dominant hemisphere: failure to put on clothing correctly ***(dressing apraxia)***, or inability to draw simple shapes ***(constructional apraxia)***.

Cortical disorders of sensation and perception

Inattention and neglect

Disorders of the non-dominant hemisphere are often difficult to recognize. There are abnormalities of perception of internal and external space. Examples are loss of way in familiar surroundings, visual inattention, or neglect (where patients fail to respond to stimuli in one side of their visual space or on one side of their body).

Cortical visual loss

This describes loss of vision stemming from damage to one (hemianopia) or both (cortical blindness) occipital lobes. See Chapter 31.

Cortical disorders of memory

Like most brain functions, memory has a modular organization, with different aspects of memory dependent on functionally and anatomically distinct brain networks. Three distinct processes are required: learning, storage, and subsequent retrieval of learned information. There is a fundamental distinction between ***explicit*** memory, which can be consciously accessed, such as long-term memory for events (***episodic*** memory) and knowledge of word meaning (***semantic*** memory) on the one hand, and ***implicit*** memory, which is not conscious: for example, how to ride a bike (Fig. 28.4). The term 'short-term' memory is frequently misunderstood and refers to ***working*** memory lasting seconds only: for example, phone numbers or digit span.

Disorders of memory can result from damage to the medial structures of the temporal lobes and their connections—the hippocampi, fornices, and mammillary bodies, as well as from damage to the thalamus and frontal lobe. The distributed anatomical basis of memory means that bilateral lesions are necessary to cause ***amnesia***. Impairment of episodic memory usually results in a temporal memory gradient: recent and new memories are mainly affected, with relative preservation of distant memories.

Memory loss (the amnestic syndrome) is frequently a symptom of dementia, especially Alzheimer's disease (see 'Alzheimer's disease' section), but also occurs as an isolated entity (Box 28.6).

Cortical disorders of executive function

Lesions of the frontal lobes may be associated with inability to plan, solve problems, focus attention, or juggle different tasks. There may be associated difficulties in regulating emotional responses or in controlling behaviour, and patients with frontal lobe lesions may be described by those who know them as having a changed personality.

CRANIAL NERVE LESIONS

See Box 28.7 and Fig. 28.5

General principles

Cranial nerve deficits may be produced by ***supranuclear***, ***nuclear***, or ***infranuclear*** lesions:

- An ***infranuclear*** or ***nuclear*** lesion of a motor nerve will produce a lower motor neuron pattern of deficit characterized by flaccid paralysis (e.g., bulbar palsy). The nucleus or tract of the nerve may be affected within the substance of the brainstem, or the cranial nerve after it has exited the brainstem.
- A ***supranuclear lesion*** produces an upper motor neuron (UMN) deficit (e.g., pseudobulbar palsy or supranuclear gaze palsy). The LMN pathways remain intact (e.g., intact vestibulo-ocular reflex [VOR] with a supranuclear gaze palsy; brisk jaw jerk with pseudobulbar palsy).

I: Olfactory nerve

This sensory nerve arises from olfactory (smell) receptors within nasal mucosa. Branches pierce the cribriform plate and synapse in the olfactory bulb. The olfactory tract passes to the olfactory cortex in the inferior temporal lobe.

Anosmia (loss of the sense of smell) is caused by head injury (shearing of olfactory neurons as they pass through the cribriform plate at the skull base) or tumours of the olfactory groove

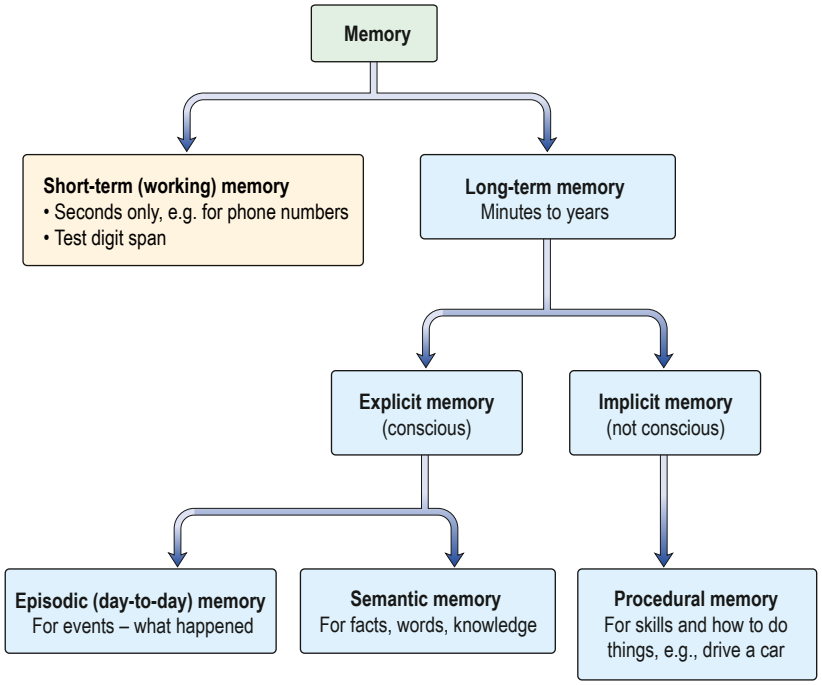


Fig. 28.4 **Types of memory.**

? Box 28.6 Causes of an amnestic syndrome[a]

- Dementia (note: multiple cognitive domains affected, not memory in isolation)
- Amnestic mild cognitive impairment (see 'Mild cognitive impairment' section)
- Alcohol—thiamine deficiency (Wernicke–Korsakoff's syndrome)
- Head injury (severe)
- Anoxic brain damage and that following carbon monoxide poisoning
- Stroke, including bilateral thalamic infarction, subarachnoid haemorrhage, diffuse small-vessel disease, and post-cardiac surgery
- Viral, paraneoplastic, and autoimmune encephalitis
- Drugs—psychotropics, anticholinergics, and solvent abuse
- Bilateral invasive tumours
- Temporal lobe epilepsy and temporal lobectomy
- Following hypoglycaemia
- Temporary amnesia: transient global amnesia, transient epileptic amnesia, post-traumatic amnesia
- Dissociative (functional/psychogenic)

[a]Amnesia must be distinguished from delirium.

i Box 28.7 Cranial nerves

Number	Name	Main clinical action
I	Olfactory	Smell
II	Optic	Vision, fields, afferent light reflex
III	Oculomotor	Eyelid elevation, eye elevation, adduction, depression in abduction, efferent pupil—light reflex
IV	Trochlear	Eye intorsion, depression in adduction
V	Trigeminal	Facial (and corneal) sensation, mastication muscles
VI	Abducens	Eye abduction
VII	Facial	Facial movement, taste fibres
VIII	Vestibulocochlear	Balance and hearing
IX	Glossopharyngeal	Sensation—soft palate, taste fibres
X	Vagus	Cough, palatal and vocal cord movements
XI	Accessory	Head turning, shoulder shrugging
XII	Hypoglossal	Tongue movement

(e.g., meningioma). Olfaction is lost or diminished after upper respiratory infections and with local disorders of the nose (this is usually temporary). Many patients with gradual-onset anosmia over many years may be unaware of the deficit: for example, in Parkinson's disease, where anosmia precedes motor symptoms by many years but is often not noticed by the patient.

Detailed smell testing is difficult in routine clinical practice and rarely performed. Adequate testing requires use of commercially available kits, such as scratch and sniff cards or odour-filled pens with forced multiple-choice identification.

II: Optic nerve and visual system

Light regulated by the pupillary aperture is converted into action potentials by retinal photoreceptors (rods and cones), which transmit information via bipolar cells to retinal ganglion cells (see Chapter 31). The lens, under control of the ciliary muscle, produces the image (inverted) on the retina. Retinal ganglion cells collectively form the optic nerve (1 in **Fig. 28.6**), and their axons decussate at the optic chiasm (2), where fibres from the nasal retina cross and join with uncrossed fibres originating in the temporal retina to form the optic tract (3). Each optic tract thus carries information from the contralateral visual hemifield.

From the lateral geniculate body, fibres pass in the optic radiation in two divisions through the parietal and temporal lobes (4 and 5) to reach the visual cortex of the occipital lobe (6 and 7), which is

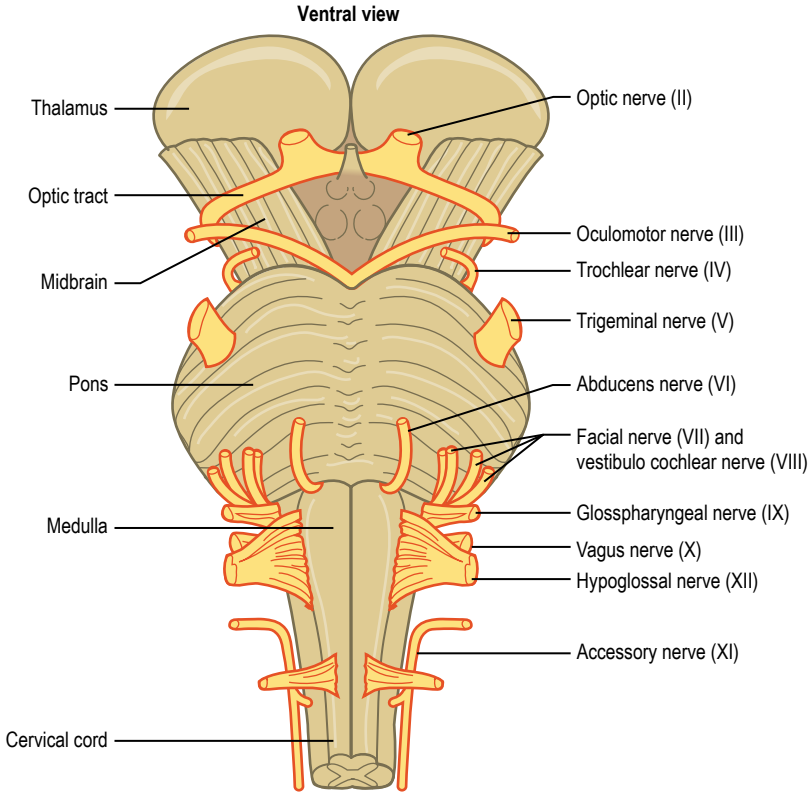


Fig. 28.5 Cranial nerves exiting the brainstem.

somatotopically organized with macular vision located at the occipital pole.

Beyond the visual cortex, visual information is further processed by neighbouring visual association areas to detect lines, orientation, shapes, movement, colour, and depth; there is even a distinct area responsible for face recognition.

Visual acuity

Visual acuity is assessed in each eye with a Snellen chart and/or Near Vision Reading Types, as described in Chapter 31.

Visual field defects

Visual fields are assessed at the bedside by confrontation—comparing the examiner's and patient's fields, one eye at a time and quadrant by quadrant. Patience and good technique are required to produce reliable results. White and red targets (traditionally, hatpins) are used to assess peripheral and central fields, respectively, although in practice a fingertip is often substituted as a cruder screening test. More detailed quantification of fields may be obtained using Goldmann (manual) or Humphrey (automated) perimetry testing.

Field defects (shown in **Fig. 28.6**) are described as hemianopic when half the field is affected and quadrantanopic when a quadrant is affected. Lesions posterior to the optic chiasm produce ***homonymous*** field defects, indicating involvement of the same part of the visual field in both eyes, as information from the two visual hemifields is separated beyond this point. Lesions damaging decussating nasal fibres at the optic chiasm cause bitemporal defects.

Retinal and local eye lesions

See Chapter 31.

Optic nerve lesions

Unilateral visual loss, commencing with a central or paracentral (off-centre) scotoma, is the hallmark of an optic nerve lesion. Because most fibres in the optic nerve subserve macular vision, lesions within the nerve disproportionately affect central vision and colour vision. A total optic nerve lesion causes unilateral blindness with loss of pupillary light reflex. Examination findings in optic neuropathy are:

- reduced acuity in the affected eye
- a scotoma (usually central)
- impaired colour vision (red desaturation)
- an afferent pupillary defect (see below)
- optic atrophy—pale disc (develops late).
  Causes are listed in **Box 28.7**.

Optic nerve head oedema

Papilloedema means swelling of the optic discs caused by raised intracranial pressure (ICP). Inflammation, infiltration, and ischaemia of the optic nerve can also cause optic disc swelling **(papillitis)**, usually unilaterally; systemic diseases and toxins can cause bilateral disc swelling. Causes of disc swelling are shown in **Box 28.8**.

Papilloedema

The earliest signs of papilloedema are disc pinkness, with blurring and heaping up of disc margins, nasal first. There is loss of spontaneous pulsation of retinal veins within the disc. The physiological cup becomes obliterated, and the disc is engorged with dilated vessels. Small haemorrhages often surround the disc.

Various conditions simulate true disc swelling. Marked hypermetropic (long-sighted) refractive errors make a disc appear pink, distant, and ill-defined; myopia will cause tilted discs that can look oval-shaped. Myelinated nerve fibres at disc margins and hyaline

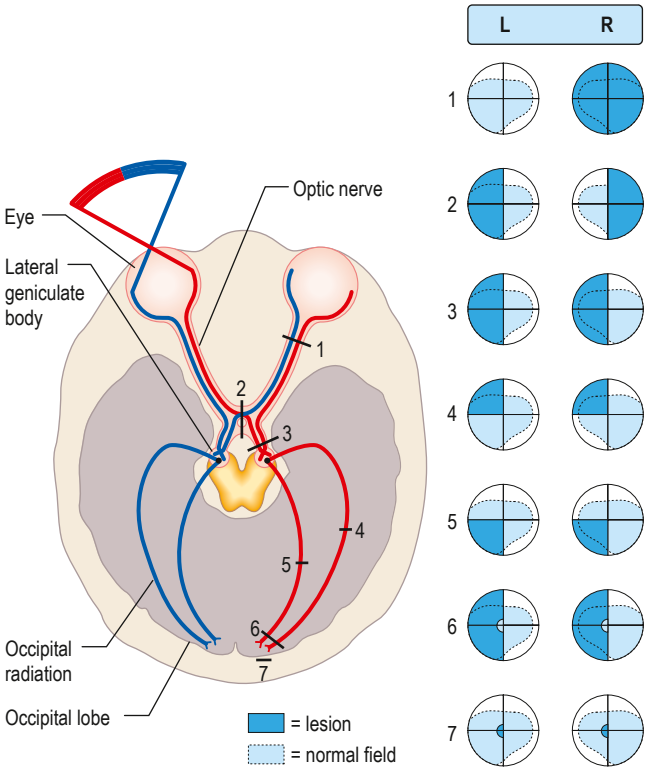


Fig. 28.6 The visual pathway. **(1)** Mononuclear field loss—complete optic nerve lesion. **(2)** Bitemporal hemianopia—chiasmal lesion. **(3)** Homonymous hemianopia—optic tract lesion. **(4)** Homonymous quadrantanopia—temporal lesion. **(5)** Homonymous quadrantanopia—parietal lesion. **(6)** Homonymous hemianopia with macular sparing—occipital cortex or optic radiation. **(7)** Homonymous hemianopia (hemiscotoma)—occipital pole lesion.

bodies (drusen; see Chapter 31) can be mistaken for disc swelling (pseudopapilloedema).

Optical coherence tomography (OCT) is useful to distinguish disc swelling from pseudopapilloedema and for monitoring.

Papilloedema produces few, if any, visual symptoms until the late stages, other than momentary visual obscurations with changes in posture. The blind spot is enlarged but this is not noticed by the patient. The underlying disease is the source of the patient's symptoms in the initial stages. However, over time, or with fulminant papilloedema, progressive and permanent constriction of visual fields occurs, ultimately culminating in optic atrophy and visual loss.

Inflammatory optic neuropathy (optic neuritis)

Optic neuritis is one of the most common causes of subacute visual loss. Symptoms may vary from a mild fogging of central vision with colour desaturation to a dense central scotoma, but very rarely complete blindness. Pain on eye movements is almost universal. Two-thirds of patients have retrobulbar neuritis, in which the optic disc appears normal, despite severe visual loss. The remainder will have anterior optic neuritis (or papillitis) with disc swelling in the acute phase. Optic atrophy develops in spite of good visual recovery, with disc pallor and detectable slowing of conduction in the anterior visual pathways on visual-evoked potentials.

A plaque of demyelination within the optic nerve is the most common cause in Western populations. Dedicated magnetic resonance imaging (MRI) of the optic nerves may show the inflammatory plaque, and imaging of the brain may show additional inflammatory lesions, which confer a higher risk of developing multiple sclerosis (MS).

? Box 28.8 Causes of optic disc swelling

- Raised intracranial pressure (papilloedema)
  - Brain tumour, abscess, or haemorrhage
  - Idiopathic intracranial hypertension
  - Hydrocephalus
  - Venous sinus thrombosis
- Optic neuropathies
  - Optic neuritis (papillitis)
  - Ischaemic optic neuropathy, e.g., giant cell arteritis
  - Toxic optic neuropathy, e.g., methanol
  - Infective optic neuropathies
- Central retinal vein occlusion
- Orbital mass lesions
- Malignant hypertension
- Metabolic causes
  - Hypercapnia, chronic hypoxia, hypocalcaemia
- Retinal vasculitis, e.g., systemic lupus erythematosus
- Disc infiltration
  - Leukaemia, sarcoidosis, optic nerve glioma

Approximately 50% of patients go on to develop MS with prolonged follow-up (see 'Multiple sclerosis' section). Recovery of visual acuity to 6/9 or better occurs in 95% of typical cases over months; recovery time is improved by high-dose intravenous or oral steroids given acutely. Bilateral, unusually severe, or sequential optic neuritis raises suspicion of neuromyelitis optica (NMO) spectrum disorders (see 'Neuromyelitis optica' section).

Optic neuritis may also be caused by infective or other inflammatory disorders, such as sarcoidosis or vasculitides (see **Box 28.9**).

Anterior ischaemic optic neuropathy

The anterior part of the optic nerve is supplied by the posterior ciliary arteries, occlusion, or hypoperfusion of which leads to infarction of all or part of the optic nerve head. There is sudden or stuttering 'altitudinal' visual loss (typically, the lower half of the visual field) with pallid disc swelling, later replaced by optic atrophy. The other eye is affected later in one-third of cases.

Individuals with small hypermetropic discs seem to be predisposed, and often there are relatively few vascular risk factors. A period of hypotension overnight frequently seems to be the trigger. Less commonly, arteritis is the cause, as in giant cell arteritis (GCA; see Chapter 32, 'Giant cell arteritis' section).

Optic atrophy

Optic atrophy means disc pallor, from loss of axons, glial proliferation, and decreased vascularity. This may eventually develop following any type of optic neuropathy of sufficient severity to damage axons extensively within the nerve, including chronic papilloedema.

Optic chiasm

Bitemporal hemianopia or quadrantanopia occurs with compression of the chiasm from above or below. Common causes are:

- pituitary tumours (see Chapter 34, 'Pituitary space-occupying lesions and tumours' section)
- meningioma
- craniopharyngioma.

Optic tract and optic radiation

Optic tract lesions (rare) cause a homonymous hemianopia (loss of the contralateral visual field in both eyes). Optic radiation lesions cause homonymous quadrantanopic defects. Temporal lobe lesions

Box 28.9 Causes of optic neuropathy

- Inflammatory (optic neuritis), e.g., demyelination, sarcoidosis, vasculitis
- Optic nerve trauma or compression, e.g., glioma, meningioma, aneurysm, bone disorders affecting orbit
- Toxic, e.g., tobacco/alcohol, ethambutol, methyl alcohol, quinine, hydroxy chloroquine, radiation
- Ischaemic optic neuropathy, e.g., giant cell arteritis
- Hereditary optic neuropathies, e.g., Leber's
- Nutritional deficiency, e.g., vitamins B_1 and B_{12}
- Infection, e.g., orbital cellulitis, syphilis, tuberculosis
- Neurodegenerative disorders, e.g., leukodystrophies
- Papilloedema and its causes (see **Box 28.8**)

(e.g., tumour, infarction) cause upper quadrantic defects, and parietal lobe lesions cause lower quadrantic defects. Middle cerebral artery territory strokes often involve both divisions of the optic radiation and result in a homonymous hemianopia (**Fig. 28.6**).

Occipital cortex

Homonymous hemianopic defects are produced by unilateral posterior cerebral artery infarction (see Chapter 29). The macular cortex (at each occipital pole) may be spared, probably because of collateral blood supply from the middle cerebral artery.

Widespread bilateral occipital lobe damage by infarction (***'top of the basilar'*** syndrome, occlusion of the rostral basilar artery), trauma, or coning causes ***cortical blindness*** (Anton syndrome). The patient cannot see but characteristically lacks insight into this; he or she may even deny it. Pupillary responses remain normal because the phylogenetically ancient pathway to the pretectal nuclei is intact (see below).

Pupils

A slight difference between the size of each pupil (up to 1 mm) is common ***(physiological anisocoria)*** and does not vary with differing light levels. The pupil tends to become smaller and irregular in old age (senile miosis); anisocoria is more pronounced. ***Convergence*** becomes sluggish with ageing.

Pupillary reactions to light and accommodation may be tested (**Fig. 28.7**). A bright torch (not an ophthalmoscope light!) should be used to test the pupillary light reaction.

Afferent pupillary defect (APD). A complete optic nerve lesion causes a dilated pupil and an APD. For a left APD:

- The pupil is unreactive to light (i.e., the direct reflex is absent).
- The consensual reflex (constriction of the right pupil when the left is illuminated) is absent. Conversely, the left pupil constricts when light is shone in the intact right eye: that is, the consensual reflex of the right eye remains intact.

Relative afferent pupillary defect (RAPD). This occurs with incomplete damage to one optic nerve relative to the other. An RAPD is a sensitive sign of optic nerve pathology and can provide evidence of an optic nerve lesion, even after recovery of vision. For a left RAPD:

- Direct and indirect reflexes are intact in each eye but differ in relative strength.
- As the pupillary light reflex responds to the average illumination detected, when the light is swung from the left to the normal right eye the strength of the afferent signal increases, and so both pupils constrict slightly via a strong consensual response. Both pupils dilate slightly when the torch is swung back to the left eye, as the amount of illumination detected by the abnormal optic nerve is lower.

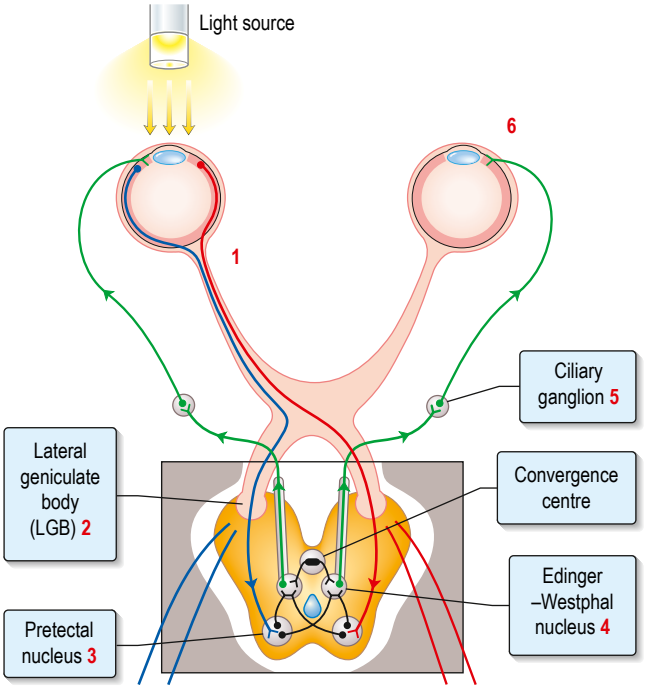


Fig. 28.7 Pupillary light reflex. *Afferent pathway:* **(1)** Light activates optic nerve axons. **(2)** Axons (some decussating at the chiasm) pass through each lateral geniculate body, and **(3)** synapse at pretectal nuclei. ***Efferent pathway:*** **(4)** Action potentials pass to Edinger–Westphal nuclei of the IIIrd nerve, then, **(5)** via parasympathetic neurons in IIIrd nerves to cause **(6)** pupil constriction.

Horner's syndrome

The sympathetic nervous supply to the eye is a three-neuron pathway originating in the hypothalamus and descending by way of the brainstem and cervical cord to the T1 nerve root, paravertebral sympathetic chain and, via the carotid artery wall, to the eye. Damage to any part of the pathway results in Horner's syndrome (**Box 28.10**). This is significant not only because it affects vision but also because it may indicate serious underlying pathology.

The clinical features of Horner's syndrome are:

- unilateral miosis (constricted pupil)
- partial ptosis
- loss of sweating on the same side (extent depending on the level of the lesion)
- possible subtle conjunctival injection and enophthalmos.

The miosis is most easily seen in low light, as the pupil cannot dilate, and may not be apparent in bright light.

Tonic pupil (Adie pupil)

This is a dilated, often irregular, pupil, and is more frequent in women; it is common and usually unilateral. There is no (or very slow) reaction to bright light and also incomplete constriction to convergence. This is due to denervation in the ciliary ganglion, of unknown cause, and has no other pathological significance. A tonic pupil is sometimes associated with diminished or absent tendon reflexes (Holmes–Adie syndrome).

Argyll Robertson pupil

Now rarely seen in clinical practice, an Argyll Robertson pupil is small and irregular; it is fixed to light but accommodates. The lesion is in the brainstem surrounding the aqueduct of Sylvius. Once

Box 28.10 Causes of Horner's syndrome

Hypothalamic, hemisphere, and brainstem lesions
- Massive cerebral infarction
- Brainstem demyelination or lateral medullary infarction

Cervical cord
- Syringomyelia and cord tumours

T1 root
- Apical lung tumour or tuberculosis
- Cervical rib
- Brachial plexus trauma or inflammation

Sympathetic chain and carotid artery in neck
- Following thyroid/laryngeal/carotid surgery
- Carotid artery dissection
- Neoplastic infiltration
- Cervical sympathectomy

Miscellaneous
- Congenital
- Cluster headache, usually transient
- Idiopathic

considered diagnostic of neurosyphilis, it is now only occasionally seen in diabetes or MS.

III, IV, VI: Oculomotor, trochlear, and abducens nerves

These cranial nerves supply the extraocular muscles and disorders commonly result in abnormal eye movements and diplopia (double vision) due to breakdown of conjugate eye movements. Diplopia may also occur with local orbital lesions or myasthenia gravis (MG).

Examination of eye movements

Pursuit (slow) eye movements and saccadic (fast) eye movements are tested separately. The examiner assesses the range of eye movements in all directions and asks the patient to report double vision. Jerky pursuit movements with saccadic intrusion (i.e., brief, fast saccades interspersed with slower pursuit movements), overshoot on saccadic movements, and nystagmus may indicate cerebellar or brainstem pathology.

Control of eye movements

Fast voluntary eye movements originate in the frontal eye fields. Fibres descend and cross in the midbrain to end in the superior colliculus, which also independently initiates reflexive saccades. The superior colliculus also receives inputs from the retina, the visual cortex (pathway concerned with tracking objects), and the inferior colliculus (tracking auditory stimuli). The superior colliculus sends fibres to centres for horizontal gaze (paramedian pontine reticular formation, PPRF) and vertical gaze (in the midbrain reticular formation).

Conjugate horizontal eye movements are coordinated from each PPRF via the medial longitudinal fasciculus (MLF; Fig. 28.8). Fibres from the PPRF pass to the ipsilateral VIth nerve nucleus (lateral rectus), which also sends interneurons through the contralateral MLF to the opposite IIIrd nerve nucleus (medial rectus subnucleus), thus linking the eyes for lateral gaze.

Abnormalities of conjugate horizontal gaze

A destructive lesion on one side allows the eyes to be driven by the intact opposite pathway. A left frontal destructive lesion (e.g., an infarct) leads to failure of conjugate lateral gaze to the right. In an acute lesion, the eyes are often deviated to the side of the lesion, past the midline, and therefore look towards the left (normal) limbs; there is usually a contralateral (i.e., right) hemiparesis. An irritative lesion (i.e., a seizure) affecting the frontal eye field will drive a gaze deviation towards the opposite side.

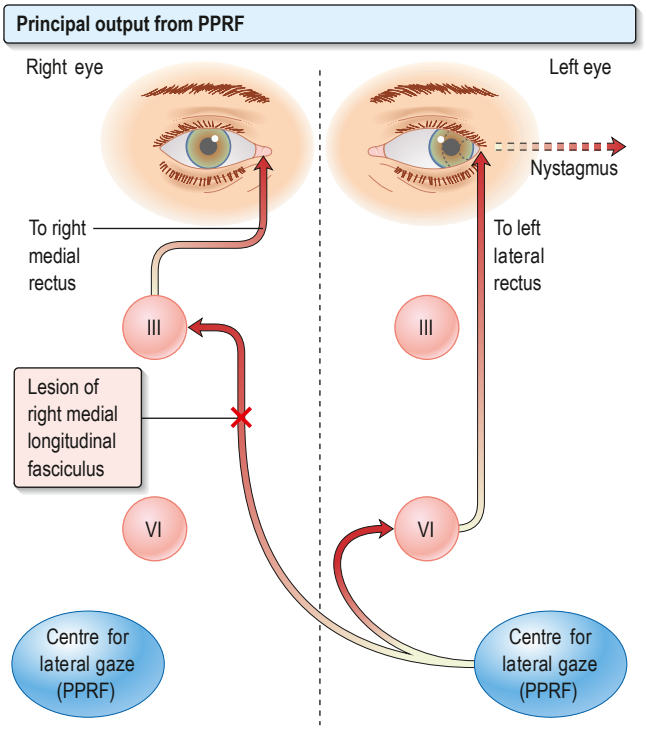


Fig. 28.8 Paramedian pontine reticular formation (PPRF) and internuclear ophthalmoplegia. Impulses from the PPRF pass via the ipsilateral VIth nerve nucleus to the lateral rectus muscle (**ab**duction), and via the medial longitudinal fasciculus to the opposite third nerve nucleus and thus to the opposite medial rectus muscle (**ad**duction). A lesion of the medial longitudinal fasciculus (X) causes failure of or slow **ad**duction in the right eye, and nystagmus in the left eye with left lateral gaze.

In the brainstem, a unilateral destructive lesion involving the PPRF leads to failure of conjugate lateral gaze *towards* that side. There is usually a contralateral hemiparesis, and lateral gaze is deviated towards the side of the paralyzed limbs.

Internuclear ophthalmoplegia

Damage to one MLF causes internuclear ophthalmoplegia (INO), a common complex brainstem eye movement disorder seen frequently in MS. In a right INO, there is a lesion of the right MLF (see Fig. 28.8). On attempted left lateral gaze, the right eye fails to **ad**duct or does so slowly in comparison to the **ab**ducting eye. The left eye develops nystagmus in **ab**duction. The side of the lesion is on the side of impaired **ad**duction. Bilateral INO is almost pathognomonic of MS. A subtle INO is more apparent on testing fast saccadic eye movements, asking the patient to look from one of the examiner's hands to the other.

One and a half syndrome

Pontine infarction involving the PPRF, VIth nerve nucleus, and MLF on one side results in an ipsilateral horizontal gaze palsy and an INO, so that abduction of the opposite eye (with nystagmus) is the only horizontal eye movement possible. Vertical gaze and convergence are preserved, as they have distinct neural control mechanisms.

Vestibulo-ocular (doll's eye) reflexes

Examination is of diagnostic value in coma (see Chapter 11, 'The unconscious patient' section) and assessment of vertigo (see 'Vertigo' section).

Abnormalities of vertical gaze

Failure of upgaze may be caused by dorsal midbrain lesions, such as pinealoma or infarct. When the pupillary light reflex fails in addition, this is called Parinaud's syndrome. Defective vertical gaze also develops in certain degenerative disorders (e.g., progressive supranuclear palsy [PSP]). Some impairment of upgaze occurs as part of normal ageing.

A supranuclear gaze palsy, either horizontal or vertical, will be overcome by testing the VOR, as the brainstem nuclei and LMN pathways subserving the reflex are intact. An impaired VOR implies a nuclear or infranuclear cause.

Nystagmus

Nystagmus is rhythmic oscillation of eye movement; it is a sign of disease of the retina, cerebellum, and/or vestibular systems and their connections. Nystagmus is either ***jerk*** or ***pendular***. Nystagmus must be sustained within binocular gaze to be of diagnostic value; a few beats at the extremes of gaze are normal.

Jerk nystagmus

Jerk nystagmus (usual in neurological disease) is a fast/slow oscillation. This is seen in vestibular, VIIIth nerve, brainstem, and cerebellar lesions. The direction of nystagmus is decided by the fast component, a reflex attempt to correct the slower, primary drift.

- ***Horizontal or rotary jerk nystagmus*** may be either of peripheral (vestibular) or central origin (VIIIth nerve, brainstem, cerebellum, and connections).
  - In peripheral lesions, nystagmus is usually acute and transient (minutes or hours) and associated with severe prostrating vertigo. Peripheral nystagmus obeys Alexander's law: the nystagmus is unidirectional, and the amplitude increases with eye movement in the direction of the fast phase, away from the side of the lesion.
  - In central lesions, nystagmus tends to be long lasting (weeks, months, or more). Vertigo caused by central lesions tends to wane after days or weeks, the nystagmus outlasting it.
  - Gaze-evoked nystagmus (left-beating on looking left and right-beating on looking right) is one of the most common types of nystagmus seen and often related to brainstem or cerebellar disorders.
  - A small amount of end-point nystagmus on far lateral gaze is normal.
- ***Vertical jerk nystagmus*** is typically caused by central lesions.
- ***Down-beat jerk nystagmus*** is a rarity caused by lesions around the foramen magnum (e.g., meningioma, cerebellar ectopia).
- ***Pendular nystagmus*** describes movements to and fro, similar in velocity and amplitude. Pendular nystagmus is usually vertical and present in all directions of gaze. The causes are generally ocular (e.g., poor visual fixation from longstanding, severe visual impairment) or congenital.

III: Oculomotor nerve lesions

The nucleus of the IIIrd nerve lies ventral to the aqueduct in the midbrain. It supplies four external ocular muscles (superior, inferior and medial recti, and inferior oblique) and levator palpebrae superioris (which lifts the eyelid), and effects parasympathetic constriction of the pupil. Causes of a IIIrd nerve lesion are listed in **Box 28.11**.

Signs of a complete IIIrd nerve palsy include:

- unilateral complete ptosis (levator weakness)
- deviation of the eye down and out (unopposed lateral rectus and superior oblique)
- a fixed and dilated pupil.

? Box 28.11 Some causes of a IIIrd nerve lesion

- Aneurysm of the posterior communicating artery
- Infarction of IIIrd nerve, e.g., diabetes, atheroma
- Coning of the temporal lobe
- Midbrain infarction
- Midbrain tumour (primary or secondary)

Patients do not complain of diplopia as the ptosis effectively covers the eye. ***Sparing of the pupil*** indicates that parasympathetic fibres are undamaged; these run in a discrete bundle on the surface of the nerve, and so the pupil is of normal size and reacts normally. Diabetic IIIrd nerve infarction is usually painless and pupil-sparing, unlike compression by a posterior communicating artery aneurysm. Differentiation of 'medical' and 'surgical' causes of a third nerve palsy based on sparing of the pupil is only reliable in the presence of complete third nerve extraocular muscle dysfunction. In the presence of a divisional or incomplete palsy, urgent brain imaging is warranted regardless of pupillary involvement.

IV: Trochlear nerve lesions

The trochlear nerve supplies the superior oblique muscle. The patient complains of torsional diplopia (two objects at an angle) when attempting to look down (e.g., descending stairs); the head is tilted away from that side. The trochlear nerve emerges from the 'wrong' (dorsal) side of the brainstem, and is the cranial nerve with the longest intracranial course; it is thus particularly susceptible to damage from traumatic brain injury (TBI), often with bilateral trochlear nerve palsies occurring.

VI: Abducens nerve lesions

The abducens nerve supplies the lateral rectus muscle (**ab**duction). Lesions cause horizontal diplopia when looking into the distance, which is maximal when looking to the side of the lesion. The eye cannot be fully abducted and an esotropia (inward eye deviation) may be visible in the primary position.

The VIth nerve has a long intracranial course. It can be damaged in the brainstem (e.g., by MS or infarction). In raised ICP (from any cause), it is compressed against the tip of the petrous temporal bone (this may be bilateral), resulting in a false localizing sign. The nerve sheath may be infiltrated by tumours, particularly nasopharyngeal carcinoma. Microvascular ischaemia of the nerve may occur in diabetes with acute onset, followed by recovery within 3 months in most cases.

Complete external ophthalmoplegia

Complete external ophthalmoplegia describes an immobile eye when IIIrd, IVth and VIth nerves are paralyzed at the orbital apex (e.g., by metastasis) or within the cavernous sinus (e.g., by sinus thrombosis or meningioma). Lesions at the orbital apex will usually be accompanied by visual loss from optic nerve involvement.

Wernicke's encephalopathy due to thiamine deficiency (see Chapter 30, 'Wernicke-Korsakoff syndrome' section) may cause a complex eye movement disorder or complete ophthalmoplegia, as may neuromuscular junction (NMJ) disorders such as myasthenia, botulism, and some myopathies or metabolic disorders.

V: Trigeminal nerve

The trigeminal is the largest cranial nerve; it is mainly sensory with a motor component to the muscles of mastication.

Sensory fibres (**Fig. 28.9**; see also **Fig. 28.14**) of the three divisions—ophthalmic (V_1), maxillary (V_2), and mandibular (V_3)—pass

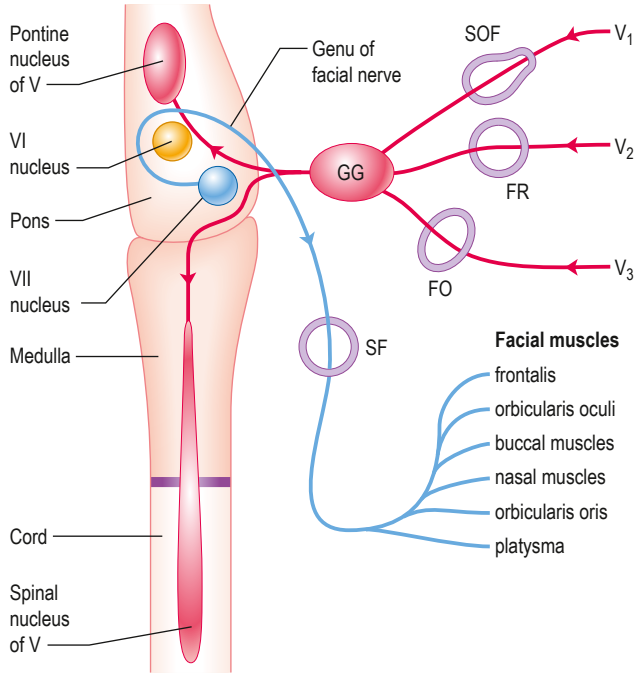


Fig. 28.9 Sensory input of fifth nerve (red) and motor output of the VIIth nerve (blue). *FO,* Foramen ovale; *FR,* foramen rotundum; *GG,* Gasserian ganglion; *SF,* stylomastoid foramen; *SOF,* superior orbital fissure.

to the trigeminal (Gasserian) ganglion at the apex of the petrous temporal bone. Ascending fibres transmitting light touch enter the fifth nucleus in the pons. Descending central fibres carrying pain and temperature form the spinal tract of V, to end in the spinal fifth nucleus that extends from the medulla into the cervical cord.

Clinical features of a Vth nerve lesion

A complete Vth nerve lesion causes unilateral sensory loss on the face, scalp anterior to the vertex, and the anterior two-thirds of the tongue and buccal mucosa; the jaw deviates to that side as the mouth opens (motor fibres). Diminution of the corneal reflex is an early and sometimes isolated sign of a Vth nerve lesion.

Aetiology

- ***Brainstem pathology*** (infarction, demyelination, or tumour) may damage the nucleus, with light touch and spinothalamic pathways sometimes being differentially involved. An 'onion-skin' pattern of sensory loss, in concentric rings, is often seen.
- ***Cerebellopontine angle tumours*** (acoustic neuroma or meningioma) may compress the nerve and also affect the VIIth and VIIIth nerves, producing facial weakness and deafness.
- ***Cavernous sinus and skull base pathology*** (tumour or infection) may affect the ganglion and proximal branches.
- ***Peripheral branches*** may be picked off individually, for example the ominous 'numb chin syndrome' seen with a breast cancer metastasis in the mandible.

Trigeminal neuralgia

Trigeminal neuralgia is discussed with facial pain (see 'Facial pain' section).

Trigeminal sensory neuropathy/neuronopathy

This causes gradually progressive facial sensory loss and tingling with normal imaging. The condition is probably heterogeneous in aetiology but may have an autoimmune basis, with inflammation of the trigeminal ganglion, occurring mainly in association with mixed and undifferentiated connective tissue disease and primary Sjögren's syndrome.

VII: Facial nerve

The VIIth nerve is largely motor, supplying muscles of facial expression. It carries sensory taste fibres from the anterior two-thirds of the tongue via the chorda tympani and supplies motor fibres to the stapedius muscle. The VIIth nerve (see **Fig. 28.9**) arises from its nucleus in the pons and leaves the skull through the stylomastoid foramen. Neurons in each VIIth nucleus supplying the upper face (principally frontalis) receive bilateral supranuclear innervation.

Unilateral facial weakness

Upper motor neuron lesions

These cause weakness of the lower part of the face on the opposite side. Frontalis is spared: normal furrowing of the brow is preserved; eye closure and blinking are largely unaffected. The earliest sign is slowing of movement of one side of the face: for example, on baring the teeth. There is sometimes relative preservation of spontaneous emotional movement (e.g., smiling) compared with voluntary movement. The most common cause of a UMN lesion is hemispheric stroke with contralateral hemiparesis affecting face, arm, and leg.

Lower motor neuron lesions

A complete unilateral LMN VIIth lesion causes weakness (ipsilateral) of all facial expression muscles. The angle of the mouth falls; unilateral dribbling develops. Frowning (frontalis) and eye closure are weak. Corneal exposure and ulceration occur if the eye does not close during sleep. Taste sensation is frequently also impaired. Sites of LMN lesions resulting in facial weakness can be differentiated by the presence of additional signs suggesting the site of the lesion:

Pons. Here the VIIth nerve loops around the VIth (abducens) nucleus (see **Fig. 28.9**), leading to a lateral rectus palsy (see 'VI: Abducens nerve lesions' section) with unilateral LMN facial weakness. When the neighbouring PPRF and corticospinal tract are involved, there is the combination of:

- LMN facial weakness
- failure of conjugate lateral gaze (towards lesion)
- contralateral hemiparesis.

Causes include pontine tumours (e.g., glioma), MS, and infarction.

Cerebellopontine angle (CPA). The neighbouring Vth, VIth, and VIIIth nerves are compressed with VII in the CPA; for example, by acoustic neuroma, meningioma, or metastasis.

Petrous temporal bone. The nerve may be damaged within the bony facial canal, within which lies the sensory geniculate ganglion (receiving taste fibres from the anterior two-thirds of the tongue via the chorda tympani). As well as LMN facial weakness, lesions in this region cause:

- loss of taste on the anterior two-thirds of the tongue
- hyperacusis (loud noise distortion—paralysis of stapedius).

Processes affecting the petrous temporal part of the facial nerve include:

- Bell's palsy (see below)
- trauma
- middle ear infection
- herpes zoster (Ramsay Hunt syndrome; see Chapter 42, 'Varicella [chickenpox] and herpes zoster [shingles]' section).

Skull base, parotid gland, and within the face. The facial nerve can be compressed by skull base tumours and in Paget's disease of bone. Branches of VII may be damaged by parotid gland tumours as the nerve traverses the parotid, and by sarcoidosis (see Chapter 20, 'Sarcoidosis' section) and trauma.

Bell's palsy

This common (37 per 100 000 incidence), idiopathic acute facial palsy is due to inflammation and swelling of the nerve within the tight petrous bone facial canal. In many cases, viral infection or herpes simplex reactivation is believed to be the cause but typically there is no established method of confirming this. Unilateral LMN facial weakness develops over 24 to 48 hours, sometimes with lost or altered taste on the tongue, and hyperacusis. Pain behind the ear is common at onset. Patients often suspect a stroke and may be very distressed. Vague altered facial sensation is often reported, although examination of facial sensation is normal.

Diagnosis is made on clinical grounds and tests are usually not required. The ear (and palate) should be examined for vesicles (see Ramsay Hunt syndrome later), hearing loss, or evidence of local pathology such as cholesteatoma or malignant otitis externa; parotid tumours should be excluded. Cases with the following features require brain imaging: slow, insidious onset; a hyperacute onset (stroke); continued worsening after 3 weeks; no improvement within 4 months. Involvement of other cranial nerves means facial weakness is not due to Bell's palsy. Lyme disease may account for one-quarter of cases of facial palsy in endemic areas and facial palsy may also be seen in HIV seroconversion.

(***Bell's phenomenon*** is the upward conjugate eye movement that occurs normally when the eyes are closed but is more visible when the orbicularis oculi muscle is weak.)

Complete, or almost complete, recovery over 4 to 12 weeks occurs in at least 85% of patients, even without specific treatment. Patients should be reassured that the prognosis is good and that the condition is unlikely to recur.

Associated inability to blink and lacrimal gland involvement may lead to exposure keratitis. Use of lubricating eye ointment is often required and patients should be advised to tape the eye closed carefully at night. For more severe facial weakness with complete inability to close the eye, early ophthalmological assessment is necessary and lateral tarsorrhaphy and/or insertion of a gold weight into the upper lid may be required until recovery occurs.

Early treatment with corticosteroids (prednisolone 60 mg for 1 week) improves outcome if started within 72 hours of onset. The latest evidence does not support use of antiviral agents.

Recovery sometimes takes up to a year if axons have to regrow rather than just remyelinate, in which case aberrant re-innervation of facial muscles (e.g., mouth twitching with blinking) is a frequent late complication. Facial physiotherapy, Botulinum toxin injections, and even plastic surgery may be helpful where recovery is not complete. Bell's palsy rarely recurs; if it does, this should prompt a search for an alternative cause.

Ramsay Hunt syndrome

This is herpes zoster (shingles) of the geniculate ganglion. There is a facial palsy (identical to Bell's) with vesicles around the external auditory meatus and/or the soft palate (sensory twigs from VII). Deafness and vertigo/unsteadiness may occur. Complete recovery is less likely than in Bell's. Antiviral treatment (aciclovir/valaciclovir) is usually given with steroids.

Bilateral facial weakness

Bilateral facial palsy is rare, accounting for less than 1% of cases of facial palsy, but is more likely than unilateral palsies to have an identifiable underlying cause. Paradoxically, bilateral weakness is often less clinically apparent than unilateral weakness, as there is no facial asymmetry.

Causes include:

- infections:
  - Lyme disease (bilateral in 25%—Bannwarth's syndrome)
  - viral: human immunodeficiency virus (HIV) seroconversion, Epstein–Barr virus
  - mastoiditis (bilateral)
  - diphtheria and botulism (rare)
- neuromuscular disorders as part of more generalized weakness:
  - Guillain–Barré syndrome (occasionally facial diplegia occurs with no other weakness)
  - myasthenia
  - myotonic dystrophy and facioscapulohumeral dystrophy
- sarcoidosis and other granulomatous disorders
- skull-base trauma and tumours
- pontine lesions, such as gliomas
- congenital and genetic causes (Gelsolin gene mutations).

Hemifacial spasm

Hemifacial spasm (HFS) is an irregular, painless unilateral spasm of facial muscles, usually occurring after middle age. It starts in the orbicularis oculi and usually progresses gradually over the years to involve other facial muscles on the same side. It varies from a mild to a severe, disfiguring spasm.

HFS is usually caused by compression of the root entry zone of the facial nerve, generally by vascular structures such as the vertebral or basilar arteries or their branches (a mechanism similar to that of trigeminal neuralgia; see 'Facial pain' section). Other lesions at or near the cerebellopontine angle, including tumours and demyelinating plaques, are the cause in approximately 1% of cases.

Occasionally, HFS may occur with ipsilateral trigeminal neuralgia, one symptom usually preceding the other: a combination called ***tic convulsif***. The paroxysms of pain and spasm occur independently. A compressive cause, such as a vascular loop or other structural lesion, is usually identified.

Management

Mild cases require no treatment. Botulinum toxin injection into affected muscles every 3 to 4 months is now the standard treatment. Drugs (e.g., carbamazepine) are of little value. Decompression of the VIIth nerve in the CPA is sometimes helpful. Surgical decompression of the facial nerve in the posterior fossa involves interposing a non-resorbable sponge between the nerve and any adjacent vascular loop identified at operation. The procedure results in complete resolution of symptoms in up to 90% of cases but is associated with a risk of facial weakness or deafness.

Other involuntary facial movements

Myokymia of orbicularis oculi is an irritating twitch, usually of the lower eyelid. It is a normal phenomenon but sometimes a cause of anxiety. More extensive facial myokymia may result from intrinsic brainstem pathology.

Tics and tardive dyskinesia frequently involve facial or perioral muscles (see 'Neuroleptics and movement disorders' section).

Blepharospasm is a form of focal dystonia affecting orbicularis oculi.

VIII: Vestibulocochlear nerve

Auditory fibres from the spiral organ of Corti within the cochlea pass to the cochlear nuclei in the pons. Fibres from these nuclei cross the midline and pass upwards via the medial lemnisci to the medial geniculate bodies and then to the temporal cortex.

Symptoms of a cochlear nerve lesion are deafness and tinnitus (see Chapter 21). Sensorineural and conductive deafness can be distinguished with tuning fork tests, for example Rinne's and Weber's (using a 512 Hz, not 128 Hz, tuning fork; see Chapter 21).

Basic investigations of cochlear lesions

- Pure tone audiometry and auditory thresholds.
- Auditory-evoked potentials (recording responses from repetitive clicks via scalp electrodes; lesion levels are determined from the response pattern).

Causes of deafness

These are listed in **Boxes 28.12** and **21.4**.

Vertigo and the vestibular system

The vestibular system of the inner ear detects head movements and has three primary functions:

- to stabilize gaze during head movements, for example looking ahead while running (the VOR)
- to control posture and balance
- to facilitate perception of orientation and motion.

Nerve impulses generated by movement of hair cells within the three semicircular canals detect head motion in the three planes (yaw/pitch/roll). Balance is maintained by integrating information from:

- the vestibular system
- the visual system
- the somatosensory system—proprioception from limbs, trunk, and neck.

The ***main symptoms*** of vestibular lesions are vertigo and loss of balance.

Vertigo

Vertigo is the illusion of movement of the subject or surroundings, typically rotatory, and should be distinguished from other types of non-specific dizziness. It can be described to patients as being similar to the sensation one feels after getting off a child's roundabout or after spinning on the spot and suddenly stopping. Some people, particularly with chronic vertigo, describe feeling like they are on the deck of a ship.

Vomiting may accompany acute vertigo. Vertigo is always made worse by head movements and patients prefer to remain still and maintain visual fixation. Walking is unsteady. Nystagmus (see 'Nystagmus' section) is the principal sign.

The dizzy patient

Complaints of 'dizziness' require careful history-taking and may signify:

- vertigo
- pre-syncopal sensations due to transient cerebral hypoperfusion, for example orthostatic light-headedness
- unsteadiness or disequilibrium, for example cerebellar or postural stability disorders
- non-specific 'dizziness', for example due to medication or anxiety and hyperventilation.

? Box 28.12 Some causes of vertigo and hearing loss

Conditions	Symptom(s)
Benign paroxysmal positional vertigo	V
Vestibular neuronitis	V
Ménière's disease	V, D
Alcohol, antiepileptic drug intoxication	V
Cerebellar lesions	V
Focal (temporal lobe) seizures	V
Migraine	V+ phonophobia
Brainstem ischaemia	V, occasionally D
Multiple sclerosis	V, occasionally D
Mumps, intrauterine rubella and congenital syphilis	D
Advancing age (presbycusis) and otosclerosis	D
Acoustic trauma	D
Congenital/genetic	D
Gentamicin, furosemide	V, D
Middle and external ear disease	V, D
Cerebellopontine angle lesions, e.g., acoustic neuroma	V, D
Carcinomatous meningitis, sarcoidosis, and tuberculous meningitis	V, D

D, Hearing loss; *V*, vertigo.

Causes of vertigo

Vertigo indicates a disturbance of the vestibular apparatus or brainstem and associated neural pathways (see **Box 28.12**). Causes:

- ***Peripheral causes*** (vestibular system) are common. Deafness and tinnitus accompanying vertigo indicate involvement of the ear or cochlear nerve (see Chapter 21).
- ***Central causes*** (brainstem and connections) are rarer. Other clinical features, such as diplopia, weakness, cerebellar signs, or cranial nerve palsies, may help localize the lesion.

Peripheral (vestibular) disorders

Vestibular disorders are fully discussed elsewhere (see Chapter 21). Attack duration and frequency, as well as trigger factors, help the clinician distinguish on history between different pathological causes. The ability to perform a Hallpike test, head impulse test, and Epley particle repositioning manoeuvre are invaluable skills for all clinicians (see Chapter 21; **Fig. 28.10** and see **Fig. 21.7**).

The most commonly encountered peripheral causes of vertigo are:

- ***benign paroxysmal positional vertigo*** (BPPV)—frequent attacks lasting seconds only, provoked by turning the head in bed (see Chapter 21). Hallpike test abnormal
- ***vestibular neuronitis***—acute onset lasting days or weeks; non-recurrent. Head impulse test will be abnormal
- ***Ménière's disease***—recurrent attacks lasting minutes or hours, usually accompanied by hearing loss, tinnitus, and a feeling of fullness in the ear
- ***trauma***—vestibular disruption following head injury; patients may also have BPPV
- ***CPA mass lesions and tumours***, such as vestibular nerve schwannomas, may cause vertigo by compressing the vestibular nerve, but rarely do so in the absence of unilateral deafness and tinnitus.

Central causes of vertigo

Vertigo may be a manifestation of brainstem or cerebellar pathology, including:

- ***infarcts*** involving the vestibular nuclei in the medulla (e.g., the lateral medullary syndrome), or the cerebellum
- ***demyelination*** involving the brainstem
- ***posterior fossa mass lesions***—such as tumours, haemorrhage or vascular malformations
- ***migrainous vertigo*** (see 'Migraine' section)—lasts hours, occurring every few weeks or months; headache may be absent
- ***drugs***—such as anticonvulsant toxicity and alcohol.

Although central vertigo occasionally occurs in isolation with brainstem or cerebellar pathology, including in a stroke, it is more typically a single component of a more complex clinical picture, associated with other symptoms or examination findings. The brainstem nuclei and tracts are tightly packed into a small space and most pathological processes affect multiple contiguous neural pathways, resulting, for example, in diplopia, eye movement disorders, cranial nerve palsies, cerebellar signs, or hemiparesis. These should be carefully sought.

Very rarely, an AICA or labyrinthine territory infarct will present with acute onset peripheral vertigo and deafness, through direct infarction of the labyrinth or LMN vestibular pathways. As always, the history is key.

Basic investigations for vestibular problems

Bedside assessment is usually sufficient to make a diagnosis in the majority of patients:

- examination of eye movements for nystagmus (see 'Nystagmus' section)
- assessment of hearing and otoscopic examination of the ear (see Chapter 21)
- head impulse (thrust) test—to assess the VOR and identify a unilateral vestibulopathy (see **Fig. 28.10**)
- Hallpike manoeuvre—a positioning test to stimulate the posterior semicircular canal and trigger an attack in BPPV (see **Fig. 21.7**).

The ***HINTS examination*** has high sensitivity and specificity for differentiating between central and peripheral causes of acute onset vertigo, but should only be used on patients who have persistent vertigo (symptomatic at the time of examination), ongoing nystagmus, and an otherwise completely normal neurological examination. The HINTS exam suggests a central cause of vertigo (and brain imaging is required) if any of the following apply:

- **H**ead **I**mpulse test normal (a corrective saccade is not generated with a central lesion; note however that bilateral corrective saccades may be central)
- **N**ystagmus—vertical or pure torsional nystagmus, or nystagmus that changes direction with gaze, is central (note if there is no nystagmus, then HINTS cannot be used)
- **T**est of **S**kew abnormal—ask the patient to look at your nose, then quickly cover each eye in turn. If there is a vertical or diagonal corrective movement as you uncover the eye, then the vertigo is likely to be central.

Of note, in the validation study the HINTS examination was performed by a neuro-ophthalmologist with specialist equipment. Therefore, it is important to exercise caution: acute onset persistent vertigo in someone with vascular risk factors, especially if they have been given vestibular sedatives that suppress the nystagmus, may warrant an MRI brain even if HINTS appears to suggest a peripheral cause.

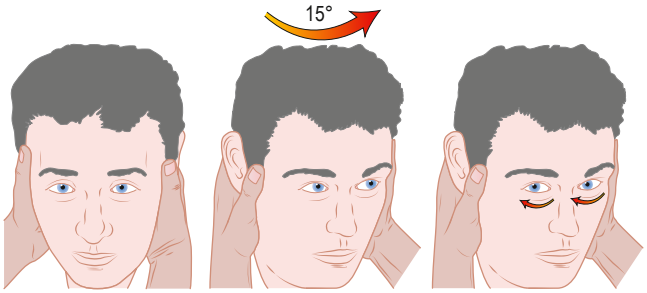


Fig. 28.10 The head impulse test. When tested on the normal side, a patient can maintain fixation on a distant object as the examiner rapidly turns the patient's head sideways through about 15 degrees. If there is unilateral vestibulopathy, the patient cannot maintain fixation on the target and can be observed to make a voluntary eye movement (a saccade) back to the target after the head impulse, as shown on the right. It is essential for the head to be turned as rapidly as possible; otherwise, smooth pursuit eye movements will compensate for the head turn.

Specialist testing is occasionally required to assess vestibular function and hearing. This includes:

- caloric testing—irrigation of the external auditory meatus with cold and then warm water to stimulate the horizontal semicircular canal and induce nystagmus; labyrinthine function is tested in each ear separately
- electro-nystagmography—to quantify and characterize nystagmus under different conditions, for example in a rotating chair
- posturography—assesses body sway on a moving platform
- pure-tone audiograms
- high-definition MRI—provides the best structural imaging of the brainstem and CPA, and is useful where a central cause of vertigo is suspected.

Vestibular neuronitis

Vestibular neuronitis is a common but poorly understood problem. It is an acute attack of isolated vertigo with nystagmus, often with vomiting, and is believed to follow viral infections. The examination findings are distinct:

- there will be unidirectional horizontal nystagmus with a torsional component, that follows Alexander's law
- the head impulse test will be abnormal in one direction.

The disturbance lasts for several days or weeks, is self-limiting and rarely recurs. The nystagmus typically settles but the head impulse test will remain abnormal. Vestibular neuronitis is sometimes followed by BPPV (see Chapter 21, 'Benign paroxysmal positional vertigo' section). Deafness is absent (except in the case of true labyrinthitis, see Chapter 21). Acute treatment is with vestibular sedatives, but these should be stopped after the first few days to allow central compensation to occur. Some patients will go on to develop chronic symptoms, and early consideration should be given to vestibular rehabilitation.

Lower cranial nerves IX, X, XI, XII

The glossopharyngeal (IX), vagus (X), and accessory (XI) nerves arise in the medulla and leave the skull through the jugular foramen. The hypoglossal (XII) nerve arises in the medulla, to leave the skull base via the hypoglossal foramen. Outside the skull, the four cranial nerves lie together, close to the carotid artery and sympathetic trunk.

Glossopharyngeal (IX)

This nerve is largely sensory, supplying sensation and taste from the posterior third of the tongue and the pharynx (afferent pathway of gag reflex). Motor fibres supply the stylopharyngeus muscle, and parasympathetic fibres supply the parotid.

Vagus (X)

The vagus is a mixed nerve, largely motor, which supplies striated muscle of the pharynx (efferent gag reflex pathway), larynx (including vocal cords via recurrent laryngeal nerves), and upper oesophagus. There are sensory fibres from the larynx. Parasympathetic fibres supply the heart and abdominal viscera.

Accessory (XI)

The accessory nerve, a complex motor nerve, supplies the trapezius and sternomastoid muscles.

Hypoglossal (XII)

The hypoglossal nerve is a motor nerve to tongue muscles.

IXth and Xth nerve lesions

Principal causes of IXth, Xth, XIth, and XIIth nerve lesions are listed in **Box 28.13**.

Isolated lesions of IXth and Xth nerves are unusual, since disease at the jugular foramen affects both nerves and sometimes XI.

A unilateral IXth nerve lesion causes diminished sensation on the same side of the pharynx and is hard to recognize in isolation. A Xth nerve palsy produces ipsilateral failure of voluntary and reflex elevation of the soft palate (which is drawn to the opposite side) and ipsilateral vocal cords.

Bilateral lesions of IXth and Xth nerves cause palatal weakness, reduced palatal sensation, an absent gag reflex, dysphonia, and choking with nasal regurgitation. ***Bulbar palsy*** is a general term describing palatal, pharyngeal, and tongue weakness of LMN or muscle origin.

Recurrent laryngeal nerve lesions. Paralysis of this branch of each vagus causes hoarseness (dysphonia) and failure of the forceful, explosive part of coughing ('bovine cough'). There is no visible palatal weakness; vocal cord paralysis is seen endoscopically. Bilateral acute lesions (e.g., postoperatively) cause respiratory obstruction—an emergency.

The left recurrent laryngeal nerve (looping beneath the aorta) is damaged more commonly than the right.

Causes of recurrent laryngeal nerve lesions include:

- mediastinal primary tumours (e.g., thymoma)
- secondary spread from bronchial carcinoma
- aortic aneurysm
- trauma or surgery of neck or thorax.

XIth nerve lesions

XIth nerve lesions cause weakness of sternomastoid (rotation of the head and neck to the opposite side) and trapezius (shoulder shrugging). Nerve section (e.g., following lymph node biopsy in the neck posterior triangle) is followed by persistent neuralgic pain.

XIIth nerve lesions

LMN lesions of XII lead to unilateral tongue weakness, wasting, and fasciculation. The protruded tongue deviates towards the weaker side. The XIIth nerve runs close to the carotid artery and carotid dissection is a common cause of acute unilateral XIIth nerve palsy. Bilateral supranuclear (UMN) lesions produce slow, limited tongue movements and a stiff tongue that cannot be protruded far. Fasciculation is absent.

Box 28.13 Principal causes of IXth, Xth, XIth, and XIIth nerve lesions

Within brainstem

- Infarction
- Syringobulbia
- Motor neuron disease (motor fibres)

Around skull base

- Nasopharyngeal carcinoma
- Glomus tumour
- Neurofibroma
- Jugular venous thrombosis
- Trauma

Within neck and nasopharynx

- Nasopharyngeal carcinoma
- Metastases
- Carotid artery dissection (XII)
- Trauma
- Lymph node biopsy in posterior triangle (XI)

Bulbar and pseudobulbar palsy

Bulbar palsy

This is LMN weakness of muscles whose cranial nerve nuclei lie in the medulla (the bulb). Paralysis of bulbar muscles is caused by disease of lower cranial nerve nuclei, lesions of IXth, Xth, and XIIth nerves (see **Box 28.13**), malfunction of their NMJs (e.g., MG, botulism), or disease of muscles themselves (e.g., dystrophies). Findings include a wasted, flaccid tongue that looks large because it is baggy; there may be fasciculations.

Pseudobulbar palsy

Describes bilateral supranuclear (UMN) lesions of lower cranial nerves producing weakness of the tongue and pharyngeal muscles. This resembles, superficially, a bulbar palsy: hence pseudobulbar. Findings are a stiff, slow, spastic tongue (not wasted but small), dysarthria, and dysphagia. Gag and palatal reflexes are preserved and the jaw jerk exaggerated. Emotional lability (inappropriate laughing or crying) often accompanies pseudobulbar palsy. Principal causes are:

- motor neuron disease (MND), often both UMN and LMN lesions (i.e., elements of both pseudobulbar and bulbar palsy)
- cerebrovascular disease, typically following multiple infarcts
- neurodegenerative disorders such as PSP (see 'Atypical parkinsonism' section)
- severe TBI
- MS, mainly as a late event.

Difficulty swallowing, dysarthria, and drooling also develop in later stages of Parkinson's disease.

Dropped head syndrome

As the name suggests, weakness of neck extensors causes neck flexion and inability to hold the head up. Seen mainly in the elderly, it is often due to isolated neck extensor myopathy of uncertain cause but may be a presenting feature of MND, myasthenia, or various myopathic disorders.

Further reading

Biousse V, Newman NJ. Ischemic optic neuropathies. *N Engl J Med* 2015; 372:2428–2436.
Kaski D, Seemungal B. The bedside assessment of vertigo. *Clin Med* 2010; 10:402–405.
Tabi YA, Husain M. Clinical assessment of parietal lobe function. *Pract Neurol* Published Online First: 16 June 2023.

DISORDERS OF MOTOR CONTROL SYSTEMS

There are three systems, each of which interacts by feedback loops with the other two, with sensory input from the reticular formation:

- ***The corticospinal*** (or ***pyramidal*****) *system*** enables purposeful, skilled, intricate, strong, and organized movements. Defective function is recognized by a distinct pattern of signs—loss of skilled voluntary movement, spasticity, and reflex change—seen, for example, in a hemiparesis, hemiplegia, or paraparesis.
- ***The extrapyramidal system*** facilitates fast, fluid movements that the corticospinal system has generated. Defective function produces slowness (bradykinesia), stiffness (rigidity), and/or disorders of movement (rest tremor, chorea, and other dyskinesias). One feature (e.g., stiffness, tremor, or chorea) will often predominate.
- ***The cerebellum*** and its connections have a role coordinating smooth and learned movement, initiated by the corticospinal system, and in posture and balance control. Cerebellar disease leads to unsteady and jerky movements (ataxia), with characteristic limb signs of past-pointing, action tremor and incoordination, gait ataxia, and/or truncal ataxia.

Corticospinal (pyramidal) system

The corticospinal tracts originate in neurons of the cortex and terminate at motor nuclei of cranial nerves and spinal cord anterior horn cells. The pathways of particular clinical significance (Fig. 28.11) congregate in the internal capsule and cross in the medulla (decussation of the pyramids), passing to the contralateral cord as the lateral corticospinal tracts. This is the pyramidal system, disease of which causes UMN lesions. 'Pyramidal' is simply a descriptive term that draws together anatomy and characteristic physical signs; it is used interchangeably with the term UMN.

A proportion of the corticospinal outflow is uncrossed (anterior corticospinal tracts). This is of no relevance in practice.

Characteristics of pyramidal lesions

Signs of an early pyramidal lesion may be minimal (Box 28.14). Weakness, spasticity, or changes in reflexes can predominate, or be present in isolation.

Pyramidal drift of an upper limb

Normally, the outstretched upper limbs are held symmetrically, when the eyes are closed. With a pyramidal lesion, when both upper limbs are held outstretched, palms uppermost, the affected limb drifts downwards and pronates. This sign, with slowing of fast finger movements, is an early one, sometimes occurring before weakness and/or reflex changes become apparent.

Weakness and loss of skilled movement

A unilateral pyramidal lesion above the decussation in the medulla causes weakness of the opposite limbs. When acute and complete, this weakness will be immediate and total: a hemiplegia—for example, following an internal capsule infarct. With time, or a slowly progressive lesion (e.g., a hemisphere glioma), a spastic hemiparesis with a characteristic (pyramidal) pattern emerges, and the characteristic posture of someone with dense pyramidal weakness will be seen: in the upper limb, flexors are stronger than extensors, whereas in the lower limb, extensors are stronger than flexors. In the presence of dense weakness such as after a stroke, the arm is held flexed at the elbow and the wrist flexed across the body. The fingers are curled up. The leg is extended and toes point down, with the foot inverted. There is also loss of skilled movement; fine finger and toe control diminishes. Wasting (except from disuse) is not a feature. When a UMN lesion is below the decussation of the pyramids—in the cervical cord, for example—hemiparesis is on the same side as the lesion.

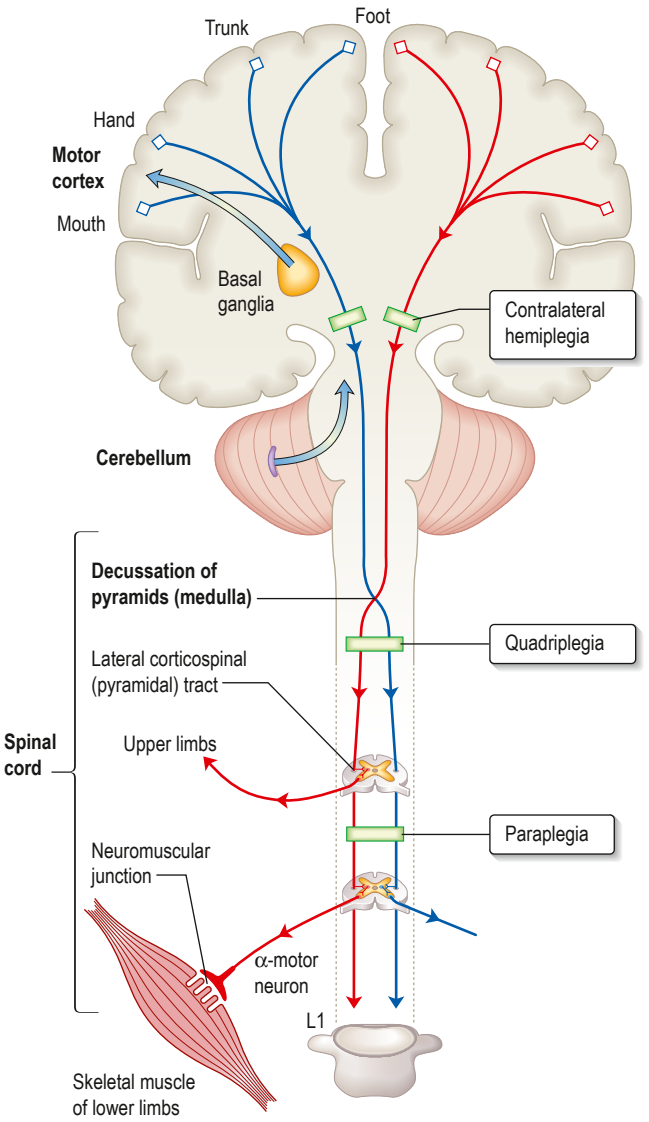


Fig. 28.11 **The motor (corticospinal, pyramidal) system.**

Box 28.14 Features of upper motor neuron lesions

- Drift of upper limb
- Weakness with characteristic pyramidal distribution
- Changes in tone: flaccid–spastic
- Exaggerated muscle stretch reflexes
- Extensor plantar response
- Loss of skilled finger/toe movements
- Loss of abdominal reflexes
- No muscle wasting
- Normal electrical excitability of muscle

Changes in tone and tendon reflexes

An acute lesion of one pyramidal tract (e.g., internal capsule stroke) causes initially flaccid paralysis with loss of tendon reflexes. Increase in tone follows, usually within several days, due to loss of inhibitory effects of the corticospinal pathways and an increase in spinal reflex activity. This increase in tone (spasticity) is most easily detectable in stronger muscles. Spasticity is characterized by sudden ***changing resistance*** to rapid passive movement—the clasp-knife effect. Relevant tendon reflexes become exaggerated; clonus may emerge.

Changes in superficial reflexes

The normal flexor plantar response becomes extensor. In a severe acute lesion, this extensor response can be elicited from a wide area of the foot. As recovery progresses, the receptive area diminishes until the lateral posterior third of the sole remains receptive to an orange-stick stimulus (the appropriate instrument). An extensor plantar response is certain when great toe dorsiflexion is accompanied by abduction of adjacent toes. Abdominal reflexes are abolished on the affected side.

Patterns of UMN disorders

There are three main patterns:

- ***Hemiparesis*** means weakness of the limbs on one side; it is usually caused by a lesion in the brain and occasionally in the cord. If the lesion affects the brain, limb weakness is commonly accompanied by facial weakness and dysarthria.
- ***Paraparesis*** means weakness of both lower limbs and is usually diagnostic of a cord lesion; bilateral medial brain lesions (e.g., parasagittal meningioma) occasionally cause paraparesis.
- ***Tetraparesis*** (also called ***quadriparesis***) means weakness of four limbs.

Hemiplegia, paraplegia, and tetraplegia (strictly) indicate *total* paralysis but are often used to describe severe weakness.

Hemiparesis

The level within the corticospinal system is recognized from particular features.

Motor cortex. Weakness and/or loss of skilled movement confined to one contralateral limb (an arm or a leg—monoparesis) or part of a limb (e.g., a clumsy hand) is typical of an isolated motor cortex lesion (e.g., an infarct or secondary neoplasm). A defect in cognitive function (e.g., aphasia) and focal epilepsy may occur.

Internal capsule. Corticospinal fibres are tightly packed in the internal capsule (about 1 cm^2); thus a small lesion causes a large deficit. A middle cerebral artery branch infarction (see Chapter 29) produces a sudden, dense, contralateral hemiplegia affecting face, arm, and leg.

Pons. A pontine lesion (e.g., an MS plaque) is rarely confined to the corticospinal tract. Adjacent structures such as VIth and VIIth nuclei, MLF, and PPRF are involved. Diplopia, facial weakness, INO, and/or a lateral gaze palsy occur with contralateral hemiparesis.

Spinal cord. A lesion of one lateral corticospinal tract (e.g., a cervical cord injury) causes an ipsilateral UMN lesion, the level being indicated by changes in reflexes (e.g., absent biceps with brisk triceps, C5/6), features of a Brown–Séquard syndrome (see 'Lesions of the sensory pathways' section) and muscle wasting at the level of the lesion.

Spastic paraparesis

Paraparesis indicates bilateral damage to corticospinal pathways, causing weakness and spasticity (or flaccid weakness in the initial phase of spinal shock after an acute cord insult). Cord compression (see 'Spinal cord compression' section) or cord diseases are the usual causes; cerebral lesions occasionally produce paraparesis. Paraparesis is a feature of many neurological conditions; finding the cause is crucial (see 'Lesions of the sensory pathways' section).

Extrapyramidal system

The extrapyramidal system is a general term for basal ganglia motor systems: that is, corpus striatum (caudate nucleus + globus pallidus + putamen), subthalamic nucleus (STN), substantia nigra, and parts of the thalamus. In basal ganglia/extrapyramidal disorders, two features (either or both) become apparent, in limbs and axial muscles:

- reduction in speed of movement ***(bradykinesia)*** or akinesia (no movement), with muscle rigidity
- involuntary ***hyperkinetic*** movements (tremor, chorea, dystonia, tics, myoclonus).

Extrapyramidal disorders are classified broadly into ***akinetic–rigid syndromes*** (see 'Movement disorders' section) where poverty of movement predominates, and ***hyperkinetic*** movement disorders where there are abnormal involuntary movements (see 'Hyperkinetic movement disorders' section).

The most common extrapyramidal disorder is Parkinson's disease.

Essential anatomy

The corpus striatum lies close to the substantia nigra, thalami, and subthalamic nuclei, lateral to the internal capsule (**Fig. 28.12**; see **Fig. 28.11**).

Function and dysfunction

Overall function of this system is modulation of cortical motor activity by a series of loops between cortex and basal ganglia (see **Fig. 28.12**). In involuntary movement disorders, there are specific changes in neurotransmitters (**Box 28.15**) rather than focal lesions seen on imaging or at postmortem.

Proposed model of principal pathways

1. ***Direct pathway*** from striatum to internal globus pallidus (GP*i*) and substantia nigra pars reticulata (SN*r*). Inhibitory ***synapse F***, GABA, and substance P.
2. ***Indirect pathway*** from striatum to external globus pallidus (GPe; inhibitory ***synapse C***, GABA, enkephalin), which projects to the STN instead of 'directly' to the thalamus (inhibitory ***synapse D***, GABA). Terminates in GP*i*–SN*r* (in excitatory ***synapse E***, glutamate).
3. ***Hyperdirect*** pathway from cortex to STN. Excitatory ***synapse J***, glutamate. STN excites GP*i*–SN*r* as in the indirect pathway.
4. **GP*i* and SN*r* to thalamus.** Inhibitory ***synapse G***, GABA.
5. **Thalamus to cortex.** Excitatory, ***synapse H***.
6. **Cortex to striatum.** Excitatory, glutamate.
7. **Dopaminergic modulation, both inhibitory and excitatory, from substantia nigra pars compacta (SNc) to striatum.** ***Synapse A***, dopamine, D_1, excitatory—activates the direct pathway; and ***synapse B***, D_2, inhibitory—inhibits the indirect pathway. The main role of SNc dopaminergic neurons is to coordinate the initiation of movement.

Ultimately, the direct pathway acts to disinhibit the thalamus which excites the motor cortex, selecting the desired motor programme, while indirect and hyperdirect pathways inhibit competing motor programmes via the thalamus. The model helps explain how basal ganglia disease can either reduce excitatory thalamo-cortical activity at synapse H—that is, movement—causing bradykinesia, or increase it, causing hyperkinetic disorders.

Parkinson's disease (PD). This is characterized by slowness, stiffness, and rest tremor (see 'Parkinsonian disorders' section). Degeneration in SNc causes loss of dopamine activity in the striatum. Dopamine is excitatory for synapse A and inhibitory for synapse B. Through the direct pathway, there is reduced activity at synapse F, leading to increased inhibitory output (G) and decreased cortical activity (H).

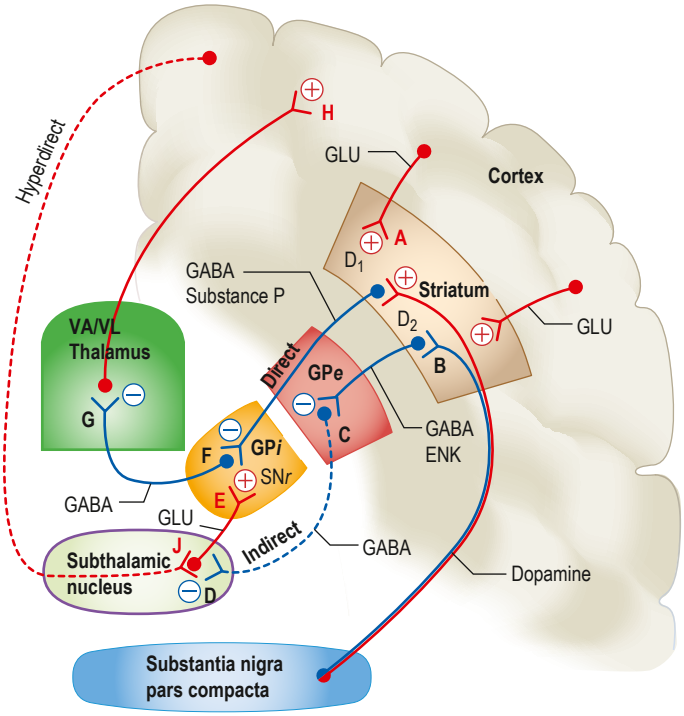


Fig. 28.12 Extrapyramidal system: connections and neurotransmitters. The *inhibitory pathways* are in blue (B, C, D, F, G) and the ***excitatory pathways*** are in red (A, E, H). *ENK,* Enkephalin; *GABA,* γ-aminobutyric acid; *GLU,* glutamate; *GPe,* external globus pallidus; *GPi,* internal globus pallidus; *SNr,* substantia nigra pars reticulata; *VA/VL,* ventral anterior and ventrolateral thalamic nuclei.

Box 28.15 Changes in neurotransmitters in Parkinson's and Huntington's diseases

Site	Neurotransmitter
Parkinson's	
Putamen	Dopamine ↓ 90% Noradrenaline (norepinephrine) ↓ 60% 5-HT ↓ 60%
Substantia nigra	Dopamine ↓ 90% GAD + GABA ↓↓
Cerebral cortex	GAD + GABA ↓↓
Huntington's	
Corpus striatum	Acetylcholine ↓↓ GABA ↓↓ Dopamine: normal GAD + GABA ↓↓

5-HT, 5-hydroxytryptamine; *GABA,* γ-aminobutyric acid; *GAD,* glutamic acid decarboxylase, the enzyme responsible for synthesizing GABA.

Also in PD, in the indirect pathway, dopamine deficiency results in disinhibition of neurons synapsing at C. This leads to reduced activity at D, and to increased activity of neurons in the STN. There is excess stimulation at synapse E, enhancing further inhibitory output of GP*m*–SN*r*.

The net effect via both pathways is to inhibit the ventral anterior (VA) and ventrolateral (VL) nuclei of the thalamus at synapse G. Cortical (motor) activity at H is thus reduced.

Levodopa helps slowness and tremor in PD (see 'Parkinsonian disorders' section) but induces unwanted dyskinesias by increasing dopamine activity at synapses A and B; it is thought to do this by reversing sequences in both direct and indirect pathways.

Hemiballismus (see 'Hyperkinetic movement disorders' section). Wild, flinging (ballistic) limb movements are caused by a lesion in the STN, typically a haemorrhage. This reduces excitatory activity at synapse E, reduces inhibition at G, with increased thalamo-cortical neuronal activity, and increases activity at H.

Cerebellum

The third system of motor control modulates coordination and learned movement patterns, rather than speed. Ataxia, that is unsteadiness, is characteristic.

The cerebellum receives afferents from:

- proprioceptive receptors (joints and muscles)
- vestibular nuclei
- basal ganglia
- the corticospinal system
- inferior olivary nuclei (relaying integrated sensory and motor information to cerebellum).

Efferents pass from the cerebellum to:

- each red nucleus
- vestibular nuclei
- basal ganglia
- the corticospinal system.

Each lateral cerebellar lobe coordinates movement of the ipsilateral limbs. The vermis (a midline structure) is concerned with maintenance of axial (midline) posture and balance.

Cerebellar lesions

Box 28.16 summarizes the main causes of cerebellar disease. Expanding lesions obstruct the aqueduct to cause hydrocephalus, with severe pressure headaches, vomiting, and papilloedema. Coning of the cerebellar tonsils (see 'Brain tumours'

Box 28.16 Principal causes of cerebellar syndromes

Tumours
- Haemangioblastoma
- Medulloblastoma
- Secondary neoplasm
- Compression by acoustic neuroma

Vascular
- Haemorrhage
- Infarction
- Arteriovenous malformation

Infection
- Abscess
- HIV
- Encephalitis

Inflammatory/immune
- Multiple sclerosis (common in MS)
- Paraneoplastic syndromes (rapidly progressive)

Developmental
- Chiari malformation
- Cerebral palsy

Toxic and metabolic
- Antiepileptic drugs
- Chronic alcohol use
- Carbon monoxide poisoning
- Lead poisoning
- Solvent misuse
- Hypothyroidism
- Hypomagnesaemia

Inherited
- Friedreich's and other spinocerebellar ataxias
- Ataxia telangiectasia

Degenerative
- Multiple system atrophy
- Prion diseases

Miscellaneous
- Hydrocephalus
- Superficial siderosis

section) through the foramen magnum leads to respiratory arrest, sometimes within minutes/hours. Rarely, tonic seizures (attacks of limb stiffness) occur.

Lateral cerebellar hemisphere lesions

A lesion within one cerebellar lobe (e.g., tumour or infarction) causes disruption of the normal sequence of movements (dyssynergia) on the same side.

Posture and gait. The outstretched arm is held still in the early stages of a cerebellar lobar lesion (cf. the drift of pyramidal lesions), but there is rebound upward overshoot when the limb is pressed downwards and released. Gait becomes broad and ataxic, faltering towards the side of the lesion.

Tremor and ataxia. Movement is imprecise in direction, force, and distance (dysmetria). Rapid alternating movements (tapping, clapping, or rotary hand movements) become disorganized (dysdiadochokinesia). Intention tremor (action tremor that increases towards the target) and past-pointing (dysmetria) are seen, but speed of fine movement is preserved (cf. extrapyramidal and pyramidal lesions).

Nystagmus. Coarse horizontal nystagmus (see 'Nystagmus' section) develops with a lateral cerebellar lobe lesion. The fast component is always towards the side of the lesion. Horizontal gaze-evoked nystagmus may indicate a bilateral or diffuse cerebellar disorder. Saccadic eye movements are dysmetric (overshoot).

Dysarthria. Halting, jerking speech (scanning speech) develops.

Other signs. Titubation—rhythmic head tremor as either forward and back (yes–yes) movements or rotary (no–no) movements—can occur, mainly when cerebellar connections are involved (e.g., in essential tremor and MS. Hypotonia (floppy limbs) and depression of reflexes (with slow, pendular reflexes) are also sometimes seen.

Midline cerebellar lesions

Cerebellar vermis lesions have dramatic effects on trunk and axial muscles. There is difficulty standing and sitting unsupported (truncal ataxia), with a broad-based, ataxic gait. Lesions of the flocculonodular region cause vertigo and vomiting with gait ataxia if they extend to the roof of the IVth ventricle.

Tremor

Tremor means a regular and sinusoidal oscillation of the limbs, head, or trunk.

Postural tremor

Everyone has a physiological tremor (often barely perceptible) of the outstretched hands at 8 to 12 Hz. This is increased with anxiety, caffeine, hyperthyroidism, and drugs (e.g., sympathomimetics, sodium valproate, lamotrigine, lithium) and occurs in mercury poisoning. A higher-amplitude, postural tremor is seen in benign essential tremor (usually quite fast at 5–12 Hz).

Intention tremor

Tremor on action that worsens towards the target, with past-pointing and accompanying incoordination of rapid alternating movement (dysdiadochokinesia), occurs in cerebellar lobe disease and with lesions of cerebellar connections. Titubation (head tremor) and nystagmus may be present.

Rest tremor

Seen typically in Parkinson's disease, this tremor is noticeably worse at rest, usually 4 to 7 Hz (often pill-rolling, between thumb and forefinger). Unlike essential tremor, parkinsonian tremor is generally unilateral for the first few years.

Other tremors

Dystonic tremor occurs after the age of 50 and is a manifestation of dystonia that is often seen without dystonic posturing. It is generally unilateral (or asymmetric), affecting hand/arm or head, and worse with posture and certain tasks such as drinking, pouring liquids, or using cutlery or a pen. It is often mistaken for parkinsonian tremor. Holmes tremor (present at rest, action, and posture) is seen following lesions of the outflow pathway from the cerebellum to the thalamus, often involving the red nucleus (e.g., infarction, MS).

Dysarthria

Dysarthria is disordered articulation—slurred speech. Language is intact; dysarthria is best considered a disorder of complex motor control rather than of language. There are various patterns of dysarthria which may be caused by lesions affecting any part of the motor control systems or the muscles required for speech. Examples include the ***gravelly 'hot potato'*** speech of pseudobulbar palsy (see 'Bulbar and pseudobulbar palsy' section), the ***jerky 'scanning'*** ataxic speech of cerebellar lesions, the hypophonic ***monotone*** dysarthria of Parkinson's, and speech in myasthenia that ***fatigues*** and dies away. Many aphasic patients are also dysarthric, and may have apraxia of speech (effortful groping towards sounds) and altered prosody (rhythm, melody, and inflection of speech).

LOWER MOTOR NEURON LESIONS

The LMN is the pathway from anterior horn cell (or cranial nerve nucleus) via a peripheral nerve to muscle motor end-plates. The motor unit consists of one anterior horn cell, its single fast-conducting axon that leaves the cord via the anterior root, and the group of muscle fibres (100–2000) supplied via the nerve. Anterior horn cell activity is modulated by impulses from:

- corticospinal tracts
- the extrapyramidal system
- the cerebellum
- sensory afferents via posterior roots.

Clinical features of lower motor neuron lesions

These are seen in voluntary muscles that depend on an intact nerve supply for both contraction and metabolic integrity. Signs follow rapidly if the LMN is interrupted (Box 28.17).

Aetiology

Examples of LMN lesions at various levels include:

- cranial nerve nuclei (e.g., in pontine stroke) and anterior horn cell (MND)
- spinal root—***radiculopathy***, for example cervical and lumbar disc protrusion (see 'Plexus and nerve root lesions' section)
- peripheral (or cranial) nerve—trauma, entrapment (see 'Peripheral nerve lesions' section) and polyneuropathy ('Polyneuropathies [peripheral neuropathy]' section).

Spinal reflex arc

Components are illustrated in Fig. 28.13. The stretch reflex is the physiological basis for all tendon reflexes. In the knee jerk, a tap on the patellar tendon activates stretch receptors in the quadriceps. Impulses in first-order sensory neurons pass directly to LMNs (L3 and L4) that contract quadriceps. Loss of a tendon reflex is caused by a lesion anywhere along the spinal reflex path. The reflex lost indicates its level (Box 28.18).

Reinforcement. Distraction of the patient's attention, clenching teeth, or pulling interlocked fingers enhances reflex activity (Jendrassik manoeuvre). Such reinforcement manoeuvres should be done before a reflex is recorded as absent.

Assessing gait

Walking is a complex activity: sensory inputs from visual, balance, and proprioceptive sensory systems are integrated by each of the three motor control systems described above, and outputs are effected by activation of muscle via LMNs.

Change in walking pattern is a common complaint (Box 28.19). The ***pattern*** of gait is valuable diagnostically, and provides pointers to the site of any neurological lesion.

Spastic gait and hemiparesis

Spasticity (see 'Spastic paraparesis' section), more pronounced in extensor muscles and with or without weakness, causes stiff, effortful, and slow walking. Toes of shoes become scuffed, catching level ground. Pace shortens; a narrow base is maintained. Clonus—involuntary extensor rhythmic leg jerking—may occur.

In a hemiparesis, when spasticity is unilateral and weakness marked, the stiff, weak leg is circumducted and drags.

Parkinson's disease: shuffling gait

Stride length shortens and, in advanced forms, the gait slows to a shuffle. Posture is stooped and arm swing reduced (initially, unilaterally). Gait becomes festinant (hurried) with short, rapid steps. There is difficulty turning quickly (count the number of steps to turn around). Eventually, gait initiation difficulty and freezing episodes (sudden involuntary halts, such as when

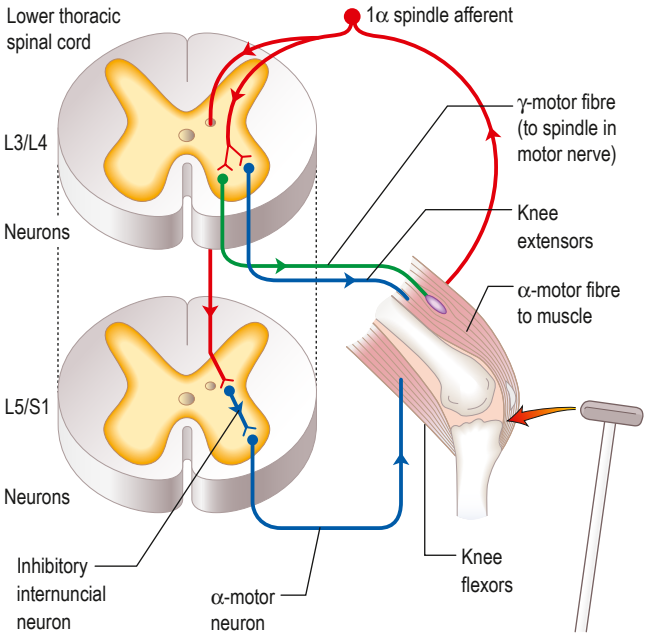


Fig. 28.13 Knee jerk: a spinal reflex arc. Sudden patellar tendon stretching generates sensory action potentials in 1α-muscle spindle afferents that synapse with γ-motor fibres to spindles and α-motor fibres. Motor action potentials cause brisk extensor muscle contraction; there is also inhibition of knee flexors.

Box 28.17 Features of lower motor neuron lesions

- Weakness
- Wasting
- Hypotonia
- Reflex loss
- Fasciculation
- Fibrillation potentials (electromyography)
- Trophic changes in skin and nails

Box 28.18 Spinal levels: recording/interpretation of tendon reflexes

Level	Reflex	Symbol	Meaning
C5–6	Supinator	0	Absent
C5–6	Biceps	±	Present with reinforcement
C7	Triceps	+	Reduced
L3–4	Knee	++	Normal
S1	Ankle	+++	Brisk
		++++	Brisk with clonus

Box 28.19 Common gait abnormalities

Cause	Gait abnormality
Stroke	Spastic/hemiparetic
Parkinson's disease	Shuffling/festinant
Cerebellar ataxia	Broad-based
Sensory neuropathy	Stamping
Position sensory loss	High-stepping
Distal weakness/foot drop	Slapping
Proximal weakness, e.g., myopathy	Waddling
Apraxia of gait, e.g., diffuse cerebrovascular disease	Incoordinated/gait initiation failure; *marche à petits pas*
Normal-pressure hydrocephalus	Hesitant
Arthritis and muscle pain	Antalgic

passing through a doorway) may develop. Falls are uncommon, except in late-stage disease, and may indicate a 'Parkinson's plus' syndrome.

Cerebellar ataxia: broad-based gait

In lateral cerebellar lobe disease (see 'Cerebellar lesions' section), stance becomes broad-based, unstable, and tremulous. Ataxia describes this incoordination. When walking, the person tends to veer to the side of the affected cerebellar lobe.

In disease of midline structures (cerebellar vermis), the trunk becomes unsteady without limb ataxia, and there is a tendency to fall backwards or sideways—truncal ataxia. Both stance and gait are affected.

Sensory ataxia: stamping gait

Peripheral sensory loss (e.g., polyneuropathy; see 'Polyneuropathies [peripheral neuropathy]' section) causes ataxia because of loss of ***proprioception*** (position sense). A broad-based, high-stepping, 'stamping' gait develops as feet are placed clumsily, relying in part on vision, so balance is worse in the dark. Romberg's test, first described in sensory ataxia of tabes dorsalis (see 'Tabes dorsalis' section), becomes positive.

Lower limb weakness: high-stepping and waddling gaits

When weakness is distal, affecting ankle dorsiflexors, such as in a common peroneal nerve palsy (see 'Common peroneal nerve palsy' section), gait becomes high-stepping to avoid tripping. The sole returns to the ground with an audible ***slap***.

Weakness of proximal leg muscles (e.g., polymyositis, muscular dystrophy) causes difficulty rising from sitting. Walking becomes a ***waddle***, the pelvis being poorly supported by each leg.

Gait apraxia

With frontal lobe disease (e.g., diffuse cerebrovascular disease, normal-pressure hydrocephalus), ***walking skills*** become disorganized, despite normal motor and sensory function when examined on the couch. The gait is shuffling with small steps ***(marche à petits pas)***, gait ignition failure and hesitancy with fear of falling. Unlike in the gait of Parkinson's disease, arm swing and posture are normal. Urinary incontinence and dementia are often present.

DISORDERS OF SENSORY PATHWAYS AND PAIN

Peripheral nerves and spinal roots

Peripheral nerves carry all modalities of sensation from either free or specialized nerve endings to dorsal roots and thence to the cord. Sensory distribution of spinal roots (dermatomes) is shown in Fig. 28.14.

Spinal cord

Dorsal columns

Axons in the dorsal columns whose cell bodies are in the ipsilateral gracile and cuneate nuclei in the medulla carry sensory modalities of vibration, joint position (proprioception), light touch, and two-point discrimination. Axons from second-order neurons then cross in the brainstem to form the medial lemniscus, passing to the thalamus (Fig. 28.15).

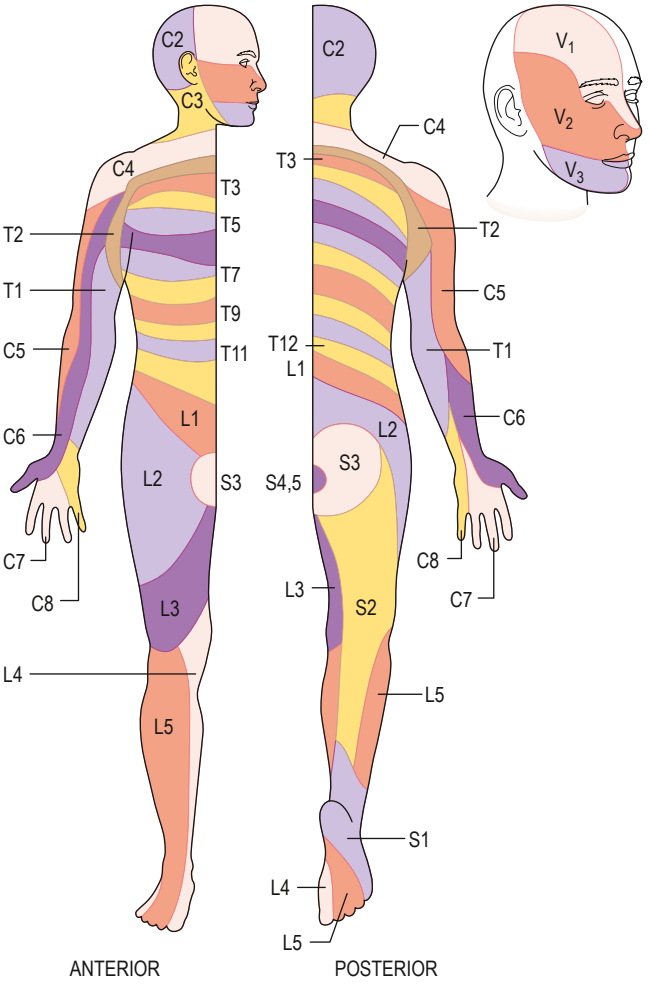


Fig. 28.14 Dermatomes of spinal roots and Vth nerve.

Spinothalamic tracts

Axons carrying pain and temperature sensation synapse in the dorsal horn of the cord, cross within the cord, and pass in the spinothalamic tracts to the thalamus and reticular formation.

Sensory cortex

Fibres from the thalamus pass to the parietal region sensory cortex (see Fig. 28.15). Connections exist between the thalamus, sensory cortex, and motor cortex.

Lesions of the sensory pathways

Altered sensation, tingling (paraesthesia), clumsiness, numbness, and pain are the principal symptoms of sensory lesions. The pattern and distribution point to the site of pathology (Fig. 28.16).

Peripheral nerve lesions

Symptoms are felt within the distribution of a peripheral nerve. Sectioning of a sensory nerve is followed by complete sensory loss. Nerve entrapment (see 'Electromyography and nerve conduction studies' section) causes numbness, pain, and tingling. Tapping the site of compression sometimes causes a sharp, electric shock-like pain in the distribution of the nerve, known as Tinel's sign, such as in carpal tunnel syndrome (CTS; see 'Carpal tunnel syndrome' section).

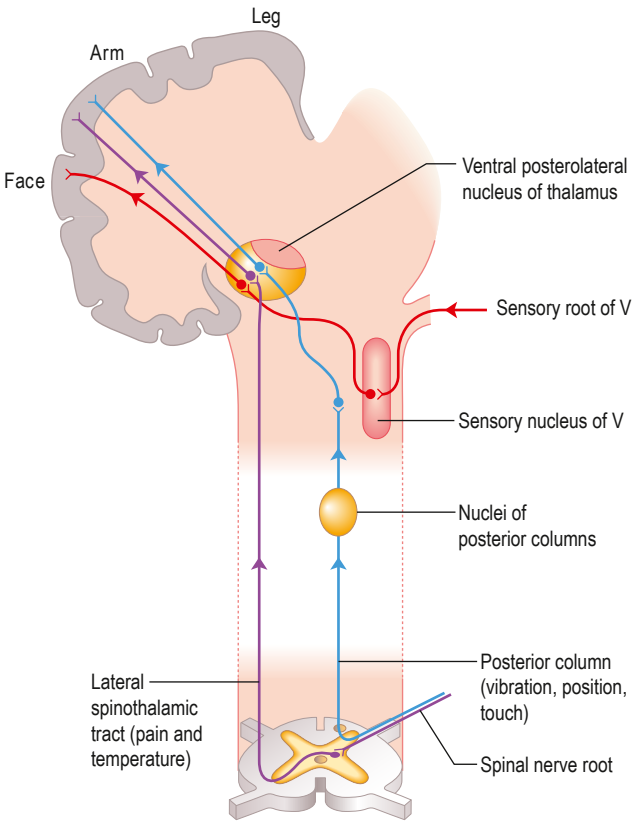


Fig. 28.15 Principal sensory pathways. Posterior columns remain uncrossed until the medulla. Spinothalamic tracts cross close to their entry into the cord.

Neuralgia

Neuralgia refers to pain, usually of great severity, in the distribution of a damaged nerve. Examples are:

- trigeminal neuralgia (see 'Facial pain' section)
- postherpetic neuralgia
- complex regional pain syndrome type II (causalgia)—chronic burning pain that occasionally follows peripheral nerve damage.

Spinal root lesions

Root pain

Pain of root compression is felt in the myotome supplied by the root, and there is also a tingling discomfort in the dermatome. The pain is worsened by manoeuvres that either stretch the root (e.g., straight-leg raising in lumbar disc prolapse) or increase pressure in the spinal subarachnoid space (coughing and straining). Cervical and lumbar disc protrusions (see 'Cervical and lumbar degeneration' section) are common causes of root lesions.

Dorsal spinal root lesions

Sectioning of a dorsal root causes loss of all modalities of sensation within a dermatome (see Fig. 28.14). However, overlap between adjacent dermatomes makes it difficult to detect anaesthesia when a single root is destroyed.

Spinal cord lesions

Dorsal column lesions

These cause:

- tingling
- electric shock-like sensations
- clumsiness
- numbness
- tight band-like sensations
- limb and gait ataxia.

Many lesions will be bilateral but these symptoms, even if lateralized, are often felt vaguely and without a clear sensory level. Position sense, vibration sense, light touch, and two-point discrimination are diminished below the lesion. Position sense loss produces a stamping gait (sensory ataxia; see 'Tabes dorsalis' section).

Lhermitte's phenomenon

Electric shock-like sensations radiate down the trunk and limbs on neck flexion. This points to a cervical cord lesion involving the dorsal columns. Lhermitte's is common in acute exacerbations of MS (see 'Multiple sclerosis' section), and also occurs in cervical myelopathy, subacute combined degeneration of the cord (see 'Vitamin deficiency neuropathies' section), radiation myelopathy, and cord compression.

Spinothalamic tract lesions

Pure spinothalamic spinal lesions cause contralateral loss of pain and temperature sensation with a clear level below the lesion. This is called ***dissociated*** sensory loss—pain and temperature are dissociated from light touch, which remains preserved. This is seen typically in syringomyelia where a cavity occupies the central cord (see 'Syringomyelia and syringobulbia' section).

The spinal level is modified by lamination of fibres within the spinothalamic tracts. Fibres from lower spinal roots lie superficially and are damaged first by compressive lesions from outside the cord. As an external compressive lesion (e.g., a mid-thoracic extradural meningioma; Fig. 28.17) enlarges, the spinal sensory level ascends as deeper fibres become involved. Conversely, a central cord lesion (e.g., a syrinx; see 'Syringomyelia and syringobulbia' section) affects deeper fibres first and the sensory loss may be *suspended*. Spinothalamic tract lesions cause loss of pain and temperature perception (e.g., painless burns). Perforating ulcers and neuropathic (Charcot) joints develop.

Spinal cord compression

Cord compression (see Fig. 28.17) causes progressive spastic paraparesis (or tetraparesis/quadriparesis) with sensory loss below the level of compression. Sphincter disturbance is common. Root pain is frequent but not invariable, felt characteristically at the level of compression. With thoracic cord compression (e.g., an extradural meningioma), pain radiates around the chest and is exacerbated by coughing and straining, as meningeal root sheaths are stretched.

Hemisection of the cord leads to the ***Brown–Séquard syndrome***. The patient complains of numbness on one side and weakness on the other. Damage affects ipsilateral spinothalamic tract (contralateral loss of pain and temperature), dorsal column (ipsilateral loss of vibration and position sense), and corticospinal tract (contralateral weakness and hyperreflexia). Paraparesis/spinal cord lesions are discussed in the 'Spastic paraparesis' section.

Pontine lesions

Since lesions (e.g., an MS plaque) lie above the decussation of the dorsal columns, and both medial lemniscus and spinothalamic tracts are close together, there is loss of all forms of sensation on the side opposite the lesion. Combinations of IIIrd, IVth, Vth, VIth, and VIIth cranial nerve nuclei are seen, and may indicate a level (see Fig. 28.15).

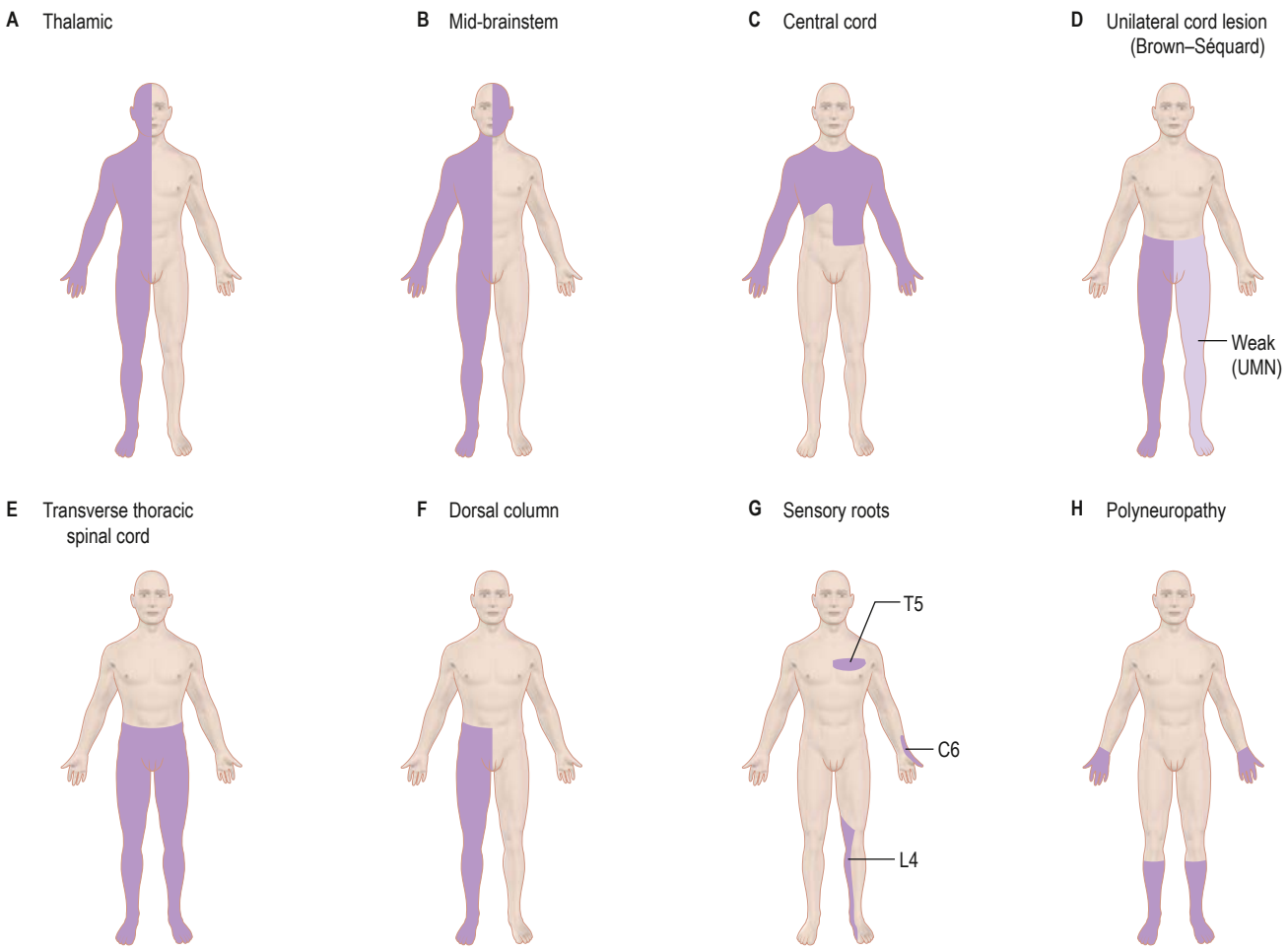


Fig. 28.16 Principal patterns of loss of sensation. (A) *Thalamic lesion:* sensory loss throughout opposite side (rare). **(B)** ***Brainstem lesion:*** contralateral sensory loss below face and ipsilateral loss on face. **(C)** ***Central cord lesion***, for example, syrinx: 'suspended' areas of loss, often asymmetrical and 'dissociated': That is, pain and temperature lost but light touch intact. **(D)** ***Hemisection of cord/ unilateral cord lesion*** —Brown–Séquard syndrome: contralateral spinothalamic (pain and temperature) loss with ipsilateral weakness and dorsal column loss below lesion. **(E)** ***Transverse cord lesion:*** loss of all modalities, including motor, below lesion. **(F)** ***Dorsal column lesion***, for example, multiple sclerosis: loss of proprioception, vibration, and light touch. **(G)** ***Individual sensory root lesions***, for example, C6, T5, L4. **(H)** ***Polyneuropathy:*** distal sensory loss. *UMN,* Upper motor neuron.

Thalamic lesions

Thalamic pain (also called ***central post-stroke pain*** or ***thalamic syndrome***) follows a small thalamic infarct. The patient has a stroke (hemiparesis and sensory loss). Weakness improves but deep-seated constant pain in the paretic limbs develops. Choreoathetotic movements occur. Secondary depression may lead to self-harm. Thalamic lesions can also cause diminished sensation alone, on the opposite side; this is less usual.

Parietal cortex lesions

Sensory loss, neglect of one side, apraxia (see 'Apraxia' section), and subtle disorders of sensation (dysgraphaesthesia, astereognosis) occur. Pain is not a feature of destructive cortical lesions. Irritative phenomena (e.g., focal sensory seizures from a parietal cortex glioma) cause tingling sensations in a limb or elsewhere.

Pain

Pain is an unpleasant, complex sensory and emotional experience. Acute pain serves a biological purpose (e.g., withdrawal) and is typically self-limiting, ceasing as healing ensues. Some forms of chronic pain (e.g., causalgia) outlast the period required for healing and may be permanent.

Essential physiology of pain

Pain perception is mediated by small-diameter myelinated A-delta and non-myelinated C fibres. Chemicals released following injury produce pain either by direct stimulation or by sensitization of nerve endings. A-delta fibres give rise to perception of sharp, immediate pain; slower-onset, more diffuse, and prolonged pain is mediated by slower-conducting C fibres.

Nociceptive (pain) impulses enter the cord via dorsal spinal roots forming synapses in the spinal cord dorsal horn (substance P and glutamate neurotransmitters). Second-order neurons then decussate to the contralateral side and ascend in the spinal cord via two main pathways: the spinothalamic tract (for pain localization) and the spinoreticular tract (mediation of the emotional component of pain; Fig. 28.18). Both pathways project via synapses in the thalamus to the sensory cortex and limbic system. There are projections to brainstem structures: the reticular formation (spinoreticular tract)

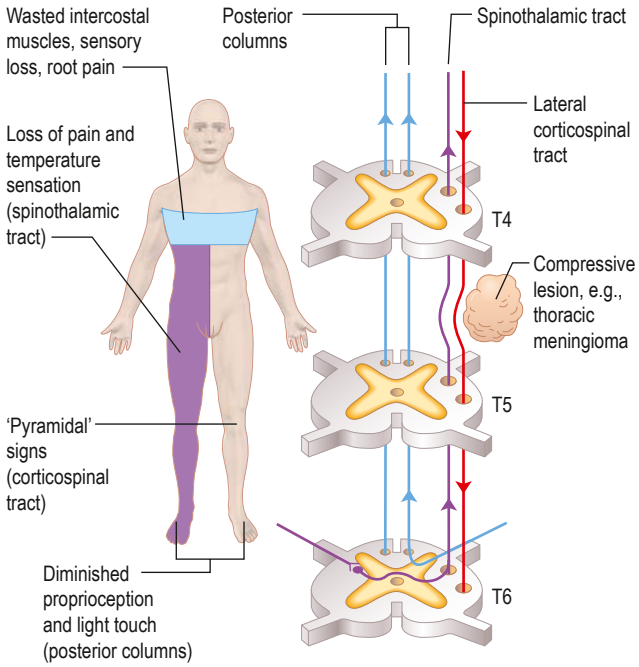


Fig. 28.17 Spinal cord compression: features and anatomy.

and periaqueductal grey (PAG) matter (spinothalamic tract), which modulates pain transmission via descending monoaminergic pathways to the spinal cord dorsal horn. The emotional and affective components of pain, mediated through cortical and limbic networks, are now recognized to be fundamental to pain perception.

Gate theory of pain

Gate theory proposes that transmission of afferent nociceptive (pain) impulses is regulated ('gated') in the spinal cord dorsal horn synapses by afferent impulses from non-nociceptive, large-diameter sensory fibres and descending pathways from the PAG and other brainstem structures. It explains the phenomenon of rubbing a painful area to ease the pain and is also the basis of transcutaneous electrical nerve stimulation (TENS) therapy. Similar gating and modulation of afferent pain signals occur in the brainstem under the influence of connections from the thalamus and limbic system.

Peripheral and central sensitization

Sensitization is an increase in the excitability of peripheral nociceptors and central pain pathways, causing normal sensory inputs such as light touch to produce abnormal responses: that is, pain (termed ***allodynia***). For example, warm water in a shower feels painfully burning on sunburnt skin and touching the scalp is painful during a migraine. This is an adaptive protective response to ensure that contact with an injured area is minimized. However, maladaptive responses to pain via this mechanism at the central or peripheral level may be the basis for chronic pain when it becomes permanent and out of proportion to the initial painful stimulus.

Plasticity and receptor changes

Pain results in neural plasticity changes and reorganization in pain pathways at spinal cord and cortical/brainstem level. These modulate pain and may be maladaptive in some cases, leading to maintenance of chronic pain. Molecular changes also occur in response to painful stimuli, including upregulation of sodium channels and receptor changes.

Types of pain

Nociceptive pain

Nociceptive pain results from physical damage to body tissues, and typically is sharp, aching, or throbbing. Nociceptive nerve fibres are irritated by inflammation, chemicals, or physical trauma. It tends to go away as the body part heals.

Neuropathic pain

Pain results directly from damage to, or dysfunction of, the pain/sensory pathways, such as in peripheral nerve damage, radiculopathy, post-stroke pain, and post-herpetic neuralgia (**Box 28.20**). It is common, affecting 6% to 8% of people. Clinically, the pain usually has an unpleasant burning, electrical, or stabbing quality, with allodynia being a frequent feature. It often becomes chronic due to peripheral and central sensitization processes.

Neurotransmitters and receptors involved in pain

These are involved in pain processing, explaining why a wide range of drug classes are helpful in treating pain.

- ***Excitatory neurotransmitters*** include substance P, glutamate, and calcitonin gene-related peptide (CGRP) in the dorsal horn of the spinal cord and peripherally.
- ***Gamma-aminobutyric acid (GABA)*** is a key inhibitory neurotransmitter.
- ***Opioid receptors***, activated by endogenous endorphins or opiate medications, are widely distributed in pain pathways, including in the spinal cord, peri-aqueductal grey matter, and other brainstem structures.
- ***Noradrenaline (norepinephrine)*** and 5-HT (serotonin) are distributed in descending regulatory pathways.
- ***The vanilloid receptor*** (TRPV1—transient receptor potential cation channel VI) on peripheral nociceptors is activated by capsaicin, the active constituent of chilli peppers.

Management of chronic pain

Chronic pain is gravely disabling, distressing, and difficult to treat. The management of chronic pain is covered in Chapter 50. Multidisciplinary pain clinics provide the best setting for long-term management.

Further reading

Cohen SP, Mao J. Neuropathic pain: mechanisms and their clinical implications. *BMJ* 2014; 348:f7656.
Magrinelli R, Zanette G, Tamburin S. Neuropathic pain: diagnosis and treatment. *Pract Neurol* 2013; 13:292–307.
O'Brien M (ed). *Aids to The Examination of The Peripheral Nervous System*. Amsterdam: Elsevier; 2022.

BLADDER CONTROL AND SEXUAL DYSFUNCTION

Changes in micturition and failure of normal sexual activity due to neurological conditions are seen in sacral, spinal cord, and cortical disease.

Essential functions and anatomy

The bladder has two functions: storage and voiding. Afferent pathways (T12–S4) respond to pressure within the bladder and sensation from the genitalia. As the bladder distends, continence is

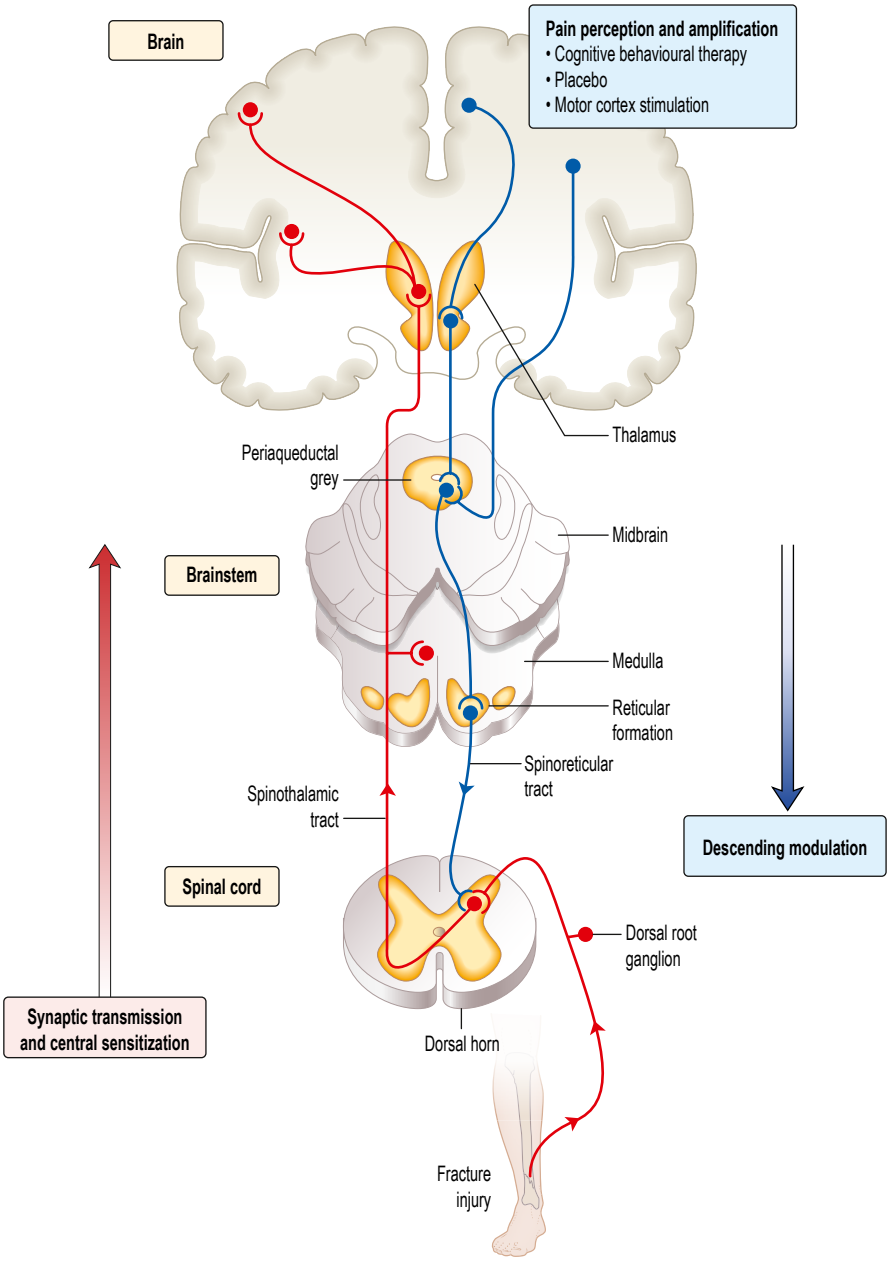


Fig. 28.18 **The spinothalamic and spinoreticular tracts in pain perception.**

maintained by suppression of parasympathetic outflow and reciprocal activation of sympathetic outflow. Both are under some voluntary control. Voiding takes place by parasympathetic activation of the detrusor, and relaxation of the internal sphincter (Box 28.21).

Cortical awareness of bladder fullness is located in the post-central gyrus, parasagittally, while initiation of micturition is in the pre-central gyrus. Voluntary control of micturition is located in the frontal cortex, parasagittally.

Neurological disorders of micturition

Urogenital tract disease is dealt with largely by urologists. Incontinence is common and easy to recognize; neurological causes are sometimes not obvious. These are:

- ***Cortical:***
  - Post-central lesions cause loss of the sense of bladder fullness.
  - Pre-central lesions cause difficulty in initiating micturition and an uninhibited bladder.
  - Frontal lesions cause socially inappropriate micturition.
- ***Spinal cord***. Bilateral UMN lesions (pyramidal tracts) cause urinary frequency and incontinence. The bladder is small and hypertonic: that is, sensitive to small changes in intravesical pressure. Detrusor-sphincter dyssynergia results in urinary retention that can cause hydronephrosis and acute or chronic kidney injury. Frontal lesions can also cause a hypertonic bladder.
- ***LMN***. Sacral lesions (conus medullaris, sacral root, and pelvic nerve—*bilateral*) and acute UMN lesions (spinal shock) cause a flaccid, atonic bladder, with retention and overflow incontinence (cauda equina; see 'Spinal stenosis' section).

Management. Assessment of both urological causes (e.g., calculi, prostatism, gynaecological problems) and potential neurological causes of incontinence is necessary. Intermittent

Box 28.20 Causes of neuropathic pain

Peripheral nerve injury
- Peripheral neuropathy—especially if small (pain) fibres are affected, e.g., diabetes, AL amyloidosis, HIV, chemotherapy (e.g., carboplatin neuropathy), vitamin deficiency, alcohol, Fabry's
- Vasculitic nerve lesions—for example, diabetic amyotrophy, vasculitic mononeuritis multiplex
- Nerve compression—carpal tunnel syndrome, meralgia paraesthetica
- Postherpetic neuralgia
- Nerve trauma or section—for example, phantom limb pain

Nerve root and plexus
- Trigeminal neuralgia
- Radiculopathy due to intervertebral disc prolapse
- Plexopathy—for example, inflammatory (brachial neuritis) or neoplastic infiltrative (e.g., breast cancer)

Central nervous system disorders
- Multiple sclerosis (pain is common in progressive forms)
- Post-stroke pain—for example, thalamic strokes
- Spinal cord lesions—intrinsic cord lesions, inflammatory or vascular (infarcts), especially
- Parkinson's disease

Box 28.21 Efferents to the bladder and genitalia

Nerve supply	Function
Parasympathetic S2–4	Bladder wall: contraction Internal sphincter: relaxation Penis/clitoris: engorgement
Sympathetic T12–L2	Bladder wall: relaxation Internal sphincter: contraction Orgasm, ejaculation
Pudendal nerves	External sphincter (skeletal muscle)

self-catheterization is used by many patients who have, for example, spinal cord lesions.

Male erectile dysfunction

Failure of penile erection often has mixed organic and psychological causes. Depression is common. Erectile dysfunction is described in Chapter 47 ('Loss of libido and erectile dysfunction' section).

INVESTIGATION OF NEUROLOGICAL DISEASE

Neurological investigations, such as brain scans, should not be a substitute for clinical evaluation through history-taking and examination. An experienced clinician will be able to make an accurate diagnosis in most patients without recourse to investigations: for example, in most individuals presenting with headache. However, cross-sectional imaging will allow confirmation of the anatomical localization of lesions and may reveal the presence of other clinically 'silent' lesions. Imaging does not always indicate the pathological nature of a visualized lesion (careful history-taking is essential in this regard) and therefore needs to be carefully interpreted in the clinical context. Neurophysiological investigations provide information about nervous system *function* and are more diagnostically helpful than imaging in some situations: for example, in patients with primary generalized epilepsies or peripheral nerve disorders.

Unnecessary or inappropriate use of investigations, particularly brain imaging, exposes patients to risk, such as radiation; is an inefficient use of resources; and may lead to identification of incidental findings that cause considerable anxiety (there is a 3% incidental abnormality rate for brain MRI scans). Clinicians should also be familiar with the range of normal in the population; for example, the high prevalence of white-matter hyperintensities on brain MRI in older people and the range of common non-specific electroencephalogram (EEG) changes.

Blood tests

Standard blood tests, such as full blood count, erythrocyte sedimentation rate (ESR), biochemistry, glucose/HbA_{1c}, C-reactive protein, creatine kinase, and protein electrophoresis, are helpful to identify underlying disorders such as diabetes, metabolic disorders, and other systemic disorders that may underlie neurological presentations (e.g., coma or peripheral neuropathy). Relevant tests are covered in disease-specific subsections.

Genetic tests

Several hundred neurogenetic disorders are now recognized and, for many of these, specific molecular genetic investigations are available for diagnostic and predictive testing (see Chapter 5). As genetic testing is complex and expensive, it should only be undertaken after senior specialist advice. Increasingly, there is a move to whole genome sequencing, usually examining virtual panels relevant to a particular phenotype. Informed patient consent is required, as a positive test may have implications for other family members. Predictive testing in non-affected, at-risk individuals should only be undertaken after detailed genetic counselling by a specialist.

Immunological tests

An increasing number of autoimmune neurological disorders are recognized, many of which are associated with specific antibody tests (see **Box 28.22**). Many of these are paraneoplastic syndromes but an increasing number of non-cancer-related autoimmune syndromes have been identified, such as forms of encephalitis or MG. In addition, monoclonal paraproteins associated with plasma cell dyscrasias may cause immune-mediated peripheral neuropathy.

Neuroimaging

See Chapter 8. Central nervous system imaging in stroke is covered in Chapter 29.

Lumbar puncture

This is explained in Chapter 11.

Neurophysiological investigations

Electroencephalography

Brain electrical activity (**Fig. 28.19**) is recorded from scalp electrodes (20 channels simultaneously). Its main use is to characterize epilepsy syndromes. Recordings are usually made between seizures (inter-ictal). Sleep-deprived EEG and 24-hour ambulatory EEG increase diagnostic sensitivity. Videotelemetry (VT) combines continuous EEG with video over several days to record the semiology and electrical characteristics of seizures. VT is invaluable diagnostically to identify seizure type where there is clinical uncertainty, and to pinpoint the seizure focus in focal epilepsies.

Epilepsy

Spikes, or spike-and-wave abnormalities, are hallmarks of epilepsy, but it is important to emphasize that patients with epilepsy often have a normal EEG between seizures (see 'Epilepsy and loss of consciousness' section).

Diffuse brain disorders

Slow-wave EEG abnormalities appear in encephalopathy (e.g., hepatic coma), encephalitis, and prion (Creutzfeldt–Jakob) diseases.

Brain death

The EEG is isoelectric (flat); EEG is not necessary to confirm brain death in many countries (see Chapter 13, 'Death by neurological criteria' section).

Box 28.22 Neurological autoantibody tests

Clinical disorder	Antibody (Ab) test
Paraneoplastic disorders: cerebellar ataxia, limbic encephalitis, opsoclonus–myoclonus, progressive encephalomyelitis with rigidity and myoclonus	Antineuronal antibodies including anti-Hu, Yo, Ma2/Ma1, CV2, Ri, Tr, amphiphysin
Teratoma-associated limbic encephalitis	NMDA Abs
Limbic encephalitis (non-paraneoplastic)	Anti-potassium channel, glycine Abs
Myasthenia gravis[a]	Acetylcholine receptor, MUSK Abs
Lambert–Eaton's myasthenic syndrome[a]	Voltage-gated calcium channel Abs
Neuromyelitis optica	Aquaporin 4 Abs
Stiff person syndrome	Anti-GAD Abs
Transverse myelitis	Anti-MOG Abs
Peripheral neuropathy	Anti-ganglioside and MAG Abs
Hashimoto's encephalopathy	Anti-TPO Abs
Dermatomyositis/ polymyositis[a]	Anti-Jo-1 and other tRNA synthetases, SRP, Mi-2, and other Abs

[a]May also be paraneoplastic. *Abs,* antibodies.

Electromyography and nerve conduction studies

Electromyography (EMG) and nerve conduction studies (NCS) are skilled investigations of nerve function that are essential for investigation of peripheral nerve, anterior horn cell, NMJ, and to a lesser extent, muscle disorders. NCS are highly sensitive in identification of large fibre peripheral neuropathy, and assessment of the nerves affected and the type of pathology (axonal loss versus demyelinating). In mononeuropathies, the anatomical site of the lesion may be identified with considerable accuracy and prognostic information obtained.

Electromyography

A concentric needle electrode is inserted into voluntary muscle. Amplified recordings, on an oscilloscope, are also heard through a speaker. The main EMG features include:

- normal interference pattern
- denervation and re-innervation changes
- myopathic and myotonic features (see 'Myotonic dystrophy [DM]' section).

Single-fibre EMG is used for disorders of NMJ transmission.

Peripheral nerve conduction

Electrical stimulation of a peripheral nerve generates a nerve impulse that may be recorded (Fig. 28.20). Four measurements are of principal value in neuropathies and nerve entrapment:

- nerve conduction velocity
- distal motor latency
- sensory action potentials (SAPs)
- compound muscle action potentials (CMAPs).

Small action potential size (SAP and CMAP) indicates loss of axons. Conduction block and slowing of conduction velocity indicate demyelinating neuropathy, as saltatory conduction is lost with damage to the myelin sheath. Repetitive nerve stimulation is used to identify NMJ conduction disorders (myasthenia).

Cerebral-evoked potentials

Visual-evoked potentials (VEPs) record the interval visual stimuli take to reach the occipital cortex, and the amplitude of response. VEPs are used to confirm previous optic neuritis (see 'Optic neuritis'

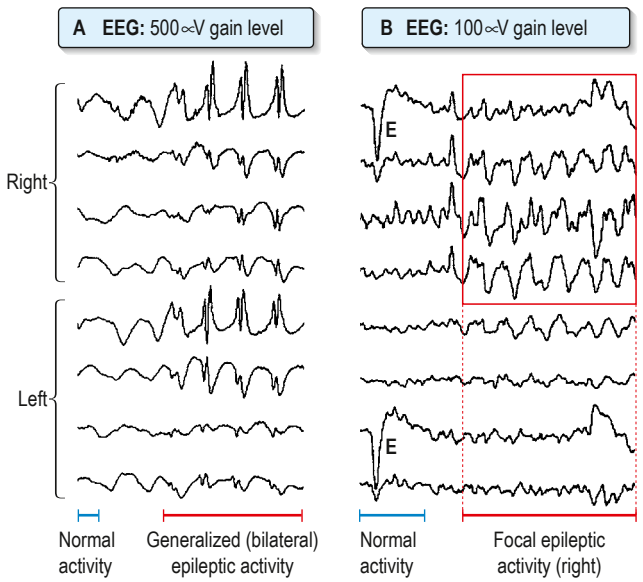


Fig. 28.19 Examples of EEG traces. **(A)** Normal activity followed by generalized epileptic spikes in all leads. **(B)** Normal activity followed by focal spikes in right leads. *E,* Eye-movement artefact.

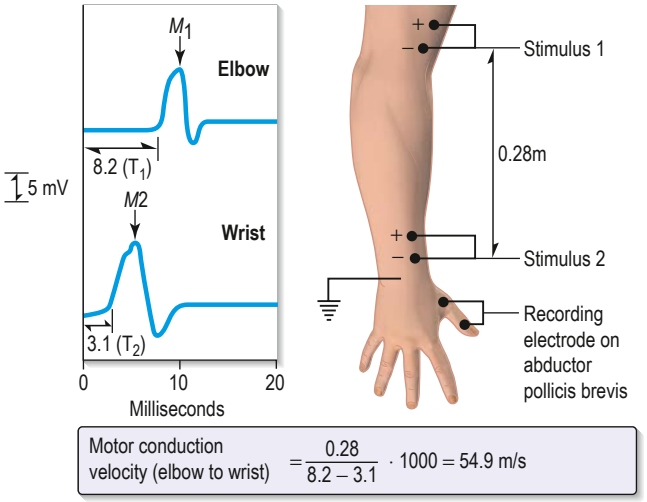


Fig. 28.20 Measurement of motor conduction velocity (MCV) in ulnar nerve. The recording electrode on abductor pollicis brevis measures the muscle action potential (M) from ulnar nerve stimulation at the elbow (stimulus 1) and wrist (stimulus 2). MCV calculation:
M_1 = muscle action potential (MAP) from stimulus 1.
M_2 = MAP from stimulus 2.
T_1 = time from elbow to recording electrode.
T_2 = time from wrist to recording electrode.

$$MCV = \frac{\text{distance (metres)}}{T_1 - T_2} \times 1000 = m/s$$

section in the 'Neuroinflammatory disorders' section); this leaves a delayed latency despite clinical recovery.

Similar techniques for auditory and somatosensory potentials (from a limb) are also used.

Biopsy

Interpretation of brain, tonsillar, muscle, and nerve histology requires specialist neuropathology services.

Brain and meninges

Brain biopsy (e.g., of a non-dominant frontal lobe) is rarely used to diagnose inflammatory and degenerative brain diseases. CT and MR stereotactic biopsies of intracranial lesions are standard procedures.

Muscle

Muscle biopsy (usually quadriceps), with light/electron microscopy and immunohistochemical analysis, is performed routinely for diagnosis of inflammatory, metabolic, and dystrophic disorders (see 'Muscle diseases' section).

Peripheral nerve

Nerve biopsy, usually of the (sensory) sural nerve or superficial sensory branch of the radial nerve, can be of value in the diagnosis of peripheral neuropathy: for example, if vasculitis, amyloid deposition, or malignant infiltration is suspected. There may be significant complications (wound healing and pain), so it should be undertaken only where all other investigations are non-diagnostic and where there is clinical evidence of a severe or progressive neuropathy.

Psychometric assessment

Psychometric testing, performed by neuropsychologists, assesses cognitive function in more detail than bedside testing during clinical examination. Decline in intellectual function, as compared with estimated pre-morbid IQ, is seen following brain injury or in dementia, for example. Individual cognitive domains, such as various types of memory, language, visuospatial function, and frontal executive function, are tested separately, as they may be differentially involved in different disorders, the pattern of involvement often being of considerable diagnostic value. Serial testing at intervals of 6 to 12 months may be needed to show clear evidence of progression in neurodegenerative disorders such as dementias.

Further reading

Morris HR, Houlden H, Polke J. Whole-genome sequencing. *Pract Neurol* 2021; 21:322–327.
Whittaker RG. SNAPs, CMAPs and F-waves: nerve conduction studies for the uninitiated. *Pract Neurol* 2012; 12:108–115.

NEUROLOGICAL CONDITIONS

Neurological conditions covered elsewhere in the book include:

- Stroke and intracerebral haemorrhage—see Chapter 29
- Coma and reduced consciousness—see Chapter 11
- Status epilepticus and acute symptomatic seizures—see Chapter 11
- Evaluation of acute headache, including intracranial haemorrhage and meningitis—see Chapter 11
- Functional neurological disorders—see Chapters 11 and 30

HEADACHE, MIGRAINE, AND FACIAL PAIN

Headache is an almost universal experience and one of the most common symptoms in medical practice. It varies from an infrequent and trivial nuisance to a pointer to serious disease. Headache symptoms are unpleasant, disabling, and common worldwide, and have a substantial economic impact because of time lost from work. The approach to acute headaches is discussed further in Chapter 11.

Mechanisms

Pain receptors are located at the base of the brain in arteries and veins, and throughout meninges, extracranial vessels, scalp, neck and facial muscles, paranasal sinuses, eyes, and teeth. Curiously, brain substance is almost devoid of pain receptors. Head pain is mediated by the Vth and IXth cranial nerves and upper cervical sensory roots.

Clinical approach to the patient with headache

In assessing patients with headache, the aim should be to make a confident diagnosis based on the history. Examination is helpful in excluding underlying medical disorders as a cause of headache, but will not distinguish between different types of primary headache.

There is an internationally agreed classification for headaches that defines all headache patterns. Headache is divided into:

- primary headache disorders such as migraine,
- secondary headaches due to underlying pathology such as raised ICP or meningitis (**Box 28.23**).

It is also useful to distinguish between episodic (recurrent) headache, single first headache episodes, and patients with chronic headache.

Box 28.23 Some causes of secondary headache

- Raised intracranial pressure, e.g., idiopathic intracranial hypertension, mass lesions
- Cerebral venous sinus thrombosis
- Cervical artery dissection
- Infections, e.g., meningitis, acute sinusitis
- Giant cell arteritis
- Intracranial haemorrhage, especially subarachnoid haemorrhage or subdural haematoma
- Low cerebrospinal fluid volume (low-pressure) headache
- Hypoxia or hypercapnia (obstructive sleep apnoea, nocturnal hypoventilation, carbon monoxide poisoning)
- Post-traumatic headache
- Cervicogenic headache
- Reversible cerebral vasoconstriction syndrome
- Cerebral vasculitis
- Acute glaucoma

In an outpatient clinic setting, most headaches will be benign. Fewer than 1% of outpatients with non-acute headache have a serious underlying cause but in the emergency department there will be a much higher prevalence of serious underlying pathology presenting with headache. New-onset severe headache in those without a previous headache history, especially in older patients (>50), requires exclusion of underlying pathology causing secondary headache.

There are some widely believed 'headache myths'. Headaches are not caused by hypertension, except rarely with malignant hypertension (see Chapter 18). Eye strain from refractive error does not cause headache and sinusitis is rarely the explanation for recurrent or chronic headache.

Taking a history for headaches

Ask about:

- Headache location (e.g., hemicranial), severity, and character (e.g., throbbing versus non-throbbing)
- Associated symptoms, e.g., nausea, photophobia, phonophobia, and motion sensitivity
- Presence of autonomic symptoms, e.g., tearing, conjunctival injection, ptosis, miosis, nasal congestion or rhinorrhoea, eyelid swelling, aural fullness, facial sweating, or swelling
- Relieving or exacerbating features, e.g., effect of posture
- Headache pattern: is headache episodic and part of a pattern of previous similar headaches? Age at onset and headache frequency
- Duration of headache episodes (helpful in distinguishing between different primary headache types)
- Triggers
- Pattern of analgesic use
- Family history of headache
- 'Red flag' symptoms (**Box 28.24**): fever (meningitis, sinusitis)
- Sudden onset, reaching maximum intensity in less than 1 minute (SAH)
- Features of raised ICP
- Jaw claudication (GCA).

Examination

Examination should include fundoscopy to look for papilloedema. In older patients, temporal arteries should be palpated for loss of pulsatility and tenderness that may be features of GCA. Fever and neck stiffness suggest meningitis. Examination is generally normal in patients with primary headache disorders.

Box 28.24 Red flag features that may indicate a secondary cause of headache

Older age (onset after age 50)
Systemic symptoms including fever, weight loss
Neurological deficits, change in cognition or personality, change in conscious level
Atypical aura (including weakness)
New onset seizures
New or recent onset headache with no prior history of similar headaches
Thunderclap onset
History of cancer, other significant systemic disease, or immunosuppression (including HIV)
Papilloedema
Valsalva headache (precipitated by coughing, exercise, sex)
Orthostatic changes in headache (worse either lying down or upright)
New headaches in pregnancy or puerperium
Painful or red eye; visual loss (other than classical migraine aura)
History of recent head injury
Substantial change in nature or pattern of headache

Investigations

Investigations, including brain imaging, do not contribute to the diagnosis of primary headache disorders. Neuroimaging is indicated only where history or examination suggests the possibility of an underlying secondary cause. Older patients with new-onset headache and those with 'red flag' symptoms (**Box 28.24**) should have brain imaging. In patients over 50 with new headache, inflammatory markers should be checked to exclude GCA.

In practice however brain imaging is frequently performed for 'reassurance'. If this is done, it is important to warn the patient of the 3% risk of incidental findings.

Primary headache disorders

Migraine

Migraine is the most common cause of episodic disabling headache (affecting 15%–20% of women and 5%–10% of men); in 90%, onset is before 40 years of age. Episodes of headache are associated with sensory sensitivity such as to light, sound, or movement, and sometimes with nausea and vomiting. There is a spectrum of severity between individuals and from one attack to another. Migraine is usually high-impact, with inability to function normally during episodes. Headache frequency in migraine varies from an occasional inconvenience to frequent headaches severely impacting on quality of life, and may transform into chronic daily headache.

Mechanisms

Genetic factors play a part in causing the neuronal hyperexcitability that is probably the biological basis of migraine. Migraine risk is usually polygenic but rare cases of familial hemiplegic migraine are channelopathies, associated with variants in genes encoding ion channels.

The pathophysiology of migraine is now thought to have a primarily neurogenic rather than vascular basis. Spreading cortical depression—a wave of neuronal depolarization followed by depressed activity slowly spreading anteriorly across the cerebral cortex from the occipital region—is thought to be the basis of the migraine aura. Activation of trigeminal pain neurons is the basis of the headache. The innervation of the large intracranial vessels and dura by the first division of the trigeminal nerve is known as the trigeminovascular system. Release of CGRP, substance P, and other vasoactive peptides, including 5-HT, by

activated trigeminovascular neurons causes painful meningeal inflammation and vasodilation. Peripheral and central sensitization of trigeminal neurons and brainstem pain pathways makes otherwise innocuous sensory stimuli (such as CSF pulsation and head movement) painful, and light and sound are perceived as uncomfortable.

Clinical features

Migraine without aura

Migraine typically starts around puberty with increasing prevalence into the fourth decade. There is a spectrum of severity and associated features, but attacks have recognizable core features (**Box 28.25**). Most migraine attacks are of sufficient severity to prevent sufferers continuing with normal activities; sleep usually helps. A 'washed-out' feeling follows the attack. The scalp may be tender to touch during episodes (allodynia) and the preference is to be still in a dark, quiet environment.

Some patients recognize changes in routine as trigger factors:

- Sleep (too little or too much).
- Stress (including letdown after a period of stress).
- Hormonal factors for women—changes in oestrogen levels, for example menstrual migraine (usually just before menses) and worsening with the oral contraceptive pill and menopause.
- Eating—skipping meals and alcohol. Contrary to popular belief, individual foods are rarely a trigger.
- Other—sensory stimuli, such as bright lights or loud sounds; physical exertion; and changes in weather patterns, for example stormy weather. Minor head injuries may trigger a worsening of migraine frequency and severity.

Migraine with aura

Approximately 25% of migraine sufferers experience focal neurological symptoms immediately preceding the headache phase in some or all attacks; this is termed migraine aura. Most never experience aura, and the presence of aura is therefore not required for a diagnosis of migraine. Aura usually evolves over 5 to 20 minutes, with symptoms changing as the wave of spreading neuronal depression moves across the surface of the cortex. It rarely lasts longer than 60 minutes and is followed immediately by the headache phase.

Visual aura is the most common type, with positive visual symptoms such as shimmering, teichopsia (zigzag lines, also called fortification spectra), and fragmentation of the image (like looking through a pane of broken glass) often accompanied by patches of loss of vision, which may move across the visual field (scotomas) and even evolve into hemianopia or tunnel vision. Positive sensory symptoms (mainly tingling), dysphasia and, rarely, loss of motor function may also occur, and may occur successively within the same episode of aura following the visual symptoms.

Box 28.25 Migraine: simplified diagnostic criteria

Headache lasting 4 hours to 3 days (untreated)

At least two of:

- Unilateral pain (may become holocranial later in attack)
- Throbbing-type pain
- Moderate to severe intensity
- Motion sensitivity (headache made worse with head movement or physical activity)

At least one of:

- Nausea/vomiting
- Photophobia/phonophobia
- Normal examination and no other cause of headache

Migraine aura usually presents no diagnostic difficulty, but problems with diagnosis may sometimes arise in men over the age of 50 who develop migraine aura for the first time without subsequent headache (sometimes referred to as acephalalgic migrainous aura). This is frequently misdiagnosed as a TIA (see Chapter 29). Distinguishing the two conditions is often difficult and relies on the characteristic evolution of symptoms over minutes and the presence of positive symptoms in aura, in contrast to TIA where symptom onset is acute and negative symptoms (visual loss as opposed to visual distortion and teichopsia) are the norm. There may also be a history of previous typical migraine aura in early adult life to help distinguish the conditions.

Migraine-related dizziness

Vertigo is now recognized as being a migrainous symptom in some individuals, with attacks lasting for hours in the context of migraine attacks. There is an overlap with what is sometimes described as basilar migraine, a poorly defined migraine subtype associated with brainstem aura-type symptoms before or during attacks, including perioral paraesthesiae, diplopia, unsteadiness and, rarely, reduced level of consciousness.

Familial hemiplegic migraine

This rare autosomal dominant disorder, a brain channelopathy, causes a hemiparesis and/or coma and headache, with recovery within 24 hours. Some patients have permanent cerebellar signs, as some forms are allelic with episodic ataxia. It is distinct from more common forms of migraine. Causative genes have been identified that cause a proportion of familial hemiplegic migraine cases, and include the α-1 subunit of the P/Q-type voltage-gated calcium channel *(CACNA1A)*, the α-2 subunit of the sodium/potassium-transporting ATPase *(ATP1A2)* and the neuronal sodium channel *(SCN1A)*.

Hemiplegic migraine (or migraine with motor aura) can also occur in a sporadic rather than familial fashion, although most of those affected will not have an identifiable brain channelopathy. The presence of new focal weakness or other atypical aura symptoms with a migraine should prompt brain imaging.

Management

General measures include:

- explanation
- avoidance of trigger factors and lifestyle modification (e.g., the importance of maintaining a regular routine, good sleep hygiene, regular meals, adequate hydration, stress management, and exercise).

Acute treatment of attacks

Analgesics, such as paracetamol 1 g, high-dose dispersible aspirin (900 mg), or an NSAID (e.g., naproxen 250–500 mg), are first-line treatment, with an antiemetic such as metoclopramide. The combined use of aspirin, paracetamol, and an antiemetic is often more effective than single agents. Antiemetics are useful even when there is no substantial nausea. Acute treatment should be taken as soon as possible after onset of headache. Patients should be aware that repeated use of analgesics leads to further headaches, and use of acute headache treatments should be limited to fewer than 12 days per month (see medication overuse headache, see 'Chronic daily headache' section).

Triptans (5-$HT_{1B/1D}$ agonists) are specific for migraine and may be effective where simple analgesics are insufficient. Sumatriptan was the first marketed; almotriptan, eletriptan, frovatriptan,

naratriptan, rizatriptan, and zolmitriptan are now available, with various routes of administration. Subcutaneous sumatriptan injection may be effective where vomiting prevents absorption of oral medication (note that nasal triptan sprays rely on gastrointestinal rather than nasal absorption). Triptans should be avoided when there is vascular disease, and like analgesics, they should not be overused.

CGRP modulates trigeminovascular pain transmission in migraine. Small molecule CGRP antagonists, known as gepants, have been shown to be effective abortive therapy for migraine. Availability of these drugs (e.g., rimegepant) is still limited but likely to increase. They may have a role in migraine prevention too, as initial data suggests that they do not cause medication overuse or rebound headache.

Status migrainosus

Severe, intractable migraine attacks lasting for longer than 72 hours, are known as status migrainosus. This is a reasonably common problem in the emergency department, and treatment options are based on limited evidence. Commonly tried treatments include intravenous fluids and dopamine receptor blockers (e.g., intravenous metoclopramide 10 mg or intramuscular prochlorperazine 12.5 mg; caution is required in young people who may develop acute dystonic reactions). Subcutaneous sumatriptan 3 to 6 mg could be given if a triptan has not already been tried for that migraine episode. Intravenous ketorolac 15 to 30 mg (an NSAID) is another option if available, and intravenous valproate 500 to 1000 mg can also be given (as a one-off dose). Parenteral corticosteroids may sometimes be used with the aim of preventing early relapse.

Migraine prevention

When migraine episodes are frequent (>4 per month, for example) and impacting on quality of life, migraine preventers should be offered. The key principles are that a period of 3 months' treatment, at an adequate dose, is usually necessary to judge efficacy, and up to 50% of patients will achieve a 50% reduction in headache frequency and severity after this time. These medications should be used in conjunction with a headache diary. Withdrawal of the medication should be attempted after 6 months because for some people the migraine frequency may be 'reset' beyond the treatment period, but the drug may need to be reintroduced if this fails. However, these medications will not be effective where ongoing analgesic overuse is an issue.

Treatment options for migraine prevention include:

- ***Tricyclics***, such as amitriptyline 10 mg, increasing weekly in 10 mg steps to 50 to 60 mg.
- ***Beta-blockers***, such as propranolol slow-release 80 to 160 mg daily.
- ***Anticonvulsants.*** Valproate (800 mg) or topiramate (100–200 mg daily) are generally the most effective options. For women of childbearing age, valproate should be avoided, and topiramate (when given for migraine) should be used with effective contraception.
- ***Candesartan.*** Up to 16 mg daily.
- ***Antidepressants***, such as venlafaxine up to 150 mg daily.
- ***Nonpharmacological treatments.*** Riboflavin 400 mg (vitamin B_2) daily has some evidence for use in migraine prevention, as does acupuncture.
- ***Botulinum toxin.*** This was recommended as a treatment for chronic migraine (see Chapter 42, 'Botulism' section). The technique involves 31 injections over the scalp and neck repeated every 3 months. Its effect is inconsistent.
- ***Greater occipital nerve blocks*** can be used for chronic migraine, occasionally on a regular basis but more commonly as a temporizing measure to allow titration of a migraine preventer medication and/or analgesia withdrawal. They are also useful in pregnancy.
- ***Pizotifen.*** This is rarely used. Flunarizine (a calcium channel blocker) and methysergide are used in refractory patients.
- ***CGRP antagonists.*** Monoclonal antibodies directed against the CGRP ligand or receptor (e.g., erenumab, fremanezumab, galcanezumab, and eptinezumab) have proven efficacy for reduction of migraine severity and frequency, but are very costly and are reserved for people who have failed to respond to multiple migraine preventers.

Tension-type headache

The exact pathogenesis of tension-type headache (TTH) remains unclear. There is overlap with migraine, and many headaches traditionally subsumed under this category probably, in fact, represent mild migraine. Since there are no diagnostic tests to separate TTH from mild migraine, it is difficult to know if the conditions are biologically distinct. In contrast to migraine, pain is usually of mild to moderate severity, bilateral and relatively featureless, with tight band sensations, pressure behind the eyes, and bursting sensations being described.

Depression is also a frequent co-morbid feature. TTH is often attributed to cervical spondylosis, refractive errors, or high blood pressure; evidence for such associations is poor.

Simple analgesics are often effective but overuse should be avoided. Physical treatments, such as massage, ice packs, and relaxation, are often recommended. Frequent or chronic TTH may respond to migraine suppression medications as above, with tricyclics often being used first line.

Trigeminal autonomic cephalalgias

The trigeminal autonomic cephalalgias are a group of primary headache disorders characterized by strictly unilateral trigeminal distribution pain (usually in the ophthalmic division of the nerve) and prominent ipsilateral autonomic features. Bilateral autonomic features are also common in migraine, but these rarer disorders should be suspected when headaches and autonomic symptoms are side-locked, always occurring and remaining on the same side of the head. Neuroimaging is mandatory for these patients, as some will have an underlying structural cause (particularly in the pituitary fossa).

Cluster headache

Cluster headache is distinct from migraine and much rarer (1 per 1000). It affects adults, mostly males aged between 20 and 40. Patients describe recurrent bouts (clusters) of excruciating unilateral retro-orbital pain with parasympathetic autonomic activation in the same eye, causing redness or tearing of the eye, nasal congestion, or even a transient Horner's syndrome. The pain is reputed to be the worst known to humans, and patients often contemplate, and sometimes commit, suicide, such is the severity of the pain. Unlike with migraine attacks, patients prefer to move about or rock rather than remain still.

Attacks are shorter than migraine, usually 30 to 90 minutes, and may occur several times per day. There is often a circadian pattern with headaches occurring at the same time of day, especially in sleep. Clusters last 1 to 2 months, with attacks most nights, before stopping completely and typically recurring a year or more later, often at the same time of year. Attacks are side-locked but they

can sometimes switch sides between 1 year and the next. Some patients unfortunately go on to develop chronic cluster headache without periodic remissions. Although the cause is not known, hypothalamic activation is seen on functional imaging studies during an attack.

Management. Analgesics are unhelpful. Subcutaneous sumatriptan is the drug of choice for acute treatment, as no other drug works quickly enough. High-flow oxygen is also used. Most prophylactic migraine drugs are unhelpful. Verapamil, lithium, and/or a short course of steroids help terminate a bout of cluster headaches.

Paroxysmal hemicrania, SUNCT and hemicrania continua

Paroxysmal hemicrania is a rare condition with similarities to cluster headache, differing mainly in that attacks are briefer (2–30 minutes) and more frequent (>5 per day, at any time of day) and do not occur in clusters. Women are more often affected than men. There is a rapid and complete response to indometacin.

Short-lasting unilateral neuralgiform headache with conjunctival injection and tearing (SUNCT) is very rare. Attacks are short, 5 seconds to 2 minutes, and very frequently occur in bouts. Distinguishing SUNCT from trigeminal neuralgia can be difficult, but there is no refractory period after a SUNCT attack and classically the trigeminal autonomic cephalalgias will arise in the distribution of the ophthalmic division of the trigeminal nerve (V_1) while trigeminal neuralgia typically affects V_2–V_3. Occasionally SUNCT attacks are so frequent that they can appear continuous and there is confusion with migraine.

Hemicrania continua is a continuous, strictly unilateral headache, sometimes with acute exacerbations and associated with ipsilateral autonomic features. Like paroxysmal hemicrania, there is a complete response to indometacin.

Other primary headache disorders

- ***Primary stabbing headache ('ice pick headache')*** involves momentary jabs or stabs of localized pain occurring either in the same spot or moving about the head. Symptoms wax and wane, and are more common in patients with other primary headache disorders, particularly migraine. Treatment is usually not needed but this type of headache responds well to indometacin.
- ***Primary cough headache*** is a sudden, sharp head pain on coughing. No underlying cause is found but intracranial pathology should be excluded. The problem often resolves spontaneously. Indometacin is the treatment of choice; LP with removal of CSF can help.
- ***Primary sex headache*** is characterized by explosive headache at or before orgasm. It often resolves spontaneously after several attacks. Investigation to exclude SAH is required after the first episode.
- ***Other varieties*** of primary headache include primary exertional headache, hypnic headache (headache triggered by sleep), nummular (coin-shaped) headache, and primary thunderclap headache.

Chronic daily headache

This is defined as headache on 15 or more days per month for at least 3 months. Up to 4% of the population are affected by daily or near-daily headache. Although there are many possible causes, including secondary headache disorders, in practice, primary headache disorders, particularly migraine, are responsible for the majority. Where migraine is the cause, the term chronic migraine is now preferred.

Overuse of analgesic medication or triptans (termed ***medication overuse headache***) is often a major factor leading to and maintaining chronicity, particularly in those with migrainous biology. Use of 15 or more doses per month of any simple analgesic, or 10 or more if triptans are being used, may eventually lead to transformation of episodic headache into chronic daily headache. Opioids are particularly problematic and will cause rebound headache or rapidly lead to medication overuse headache in susceptible individuals; opioids should be avoided entirely for headache treatment (and patients should be warned that some over-the-counter migraine preparations contain opioids).

Explanation that medication overuse is part of the problem is essential to help patients withdraw from or substantially reduce analgesic intake. This is a difficult process for many patients, especially as there may be a period of transient rebound worsening of headache after withdrawal. Concurrent introduction of migraine suppression medication (see above) may help withdrawal but will not be effective if patients cannot withdraw from frequent analgesic use. Occasionally, hospital admission for analgesic withdrawal with parenteral administration of dihydroergotamine is required.

Secondary headache disorders

Raised intracranial pressure headache

Any headache present on waking and made worse by coughing, straining, or sneezing may be due to raised intracranial pressure (ICP) caused by a mass lesion, hydrocephalus, or venous sinus thrombosis. Vomiting often accompanies pressure headaches. Visual obscurations (momentary bilateral visual loss with bending or coughing) are characteristic and seen in the presence of papilloedema. Occasionally, where ICP rises quickly, papilloedema may not be present.

Neuroimaging is mandatory if raised ICP is suspected. Where no mass lesion, venous sinus thrombosis, or hydrocephalus is detected on imaging in the presence of papilloedema, idiopathic intracranial hypertension may be the cause, and LP is performed to measure CSF opening pressure.

Idiopathic intracranial hypertension

Idiopathic intracranial hypertension (IIH) probably results from reduced CSF resorption. IIH typically develops in younger, overweight female patients. Headaches and transient visual obscurations due to the florid papilloedema are the presenting features. A VIth nerve palsy may develop—a false localizing sign (see 'VI: Abducens nerve lesions' section). CSF pressure is very elevated, with normal constituents. Brain imaging is normal, although ventricles may be small and appear 'slit-like'.

Various drugs, such as tetracyclines and vitamin A supplements, have been implicated. Other causes of papilloedema (or pseudopapilloedema) should be excluded. Sagittal sinus thrombosis can cause a similar picture and should always be looked for on MR or CT venography.

IIH is usually self-limiting. However, optic nerve damage can result from longstanding or fulminant papilloedema with progressive loss of peripheral visual fields. Suspected IIH with visual loss should be managed without delay, with an early lumbar puncture for both diagnosis and therapeutic pressure-lowering. Regular monitoring of visual fields with perimetry is essential. Repeated LP, acetazolamide, thiazide diuretics, and topiramate are used to reduce

CSF production. Weight reduction is helpful. Ventriculoperitoneal shunt insertion or optic nerve sheath fenestration to protect vision is sometimes necessary.

Low-CSF-volume (low-pressure) headache

Although most often seen following dural puncture, CSF leaks may occur spontaneously, leading to postural headache, worse on standing or sitting, and relieved completely by lying flat. The patient may give a history of vigorous Valsalva, straining, or coughing just prior to onset. Pachymeningeal enhancement may be seen on MRI but is not always present. LP is generally avoided for obvious reasons but may reveal low opening pressure. Conservative measures that are sometimes effective include bed rest, caffeine (intravenous or oral), and good hydration. The site of the leak is usually within the spine; thus treatment consists of injection of autologous blood into the spinal epidural space to seal the leak (a 'blood patch'), or occasionally surgical repair of the dural tear.

Post-traumatic headache

Pre-existing migraine may worsen following head injury. *De novo* headache sometimes follows a minor head injury but post-traumatic headache is an ill-defined entity. Improvement over a few weeks is the norm, but where litigation is ongoing, symptoms can persist for long periods. Subdural haematoma and low-pressure headache need to be considered as a possible cause.

Other secondary headache disorders

See **Box 28.23**. It is worth considering medical causes of headaches, including obstructive sleep apnoea (causing morning headaches), and ocular causes (posterior scleritis or subacute angle-closure glaucoma occasionally present as 'headache' rather than eye pain).

New onset, sudden headaches should always prompt consideration of the following emergencies which require immediate specialist investigation and management:

- Subarachnoid haemorrhage (Chapter 11, 'Subarachnoid haemorrhage' section)—classically a 'thunderclap' onset
- Meningitis (Chapter 11, 'Meningitis' section)—when associated with vomiting, rash, and sepsis
- Intracerebral haemorrhage (Chapter 29, 'Intracerebral haemorrhage' section)—when associated with sudden onset focal neurological signs
- Cervical artery dissection (Chapter 29, 'Carotid and vertebral artery dissection' section)—classically causes neck pain radiating upwards; there may be a Horner's syndrome or other focal neurological signs
- GCA (Chapter 32, 'Giant cell arteritis' section)—when associated with scalp tenderness, visual loss, and raised inflammatory markers.

Facial pain

The face has many pain-sensitive structures: teeth, gums, sinuses, temporomandibular joints, jaw, and eyes. Dental causes are common and should always be considered. Facial pain is also caused by specific neurological conditions.

Trigeminal neuralgia

Trigeminal neuralgia typically starts in the sixth and seventh decades; hypertension is the main risk factor. Compression of the trigeminal nerve at or near the pons by an ectatic vascular loop is the usual cause. High-resolution MRI studies may demonstrate the vascular loop in contact with the nerve in a high proportion of cases. In younger patients, MS or a cerebellopontine angle tumour (acoustic schwannomas, meningiomas, epidermoids) is more likely to be the cause.

Clinical features

Paroxysms of knife-like or electric shock-like pain, lasting seconds, occur in the distribution of the Vth nerve. Pain tends to commence in the mandibular division (V_3) but may spread over time to involve the maxillary (V_2) and, occasionally, the ophthalmic divisions (V_1). Bilateral trigeminal neuralgia is rare (3%) and usually due to intrinsic brainstem pathology, such as demyelination. Episodes occur many times a day with a refractory period after each. They may be brought on by stimulation of one or more trigger zones in the face. Washing, shaving, a cold wind, and chewing are examples of trivial stimuli that provoke pain. The face may be screwed up in agony. Spontaneous remissions last months or years before (almost invariable) recurrence. There are no signs of Vth nerve dysfunction on examination.

Management

Carbamazepine (600–1200 mg daily) reduces the severity of attacks in the majority. Oxcarbazepine, lamotrigine and gabapentin are also used. If drugs fail or are not tolerated, a number of surgical options are available that, in general, are more effective than medical treatments. Percutaneous radiofrequency selective ablation of the trigeminal ganglion is performed as a day-case procedure; recurrence may occur after an average of 2 years. Microvascular decompression of the nerve in the posterior fossa has a high long-term success rate (approximately 90%).

Atypical facial pain

Facial pain differs from trigeminal neuralgia in quality and distribution, and trigger points are absent. The condition is probably heterogeneous in aetiology but is believed by some (on little evidence) to be a somatic manifestation of depression. Tricyclic antidepressants and drugs used in neuropathic pain are sometimes helpful.

Other causes of facial pain

Facial pain occurs in the trigeminal autonomic cephalgias (see above), occasionally in migraine and in carotid dissection.

Giant cell arteritis (temporal arteritis)

This is a granulomatous large-vessel arteritis seen almost exclusively in people over 50 with a poorly defined aetiology (see Chapter 32, 'Giant cell arteritis' section).

Clinical features

- ***Headache*.** This is almost invariable in GCA. Pain develops over inflamed superficial temporal and/or occipital arteries. Touching the skin over an inflamed vessel (e.g., when combing the hair) causes pain. Arterial pulsation is soon lost; the artery becomes hard, tortuous, and thickened. The scalp over inflamed vessels may become red. Rarely, gangrenous patches appear.
- ***Facial pain*.** Pain in the face, jaw, and mouth is caused by inflammation of facial, maxillary, and lingual branches of the external carotid artery in GCA. Pain is characteristically worse on eating (jaw claudication). Mouth opening and protruding the tongue become difficult. A painful, ischaemic tongue occurs rarely.
- ***Visual problems*.** Visual loss from arterial inflammation and occlusion occurs in 25% of untreated cases. Posterior ciliary artery occlusion causes anterior ischaemic optic neuropathy (AION) in three-quarters of these. The disc is swollen and pale; retinal branch vessels usually remain normal. The majority of cases of AION are non-arteritic but inflammatory markers must always

be checked and GCA considered; arteritic AION carries a worse prognosis. Other mechanisms of visual loss in GCA are central retinal artery occlusion (causing disc pallor and visible retinal ischaemia), cilioretinal artery occlusion, and posterior ischaemic optic neuropathy. All of these entities cause sudden monocular visual loss (partial or complete). Amaurosis fugax (see Chapter 29, 'Amaurosis fugax' section) may precede permanent blindness, and bilateral blindness may ensue, with the second eye being affected 1 to 2 weeks after the first.

Diagnosis and management

See Chapter 32 ('Giant cell arteritis' section).

Further reading

Headache Classification Committee of the International Headache Society. *The International Classification of Headache Disorders*, 3rd edn. *Cephalalgia* 2018; 38:1–211.
http://www.bash.org.uk *British Association for the Study of Headache guidelines on headache diagnosis and management*.

EPILEPSY AND LOSS OF CONSCIOUSNESS

Epilepsy

An *epileptic seizure* can be defined as a sudden synchronous or excessive discharge of cerebral neurons causing symptoms or signs that are apparent either to the patient or to an observer. For example, a limited discharge affecting only part of the cortex may cause a subjective aura apparent only to the patient, or a generalized seizure may cause a convulsion witnessed by an observer of which the patient may be unaware. This definition excludes disorders such as migrainous aura that are more gradual in onset, and EEG discharges that do not have a clinical correlate.

The term *epilepsy* describes an ongoing liability to recurrent epileptic seizures. The emergency management of seizures is covered in Chapter 11.

Epidemiology

Epilepsy is common. Its population prevalence is 0.7% to 0.8% (higher in low and middle income countries). Approximately 500 000 people in the UK have epilepsy. The incidence of epilepsy is age-dependent; it is highest at the extremes of life, most cases starting before the age of 20 or after the age of 60. The cumulative incidence (lifetime risk) of epilepsy is over 3% and the lifetime risk of having a single seizure is 5%. The fact that the prevalence is much lower than the cumulative incidence in part reflects the fact that epilepsy often goes into remission.

Classification

The International League Against Epilepsy (ILAE) Commission on Classification and Terminology has published a classification system that allows for diagnosis at three levels (seizure types, epilepsy types, and epilepsy syndromes). It may not be possible in every case to make a diagnosis at all three levels, but accurate characterization of epilepsy allows for more effective intervention, helps with seizure risk management, and may identify an aetiology.

Classification of seizure types

Seizures are divided by clinical pattern into three main groups by seizure onset (**Box 28.26** and **Fig. 28.21**): ***generalized*** seizures, ***focal*** seizures, and ***unknown onset*** seizures.

- ***Generalized seizures*** arise at some point within and rapidly engage bilateral brain networks. They are usually associated with loss or impairment of consciousness (except in generalized myoclonic seizures). Generalized seizures may be:

Box 28.26 Classification of seizures

Generalized onset seizure	
Motor	**Non-motor (absence)**
Tonic–clonic Myoclonic Other: Clonic, tonic, atonic, myoclonic–tonic–clonic, myoclonic-atonic, epileptic spasms	Typical absence Atypical absence Other: Myoclonic, eyelid myoclonia with absence
Focal onset seizure	
Motor onset	**Non-motor onset**
Aware, impaired awareness (or unknown)	Aware, impaired awareness (or unknown)
Automatisms Clonic Hyperkinetic Myoclonic Tonic Other: Atonic, epileptic spasms	Autonomic Behaviour arrest (only if behaviour arrest throughout seizure) Cognitive Emotional Sensory
Focal to bilateral tonic–clonic	
Unknown onset seizure	
Motor	**Non-motor**
Tonic–clonic Epileptic spasms	Behaviour arrest

(Adapted from Fisher RS, Cross' JH, D'Souza C et al. Instruction manual for the ILAE 2017 operational classification of seizure types. Epilepsia *2017; 58:531-42. 'Fine print' seizure types mainly of interest to specialists are listed under 'other'.)*

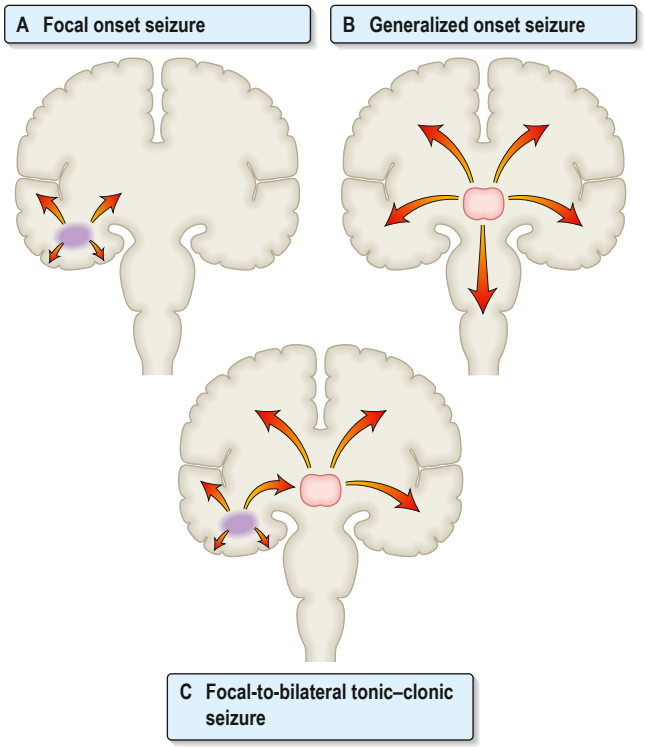


Fig. 28.21 Seizure types.

  - ***Motor***, for example generalized onset tonic-clonic seizure. The term 'generalized' is only used now when the seizure is reasonably believed to be generalized at onset (otherwise use the term 'bilateral tonic-clonic seizure')
  - ***Absences***
  - ***Generalized myoclonic seizures***
- A ***focal seizure*** is caused by electrical discharges originating in brain networks in one hemisphere. Focal seizures are further characterized based on:
  - ***Level of awareness*** (aware, impaired awareness, or unknown).
  - ***Motor or non-motor*** features at onset (automatisms, clonic movements, cognitive or emotional symptoms, etc). In general we are most interested in the features at onset as they correlate best (but not perfectly) with the onset zone of the epileptic activity. This can have implications for aetiological work-up and epilepsy surgery.
  - Where possible additional descriptors will be applied to clarify the nature of the cardinal features at onset, and subsequent features.
  - Laterality is important to identify.
  - A clinician may therefore describe a seizure, for example, as a '*focal impaired awareness autonomic seizure, with epigastric rising sensation, and right hand automatisms*'.
- ***Focal-to-bilateral tonic-clonic seizures*** were previously referred to as 'secondary generalized seizures'. These result from the spread of a focal onset seizure to involve both hemispheres. The patient may remember the initial focal seizure before losing consciousness, in which case this is called an ***aura***, and the focal onset can be inferred from the semiology; sometimes, however, the spread of electrical activity is so rapid that the patient does not experience any warning (in which case the onset will often be unknown).
- Many people will have ***unknown onset bilateral tonic-clonic seizures***, at least initially, before further information clarifies the situation. A useful rule of thumb is that seizures that occur upon waking, or in the morning, are more likely to be generalized onset seizures while most seizures arising from deep sleep (2–3 a.m.) will prove to be focal onset seizures.

Generalized seizure types

Typical absence seizures ('petit mal')

Epilepsy causing absences almost invariably begins in childhood. There is loss of awareness and a vacant expression for less than 10 seconds before returning abruptly to normal and continuing as though nothing had happened. Apart from slight fluttering of the eyelids, there are no motor manifestations. Each attack is accompanied by 3 Hz spike-and-wave EEG activity (see Fig. 28.19). Patients often do not realize they have had an attack but may have many per day; they may be identified as 'daydreaming' in class. Children with absence seizures may go on to develop generalized tonic-clonic seizures. Absences are never due to acquired lesions such as tumours; they are a manifestation of generalized epilepsies (typical absences are associated with the genetic generalized epilepsies [GGEs]). It is a common mistake to refer elderly patients having vacant spells as suspected new-onset absences: if these are seizures, they will be focal onset impaired awareness seizures (often arising from the temporal lobe).

Generalized onset tonic–clonic seizures ('grand mal')

Prodrome. There is no warning before generalized onset tonic–clonic seizures. When an aura is described, consider a *focal to bilateral tonic–clonic* seizure.

Tonic–clonic phase. An initial tonic stiffening is followed by the clonic phase with synchronous jerking of the limbs, reducing in frequency over about 2 minutes until the convulsion stops. The patient may utter an initial cry and then falls, sometimes suffering serious injury. The eyes remain open and the tongue is often bitten. There may be incontinence of urine or faeces.

Post-ictal phase. A period of flaccid unresponsiveness is followed by gradual return of awareness with confusion and drowsiness lasting 15 minutes to an hour or longer. Headache is common after a tonic-clonic seizure.

Myoclonic, tonic, and atonic seizures

Myoclonic seizures or 'jerks' take the form of momentary brief contractions of a muscle or muscle groups: for example, causing a sudden involuntary twitch of a finger or hand. They are common in GGEs. ***Tonic seizures*** consist of stiffening of the body with loss of consciousness, not followed by jerking. ***Atonic seizures*** involve a sudden collapse with loss of muscle tone and consciousness.

Focal seizure types

Focal aware seizures

These used to be called simple partial seizures. Awareness must be fully retained throughout, and the patient remembers the entire event (which could consist of motor phenomena such as focal clonic or myoclonic movements, or non-motor features such as spreading sensory disturbance).

Focal impaired awareness seizures

These usually arise from the temporal lobe (60%) or the frontal lobe. There is often a preceding aura followed by a period of complete or partial loss of awareness of surroundings, lasting for 1 to 2 minutes on average (as opposed to 10 seconds in absence seizures), which the patient generally does not remember subsequently. If awareness is impaired at any point, it is termed a focal impaired awareness seizure. The attack may be followed by a short period of post-ictal confusion or may develop into a *bilateral tonic–clonic* seizure.

Focal non-motor seizures

The nature of the initial symptoms helps to localize the seizure focus. Temporal lobe non-motor auras include cognitive symptoms such as déjà vu or jamais vu; emotional features such as fear (may be mistaken for panic attacks); sensory aura with olfactory, gustatory, or auditory hallucinations or vertigo; and autonomic symptoms such as abdominal rising sensation or nausea. Occipital lobe auras include visual phenomena such as zigzag lines and coloured scotomas. Some auras are vague or hard for patients to describe but they are generally stereotyped: that is, the same on each occasion.

Focal motor seizures

Motor phenomena include simple clonic movements, which often originate within the motor cortex. Jerking typically begins on one side of the mouth or in one hand, sometimes spreading to involve the entire side. This visible spread of activity is called the ***Jacksonian march*** of a seizure, and correlates with spread of seizure activity along the motor homunculus. Local temporary paralysis of the limbs affected sometimes follows: ***Todd's paralysis***. Automatisms are common in temporal lobe seizures. These are semi-purposeful stereotyped motions such as lip-smacking, fidgeting, or picking motions, or more complex motor behaviours such as walking in a

circle or undressing. Tonic posturing and the direction of any head turn are very important to note: frontal lobe seizures may cause early forced head and eye version away from the site of seizure onset, and involvement of the supplementary motor area results in asymmetrical tonic posturing of the limbs ('figure of 4' or 'fencing' posture) with extension of the contralateral arm. Unusual hyperkinetic seizures may appear bizarre (e.g., bicycling movements) but can arise from the frontal lobes.

Many focal seizures will have a combination of motor and non-motor features. For instance a mesial temporal lobe seizure may cause déjà vu and an epigastric rising sensation, followed by behavioural arrest, automatisms, and impaired awareness.

Epilepsia partialis continua

This refers to continuous focal jerking with retained awareness that continues for long periods of time (at least an hour, but frequently for days and sometimes even years). It usually arises from contralateral motor cortex, and the face and hand are most commonly involved. The affected limb will often be weak (a Todd's paresis) and there is a high chance of finding a structural cause.

Classification of epilepsy type

Once a seizure type has been identified, a decision needs to be made about whether there is an ongoing liability to seizures (by definition epilepsy), or whether the seizure risk was transient. This can be the case either with a first seizure (see below), as 50% of people with a first seizure will not go on to have another one, or if the patient has had *acute symptomatic seizures*, secondary to an acute event which may or may not lead to long-term epilepsy. If epilepsy is diagnosed, classifying the epilepsy type is usually simple:

- ***Generalized epilepsy***, where all the seizures originate at some point within, or rapidly engage, bilateral brain networks. The structures involved can be subcortical or cortical.
- ***Focal epilepsy***, where all the seizures are believed to arise from networks within one hemisphere. Evidence for this may be clinical (based on seizure semiology) or from EEG findings.
- ***Generalized and focal epilepsy*****.** Several severe childhood-onset epilepsy syndromes, such as Dravet and Lennox-Gastaut syndrome, that lead to multiple seizure types, can cause a mix of generalized and focal seizures.
- ***Unknown.***

Aetiology of epilepsy, and epilepsy syndromes

A precise epilepsy syndrome can only be identified for a subset of patients, but in all cases it is worth thinking about aetiology of epilepsy. Six aetiological categories are recognized in the ILAE classification (**Box 28.27**).

The range of causes of epilepsy is different at different ages and in different countries.

- ***Children and teenagers:*** genetic, perinatal structural, and metabolic disorders predominate
- ***Adults:*** hippocampal sclerosis is one of the most important causes of epilepsy; TBI is common in a younger adult population
- ***Internationally***, infections are the most common cause of epilepsy, in particular neurocysticercosis which accounts for 30% of cases of epilepsy in endemic areas
- ***Older ages (>60 years):*** vascular disease, dementia, and mass lesions such as neoplasms are important.

It is often necessary to distinguish between the cause of a seizure, and the aetiology of epilepsy. Virtually any insult to the brain can provoke a seizure (as opposed to causing an ongoing liability to seizures, or epilepsy).

? Box 28.27 Causes of epilepsy

- Genetic
  - Brain channelopathies (e.g., Dravet syndrome)
  - Genetic generalized epilepsies, e.g., juvenile myoclonic epilepsy
- Structural
  - Hippocampal sclerosis
  - Traumatic brain injury, brain surgery
  - Birth trauma
  - Developmental, e.g., neuronal migration abnormalities, cortical dysplasia
  - Intracranial mass lesions, e.g., tumour
  - Vascular, e.g., cerebral infarction, arteriovenous malformation, venous sinus thrombosis
  - Neurodegenerative disorders, e.g., Alzheimer's
- Metabolic
  - Leigh syndrome (mitochondrial disease)
  - Glucose transporter deficiency
  - Chronic alcohol use (note that a single episode of alcohol withdrawal is a cause of *seizures* but not of *epilepsy*)
- Infectious
  - Viral encephalitis
  - Cerebral tuberculosis
  - HIV
  - Neurocysticercosis
- Immune
  - NMDA receptor antibody
  - LGI-1 antibody-associated limbic encephalitis
- Unknown
  - One-third of all cases are of unknown aetiology

Epilepsy syndrome

An epilepsy syndrome can be assigned in a proportion of cases, where the seizure types and associated features, plus often characteristic EEG and/or imaging abnormalities, form part of a recognizable, distinct seizure disorder (e.g., juvenile myoclonic epilepsy, mesial temporal lobe epilepsy with hippocampal sclerosis). These often have prognostic, aetiological, or treatment implications. Some epilepsy syndromes have a single identifiable cause (e.g., Dravet syndrome, caused by variants in the *SCN1A* gene) while others will be a manifestation of multiple causes (e.g., Lennox-Gastaut syndrome, with multiple seizure types from early childhood, can be caused by various structural brain abnormalities or presumed genetic aetiologies). Many people with epilepsy, especially those with focal epilepsy with onset in adulthood, will not fall into a well-defined epilepsy syndrome.

Aetiology of generalized epilepsy

Presenting in childhood and early adult life, genetic generalized epilepsies (GGEs), account for up to 20% of all patients with epilepsy. The cause is thought to be polygenic with complex inheritance. This does not always mean that these epilepsies will run in the family, and so the term 'idiopathic generalized epilepsy' is often used synonymously with GGE. The brain is structurally normal but abnormalities of ion channels influencing neuronal firing, abnormalities of neurotransmitter release, and synaptic connections are probably the underlying molecular pathological substrates. They include:

- ***Childhood absence epilepsy:*** absence seizures. Spontaneous remission by age 18 is usual.

- ***Juvenile myoclonic epilepsy*** (JME; Box 28.28): this accounts for 10% of all epilepsy patients. Typically, myoclonic jerks start in the teenage years (and are usually ignored by the patient; ask about jerks when taking an epilepsy history), followed by generalized tonic–clonic seizures that bring the patient to medical attention. One-third of patients also have absences. Seizures and jerks often occur in the morning after waking. Lack of sleep, alcohol, and strobe or flickering lights are seizure triggers in JME. JME usually responds reasonably well to treatment, is associated with characteristic EEG abnormalities (3–6 Hz generalized polyspike and wave discharges), and requires life-long treatment.

Aetiology of focal seizures and focal epilepsy

Almost any process disrupting the cortical grey matter can cause acute symptomatic seizures, or go on to cause epilepsy. Seizures arise from the affected area of cortex, with or without *bilateral tonic–clonic* seizures (these may obscure the focal onset). A focal seizure onset often indicates a structural cause and detailed imaging is required to identify this. In general, the response to treatment is less good than with GGE.

Genetic

Many single-gene epilepsy disorders are due to mutations of voltage- or ligand-gated neuronal ion channels (channelopathies). Autosomal dominant nocturnal frontal lobe epilepsy is a familial disorder causing brief nocturnal frontal lobe seizures, often with onset in childhood. In a minority of cases, it is caused by variants in different subunits in the neuronal nicotinic acetylcholine receptor gene. However, monogenic epilepsy syndromes are relatively uncommon in clinical practice, especially compared to some other fields of neurology. Over 200 multisystemic genetic disorders, such as tuberous sclerosis and mitochondrial cytopathies, include epilepsy among their core features.

Structural

Hippocampal sclerosis. Hippocampal sclerosis (damage with scarring and atrophy of the hippocampus and surrounding cortex) is the main pathological substrate causing temporal lobe epilepsy and the leading cause of focal epilepsy in adults. Childhood febrile convulsions are the main risk factor. Hippocampal sclerosis is usually visible on MRI. Seizures can be quite refractory to treatment, in which case surgical resection of the damaged temporal lobe may be an option.

Traumatic brain injury. This may cause epilepsy, sometimes years after the event. The risk is not increased after mild injury (loss of consciousness or post-traumatic amnesia [PTA] <30 min). Depressed skull fracture, penetrating injury, and intracranial haemorrhage increase risk significantly. Early post-traumatic seizures (in the first week) are common and not necessarily indicative of long-term epilepsy risk.

Brain surgery. This is followed by seizures in up to 17% of cases. Prophylactic anticonvulsant use after surgery is not recommended.

Perinatal brain injury and cerebral palsy. Periventricular leukomalacia and brain haemorrhage associated with prematurity and fetal hypoxia, may cause early-onset epilepsy. One-third of children with cerebral palsy have epilepsy.

Developmental disorders. Neuronal migration defects during brain development, dysplastic areas of cerebral cortex, and hamartomas contribute to seizures in both infancy and adult life.

Brain tumours and other mass lesions. Mass lesions involving the cortex cause epilepsy. Seizures are one of the most common presenting features of brain tumours. Brain tumours cause 6% of cases of adult-onset epilepsy.

Vascular disorders. Stroke and small-vessel cerebrovascular disease is the most common cause of epilepsy after the age of 60. Cortical venous thrombosis or venous sinus thrombosis will often lead to epilepsy. Cavernous haemangiomas (cavernomas) usually present with epilepsy (Fig. 28.22); other vascular malformations (such as arteriovenous malformations [AVMs]) can also cause seizures.

Posterior reversible encephalopathy syndrome. This syndrome is caused by rapid rises in blood pressure and is characterized by the acute onset of headache, visual disturbance, altered mental status, and seizures. It may be seen in eclampsia, renal disease, autoimmune diseases (e.g., systemic lupus erythematosus [SLE]), or as a side-effect of certain medications. White matter changes are seen on CNS imaging, primarily affecting the occipital lobes; management is of hypertension and the underlying disorder; often carried out in a critical care environment. It is unusual for posterior reversible encephalopathy syndrome (PRES) to lead to chronic epilepsy although occasionally long-term sequelae are seen.

Neurodegenerative disorders involving the cerebral cortex, such as Alzheimer's disease, are associated with an increased risk of epilepsy.

Metabolic disorders

Genetic metabolic disorders, such as mitochondrial Leigh syndrome, or glucose transporter deficiency, can cause epilepsy. Transient metabolic abnormalities (hyponatraemia, hypoglycaemia, uraemia) mostly cause acute symptomatic seizures rather than epilepsy, but if severe or persistent enough to cause a brain injury, epilepsy may

Box 28.28 Juvenile myoclonic epilepsy

- 10% of all epilepsy patients
- Starts in teenage years
- Clinical features:
  - Myoclonic jerks
  - Generalized tonic–clonic seizures
  - Absences in one-third
  - Triggers: sleep deprivation, alcohol, strobe lighting
- Abnormal EEG
- Good response to treatment
- Requires life-long treatment

Fig. 28.22 Magnetic resonance T2-weighted image: frontal cavernoma *(arrowed)*.

ensue. Chronic alcohol use is a common cause of epilepsy. Alcohol withdrawal itself will cause acute symptomatic seizures, but over a long period of time, chronic alcohol use or repeated withdrawals will predispose to epilepsy and an ongoing seizure risk. Alcohol-induced hypoglycaemia and head injury also lead to seizures.

Infection

Seizures are often the presenting feature of encephalitis, cerebral abscess, and tuberculomas. They also occur in chronic meningitis (e.g., tuberculosis) and may rarely be the first sign of acute bacterial meningitis. Neurocysticercosis is a major cause of seizures in countries where the pork tapeworm is endemic, such as India and South America, where it causes up to 30% of all cases of epilepsy. HIV and complications of immunosuppression, such as cerebral toxoplasmosis, may also lead to seizures.

Immunological disorders

Autoimmune antibody-mediated encephalitis, such as that due to antibodies against potassium channels or NMDA or glycine receptors, typically presents with seizures, as may paraneoplastic limbic encephalitis (PLE).

Aetiology of (combined) generalized and focal epilepsy

Epilepsy caused by diffuse structural, genetic, or metabolic aetiologies can sometimes cause a mixture of generalized and focal seizures. These include some childhood-onset severe epilepsy syndromes.

Dravet syndrome, also known as severe myoclonic epilepsy of infancy, is one of the more common single-gene epilepsy disorders. It is a *channelopathy* caused by variants in the sodium channel gene, *SCN1A*; it presents in the first year of life with a mixture of generalized and focal onset seizures. The seizures are intractable and an *epileptic encephalopathy* develops, with cognitive and behavioural regression occurring partly as a result of the seizures themselves.

Lennox-Gastaut syndrome is another epileptic encephalopathy, with multiple causes, resulting in multiple seizure types (but particularly tonic seizures in sleep), cognitive decline and diffuse slow spike-and-wave and paroxysms of fast activity on EEG.

Diagnosis of the first fit and investigations

The diagnosis of a seizure is essentially a clinical one, based on taking a history from the patient and any witnesses (**Boxes 28.29 and 28.30**). It is a sad fact that seizures that happen in hospital are usually the most poorly described. It is useful for any witnesses (including healthcare professionals) to make a note of exactly what they saw, in plain language (avoid terms such as 'GTCS'). The usefulness of a video recording cannot be over-emphasized but the patient must have given consent for this in advance. Investigations have a limited role in distinguishing between a seizure and other causes of a blackout or attack (see Chapter 11, 'Dizziness and blackouts' section).

The majority of patients referred to a first fit clinic have not had a seizure. The most common error is to misdiagnose a syncopal blackout for a seizure.

Which investigations are needed?

Basic investigations in patients presenting with seizures (**Box 28.31**) are covered in Chapter 11.

Electroencephalography

EEG is most useful to categorize epilepsy and understand its cause, rather than as a means of confirming a doubtful diagnosis

Box 28.29 Diagnosis of a seizure: taking a history after an episode of loss of consciousness

Witness account is crucial.

- What happened:
  - Before
    - Aura—unusual sensations, taste or smell; rising sensation in the stomach; déjà vu or jamais vu; strong emotion
    - Pre-syncopal prodrome—feeling lightheaded, hot, black spots in the vision
  - During
    - Change in muscle tone (floppy/stiff)
    - Turn of the head or forced limb posture
    - Confusion/vacant appearance/loss of consciousness
    - Change in breathing pattern
    - Change in facial colour—pallor (syncope) or cyanosis (seizure)
    - Eyes open or closed; gaze deviation
    - Abnormal movements. Jerking/twitching/tremor/flailing. Both sides of the body or just one? Did it start in one place and travel? Were the movements rhythmical, and synchronous?
    - Automatisms (fiddling, picking, grimacing/lip-smacking)
    - Vocalizations
    - Incontinence
    - Tongue-biting (which part of the tongue, how severe?)
    - Duration of seizure
  - After
    - Post-ictal confusion and sleepiness
    - Rapid recovery in syncope
- Circumstances
  - Sleep (what time of night)/wakefulness
  - Position (lying/standing)
  - Seizure triggers? Sleep deprivation, alcohol binge, or drugs
  - Syncope triggers? Pain, heat, prolonged standing
- Epilepsy risk factors?
  - Childhood febrile convulsions
  - Significant head injury
  - Meningitis or encephalitis
  - Family history of epilepsy
- Previous unrecognized seizures?
  - Myoclonic jerks
  - Absences
  - Auras (focal seizures)
- Alcohol excess?
- Medication lowering seizure threshold?
- Hold a driving licence?

Box 28.30 Useful points when diagnosing epilepsy

- ***Diagnosis is a clinical one.*** Tests such as EEG have only a limited role in distinguishing between epilepsy and other attack types (unless an event is captured during the recording).
- ***Poor clinical discriminators between types of blackout:***
  - Urinary incontinence—may occur in both seizure and syncope
  - Injuries—may occur in both seizure and syncope (but less common in dissociative seizures)
- ***Good discriminators between types of blackout:***
  - Prolonged recovery period (seizure)—what is the first thing they remember after the seizure? If they woke up in hospital or with the paramedics around them often that indicates a prolonged recovery.
  - Bitten tongue—side of tongue usually (seizure); biting of the tip of the tongue is less specific
  - Colour change—pallor (syncope), cyanosis (seizure)
- ***Stereotyped attacks are usually due to epilepsy***

of epilepsy. EEG has a high false-negative rate in epilepsy (20%–40% even with provocation measures and awake and sleep recordings) and a low false-positive rate (1% of people without epilepsy have true epileptiform changes on EEG; however, other nonspecific abnormalities are common and should not be overinterpreted).

- ***EEG abnormalities in epilepsy:*** focal cortical spikes (e.g., over a temporal lobe) or generalized spike-and-wave activity (in GGE). Epileptic activity is continuous in status epilepticus.
- ***Sleep recordings or 24-hour ambulatory EEG*** increase sensitivity when routine EEG is normal.
- ***Continuous EEG videotelemetry*** is helpful for diagnosis in attacks of uncertain cause.

Brain imaging

MRI is indicated in most patients after a first seizure, particularly in focal-onset seizures and in older patients where the chance of a focal brain lesion is greatest. In patients below the age of 30 with a definite electroclinical diagnosis of GGE, brain imaging is not essential.

Box 28.31 Checklist after a first seizure

- Perform tests:
  - Blood tests and ECG
  - EEG
  - CT +/– MRI brain (most patients—see text)
- Avoid or remove precipitants: drugs, alcohol, sleep deprivation
- Give advice on safety: for example, avoid swimming, baths, working at heights
- Stop driving and ask the patient to inform the Driver and Vehicle Licensing Authority (DVLA; in the UK)
- Discuss recurrence risk and treatment

Recurrence risk after a first fit

About one-third of people will have a second seizure within 5 years. The risk of seizure recurrence is significantly increased by features of GGE on EEG, abnormal neurological examination, a first seizure that occurs in sleep, and the presence of structural brain lesions.

Diagnosing epilepsy

Epilepsy is diagnosed when:

- At least two unprovoked seizures occur, more than 24 hours apart
- Only one unprovoked seizure has occurred, but the risk of recurrence is similar to the risk after two seizures (suggested to be ≥60%). This will be the case when there is an established structural brain abnormality (e.g., an old stroke), or when the EEG shows clear epileptiform features.

Management

Acute management

The acute management of seizures, including ongoing continuous seizure activity (status epilepticus) is covered in Chapter 11.

Anti-seizure medications

Anti-seizure medications (ASMs; **Box 28.32**) are indicated when there is a firm clinical diagnosis of epilepsy and a substantial risk of recurrent seizures. Some general principles apply:

- Introduce ASMs at low dose and slowly titrate upwards until the seizures are controlled or side-effects become unacceptable. In emergency settings, rapid loading and faster titration may be appropriate.
- Aim for monotherapy; 70% of patients will have good seizure control with a single AED.

Box 28.32 Antiepileptic drugs and common seizure types

	Generalized epilepsy, with primarily GTCS	Focal epilepsy	Myoclonic seizures	Absence seizures
First-line	**Sodium valproate**[a] **Lamotrigine**[b] Levetiracetam	**Lamotrigine**[b] Levetiracetam Oxcarbazepine Carbamazepine	Sodium valproate[a] Levetiracetam	Ethosuximide Sodium valproate[a] Lamotrigine[b]
Second line and/or add-ons	Zonisamide Lacosamide Topiramate Perampanel Brivaracetam Clobazam Clonazepam Phenobarbital Primidone Phenytoin[c]	Lacosamide Zonisamide Topiramate Sodium valproate[a] Eslicarbazepine Clobazam Brivaracetam Cenobamate Perampanel Gabapentin Pregabalin Tiagabine Vigabatrin Phenobarbital Phenytoin[c]	Brivaracetam Topiramate Clonazepam Clobazam Lamotrigine[b] Piracetam Phenobarbital Lacosamide	Levetiracetam
May worsen attacks			Carbamazepine Ox/Eslicarbazepine Gabapentin Phenytoin[c] Vigabatrin Tiagabine	Carbamazepine Ox/Eslicarbazepine Phenytoin[c] Vigabatrin Tiagabine

[a]Sodium valproate should be avoided in women of childbearing age.
[b]Lamotrigine can also occasionally worsen myoclonic seizures, but nonetheless it is often used first-line for juvenile myoclonic epilepsy, especially in women of childbearing age.
[c]Phenytoin can be used for generalized epilepsies but sometimes has the potential to worsen generalized-onset seizure types.

- If seizures are not controlled with the first ASM, gradually introduce a second agent and then slowly withdraw the first ASM. If the patient is still not seizure-free, then combination therapy is required.
- Epilepsy is one of the few disorders where non-generic ('brand name') prescribing is justified to ensure consistent drug levels.
- Routine monitoring of ASM levels is not needed and should be reserved for assessing compliance and toxicity.
- Many of the older ASMs are enzyme inducers and will interact with many other medications, and with each other (phenytoin, phenobarbital, carbamazepine). Sodium valproate is a potent hepatic enzyme inhibitor. It particularly increases drug levels of lamotrigine, so if sodium valproate is added to lamotrigine caution must be taken to avoid toxicity. Carbapenem antibiotics interact with sodium valproate, profoundly lowering serum levels of valproate. New-generation ASMs have fewer interactions.
- Phenytoin is no longer considered a first-line ASM; it is now principally used in emergency control of seizures (see status epilepticus, Chapter 11, 'Status epilepticus' section)
- There is high-quality evidence from the Standard and New Antiepileptic Drugs (SANAD) trials to suggest that lamotrigine should be the first-line drug of choice in patients with focal epilepsy. The disadvantage of lamotrigine is that it must be titrated up slowly, or the risk of Stevens–Johnson syndrome increases.
- Sodium valproate was shown to be most effective in generalized epilepsies but should be avoided in women of childbearing age, for whom lamotrigine is the better choice. In children born to women taking valproate in pregnancy, around 10% develop birth defects, and up to 40% will have problems with learning or development. There are also now emerging concerns about the use of sodium valproate in men who may conceive children, although the evidence is limited and practice internationally remains variable.
- Levetiracetam is increasingly used in most types of epilepsy, and has advantages when rapid titration is desired, but it is not as free of side-effects as many clinicians believe (primarily effects on mood and occasionally psychosis).

Unwanted effects of drugs

Intoxication with most ASM causes unsteadiness, nystagmus, and drowsiness. For some agents, drug intoxication can provoke more seizures. Side-effects are more common with multiple ASM usage. Skin rashes are seen particularly with lamotrigine, carbamazepine, and phenytoin; Stevens–Johnson syndrome may occur, and patients must be counselled about this. A wide variety of idiosyncratic drug reactions may occur, such as blood dyscrasias or hyponatraemia with carbamazepine.

Epilepsy and driving

Patients should be asked to stop driving after a seizure and to inform the regulatory authorities if they hold a driving licence. After a seizure, a temporary driving ban until seizure-free is usual but regulations vary from country to country. Many driving regulatory bodies also suggest refraining from driving while withdrawing from AEDs.

Lifestyle and safety

People with epilepsy should be encouraged to lead lives as unrestricted as reasonably possible, though observing simple safety measures such as avoiding swimming and dangerous sports like rock climbing. Advice at home includes leaving bathroom and lavatory doors unlocked and taking showers rather than baths. Epilepsy triggers, such as sleep deprivation, excess alcohol, and drugs, should be avoided, as should strobe lighting where there is EEG evidence of a photo-paroxysmal response (seen in a minority of patients).

Drug withdrawal

Withdrawal of ASMs should be considered after a seizure-free period of at least 2 to 3 years, depending on the aetiology of epilepsy (and many recommend 5 years of seizure-freedom). There is a 50% seizure recurrence rate after withdrawal, so detailed discussion and explanation are essential.

Refractory epilepsy

- Seizures may persist despite treatment, especially with temporal lobe epilepsy. About 20% to 40% of patients with epilepsy will be treatment-refractory.
- Re-evaluate the diagnosis.
- Consider concordance.
- Combine ASMs and use the maximum tolerated dose.
- Refer to a specialist unit for consideration of epilepsy surgery.
- Other non-pharmacological treatments, such as vagal nerve stimulation and the ketogenic low-carbohydrate diet, may sometimes be useful.

Epilepsy surgery

Anterior temporal lobectomy will result in seizure freedom in 50% to 70% of selected patients with uncontrolled seizures caused by hippocampal sclerosis (defined by imaging and confirmed by EEG). Other patients with refractory focal epilepsy in whom a localized seizure onset zone can be identified can be considered for epilepsy surgery.

Other causes of blackouts

The distinction between a fit (seizure), a faint (syncope), a dissociative (non-epileptic) attack, or another type of 'funny turn' is primarily a clinical one, dependent on the history and an eyewitness account (**Boxes 28.29, 28.30, and 28.33**).

Syncope

Syncope is discussed in detail in Chapter 11.

Dissociative (non-epileptic) attacks

These are functional (or 'dissociative') seizure-like attacks that are not caused by abnormal electrical activity in the brain. They are a common problem in clinical practice, and can be difficult to diagnose and manage. Many patients are misdiagnosed with epilepsy. While in an emergency setting, it is safer to err on the side of treating as epilepsy if there is doubt or if there is no established diagnosis, over time patients are exposed to significant iatrogenic risks unless they are given an accurate diagnosis. Moreover, a long delay to diagnosis prevents access to the right treatments and confers a substantially worse prognosis.

Multiple terms are currently used, including non-epileptic attack disorder (NEAD; this is acceptable but defines the condition by what it is not, and as such can leave the patient with the impression that they have not been given a diagnosis). However, outdated and judgement-laden terms such as 'pseudoseizures' and 'hysterical seizures' should be avoided. Dissociative attacks are not consciously produced or 'faked', and it is vital that this is understood by the medical team and explained to the patient.

Box 28.33 Causes of blackouts and 'funny turns'

- Epilepsy
- Syncope:
  - Neurocardiogenic syncope (vasovagal)
  - Cardiac syncope (Stokes–Adams attacks)
  - Micturition syncope
  - Cough syncope
  - Postural hypotension
  - Carotid sinus syncope
- Non-epileptic attacks (dissociative seizures)
- Panic attacks and hyperventilation
- Hypoglycaemia
- Hydrocephalic attacks
- Basilar migraine
- Severe vertigo (e.g., Ménière's disease)
- Cataplexy, narcolepsy, sleep paralysis

The gold standard for diagnosis of non-epileptic attacks is a video-EEG recording, capturing a habitual attack and demonstrating that it is not accompanied by electrographic seizure activity. This may not be practically possible, especially as attacks can be infrequent, in which case an expert review of a video recording has reasonable diagnostic accuracy. There is no single reliable diagnostic feature to differentiate between epileptic and dissociative attacks, and some focal seizures such as those arising from the frontal lobe have very unusual features that can be a trap for the unwary.

Epileptic seizures however fall into a limited range of semiologies, and it is deviation from this pattern that raises the suspicion of dissociative attacks. Some indicators of a dissociative seizure include: a gradual onset with a waxing and waning course; eyes tightly closed; asynchronous bilateral limb movements; side-to-side head movements or pelvic thrusting; violent flailing movements of the limbs; swift recovery; subsequent recall for a period of unresponsiveness. Urinary incontinence and biting of the tip of the tongue (but less commonly severe biting involving the sides of the tongue) can be seen in NEAD (**Box 28.33**). Dissociative attacks are often more prolonged than epileptic seizure but this fact should not prevent the treatment of *status epilepticus* unless there is a high level of certainty around the diagnosis.

Skilful explanation of the diagnosis can be of significant therapeutic benefit. Dissociative attacks are events where the brain 'shuts down' or 'reboots', usually in response to being overwhelmed. Often the trigger, which can be emotional or physical (e.g., pain), may be difficult to identify; however, the brain can develop an automatic subconscious response over time where it goes into a dissociative state. Psychiatric co-morbidity may or may not be present, but where it coexists with NEAD it should be treated. Psychological therapies including cognitive behavioural therapy and mindfulness techniques such as grounding, and relaxation can be useful tools to treat dissociative attacks. There is further discussion of functional neurological disorders in the Acute medicine (Chapter 11) and Psychiatry (Chapter 30) chapters.

Sleep disorders

Myoclonic jerks when falling asleep are a normal phenomenon. Seizures may occur predominantly or solely during sleep.

Narcolepsy and cataplexy

Narcolepsy is caused by abnormalities of the brain neurotransmitter hypocretin (orexin), which is a regulator of sleep. CSF levels are usually low, thought in most cases to be due to autoimmune damage to the hypothalamic cells secreting the neurotransmitter. Narcolepsy is strongly associated with human leucocyte antigen (HLA)-DR2 and HLA-DQBl*0602 antigens. The prevalence is estimated at 30 to 50/100 000.

There are four main clinical features but not all patients have the full tetrad:

- ***Excessive daytime sleepiness (EDS).*** This is the usual presenting symptom and the main cause of disability. Patients have frequent irresistible sleep attacks during the day, often in inappropriate circumstances, for example during meals or conversations or while driving. EDS may be quantified with the Epworth Sleepiness Scale. Night-time sleep may be disrupted and, paradoxically, insomnia may occur.
- ***Cataplexy.*** This is sudden loss of muscle tone leading to head droop or even falling with intact awareness. Attacks are often set off by sudden surprise or emotion, for example laughter.
- ***Hypnagogic/hypnopompic hallucinations.*** These terms refer to dream-like hallucinations occurring while falling asleep or waking from sleep; these are often frightening.
- ***Sleep paralysis.*** A brief paralysis on waking or while falling asleep is due to intrusion of rapid eye movement (REM) atonia into wakefulness. This occasionally occurs in people without narcolepsy.

Diagnosis and management

Multiple sleep latency testing demonstrating a rapid transition from wakefulness to sleep and a short time to onset of REM sleep confirms the diagnosis.

Good sleep hygiene advice is necessary. Modafinil, dexamfetamine, and methylphenidate are used to treat EDS, often with only partial response. Antidepressants, particularly venlafaxine and clomipramine, can improve cataplexy. Sodium oxybate is also used and has effects against both narcolepsy and cataplexy.

Parasomnias

Disruptive motor or verbal behaviours occurring during sleep are divided into REM and non-REM parasomnias, depending on which stage of sleep they arise in. Non-REM parasomnias include sleepwalking, night terrors, and confusional arousals. REM parasomnias include recurrent isolated sleep paralysis (intrusion of REM-atonia into wakefulness) and REM sleep behaviour disorder (RBD), in which the mechanisms of REM sleep atonia fail and so dreams are enacted. RBD may be an early feature of Parkinson's disease.

Obstructive sleep apnoea

See Chapter 20 ('Obstructive sleep apnoea' section).

Restless leg syndrome (Willis–Ekbom disease)

Also known as Wittmaack–Ekbom's syndrome, this affects about 10% of adults, both women and men, and is also seen in children. Most people do not seek medical advice but it can be quite severe in 2% to 3%. The syndrome is characterized by an unpleasant sensation of 'wanting to move' the legs with throbbing or pulling, and usually occurs when the person is resting; it can be totally or partially relieved by stretching or walking. It occurs in the evenings or nighttime and appears to be more frequent in the elderly. The condition is usually idiopathic and there is a familial tendency. Low iron stores are a frequent association, and iron replacement can be very helpful. It has also been associated with pregnancy, neuropathy, spinal cord disease, and uraemia. Some drugs have been implicated.

For moderate to severe restless legs syndrome, gabapentin or pregabalin are now considered first-line, ahead of dopamine agonists. Some suggest a decreased use of alcohol or caffeine.

Further reading

Fisher RS, Cross, JH, French JA et al. Instruction manual for the ILAE 2017 operational classification of seizure types. *Epilepsia* 2017; 58:531–542.
Harris L, Angus-Leppan H. Epilepsy update: diagnosis, classification and management. *Medicine* 2023; 51:545–551.
Moshé SL, Perucca E, Ryvlin P et al. Epilepsy: new advances. *Lancet* 2015; 385:884–898.
Scammell TE. Narcolepsy. *N Engl J Med* 2015; 373:2054–2662.
http://www.epilepsysociety.org.uk. *National Society for Epilepsy, UK patient support group.*
http://www.narcolepsy.org.uk. *Narcolepsy UK.*

MOVEMENT DISORDERS

Disorders of movement are divided broadly into two categories:

- ***hypokinetic***—characterized by slowed movements with increased tone (parkinsonism)
- ***hyperkinetic***—excessive involuntary movements.

Both types may coexist; for example, in Parkinson's disease, where there are both slowed movements and tremor. Many of these disorders (not all) relate to dysfunction of the basal ganglia.

Parkinsonian disorders

Idiopathic Parkinson's disease

In 1817, James Parkinson, a physician in Hoxton, London, published *The Shaking Palsy*, describing this common worldwide condition that has a prevalence of 1/1000. Parkinson's disease (PD) is clinically and pathologically distinct from other parkinsonian syndromes.

Aetiology

The causes of idiopathic PD are still not fully understood. The relatively uniform worldwide prevalence suggests that a single environmental agent is not responsible. There may be multiple interacting risk factors, including genetic susceptibility.

Age and gender

Prevalence increases sharply with age, particularly over 70 years, with prevalence of 1 in 200 over age 80. Ageing changes are likely to be a factor in causation. Prevalence is higher in men (1.5:1 male to female).

Environmental factors

Epidemiological studies consistently show a small increased risk with rural living and drinking well water. Pesticide exposure has been implicated and pesticide-induced rodent models of PD exist, which increases biological plausibility of a link. The chemical compound methyl-phenyl tetrahydropyridine (MPTP), a potent mitochondrial toxin, causes severe parkinsonism, leading to suggestions that oxidative stress may be a factor leading to neuronal cell death in idiopathic PD. Studies consistently show that non-smokers have a higher risk of PD than smokers (even after controlling for shorter life expectancy in smokers), an observation that is difficult to explain.

Genetic factors

Idiopathic PD is not usually familial but twin studies show that there is a significant genetic component in early-onset PD (onset before 40). Several genetic loci for monogenic forms of PD have now been identified (**Box 28.34**), including autosomal dominant, autosomal recessive, and X-linked forms of PD. Most of these are rare but together they account for a reasonable proportion of early-onset and familial PD, and a small proportion (perhaps 1%–2%), of sporadic late-onset cases. The main significance of the monogenic forms of PD is that they provide insights into the pathophysiological mechanisms underlying PD that may be relevant to sporadic cases.

Aside from the rare disease-causing variants, polymorphisms in these and other genes may, in combination, constitute a susceptibility to PD that can be triggered by environmental factors or the ageing process. In particular, heterozygous variants in the glucocerebrosidase 1 *(GBA1)* gene are an important risk factor for PD (homozygous variants cause Gaucher disease).

Box 28.34 Selected Parkinson's disease (PD) genes

Gene	Protein	Inheritance	Comments
SNCA	α-Synuclein	AD	Rare but a major protein in Lewy bodies
LRRK2	Dardarin (a cytoplasmic kinase)	AD	The most common form of monogenic PD. Phenotype almost identical to sporadic PD. Found in 1% of apparently sporadic PD patients. High frequency in Jewish and North African Arab patients
PRKN	Parkin	AR	Responsible for most cases of juvenile PD and 20% of early-onset PD cases
PINK1	PTEN-induced kinase 1	AR	Rare. Protein involved in mitochondrial function

AD, Autosomal dominant; *AR,* autosomal recessive.

Pathology

The pathological hallmarks of PD are the presence of neuronal inclusions called Lewy bodies (see 'Dementia with Lewy bodies and Parkinson's disease dementia' section) and loss of the dopaminergic neurons from the pars compacta of the substantia nigra in the midbrain that project to the striatum of the basal ganglia (see **Fig. 28.12**). Lewy bodies contain tangles of α-synuclein and ubiquitin, and become gradually more widespread as the condition progresses, spreading from the lower brainstem and the olfactory bulb early in the disease, to the midbrain and then eventually into the cortex. Degeneration also occurs in other basal ganglia nuclei. The extent of nigrostriatal dopaminergic cell loss correlates with the degree of akinesia. Interestingly, α-synuclein deposition is also found in the sympathetic ganglia, the enteric nervous system, the GI tract itself, and even in the heart and skin. There is considerable interest in the involvement of the gut-brain axis in the pathogenesis of PD.

Clinical features

PD almost always presents with the typical motor symptoms of tremor and slowness of movement but it is likely that the pathological process starts many years before these symptoms develop. By the time of first presentation, on average 70% of dopaminergic nigrostriatal cells have already been lost.

Prodromal pre-motor symptoms

Patients develop a variety of non-specific non-motor symptoms during the approximately 7 years, sometimes longer, before the motor symptoms become manifest. These include:

- anosmia (present in 90%)—the olfactory bulb is one of the first structures to be affected
- constipation
- depression/anxiety (50%)
- aches and pains
- REM sleep behaviour disorder
- autonomic features—urinary urgency, hypotension
- restless legs syndrome.

Motor symptoms

These develop slowly and insidiously, and are often initially attributed to 'old age' by patients. The core motor features of PD are:

- bradykinesia
- tremor
- rigidity
- postural and gait disturbance.

Slowness causes difficulty rising from a chair or getting into or out of bed. Writing becomes small (micrographia) and spidery, tending to tail off. Relatives often notice other features: slowness and an impassive face. Idiopathic PD is almost always more prominent initially on one side. The diagnosis is usually evident from the overall appearance.

Bradykinesia

Poverty (akinesia) and slowing (bradykinesia) of movement are the cardinal clinical features of Parkinson's disease and the main cause of disability. The two terms are often used interchangeably. What distinguishes bradykinesia from slowness of movement from other causes is a progressive ***decrement*** in amplitude and speed of repetitive movements, such as opening and closing the hand or finger tapping.

There is difficulty initiating movement. The upper limb is usually affected first and symptoms are almost always unilateral for the first years. Rapid dexterous movements are impaired, causing difficulty writing (micrographia) and doing up buttons and zips, for example. Facial immobility gives a mask-like semblance of depression. Frequency of spontaneous blinking diminishes, producing a ***serpentine stare***.

Tremor

Tremor is the presenting symptom in 70% of patients. It typically starts in the fingers or hand (3–6 Hz) and is unilateral initially, spreading later to the leg on the same side and, after some years, to the opposite side. The tremor is present at rest, and reduces or stops completely when the hand is in motion. It is often described as ***pill-rolling*** because the patient appears to be rolling something between thumb and forefinger. As with most tremors, it is made worse by emotion or stress.

Rigidity

This is a sign rather than a symptom. Stiffness on passive limb movement is described as 'lead pipe', as it is present throughout the range of movement and, unlike spasticity, is not dependent on speed of movement. When stiffness occurs with tremor (not always visible), a ratchet-like jerkiness is felt, described as 'cogwheel' rigidity.

Postural and gait changes

A stooped posture is characteristic. Gait gradually becomes shuffling with small stride length, slow turns, freezing, and reduced arm swing. Postural stability eventually deteriorates, leading to falls, but this is a late-stage feature that should arouse suspicion of an alternative diagnosis if present during the first 5 years.

Speech and swallowing

Speech becomes quiet, indistinct, and flat. Drooling may be an embarrassing problem and swallowing difficulty is a late feature that may eventually lead to aspiration pneumonia as a terminal event.

Cognitive and psychiatric changes

Cognitive impairment is now recognized to be common in late-stage PD (80%) and may develop into dementia. Visual hallucinations and psychosis are not uncommon, and may herald evolving cognitive decline. Cholinesterase inhibitors may be helpful.

Depression is common, probably due to involvement of serotonergic and adrenergic systems, and is a cause of reduced quality of life in PD. Anxiety is also co-morbid with PD.

Clinical evolution of PD

Most patients respond well to treatment and there is generally a period of several years in which symptoms are well controlled with relatively little disability. Response to dopaminergic drugs is never lost but treatment-related fluctuations may develop (see below), which can be limiting, especially for patients with early age at onset. Eventually, usually by the mid-seventies, late-stage, treatment-unresponsive features, such as cognitive impairment, swallowing difficulty, loss of postural stability, and falls, start to emerge.

The rate of progression is very variable, with a benign form running over several decades. Usually, the course is over 10 to 20 years, death resulting from bronchopneumonia and immobility.

Diagnosis

There is no laboratory test; diagnosis is made by recognizing physical signs and distinguishing idiopathic PD from other parkinsonian syndromes. Patients with suspected PD should be referred to a specialist without initiation of treatment.

MRI is normal and not necessary in typical cases; it can provide clues about atypical parkinsonian disorders. Dopamine transporter imaging using SPECT (DaT scan) makes use of a radiolabelled ligand binding to dopaminergic terminals to assess the extent of nigrostriatal cell loss. It may occasionally be needed to distinguish PD from other causes of tremor or from drug-induced parkinsonism, but it cannot discriminate between PD and other akinetic–rigid syndromes. Cardiac MIBG scintigraphy demonstrates postganglionic cardiac sympathetic denervation in idiopathic Parkinson's disease and Dementia with Lewy Bodies (DLB), but not in other neurodegenerative disorders.

Management

Education about the condition is necessary and physical activity is beneficial and should be encouraged. Dopamine replacement with levodopa or a dopamine agonist improves motor symptoms and is the basis of pharmacological therapy. Treatment of non-motor symptoms, such as depression, constipation, pain, and sleep disorders, is also necessary and significantly improves quality of life.

Dopamine replacement may not always be needed in early-stage PD and is only started when symptoms begin to cause disability. The mechanism of action of drugs in PD is shown in Fig. 28.23.

Levodopa

Levodopa remains the most effective form of treatment and all patients with PD will eventually need it. It is combined with a dopa

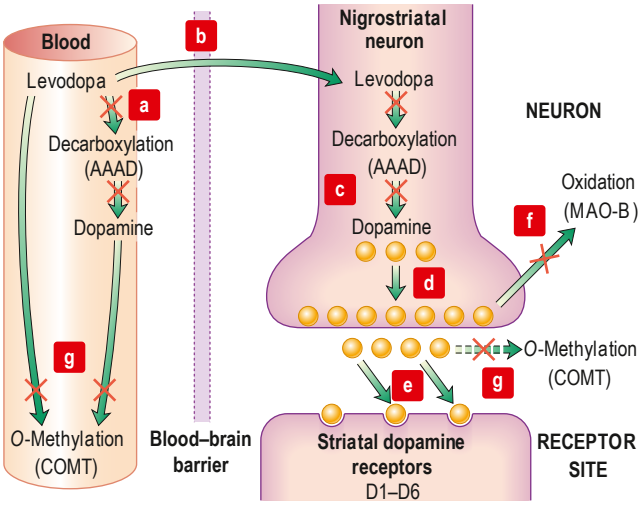


Fig. 28.23 Drugs in Parkinson's disease. Levodopa crosses the blood–brain barrier, enters the nigrostriatal neuron, and is converted to dopamine. **(a)** Carbidopa and benserazide reduce peripheral conversion of levodopa to dopamine, thus reducing the side-effects of circulating dopamine. **(b)** Dietary amino acids from high-protein meals can inhibit active transport across the blood–brain barrier by competing with levodopa. **(c)** Levodopa is converted (aromatic amino acid decarboxylase, AAAD) to dopamine. **(d)** Amantadine enhances dopamine release. **(e)** Dopamine agonists react with dopamine receptors. **(f)** Monoamine oxidase (MAO)-B inhibitors block dopamine breakdown. **(g)** Catechol-*O*-methyl transferase (COMT) inhibitors prolong dopamine activity by blocking breakdown.

decarboxylase inhibitor—benserazide (co-beneldopa) or carbidopa (co-careldopa)—to reduce the peripheral adverse effects (e.g., nausea and hypotension); 50 mg of levodopa (e.g., co-careldopa 62.5 mg) three times daily, increasing after 1 week to 100 mg three times daily, is a typical starting dose. The response is often dramatic.

Dopamine agonists

Dopamine agonists (DAs) may be used in combination with levodopa or as initial monotherapy in younger patients (below age 65–70) with mild to moderate impairment. Although less efficacious in symptom control than levodopa and generally less well tolerated, DAs are associated with fewer motor complications over a 5-year period. Non-ergot DAs (pramipexole and ropinirole or rotigotine via transdermal patch) are used in preference to ergot-derived drugs, which may be associated with fibrotic reactions, including cardiac valvular fibrosis.

Both DAs and levodopa can cause nausea, which abates with time. This is reduced by taking levodopa with meals upon initiation, but once established it should be taken on an empty stomach to maximize absorption. Domperidone can be used for short periods if required; the risk of cardiac side-effects precludes its routine use in all patients. Trimethobenzamide is an alternative in countries where domperidone is not available. Other antiemetics should not be used, as they may worsen symptoms by blocking central dopamine receptors.

Other drugs used in PD

- ***Monoamine oxidase (MAO)-B inhibitors*** (e.g., rasagiline, selegiline, and safinamide) reduce catabolism of dopamine in brain. They have a mild symptomatic effect, and the first two agents may be used in selected mild cases as monotherapy (or as adjuncts to levodopa). Safinamide is most often used as add-on therapy with levodopa to help with motor fluctuations.
- ***Amantadine*** has a modest anti-parkinsonian effect but is used mainly to improve dyskinesias in advanced disease.
- ***Antimuscarinics*** (e.g., ***orphenadrine***, ***procyclidine***, ***trihexyphenidyl***) may help tremor but are rarely used in PD except in younger patients. They have a high propensity to cause confusion and cognitive impairment in older patients.
- ***Apomorphine*** is a potent, short-acting DA administered subcutaneously by an autoinjector pen as intermittent 'rescue' injection for off periods or by continuous infusion pump. It is used in advanced PD.

Long-term response to treatment

As the disease progresses, medical therapy for PD becomes more difficult as higher doses of dopamine replacement therapy are required and response becomes more unpredictable with the development of motor fluctuations and dyskinesias; response to dopaminergic drugs is never lost, however.

Approximately 10% of patients per year develop motor complications in the form of 'wearing off' (the duration of effect of individual doses of levodopa becomes progressively shorter), dyskinesias (involuntary choreiform movements) and, eventually, 'on/off' phenomena (sudden, unpredictable transitions from mobile to immobile). Eventually, patients may alternate between the 'on' state with dopamine-induced dyskinesias and periods of complete immobility ('off').

Management of the motor complications of treatment represents one of the greatest challenges in the management of PD. Strategies include:

- dose fractionation of levodopa—increasing dose frequency
- addition of a catechol-*O*-methyl transferase (COMT) inhibitor such as ***entacapone*** or ***opicapone*** to prolong duration of action; entacapone is also available as a combined preparation with levodopa and carbidopa
- slow-release levodopa—mostly used for overnight symptoms, as absorption is erratic and difficult to predict, limiting effectiveness in control of daytime symptoms
- avoidance of protein-rich meals (which impair levodopa absorption) and taking doses at least 1 hour before or 2 hours after meals
- apomorphine continuous subcutaneous infusion (see above)
- deep brain stimulation (DBS) and levodopa intestinal gel (see below).

Deep brain stimulation

Stereotactic insertion of electrodes into the brain has proved to be a major therapeutic advance in selected patients (usually under age 70) with disabling dyskinesias and motor fluctuations not adequately controlled with medical therapy. Targets include:

- the STN—response similar to levodopa with reduction in dyskinesia
- the globus pallidus—improves dyskinesia but levodopa is still required for motor symptoms
- the thalamus—for tremor only.

Dopaminergic drugs can be reduced (but not withdrawn) after DBS. Approximately 10% of patients will be candidates at some point, usually patients with younger onset, as they have a higher rate of motor complications such as dyskinesia. There is a trend towards earlier use of DBS before motor complications become severe.

Levodopa intestinal gel infusion

Continuous infusion of this gel into the small intestine via a jejunostomy using a patient-operated pump is effective for selected patients with severe motor complications.

Tissue transplantation

Transplantation of embryonic mesencephalic dopaminergic cells directly into the putamen has produced mixed results but is potentially promising; research is ongoing to refine the technique. Stem cell and gene therapy approaches are in development.

Physiotherapy, occupational therapy, and physical aids

Physiotherapy, occupational therapy, and speech therapy all have a role to play in managing PD and reducing disability, speech and swallowing problems, and falls. Walking aids are often a hindrance early on, but later a frame or a tripod may help. A variety of external cueing techniques may help with freezing. Exercise is of proven benefit for patients with PD.

Other akinetic–rigid syndromes

Drug-induced parkinsonism

Dopamine-blocking or depleting drugs, particularly antipsychotics (with the exception of clozapine), can induce parkinsonism. These drugs also worsen symptoms in patients who already have PD, and may precipitate symptom onset in elderly patients in the presymptomatic phase. Antimuscarinic drugs reduce these symptoms, although tardive dyskinesia may be made worse.

Atypical parkinsonism

A number of neurodegenerative disorders affect the basal ganglia, causing prominent parkinsonism as part of the clinical picture; they may be mistaken for idiopathic PD in the early stages. These include:

- ***Multiple system atrophy (MSA).*** Autonomic and cerebellar symptoms occur in addition to parkinsonism. Either parkinsonism or cerebellar symptoms predominate; many patients have RBD. Stridor and sleep-disordered breathing can be dangerous features. Pathologically, α-synuclein-positive glial cytoplasmic inclusions are found in the basal ganglia, cerebellum, and motor cortex. Patients do not respond to levodopa therapy. Management is symptomatic and they have a poorer prognosis than patients with Parkinson's disease.
- ***Progressive supranuclear palsy*** (Steele–Richardson–Olszewski syndrome). This causes parkinsonism, postural instability with early falls, vertical supranuclear gaze palsy, pseudobulbar palsy, and dementia. Tau deposition is seen pathologically in the substantia nigra, STN, and midbrain ***(tauopathy).*** There is considerable clinical and pathological overlap between PSP and frontotemporal dementia (behavioural variant frontotemporal dementia, and non-fluent variant primary progressive aphasia; see 'Frontotemporal dementia' section).
- ***Corticobasal degeneration (CBD).*** Another tauopathy, in which alien limb phenomena, asymmetrical parkinsonism, apraxia, myoclonus, and dementia occur. Strictly speaking CBD is a postmortem diagnosis; the ***corticobasal syndrome*** can be associated with neuropathological features of CBD, PSP, or even Alzheimer's disease.

These disorders are relentlessly progressive; although they sometimes respond to levodopa, they usually cause death within a decade. 'Red flag' symptoms suggesting one of these disorders include:

- symmetrical presentation and absence of tremor
- levodopa unresponsiveness (or poor response)
- early falls (within first year)
- additional neurological features.

Hyperkinetic movement disorders

There are five hyperkinetic movement disorders, which can sometimes be difficult to separate from one another and may occur in combination.

- ***Tremor***—rhythmic sinusoidal oscillation of a body part.
- ***Chorea***—rapid, irregular movements flowing from one body part to another ('dance-like'); milder forms of chorea appear fidgety. ***Ballismus*** is a large amplitude, proximal flinging movement linked to chorea.
- ***Myoclonus***—brief, electric shock-like jerks.
- ***Tics***—stereotyped movements or vocalizations (may be temporarily suppressed).
- ***Dystonia***—sustained muscle spasms causing twisting movements and abnormal postures.

Dyskinesia encompasses a range of involuntary movements, including chorea, akathisia, and dystonia, but is often the preferred term when the movement disorder is drug-induced.

Essential tremor

This common condition, often inherited as an autosomal dominant trait, causes a bilateral, fast, low-amplitude tremor, mainly in the upper limbs. The head and voice are occasionally involved. Tremor is postural, such as when holding a glass or cutlery. Essential tremor occurs at any age. Tremor is slowly progressive but rarely produces severe disability. There may be a cerebellar-type action tremor component. Anxiety exacerbates the tremor, and alcohol suppresses it.

Treatment is often unnecessary or unsatisfactory. Many patients are reassured to find they do not have PD, with which essential tremor is often confused. Beta-blockers (propranolol), primidone or gabapentin may help. Sympathomimetics (e.g., salbutamol) make all tremors worse. Stereotactic thalamotomy and thalamic DBS are used in severe cases.

Chorea

There is a wide variety of possible causes of chorea. These include:

- systemic disease—hyperosmolar hyperglycaemic state, thyrotoxicosis, SLE, antiphospholipid syndrome, polycythaemia vera
- genetic disorders—Huntington's disease and genetic phenocopies, neuroacanthocytosis, benign hereditary chorea
- structural and vascular disorders affecting the basal ganglia
- drugs (e.g., levodopa, antipsychotics, and the oral contraceptive pill)
- post-infectious (Sydenham's chorea)—following months after streptococcal infection or as part of acute rheumatic fever
- pregnancy (chorea gravidarum).

Treatment is of the underlying cause but dopamine-blocking drugs, such as olanzapine and risperidone, and dopamine-depleting drugs (tetrabenazine), reduce chorea (as the prototypical *excessive* movement condition, the treatment is the *opposite* of PD).

Huntington's disease

Huntington's disease is a cause of chorea, usually presenting in middle life, initially with subtle 'fidgetiness' followed by development of progressive psychiatric and cognitive symptoms.

Oculomotor apraxia is an early clinical sign (delayed saccadic initiation).

Prevalence worldwide is about 5/100 000. HD is due to a CAG trinucleotide repeat expansion, which forms the basis of the diagnostic test. This results in translation of an expanded polyglutamine repeat sequence in huntingtin, the protein gene product, the function of which is unclear. The expansion is thought to be a toxic 'gain-of-function' mutation. Most adult-onset patients have 36 to 55 repeats and there is an inverse relationship between repeat length and age at onset, with juvenile-onset patients having over 60 repeats. Expansion of the unstable CAG repeat during meiosis, particularly spermatogenesis, is the molecular basis for the phenomenon of anticipation (a tendency for successive generations to have earlier onset and more severe disease), particularly when inherited from the father.

HD is inherited in an autosomal dominant manner with complete penetrance (all gene carriers will develop the disease eventually). Sufferers are often unaware of previous family history, owing to historical family breakdown. There is no disease-modifying treatment at present, although chorea can improve with treatment, such as risperidone or tetrabenazine, but progressive neurodegeneration leads to dementia and ultimately death after 10 to 20 years. Patients with small or intermediate-range expansions may present in old age with isolated chorea.

Absence of treatment results in a low take-up rate for presymptomatic testing in at-risk individuals. Test centres have protocols for counselling families and addressing ethical issues. Much research is being conducted into disease-modifying therapies, and an antisense oligonucleotide has been shown to engage with the target and lower concentrations of mutant huntingtin protein in CSF, but larger trials were halted early because of worsening clinical outcomes or failure to lower mutant HTT.

Hemiballismus

Hemiballismus describes violent swinging movements of one side, usually caused by infarction or haemorrhage in the contralateral STN.

Acute hemichorea–hemiballism also occurs after diabetic hyperosmolar hyperglycaemia, with signal change seen in the contralateral striatum on CT or MRI; it is thought to be due to osmotic shifts causing myelinolysis.

Myoclonus

Myoclonus refers to a brief, shock-like muscle contraction (or a brief loss of muscle tone—negative myoclonus). The word may indicate a movement disorder or an epileptic phenomenon. Myoclonus can arise from virtually every part of the nervous system, from the cortex to the peripheral nerve.

Cortical myoclonus is usually distal (hands and fingers especially) and stimulus sensitive (spontaneous but also triggered by touch or loud noises). It is caused by a wide variety of pathologies affecting the cerebral cortex; ***spinal and brainstem myoclonus*** are caused by localized lesions affecting these structures.

Primary myoclonus

Physiological myoclonus. Nocturnal myoclonus (hypnic jerks) consists of sudden jerks (often with a feeling of falling) on dropping off to sleep or waking; it is common and not pathological. Hiccups are a form of myoclonus affecting the diaphragm. The startle response is also a form of brainstem myoclonus.

Myoclonus-dystonia (DYT11). Myoclonic 'lightning jerks', often with dystonia, are inherited as a rare autosomal dominant disorder due to mutations in the ε-sarcoglycan gene.

Myoclonus in epilepsy

Myoclonic jerks occur in several forms of epilepsy, either as a primary seizure type, or as a fragment of epilepsy (see 'Epilepsy and loss of consciousness' section). An antiepileptic drug, such as valproate, may be helpful.

Progressive myoclonic epilepsy–ataxia syndromes

These rare conditions include genetic and metabolic disorders in which myoclonus accompanies progressive epilepsy, cognitive decline, and/or ataxia. ***Lafora body disease***, ***neuronal ceroid lipofuscinosis***, and ***Unverricht–Lundborg*** disease are examples.

Secondary myoclonus

Myoclonus may be seen in a wide variety of metabolic disorders, including hepatic and renal failure (asterixis—negative myoclonus), as part of several dementias and neurodegenerative disorders (e.g., Alzheimer's disease), and encephalitis.

Post-anoxic myoclonus can follow severe cerebral anoxia. A distinction must be drawn between early post-anoxic myoclonic status epilepticus, which is a very poor prognostic sign following hypoxic–ischaemic brain injury (HIBI), and delayed post-hypoxic Lance Adams myoclonus, which arises when the patient is awake and causes disabling action myoclonus but does not portend death.

Tics

Tics are common (15% lifetime prevalence), brief, stereotyped movements usually affecting the face or neck but which may involve any body part; they include vocal tics. Unlike other movement disorders, they may be transiently suppressed, leading to a build-up of anxiety and overflow after release.

Simple transient tics (e.g., blinking, sniffing, or facial grimacing) are common in childhood but may persist. Adult-onset tics are rare and usually have a secondary cause, or are functional. The borderland between normal and pathological is vague.

Tourette's syndrome

The most common cause of tics, characterized by multiple motor tics and at least one vocal tic, starts in childhood and persists longer than a year. Boys are affected more often than girls in a 3:1 ratio. Behavioural problems, including attention-deficit/hyperactivity disorder (ADHD) and obsessive–compulsive disorder (OCD), are common and may sometimes be the major cause of disability. There is sometimes explosive barking and grunting of obscenities (coprolalia) and gestures (copropraxia) or echolalia (copying what other people say). Many affected individuals never come to medical attention. The cause is not known but it may be a complex problem with changes in various neurotransmitters.

Dystonias

Dystonia (Box 28.35) is most usefully classified by aetiology into:

- ***primary dystonias***—where dystonia is the only, or main, clinical manifestation (usually genetic, or idiopathic)
- ***secondary dystonia***—due to brain injury, cerebral palsy, or drugs, for example
- ***hereditary degenerative dystonia***—part of a wider neurodegenerative disorder
- ***paroxysmal dystonias***—rare, mostly genetic, attacks of sudden involuntary movements with elements of dystonia and chorea.

Primary dystonias

Young-onset

DYT1 dystonia: mutations in the *DYT1* gene locus, seen particularly in the Ashkenazi Jewish population, cause limb-onset dystonia

Box 28.35 A classification of dystonias

- Generalized dystonia
- Primary torsion dystonia (PTD)
- Dopamine-responsive dystonia (DRD)
- Drug-induced dystonia (e.g., metoclopramide)
- Symptomatic dystonia (e.g., after encephalitis lethargica or in Wilson's disease)
- Paroxysmal dystonia (very rare, familial, with marked fluctuation)
- Focal dystonia
- Spasmodic torticollis
- Writer's cramp
- Oromandibular dystonia
- Blepharospasm
- Hemiplegic dystonia, e.g., following stroke
- Multiple sclerosis—rare

(usually foot), before the age of 28. Penetrance is low (autosomal dominant) and phenotype very variable, but the disorder often spreads over years to become generalized, and can result in severe disability. Cognitive function remains normal. The condition is rare and the definitive form of treatment for severe cases is DBS.

Dopa-responsive dystonia: in this rare disorder, dystonia is completely abolished by small doses of levodopa. Typically, dystonic walking begins in childhood and may resemble a spastic paraparesis or even present as cerebral palsy. Dominantly inherited mutations in the *GTP* cyclohydrolase gene on chromosome 14q26.3 (necessary for synthesis of a co-factor—tetrahydrobiopterin—needed for dopamine synthesis) lead to brain dopamine deficiency. Patients with dystonic gaits are sometimes given test doses of levodopa.

Adult onset

This is the most common type of primary dystonia. Onset is usually around 55 and dystonia is usually focal (restricted to one body part), particularly affecting the head and neck, unlike *DYT1* dystonia. Various patterns are recognized.

Torticollis—dystonic spasms gradually develop in neck muscles, causing the head to turn (torticollis) or to be drawn backwards (retrocollis). There may also be a jerky head tremor. A gentle touch with a fingertip at a specific site may relieve the spasm temporarily (sensory trick or 'geste antagoniste').

Writer's cramp and task-specific dystonias—there is a specific inability to perform a previously highly developed, repetitive, skilled movement, such as writing. The movement provokes dystonic posturing. Other functions of the hand remain normal. Overuse may lead to task-specific dystonias in certain occupations, such as musicians, typists, and even golfers.

Blepharospasm and oromandibular dystonia—these consist of spasms of forced blinking or involuntary movement of the mouth and tongue (e.g., lip-smacking and protrusion of the tongue and jaw). Speech may be affected.

Neuroleptics and movement disorders

Neuroleptics (antipsychotic drugs used to treat schizophrenia) and related drugs used as antiemetics (e.g., metoclopramide) can cause a variety of movement disorders:

- ***Akathisia.*** This is a restless, repetitive, and irresistible need to move.
- ***Parkinsonism.*** This is due to D_1 and D_2 dopamine receptor blockade (see above).
- ***Acute dystonic reactions.*** Spasmodic torticollis, trismus, and oculogyric crises (episodes of sustained upward gaze) develop, dramatically and unpredictably, after single doses.
- ***Tardive dyskinesia.*** These choreiform movements, often affecting the face and mouth with mouthing and lip-smacking grimaces, occur after several years of neuroleptic use. They often become temporarily worse when the drug is stopped or the dose reduced. Even if treatment ceases, resolution seldom follows. Atypical neuroleptics are less prone to cause this complication. Procyclidine, often prescribed for neuroleptic-induced parkinsonism, can make tardive dyskinesia more pronounced.

Wilson's disease

This rare and treatable disorder of copper metabolism is inherited as an autosomal recessive trait. Copper deposition occurs in the basal ganglia, cornea, and liver (see Chapter 23, 'Wilson's disease [progressive hepatolenticular degeneration]' section), where it can cause cirrhosis. Copper-coloured Kayser-Fleischer rings in the cornea are pathognomonic. All patients below age 50 with any hyperkinetic movement disorder or with liver cirrhosis should be screened for Wilson's disease (serum copper and caeruloplasmin should be checked). Intellectual impairment develops. Neurological damage is reversible with early treatment. Diagnosis and treatment with the chelating agent penicillamine are discussed in Chapter 23 ('Wilson's disease [progressive hepatolenticular degeneration]' section).

Further reading

Kalia LV, Lang AE. Parkinson's disease. *Lancet* 2015; 386:896–912.
Obeso JA, Rodriguez-Oroz MC, Stamelou M et al. Expanding the universe of disorders of basal ganglion. *Lancet* 2014; 384:523–531.
Schapira AH, Olanow CW, Greenamyre JT et al. Slowing of neurodegeneration in Parkinson's disease and Huntington's disease: future therapeutic perspectives. *Lancet* 2014; 384:545–555.
Stoessi AJ, Leherly S, Strafella AD. Imaging insights into basal ganglion function, Parkinson's disease and dystonia. *Lancet* 2014; 384:532–544.

NEUROINFLAMMATORY DISORDERS

MS is by far the most common neuroinflammatory disorder in Western populations. Other CNS inflammatory conditions include post-infectious disorders such as acute disseminated encephalomyelitis (ADEM) and transverse myelitis (TM), distinct autoimmune disorders such as NMO and multisystem inflammatory disorders such as sarcoidosis and Behçet's disease.

Multiple sclerosis

Debates about the causes and mechanisms of MS still rage, but inflammation, myelin destruction, and axonal loss combine to generate neurological injury. Multiple plaques of demyelination are found throughout the brain and spinal cord, occurring sporadically over years (dissemination in space and time is crucial for diagnosis).

MS is a major cause of disability in young adults but several disease-modifying therapies are now available.

Epidemiology

- ***Prevalence.*** MS is a common neurological disorder with over 2 800 000 cases worldwide. In the UK, the prevalence is 1.2/1000. Approximately 80 000 people in the UK have MS.
- ***Gender.*** Women outnumber men by 2:1. There is evidence that this ratio is widening, with an increasing proportion of women being affected (partly accounted for by the increase in women smoking).
- ***Age.*** Presentation is typically between 20 and 40 years of age. Presentation after age 60 is rare, although diagnosis may sometimes be much delayed, occurring years after the initial symptoms.

Prevalence varies widely in different geographical regions and ethnic groups. This probably reflects both genetic (see below) and environmental influences in pathogenesis. MS is much more common in white populations and with increasing distance from the equator. Even within the UK, there is a north–south divide, prevalence being higher in Scotland than in southern England. Migration studies show that children moving from a low-risk to a high-risk area (e.g., the UK) develop a higher risk of MS, similar to the population of the country to which they migrate, indicating that environmental factors are a factor in pathogenesis.

Certain other autoimmune disorders occur with increased frequency in patients with MS and their relatives, indicating a genetic predisposition to autoimmunity.

Aetiology and pathogenesis

In its early stages, MS is a primarily immune-mediated disorder, with T cells, B cells, and macrophages interacting to cause focal demyelination; eventually neurodegeneration dominates. Traditionally viewed as a disease of white matter, grey matter involvement is increasingly recognized. The aetiology of MS is complex and not yet fully understood.

Genetic susceptibility

Multiple genes interact to confer an increased risk of MS, giving a complex polygenic inheritance pattern. Genetic differences between different populations probably account for part of the observed variation in MS incidence around the world.

Family studies show that there is a much-increased risk of MS in first-degree relatives of affected patients (approximately 5% lifetime risk of developing MS). Twin studies confirm a major genetic component to susceptibility, with 30% of monozygotic twins being concordant for MS in the UK and Canada versus 5% of dizygotic twins. However, even the twin concordance rate is much lower in southern latitudes, pointing to a gene-environment interaction.

Genes. Over 200 polymorphisms have been identified as conferring an increased risk of MS; many are close to genes encoding proteins involved in immune function and regulation, including the human leukocyte antigen (HLA) complex. The biggest risk is associated with the HLA-DRB1*15:01 allele, which appears to be associated with an earlier disease onset.

Environmental factors

Migration studies (see above) and twin studies indicate that environmental factors play a role in the development of MS but these factors are still largely unknown. Viral infections can precipitate MS relapses, and exposure to infectious agents at critical times in development may trigger MS in genetically susceptible individuals. There is evidence that exposure to Epstein–Barr virus (EBV) may be linked to MS; EBV seropositivity is higher in patients with MS than in the general population. Another theory states that exposure to infectious agents in childhood may reduce the risk of developing MS and other autoimmune disorders (the 'hygiene hypothesis'). There is good evidence that low levels of vitamin D and/or lack of sunlight exposure increase risk of MS, as does smoking tobacco.

Pathology

Plaques of demyelination, 2 to 10 mm in size, are the cardinal features (Fig. 28.24). Plaques occur in a perivenular distribution and arise anywhere in CNS white matter, but there is a predilection for distinct CNS sites: optic nerves, the juxtacortical and periventricular region, the corpus callosum, the brainstem and its cerebellar connections, and the cervical cord (corticospinal tracts and dorsal columns). MRI studies show that most inflammatory plaques are asymptomatic. Recent advances in pathology and MRI techniques show that the grey matter of the cortex is also affected from an early stage in MS, and normal-appearing grey and white matter can be diffusely damaged. Peripheral myelinated nerves are not directly affected in MS.

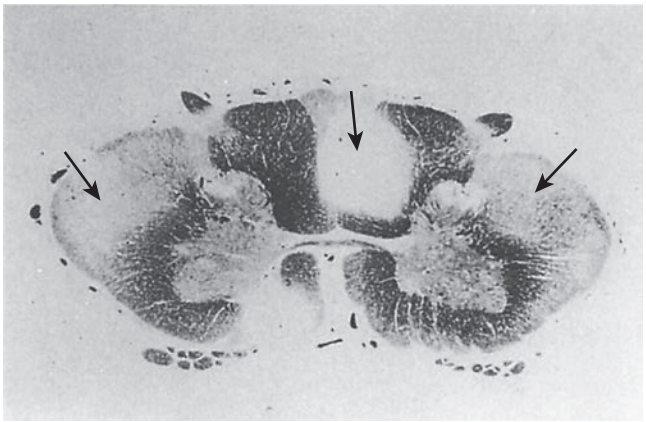


Fig. 28.24 Multiple sclerosis (MS). Cross-section of spinal cord showing MS plaques in posterior column and lateral corticospinal tracts *(arrowed)*. *(Courtesy of the late Professor Ian Macdonald.)*

Acute relapses are caused by focal inflammation, with a predominant T-cell infiltrate, causing oligodendrocyte damage and hence demyelination. Recovery follows as inflammation subsides and remyelination occurs. As the disease progresses, microglial activation and secondary permanent axonal destruction occur. Neurodegeneration is the pathological basis of the progressive disability seen in progressive forms of MS. The extent of grey matter damage correlates with the severity of disability and cognitive involvement. The exact relationship between the inflammatory lesions seen in early relapsing–remitting forms of MS and the progressive axonal loss of chronic forms of MS is disputed.

Clinical features

No single group of signs or symptoms is diagnostic. A wide variety of possible symptoms may occur, depending on the anatomical site of the lesions; MS has been described as a 'great imitator'. The clinical time course of attacks and tempo of evolution of symptoms are as helpful as the symptoms themselves in making the diagnosis of MS.

Types of MS

MS is divided into four clinical phenotypes (Fig. 28.25).

- ***Clinically isolated syndrome (CIS).*** See below. A neurological episode that is typical of MS, without prior attacks. The time course is characteristic: onset over days followed by a plateau for several weeks, and then typically recovery, either partial or (ostensibly) complete, over weeks.
- ***Relapsing–remitting MS (RRMS).*** Symptoms occur in attacks (relapses), disseminated in space and time. On average, patients have one relapse per year but occasionally many years may separate relapses (so-called benign MS—10% of patients, but the possibility of eventual severe worsening is not excluded). For every clinical relapse, there will be about 10 subclinical lesions noted on MRI brain. Neuronal damage accrues even when recovery seems complete, and eventually will translate into mounting disability.

- ***Secondary progressive MS (SPMS).*** This stage of MS consists of gradually worsening disability progressing slowly over years. On average, people with MS reach a progressive phase 10 to 20 years after onset. The transition is not sharp and there will be a gradual change from a predominance of (inflammatory) relapses to a predominance of progression; there can be movement from progression to relapses and back again. Patients with SPMS can be subcategorized as *active* (with clinical relapses or new MRI lesions) or not active. Patients with active SPMS may benefit from disease-modifying treatment.
- ***Primary progressive MS (PPMS)*** (10%–15%). This pattern is characterized by gradually worsening disability from the onset. Some patients may have superimposed relapses or evidence of new inflammatory MRI lesions, denoting *active* disease (the term 'relapsing-progressive MS' is no longer used). It typically presents later and is associated with fewer inflammatory changes on MRI.

In reality the relapsing and progressive forms of the disease exist along a continuous spectrum. There is increasing evidence that progression begins early in the disease course, and inflammatory relapses may re-emerge even after a period of progression independent of relapse activity. There is a move towards using a more flexible phenotypic classification, with the above categories modified by a description of inflammatory *activity* or inactivity (relapses or new MRI lesions), and the presence or absence of *progression*.

Severity is described by various terms, the most important being 'highly active MS' (unchanged or increased relapse rate or ongoing severe relapses despite treatment; or one relapse in the previous year and MRI evidence of disease activity) and 'rapidly evolving severe MS' (two disabling relapses within the last year and MRI evidence of significant disease activity).

Clinical presentations

Three characteristic common presentations of MS are optic neuropathy (neuritis), brainstem demyelination, and spinal cord lesions.

Optic neuritis

See 'Inflammatory optic neuropathy (optic neuritis)' section.

Brainstem demyelination

A relapse affecting the brainstem causes combinations of diplopia, vertigo, facial numbness/weakness, dysarthria, and dysphagia.

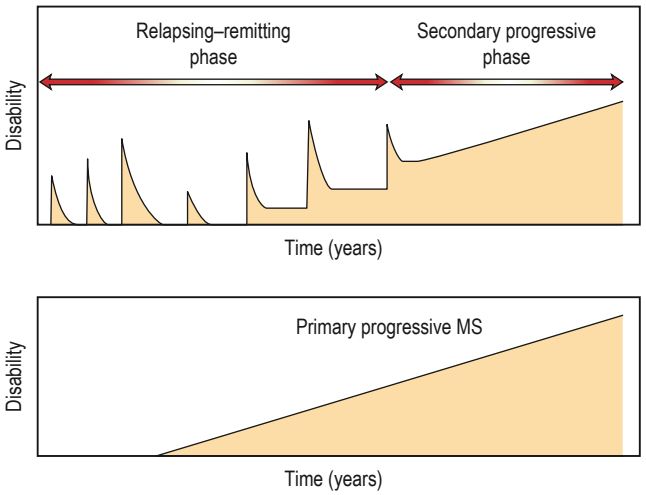


Fig. 28.25 Clinical patterns of multiple sclerosis.

Pyramidal signs in the limbs occur when the corticospinal tracts are involved. A typical picture is sudden diplopia, and vertigo with nystagmus, but without tinnitus or deafness. Bilateral INO (see 'Internuclear ophthalmoplegia' section) is virtually pathognomonic of MS.

Spinal cord lesions

Paraparesis developing over days or weeks is a typical result of a plaque in the cervical or thoracic cord, causing difficulty in walking and limb numbness with tingling, often asymmetric. Lhermitte's sign may be present (see 'Lhermitte's phenomenon' section). Sometimes, the arms are also involved in high cervical cord lesions. A tight band sensation around the abdomen or chest is common with thoracic cord lesions (the 'MS hug').

Common symptoms in MS

Disability and neurological impairments accumulate gradually over the years. Several symptoms are common and many can be improved with symptomatic treatments (Box 28.36).

Box 28.36 Symptomatic treatments in multiple sclerosis

Urinary symptoms
- Antimuscarinics for overactive bladder e.g., oxybutynin, tolterodine, solifenacin, trospium
- Alpha-blockers for urinary retention
- Desmopressin spray for nocturia
- Intermittent self-catheterization (ISC)
- Botulinum toxin type A intravesical injections (usually also require ISC)
- Sacral nerve stimulator
- Bladder training exercises
- Indwelling catheter

Spasticity
- Self-management, including stretching, physiotherapy, splinting
- Skeletal muscle relaxants: baclofen, tizanidine, clonazepam
- Gabapentin
- Botulinum toxin type A (electromyelography-guided injections) for focal spasticity
- Cannabis: oromucosal spray
- Intrathecal baclofen pump
- Intrathecal phenol—destructive procedure in advanced disease with paraplegia

Pain
- As for treatment of neuropathic pain
- Trigeminal neuralgia—carbamazepine, lamotrigine
- Lhermitte's—carbamazepine

Depression
- Cognitive behavioural therapy
- Antidepressants

Impaired mobility
- Physiotherapy ± walking aids
- Treat spasticity
- Functional electrical stimulation
- Fampridine—selected patients

Erectile dysfunction
- Sildenafil, tadalafil, or vardenafil

Tremor
- Beta-adrenoceptor-blocking drugs
- Botulinum toxin type A injections (head or arms)
- Deep brain stimulation for severe Holmes tremor

Fatigue
- Fatigue management programme, treat depression
- Amantadine, modafinil (often ineffective)

- Visual changes
- Sensory symptoms—often unusual, for example a sensation of water trickling down the skin. Sensory symptoms are the presenting feature in 40% of patients. Reduced vibration sensation and proprioception in the feet are among the most common abnormalities on examination, but examination may be normal despite significant sensory symptoms
- Clumsy/useless hand or limb—due to loss of proprioception (often a dorsal column spinal plaque)
- Unsteadiness or ataxia
- Weakness—can have a give-way character in MS
- Urinary symptoms—bladder hyper-reflexia causing urinary urgency and frequency, and/or urinary retention. Urinary tract infections are very common in people with MS
- Pain—neuropathic pain is common
- Fatigue—a common and often debilitating symptom, which can occur in patients with otherwise mild disease
- Spasticity
- Depression
- Cognitive impairment
- Sexual dysfunction
- Temperature sensitivity—temporary worsening of pre-existing symptoms with increases in body temperature, for example after exercise or a hot bath, is known as Uhthoff's phenomenon.

Unusual presentations

Epilepsy and trigeminal neuralgia (see 'Trigeminal neuralgia' section) occur more commonly in MS patients than in the general population. Paroxysmal symptoms, such as tonic spasms, are sometimes seen and are thought to represent ephaptic transmission along damaged axons.

Late-stage MS

Late MS causes severe disability with spastic tetraparesis, ataxia, optic atrophy, nystagmus, brainstem signs (e.g., bilateral INO), pseudobulbar palsy, and urinary incontinence. Cognitive impairment, often with frontal lobe features, can be prominent.

Diagnosis

Few other neurological diseases have a similar relapsing and remitting course. The diagnosis of MS requires two or more attacks affecting different parts of the CNS: that is, dissemination in time and space, and exclusion of other possible causes. History and support from investigations, particularly MRI, make the diagnosis. The 2017 McDonald criteria formalize the diagnostic features and can be useful in clinical practice when only one clinical attack has occurred, but there is ancillary evidence of dissemination in space and time (by MRI and/or CSF criteria). New revisions to the McDonald criteria are expected to be published in full in 2025.

When taking a history at the time of initial presentation, it is essential to ask about previous episodes of neurological symptoms, often years previously, that may represent episodes of unrecognized demyelination: for example, a severe episode of vertigo lasting weeks or loss of vision in one eye that gradually recovered.

Investigations

The purpose of investigations is to provide supportive evidence of dissemination in time and space (i.e., to show scattered demyelinating lesions that evolve over time), to exclude other diseases and to provide evidence of immunological disturbance.

- MRI of brain and cord is the definitive investigation, as it demonstrates areas of demyelination with high sensitivity. Multiple scattered plaques are usually seen (Fig. 28.26A), typically oval in shape, up to 2 cm in diameter, and often orientated perpendicular to the lateral ventricles. Acute lesions show gadolinium enhancement for 6 to 8 weeks. Occasionally, large 'tumefactive' (swollen) lesions are seen (Fig. 28.26B).
- Although a sensitive technique to demonstrate plaques (normal MRI in MS is possible but distinctly rare), it is limited by lower specificity. Over the age of 50, small ischaemic lesions may be difficult to distinguish from demyelination, and in younger patients other neuroinflammatory disorders such as sarcoidosis, Behçet syndrome, and vasculitis may produce similar imaging appearances. The presence of spinal cord lesions is quite specific for inflammatory disorders such as MS rather than ischaemic lesions, so cord imaging is often useful where there is diagnostic difficulty.
- Plaques are rarely visible on CT.
- CSF examination is often unnecessary with suggestive MRI and a compatible clinical picture. CSF analysis shows oligoclonal IgG bands in over 90% of cases but these are not specific for MS. The presence of unpaired oligoclonal bands, in CSF but not serum, provides evidence for dissemination in time as per the McDonald criteria. The CSF cell count may be raised during an acute relapse (5–60 mononuclear cells/mm^3).
- Evoked responses, for example visual-evoked responses in optic nerve lesions, may demonstrate clinically silent lesions. However, since the advent of MRI, they are less used in diagnosis.
- Blood tests are used to exclude other inflammatory disorders such as sarcoidosis or SLE, or other causes of paraparesis, for example adrenoleukodystrophy, HIV, human T-cell lymphotropic virus 1 (HTLV-1), and vitamin B_{12} deficiency.

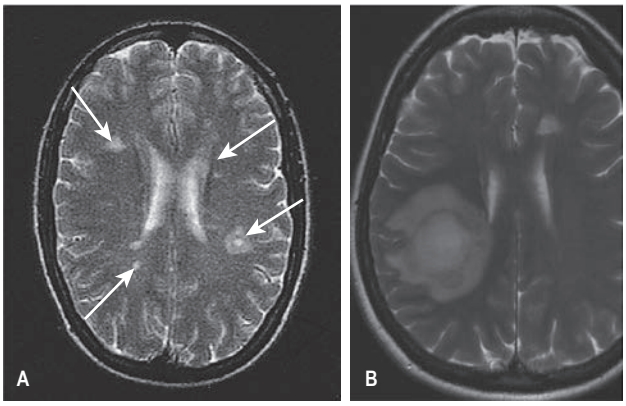


Fig. 28.26 Multiple sclerosis. (A) T2-weighted magnetic resonance image. Brain plaques are shown *(arrowed)*. Typical lesions are oval in shape, up to 2 cm in diameter, and often orientated perpendicular to the lateral ventricles. **(B)** Tumefactive multiple sclerosis lesion.

The clinically isolated syndrome

A first-ever episode of neurological symptoms suggestive of neuroinflammation is termed a 'clinically isolated syndrome' (CIS). A second clinical event indicative of a new lesion in a different anatomical location allows the diagnosis of MS to be confirmed. Alternatively, MS can still be diagnosed after a single attack, as per the 2017 McDonald criteria, if ancillary evidence is available to demonstrate dissemination in space and time.

- *Dissemination in space* can be demonstrated on MRI if lesions are present in at least two of the four MS-typical regions of the nervous system (periventricular, juxtacortical, infratentorial, and spinal cord).

- *Dissemination in time* can be inferred from the presence of both gadolinium-enhancing (acute) and non-enhancing (old) lesions, or by the appearance of new T2 hyperintense lesions on an interval scan.

Fewer people are diagnosed with CIS nowadays because the 2017 criteria have broadened the definition of MS. Occasionally patients with CIS who are felt to be at high risk of conversion to MS may be offered disease-modifying treatment from the time of first presentation.

The radiologically isolated syndrome

A radiologically isolated syndrome (RIS) is when MRI findings are highly suggestive of MS, with dissemination in space, but the patient has never had a clinical attack suggestive of MS. MS cannot be diagnosed unless and until a clinical event occurs; surveillance is usually suggested.

Management

The treatment landscape of MS has been transformed over the last decade. We now have highly effective therapies, that have high rates of success in relapse prevention. Although long-term data are limited, there is a growing body of evidence that early intensive treatment delays the onset of disability and slows secondary progression. The majority of treatments are still directed at relapsing–remitting MS, but there is now evidence of efficacy for some agents in active secondary progressive MS and early primary progressive MS.

General measures

Education, provision of appropriate written materials, and support from a multidisciplinary team, including an MS nurse specialist, are essentials. Physiotherapy and occupational therapy are helpful where there is persisting impairment between relapses. Infections should be treated early, as they may precipitate relapses or lead to worsening of existing symptoms (pseudo-relapses); urinary tract infections are particularly problematic for people with MS. Immunizations are safe and recommended (but not live vaccines if on disease-modifying drugs).

Acute relapses

Short courses of steroids, such as i.v. methylprednisolone 1 g/day for 3 days or high-dose oral steroids, are used for severe relapses. They speed recovery but do not influence long-term outcome.

Disease-modifying treatments

A 'watch-and-wait' strategy used to be common practice but is now rare. Newly diagnosed patients with RRMS (including those with a single clinical episode who meet McDonald criteria) should receive disease-modifying treatment (DMT, **Box 28.37**).

- ***Immunomodulatory drugs***, include the older generation of injectables (β-interferon, glatiramer acetate) and some oral agents (teriflunomide, dimethyl fumarate). These agents are of moderate efficacy. They reduce relapse rate by about one-third; the efficacy of dimethyl fumarate is slightly higher. The toxicity profile is better than with the newer high-efficacy agents, and they may be preferred for selected patients for this reason.
- ***Immunosuppressant drugs*** include oral agents and monoclonal antibodies.
  - *Oral:* Sphingosine 1-phosphate receptor modulators (fingolimod, siponimod) sequester lymphocytes in lymph nodes. These agents are of higher efficacy than the older generation injectables.
  - *Monoclonal antibodies* for MS are all of very high efficacy, reducing relapse rate by up to 70%. Natalizumab inhibits migration of leukocytes into the CNS. In the UK, it is approved for *rapidly evolving severe MS*. Patients who are seropositive for JC virus antibodies are at risk of developing the much-feared complication of progressive multifocal leukoencephalopathy (see Chapter 42, 'Progressive multifocal leucoencephalopathy' section). Ocrelizumab (intravenous) and ofatumumab (subcutaneous) are anti-CD20 monoclonal antibodies that deplete B-cells, binding to sites distinct from rituximab (also used for MS occasionally, off-label). The efficacy of these agents has challenged the conventional 'T-cell-driven' view of MS. In general, treatment with these agents is continued long term, with attendant risks of long-term immunosuppression, and indications for stopping are unclear.
- ***Immune reconstitution/induction therapies*** are given as short courses with the aim of 'resetting' the immune system. Alemtuzumab is a humanized anti-CD52 monoclonal antibody that non-selectively depletes white cells. It has very high efficacy, but secondary autoimmunity limits its use to the more severe cases. Cladribine, given orally, is a purine antimetabolite that selectively depletes T- and B-cells. It is of 'high' (rather than 'very high') efficacy. Autologous haematopoietic stem cell transplantation (AHSCT) is available for a small number of patients with highly active disease who have not responded to other treatments.
- ***Progressive disease:*** in the UK, historically DMTs have only been offered to ambulant patients with RRMS, but recent developments have expanded the options for patients with progressive disease. Siponimod is licensed for use in active SPMS, and ocrelizumab for early PPMS. Interferon β-1b is also licensed for active SPMS (with relapses), and can be used for this purpose when siponimod is contraindicated.

Eligibility criteria in the UK are complex and ever-changing, and access to different drugs varies across the world. The stepwise approach, of lower efficacy treatments first followed by more potent drugs if there is breakthrough disease, is now discouraged in favour of starting with intensive treatment from diagnosis, especially in highly active disease. Cost is a major barrier to treatment in the developing world.

Pregnancy does not affect long-term outcomes in MS, although there is an increase in relapse rate in the first few months postpartum. Effective therapy for MS should not be delayed because women may want families in the future; preconceptual counselling is vital, and an individualized plan is required for women trying to conceive. Several but not all DMTs need to be avoided in pregnancy (beta interferon and glatiramer acetate are safe; continuation of natalizumab can be considered as there is a high risk of rebound if stopped).

Other drugs and symptomatic therapies

- ***Vitamin D.*** There is some evidence that vitamin D supplementation may be beneficial in MS but this is still controversial. Some specialists advocate treatment with up to 4000 IU daily.
- ***Fampridine*** (4-aminopyridine, an oral potassium-channel blocker). This has been shown to improve walking significantly in selected patients with significant MS-related disability.
- ***Other symptomatic treatments.*** Symptomatic treatments are available for most complications of MS and significantly improve quality of life (see **Box 28.36**). Urinary urgency and frequency, pain, spasticity, and depression are all common and treatable.

Prognosis

The clinical course of MS is unpredictable; a high MR lesion load at initial presentation, high relapse rate, early onset of progression, male gender, and late presentation are poor prognostic features but

Box 28.37 Disease-modifying drugs used in multiple sclerosis

Drug	Mode of action	Administration	Efficacy	Reduction in relapse rate	Effect on disability	Adverse effects/comments
Immunomodulatory						
β-interferon (IFN-β1b and 1a)	Pleiotropic immune effects	s.c. alternate days or i.m. weekly	Moderate	33%	?	Flu-like symptoms
Glatiramer acetate	Pleiotropic immune effects	s.c. daily	Moderate	33%	?	Flu-like symptoms
Teriflunomide	Blocks proliferation of activated lymphocytes	Oral daily	Moderate	30%	?+	Abnormal liver enzymes
Dimethyl fumarate	Pleiotropic immune effects	Oral twice daily	High	50%	?+	Good safety record in psoriasis; well tolerated but one case of PML
Immunosuppressant						
Fingolimod	Sphingosine-1-phosphate receptor (S1-PR) ligand. Prevents T cells leaving lymph nodes	Oral daily	High	60%	?+	Bradycardia, OIs, PML
Siponimod	Sphingosine-1-phosphate receptor (S1-PR) ligand. Prevents T cells leaving lymph nodes	Oral daily	High	50%	+	*Used in active SPMS*
Natalizumab	mAb, blocks α-4 integrin vascular adhesion molecule. Prevents T cells entering CNS	i.v. monthly	Very high	68%	+	Rarely, PML (fatal) Hypersensitivity reactions
Ocrelizumab	mAb, anti-CD20 Depletes B cells	i.v. every 6 months	Very high	50% compared to IFNβ	+	Infusion reactions, OIs, hypogammaglobulinaemia with prolonged use *Licensed for RRMS and early PPMS*
Ofatumumab	mAb, anti-CD20 Depletes B cells	s.c. monthly	Very high	50% compared to teriflunomide	+	OIs, hypogammoglobulinaemia with prolonged use
Immune reconstitution/induction therapies						
Alemtuzumab	mAb, anti-CD52 Depletes T >B cells	i.v. once and repeat at 1 year	Very high	50% compared to IFNβ	+	Autoimmune disorders: Graves' disease and ITP, OIs
Cladribine	Purine analogue, antimetabolite, selective T- and B-cell depletion	Oral 8–10 days months 1–2, repeated at 1 year	High	58%	+	Lymphopaenia, OIs
Autologous haematopoietic stem cell transplantation			Very high	70% disease-free survival at 3 years	++	Dangers of induction chemotherapy

ITP, Immune thrombocytopenic purpura; *mAb,* monoclonal antibody; *OIs,* opportunistic infections; *PML,* progressive multifocal leukoencephalopathy.

not invariably so. There is wide variation in severity, but overall progression is slow. Many patients continue to live independent, productive lives; a minority become severely disabled. Life expectancy is reduced by 7 years on average. The precise effect of DMTs on the long-term clinical course remains unclear.

Other neuroinflammatory disorders

Transverse myelitis

Transverse myelitis (TM) is an acute inflammatory disorder affecting the spinal cord with cord swelling and loss of function. Typically, one or two spinal segments are affected with part or all of the cord area at that level involved (*transverse* indicates involvement of the whole cross-section of the cord at the affected level). *Longitudinally extensive transverse myelitis* (LETM) extends over three or more vertebral segments and raises the possibility of more unusual causes. Clinically, a myelopathy evolves over days, and recovery (often partial) follows over weeks or months. MRI shows cord swelling and oedema with gadolinium enhancement at the affected level(s). Usually, images are also taken of the brain to look for evidence of demyelination. CSF may be inflammatory with an excess of lymphocytes. The differential diagnosis

includes other causes of acute myelopathy (see 'Other spinal cord disorders' section). Causes include:

- ***Para- or post-infectious autoimmune inflammatory response.*** This is the most common cause and may follow viral infection or immunization, for example.
- ***Systemic inflammatory disorders***, for example, SLE, Sjögren's, sarcoidosis.
- ***Infection.*** This may be caused by viruses—for example, herpesviruses, such as varicella zoster or EBV, HIV, HTLV-1 and 2 (tropical spastic paraparesis); mycobacteria, e.g., tuberculosis; bacteria—e.g., syphilis and Lyme disease; or flatworms—e.g., schistosomiasis. In recent years outbreaks of *acute flaccid myelitis* (AFM), reminiscent clinically of acute paralytic polio, have been linked to outbreaks of viruses, particularly enterovirus D68, although evidence of direct viral injury is lacking. The clinical course of AFM is more rapid and aggressive than classical TM, with prominent anterior horn cell involvement from grey matter damage.
- ***Multiple sclerosis.*** TM may occur as part of a relapse or be a presenting feature of MS. If the brain MRI is normal, the risk of later development of MS is around 20%; in those that do develop MS with a normal brain MRI, prognosis is better than average.
- ***Neuromyelitis optica spectrum disorder*** and ***myelin oligodendrocyte glycoprotein antibody-associated disease (MOGAD)***—see below.
- ***Acute disseminated encephalomyelitis (ADEM).*** TM can be seen as part of ADEM—see below.

Treatment is usually with high-dose steroids or other immunomodulatory treatments (including plasma exchange and IVIG), or antimicrobial therapy in the case of specific infections. Involvement of the high cervical spine, particularly in longitudinally extensive cases or AFM, can necessitate intubation and ventilation. Specialist rehabilitation after an attack of TM is essential.

Neuromyelitis optica

This is a distinct inflammatory relapsing demyelinating disorder, previously thought to be a variant of MS. It is characterized by longitudinally extensive TM (>3 segments) and bilateral or recurrent optic neuritis. Brainstem involvement can occur, and the phenotype is expanding. Limited forms are seen, such as recurrent myelitis only. Serum antibodies to aquaporin-4 (AQP4) water channels account for the majority of cases; a proportion of the remainder have antibodies to myelin oligodendrocyte glycoprotein (MOG, see below). It is rarer than MS but there is a higher incidence in Asian and African Caribbean people.

Treatment with steroids and immunosuppressive drugs is essential from the time of diagnosis, as the relapse rate is high and relapses are much more disabling than in MS.

Acute disseminated encephalomyelitis

Acute disseminated encephalomyelitis is usually post-infectious, following infections such as measles, mycoplasma, mumps, and rubella, and occasionally 1 to 2 weeks after immunization. There is a monophasic illness with the acute appearance of multifocal neurological symptoms, often with encephalopathy. Imaging shows brain, brainstem, and often spinal cord inflammatory demyelinating lesions, which can become confluent; there is often swelling and there may be a haemorrhagic component. ADEM is caused by an immune-mediated host response to infection and occurs principally in children and young adults. Treatment is with steroids, moving on to plasma exchange or IVIG, and symptomatic measures such as anticonvulsants. Mild cases recover completely, but fulminant cases can be fatal, and survivors often have neurological sequelae.

Myelin oligodendrocyte glycoprotein antibody-associated disease

A recently recognized entity, that presents as a number of different inflammatory demyelinating CNS syndromes, linked to the presence of MOG antibodies. MOGAD may have a relapsing or monophasic course. Clinical syndromes include bilateral and recurrent optic neuritis, TM (often with involvement of the conus), and ADEM. MOG antibodies were first recognized in children with ADEM, and the presence of the antibody makes recurrence more likely. MOGAD can closely mimic NMO, and it can also present as a unilateral cortical encephalitis with headache, seizures, and focal neurology.

Autoimmune encephalitis

This group of disorders is increasingly being recognized. Autoantibodies directed against neuronal epitopes cause a subacute encephalitic illness: limbic encephalitis or panencephalitis. Limbic encephalitis presents over weeks or months with memory impairment, confusion, psychiatric disturbance, and seizures—usually temporal lobe seizures, reflecting involvement of the hippocampus and mesial temporal lobes.

Paraneoplastic limbic encephalitis (PLE). PLE is seen particularly with small-cell lung cancer and testicular tumours, and is associated with a variety of antibodies, including anti-Hu and anti-Ma2. Antibodies can be detected in 60% of cases. MRI usually shows a hippocampal high signal. PLE precedes the diagnosis of cancer in most cases and should prompt investigation to identify the tumour.

Anti-Voltage-gated potassium channel (VGKC) complex limbic encephalitis. Antibodies to the VGKC complex produce a variety of disorders, including a characteristic limbic encephalitis in association with antibodies to the leucine-rich glioma-inactivated 1 (LGI1) protein. Affected patients may present with typical faciobrachial dystonic seizures, in combination with confusion, agitation, and hyponatraemia. This usually occurs in patients over 50 years of age but is rarely associated with cancer (thymoma). Treatment is with high-dose steroids. Neuromyotonia and peripheral nerve hyperexcitability syndromes may also be seen with antibodies to contactin-associated protein 2 (CASPR2), another subunit of the VGKC-complex.

Anti-NMDA receptor antibody encephalitis. This presents as a panencephalitis characterized by complex psychiatric features, behavioural change, sleep disturbances, seizures, dysautonomia, and a prominent movement disorder, with orofacial and upper limb dyskinesias. Coma may follow. Patients are usually younger, and many have teratomas, such as ovarian.

Patients may respond to immunotherapy: steroids, intravenous immunoglobulin, and plasma exchange initially, followed by rituximab or cyclophosphamide. PLE responds less well to treatment, although treatment directed at any underlying tumour offers the best hope.

Neuroinflammation in systemic disease

A number of systemic inflammatory diseases can involve both the central and peripheral nervous system.

Neurosarcoidosis

Neurosarcoidosis, with or without systemic sarcoidosis, causes chronic meningoencephalitis, TM, cranial nerve palsies—particularly bilateral VIIth nerve lesions, polyneuropathy, and myopathy (see Chapter 20, 'Sarcoidosis' section). Diagnosis can be difficult if disease is confined to the CNS.

Behçet's disease

Behçet's principal features (see Chapter 32, 'Behçet's disease' section) are recurrent oral and/or genital ulceration, inflammatory ocular disease (uveitis; see Chapter 31, 'Uveitis' section), and neurological syndromes. Brainstem and cord lesions, aseptic meningitis, encephalitis, and cerebral venous thrombosis occur. There is a predilection for ethnic groups along the ancient 'Silk Road'—Turkey, the Middle East, and Asia. Behçet's is associated with the HLA-B51 allele.

Systemic lupus erythematosus

Neurological involvement in SLE takes protean forms, and 20% to 40% of patients with SLE may have some form of nervous system involvement. Manifestations include encephalopathy, neuropsychiatric illness, optic neuritis, aseptic meningitis, movement disorders, myelitis, and peripheral neuropathy. The underlying pathophysiology can include neuroinflammation (with antibody-, cell-, and cytokine-mediated mechanisms) and vascular disease (vasculitis is less common than thromboembolic disease, e.g., with antiphospholipid antibodies). Systemic complications of SLE (e.g., renal disease) can cause secondary neurological effects (e.g., PRES or metabolic encephalopathy).

Further reading

Dobson R and Giovannoni G. Multiple sclerosis—a review. *Eur J Neurol* 2019; 26: 27–40.
Frohman EM, Wingerchuk DM. Clinical practice. Transverse myelitis. *N Engl J Med* 2010; 363:564–572.
Thompson AJ, Banwell BL, Barkhof F et al. Diagnosis of multiple sclerosis: 2017 revisions to the McDonald criteria. *Lancet Neurol* 2018; 17:162–173.
Treatment Algorithm for Multiple Sclerosis Disease-Modifying Therapies. NHS England 2018.
https://mstrust.org.uk. *MS Trust (UK national charity)*.

NERVOUS SYSTEM INFECTION

Meningitis

Bacterial meningitis is a medical emergency requiring immediate intervention. It is covered in Chapter 11 ('Meningitis' section).

Chronic meningitis

Tuberculous meningitis (TBM) and ***cryptococcal meningitis*** typically commence with vague headache, lassitude, anorexia, and vomiting. Acute meningitis can occur but is unusual. Meningitic signs often take some weeks to develop. Drowsiness, focal signs (e.g., diplopia, papilloedema, hemiparesis, myelopathy), and seizures are common. Syphilis, sarcoidosis, and Behçet's also cause chronic meningitis. In some cases of chronic meningitis, an organism or underlying cause is never identified.

Investigation and management of tuberculous meningitis

TBM is a common and serious disease worldwide. Brain imaging, usually with MRI, may show meningeal enhancement, hydrocephalus, and tuberculomas, although it may remain normal. CSF pressure is raised, and microscopy reveals a high lymphocyte count with high protein and low glucose. In many cases, the sparse tuberculous organisms cannot be seen on staining and PCR testing should be performed, although results may be falsely negative. Repeated large-volume CSF examination is often necessary and it will be some weeks before cultures are confirmatory. Treatment with anti-tuberculosis drugs (see Chapter 20, 'Tuberculosis' section)—rifampicin, isoniazid, and pyrazinamide—must commence on a presumptive basis and continue for at least 9 months. Ethambutol is avoided in many cases of TBM because of its eye complications; a fluoroquinolone is typically used in its stead. Adjuvant corticosteroids, such as dexamethasone 0.4 mg/kg/day, are recommended and shown to reduce mortality; in complicated cases, an extended taper is often needed. Relapses and complications (e.g., seizures, hydrocephalus, infectious vasculitis) are common in TBM. The mortality remains over 50%, even with early treatment.

Malignant meningitis

Malignant cells can cause a subacute or chronic, non-infective, meningitic process. A meningitic syndrome, cranial nerve lesions, paraparesis, and root lesions are seen, often in confusing and fluctuating patterns. CSF cytology may demonstrate malignant cells but sensitivity is low, even with rapid analysis of large volumes (at least 10 mL) of fresh CSF. Up to three LPs may be needed to reach 90% sensitivity. In general, the prognosis is 3 to 4 months but occasionally intrathecal or systemic treatment can be helpful.

Cells in a sterile CSF (CSF pleocytosis)

A raised CSF cell count is sometimes present without an evident infecting organism in various conditions listed in **Box 28.38**.

Encephalitis

Encephalitis means acute inflammation of brain parenchyma. The commonest cause is viral infection, when fever (90%) and meningism are usual; in contrast to meningitis, however, the clinical picture is dominated by inflammation of brain parenchyma. Personality and behavioural change is a common early manifestation, which progresses to a reduced level of consciousness and even coma. Seizures (focal and generalized) are very common and focal neurological deficits, such as speech disturbance, often occur (especially in herpes simplex encephalitis, which is a medical emergency—see Chapter 11).

Autoimmune encephalitis is at least as common as infective encephalitis in the developed world, and is covered in the 'Autoimmune encephalitis' section.

? Box 28.38 Causes of sterile CSF pleocytosis

Infective
- Partially treated bacterial meningitis
- Viral meningitis
- Encephalitis, including HIV
- Tuberculous or fungal meningitis
- Intracranial abscess
- Parameningeal foci, e.g., paranasal sinus
- Syphilis
- Cerebral malaria
- Rarities, e.g., Lyme disease, Brucellosis, infective endocarditis, Mollaret meningitis

Non-infective
- Neoplastic meningitis
- Cerebral venous thrombosis
- Drugs (e.g., IVIG, co-trimoxazole)
- Systemic lupus erythematosus
- CNS inflammatory diseases including ADEM (or other post-infectious encephalitis), MS
- Rarities, e.g., CNS vasculitis, sarcoidosis, Behçet's syndrome, Vogt-Koyanagi-Harada syndrome

HIV and neurology

HIV-infected individuals frequently present with or develop neurological conditions. The HIV itself is directly neuroinvasive and neurovirulent. Immunosuppression leads to indolent, atypical clinical patterns (see Chapter 44, 'Neurological disease' section). HIV patients also have a high incidence of stroke. The pattern of disease has changed significantly with anti-retroviral (ART) therapy.

CNS and peripheral nerve disease in HIV

HIV seroconversion can cause meningitis, encephalitis, Guillain–Barré syndrome, and Bell's palsy.

Chronic meningitis occurs with fungi (e.g., Cryptococcus *neoformans* or *Aspergillus*), tuberculosis, *Listeria*, coliforms, or other organisms. Raised CSF pressure is common in cryptococcal meningitis.

AIDS–dementia complex (ADC). A progressive, HIV-related dementia, sometimes with cerebellar signs, is still seen where ART is unavailable.

Encephalitis and brain abscess. Toxoplasma, cytomegalovirus, herpes simplex, and other organisms cause severe encephalitis. Multiple brain abscesses develop in HIV infection, usually due to toxoplasmosis.

CNS lymphoma. This is typically fatal (see Chapter 39, 'Primary central nervous system lymphoma' section).

Progressive multifocal leukoencephalopathy (PML) is due to JC virus and occurs with very low CD4 counts (see Chapter 44, 'Neurological disease' section).

Spinal vacuolar myelopathy. This occurs in advanced disease.

Neurosyphilis is seen in all its forms in HIV.

Peripheral nerve disease. HIV-related peripheral neuropathy is common (70%) and can be difficult to distinguish from the effects of certain ART toxic to peripheral nerves. A polyradiculitis is frequently seen with CMV and other opportunistic infections.

Other infections and post-infectious inflammatory conditions

Other infections involving the CNS are discussed in Chapter 42: for example, rabies, tetanus, botulism, Lyme disease, and leprosy.

Herpes zoster (shingles)

This is caused by reactivation of varicella zoster virus (VZV), usually within dorsal root ganglia. Primary infection with VZV causes chickenpox, following which the virus remains latent in sensory ganglia. Development of shingles may indicate a decline in cell-mediated immunity, such as that due to age or malignancy.

Clinical patterns and complications

Dermatomal shingles. See Chapter 45 ('Herpes zoster [shingles]' section).

Postherpetic neuralgia. This is defined as pain lasting for more than 4 months after developing shingles; it occurs in 10% of patients (often elderly). Burning, intractable pain responds poorly to analgesics. Response to treatment is unsatisfactory but there is a trend towards gradual recovery over 2 years. Amitriptyline or gabapentin is commonly used, and topical lidocaine patches may help.

Cranial nerve involvement. Only cranial nerves with sensory fibres are affected, particularly the trigeminal and facial nerves. Ophthalmic herpes is due to involvement of V_1. This can lead to corneal scarring and secondary panophthalmitis. Involvement of the geniculate ganglion of the facial nerve, causing facial palsy with associated painful shingles rash, is also called Ramsay Hunt syndrome.

Myelitis. This may occur in the context of shingles, when the inflammatory process spreads from the dorsal root ganglion to the adjacent spinal cord.

Immunization. Older adults can be vaccinated against herpes zoster (even those who have had shingles previously), as it boosts immunity against VZV and reduces the incidence of shingles by about 50%.

Neurosyphilis

Many neurological symptoms occur, sometimes mixed (see also syphilis; Chapter 43).

Early neurosyphilis

Asymptomatic neurosyphilis

This means positive CSF serology without signs.

Meningovascular syphilis

This causes:

- subacute meningitis with cranial nerve palsies and papilloedema
- a gumma—a chronic expanding intracranial mass
- paraparesis—a spinal meningovasculitis.

Late neurosyphilis

Tabes dorsalis

Demyelination in dorsal roots causes a complex deafferentation syndrome. The elements of tabes are:

- lightning pains
- ataxia, stamping gait, reflex/sensory loss, wasting
- neuropathic (Charcot) joints
- Argyll Robertson pupils (see 'Argyll Robertson pupil' section)
- ptosis and optic atrophy.

'General paresis of the insane'

The grandiose title describes dementia and weakness. The dementia component is typically similar to that of Alzheimer's (see 'Alzheimer's disease' section). Progressive cognitive decline, seizures, brisk reflexes, extensor plantar reflexes, and tremor develop. Death follows within 3 years. Argyll Robertson pupils are a usual finding. GPI and tabes are rarities in the UK.

Other forms of neurosyphilis

In ***congenital*** neurosyphilis (acquired *in utero*), features of combined tabes and GPI develop in childhood: taboparesis.

Management

Penicillin is the treatment of choice for all stages of syphilis. Cytokine-mediated febrile (Jarisch–Herxheimer) reactions can occur; steroid cover may be given with penicillin. Established neurological disease is arrested but not reversed by treatment.

Neurocysticercosis

The pork tapeworm, *Taenia solium*, is endemic in Latin America, Africa, India, and much of South-east Asia (see Chapter 42, '*Taenia solium* and cysticercosis' section). Epilepsy is the most common clinical manifestation of neurocysticercosis, and the infection is one of the most common causes of epilepsy in endemic countries. Most infected people remain asymptomatic.

Brain CT and MRI show ring-enhancing lesions with surrounding oedema when the cyst dies, and later calcification. Multiple

cysts are often seen in both brain and skeletal muscle. Serological tests indicate infection but not activity. Biopsy of a lesion is rarely necessary. Management is primarily the control of seizures, and the anthelminthic agent, albendazole, is given in the presence of viable (non-calcified) or degenerating cysts (usually with steroid cover).

Subacute sclerosing panencephalitis

Persistence of measles antigen in the CNS is associated with this rare late sequela of measles. Progressive mental deterioration, fits, myoclonus, and pyramidal signs develop, typically in a child. Diagnosis is made by high measles antibody titre in blood and CSF. Measles immunization protects against subacute sclerosing panencephalitis (SSPE), which has now been almost eliminated in the UK.

Progressive rubella encephalitis

Some 10 years after primary rubella infection, this causes progressive cognitive impairment, fits, optic atrophy, cerebellar, and pyramidal signs. Antibody to rubella viral antigen is produced locally within the CNS. It is even rarer than SSPE.

Mollaret's meningitis

Recurrent, self-limiting episodes of aseptic meningitis (i.e., no bacterial cause is found) occur over many years. Herpes simplex virus type 2 has been shown to be the cause in most cases.

Whipple's disease

CNS Whipple's disease, due to *Tropheryma whipplei* infection, is characterized by neurological symptoms, such as myoclonus, dementia, and supranuclear ophthalmoplegia (see Chapter 22,

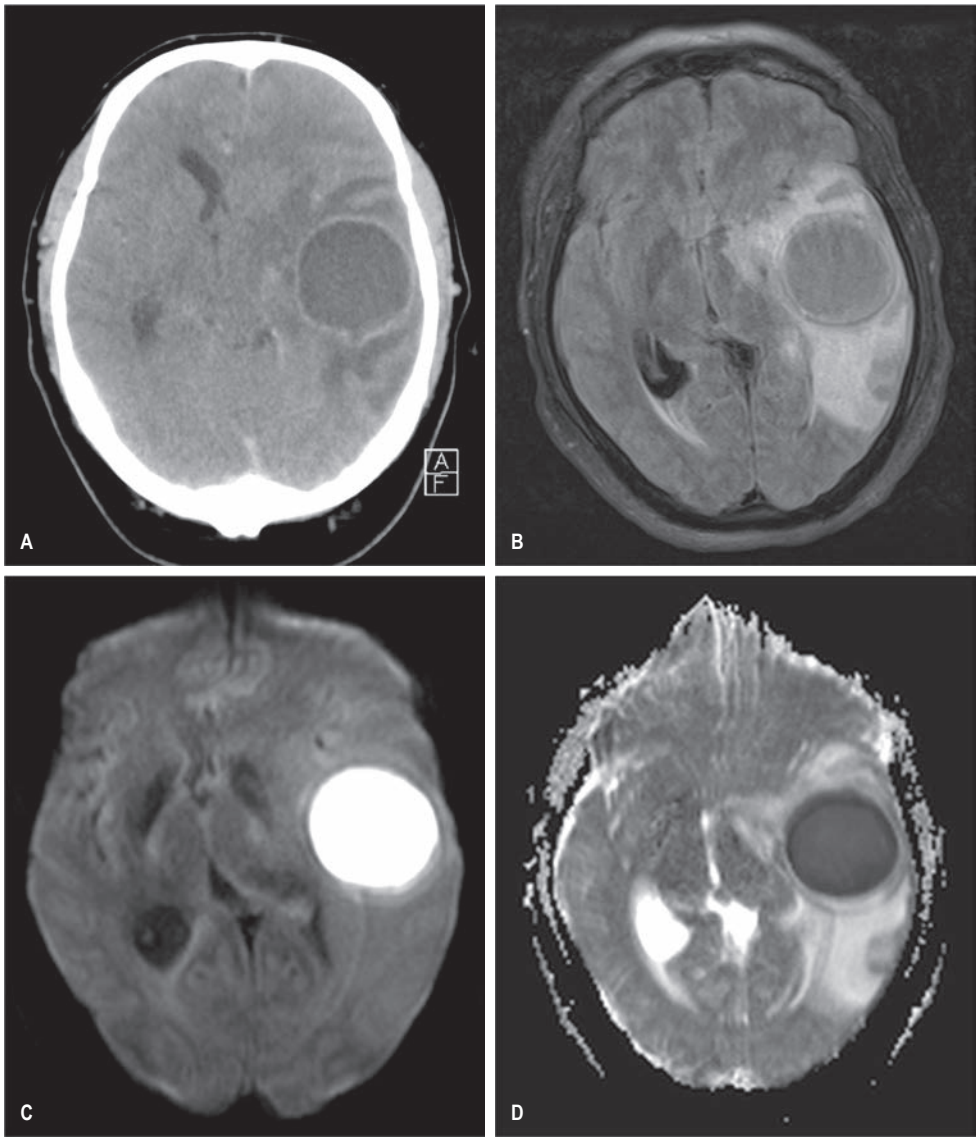


Fig. 28.27 Pyogenic cerebral abscess. **(A)** Axial post-contrast CT demonstrating a ring-enhancing lesion in the left temporoparietal region. **(B)** Subsequent T2-weighted fluid suppressed (FLAIR) images show the extent of surrounding oedema. MRI is also useful to characterize the lesion further. **(C)** Here, diffusion-weighted and **(D)** ADC map confirm restricted diffusion within the lesion, which is characteristic of intracranial pyogenic abscess.

'Whipple's disease' section). Diagnosis of CNS involvement is made by CSF PCR or brain biopsy.

Brain and spinal abscesses

Brain abscess

Focal bacterial infection behaves as any expanding mass. The typical bacteria found are *Streptococcus anginosus* and *Bacteroides* species (from the paranasal sinuses and teeth), and staphylococci (associated with penetrating trauma). Mixed infections are common. Multiple abscesses may develop, particularly in HIV infection. Fungi also cause brain abscesses. A parameningeal infective focus (e.g., ear, nose, paranasal sinus, skull fracture) or a distant source of infection (e.g., lung, heart, abdomen) may be present. Frequently, no underlying cause is found. An abscess is more than ten times rarer than a brain tumour in the UK.

Clinical features and management

Headache, focal signs (e.g., hemiparesis, aphasia, hemianopia), epilepsy, and raised ICP develop. Fever, leucocytosis, and raised inflammatory markers are common, but not invariable.

Urgent imaging is essential. MRI shows a ring-enhancing mass, usually with considerable surrounding oedema, and central restricted diffusion (**Fig. 28.27**). The search for a focus of infection should include a detailed examination of the skull, ears, paranasal sinuses and teeth, and distant sites such as heart and abdomen. LP is potentially dangerous and the risks and benefits should be weighed up; often CSF if obtained will be bland as the abscess is walled-off. Neurosurgical aspiration with stereotactic guidance allows the infective organism to be identified. Treatment is with high-dose antibiotics and, sometimes, surgical resection/decompression. Despite treatment, mortality remains high at approximately 25%. Epilepsy is common in survivors.

Brain tuberculoma

Tuberculosis causes chronic caseating intracranial granulomatous masses: tuberculomas. These are the most common intracranial masses in countries where tuberculosis is common, such as India. Brain tuberculomas either present as mass lesions *de novo* or develop during TBM; they are also found as symptomless intracranial calcification on imaging. Spinal cord tuberculomas also occur. Treatment is described in Chapter 20 ('Management' section in the 'Tuberculosis' section).

Subdural empyema and intracranial epidural abscess

Intracranial subdural empyema is a collection of subdural pus, usually secondary to local skull or middle ear infection, or following neurosurgery. Subdural pus is intensely epileptogenic and patients may present with intractable seizures and fever. Imaging is diagnostic.

In ***intracranial epidural abscess***, a layer of pus, 1 to 3 mm thick, tracks along the epidural space, causing sequential cranial nerve palsies. There is usually local infection: in the middle ear, for example. MRI shows the collection; CT is typically normal. Drainage is required, as are antibiotics.

Spinal epidural abscess

Staphylococcus aureus is the usual organism, reaching the spine via the bloodstream: for example, from a boil. Fever and back pain are followed by paraparesis and/or root lesions. Emergency imaging and antibiotics are essential and surgical decompression is often necessary.

Box 28.39 Common brain tumours

Tumour	Approximate frequency
Metastases	50%
Bronchus Breast Stomach Prostate Thyroid Kidney	
Primary malignant tumours of neuroepithelial tissues	15%
Adult-type diffuse gliomas Astrocytoma, IDH-mutant Oligodendroglioma, IDH mutant, and 1p/19q codeleted Glioblastoma, IDH-wildtype Ependymoma Primary cerebral lymphoma Medulloblastoma	
Benign tumours	35%
Meningioma Schwannoma Pituitary adenoma	

IDH, Isocitrate dehydrogenase.

Further reading

Brouwer MC, Tunkel GM, McKhann HGM. Brain abscess. *N Engl J Med* 2014; 371:447–456.
van de Beek D, Brouwer MC, Thwaites GE et al. Advances in treatment of bacterial meningitis. *Lancet* 2012; 380:1693–1702.
http://www.meningitisnow.org/. *Formed in 2013 by merger of Meningitis UK and the Meningitis Trust*.

BRAIN TUMOURS

Primary intracranial tumours account for 3% of all cancers. The most common tumours are outlined in **Box 28.39**, with ***cerebral metastases*** being more common than primary intracranial tumours (**Fig. 28.28**). Symptomless meningiomas (benign) are found quite commonly on imaging or at autopsy. The most recent 5th edition of the WHO classification of CNS tumours utilizes a combination of histological and molecular features to define tumour type and grade.

Gliomas

These tumours of neuroepithelial origin are usually seen within the hemispheres, but occasionally in the cerebellum, brainstem, or cord (**Fig. 28.29**). Their cause is unknown in most cases, and the main risk factors are rare genetic risk syndromes (e.g., neurofibromatosis) and exposure to ionizing radiation. They tend to spread by direct extension, virtually never metastasizing outside the CNS.

- ***Astrocytomas*** are gliomas that arise from astrocytes. They may be diffuse or circumscribed (e.g., pilocytic astrocytoma, a slow-growing and benign tumour typically seen in childhood, often in the cerebellum). Adult-type diffuse gliomas are classified first by isocitrate dehydrogenase (IDH) status (mutated or wild-type) and then by further molecular and histological features. IDH mutations indicate that a tumour is slower-growing, and when found in a high-grade tumour may indicate that the tumour has transformed from a low-grade lesion; an IDH-wildtype diffuse astrocytoma is always grade 4 and termed a ***glioblastoma***. Grade 4 glioblastoma causes death within several months.

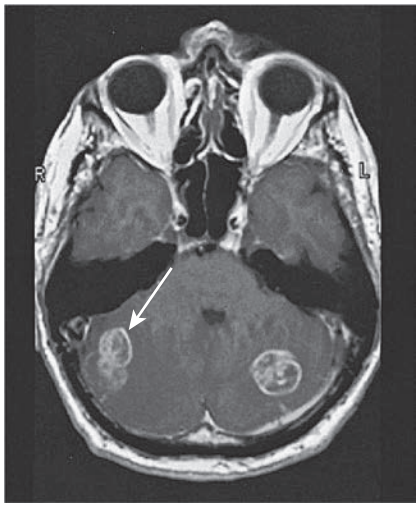


Fig. 28.28 **Bilateral cerebellar metastases *(arrowed)*: T1-weighted MRI.**

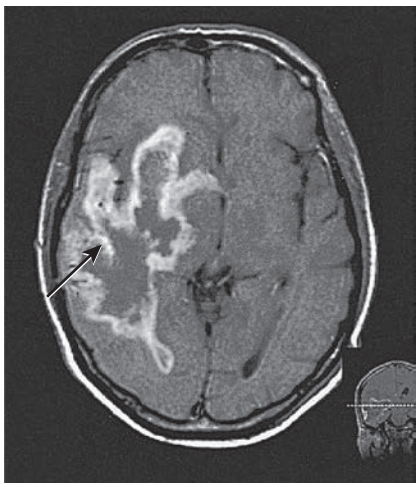


Fig. 28.29 **Cerebral glioblastoma: T1-weighted MRI.**

- ***Oligodendrogliomas*** arise from oligodendrocytes. They grow slowly, usually over several decades. Calcification is common. They are IDH-mutated and are distinguished from astrocytomas by the presence of the chromosomal 1p/19q co-deletion.

Meningiomas

These benign tumours (Fig. 28.30) arise from the arachnoid and may grow to a large size, usually over years. Those close to the skull erode bone or cause local hyperostosis. They often occur along the intracranial venous sinuses, which they may invade. They are unusual below the tentorium. Common sites are the parasagittal region, sphenoidal ridge, subfrontal region, pituitary fossa, and skull base.

Schwannomas

These solid benign tumours arise from Schwann cells and occur principally in the cerebellopontine angle, where they arise from the VIIIth nerve sheath (vestibular schwannoma, also known as acoustic neuroma; see Chapter 21, 'Vestibular schwannoma' section). They may be bilateral in neurofibromatosis type 2 (see 'Neurofibromatosis type 2' section).

Other neoplasms

Other less common neoplasms include cerebellar haemangioblastomas, ependymomas of the IVth ventricle, colloid cysts of the IIIrd ventricle, pinealomas, chordomas of the skull base, glomus tumours of the jugular bulb, medulloblastomas (a cerebellar childhood tumour), craniopharyngiomas and primary CNS lymphomas (Chapter 39, 'Primary central nervous system lymphoma' section). For pituitary tumours, see Chapter 34 ('Pituitary space–occupying lesions and tumours' section).

Clinical features

Mass lesions within the brain produce symptoms and signs by three mechanisms:

- direct effect—the brain is infiltrated and local function is impaired, producing focal neurological signs
- secondary effects of raised ICP and shift of intracranial contents (e.g., papilloedema, vomiting, headache)
- provocation of generalized and/or focal seizures.

Although neoplasms, either secondary or primary, are the most common mass lesions in the UK, cerebral abscess, tuberculoma, neurocysticercosis, and subdural and intracranial haematomas can also produce features that are clinically similar.

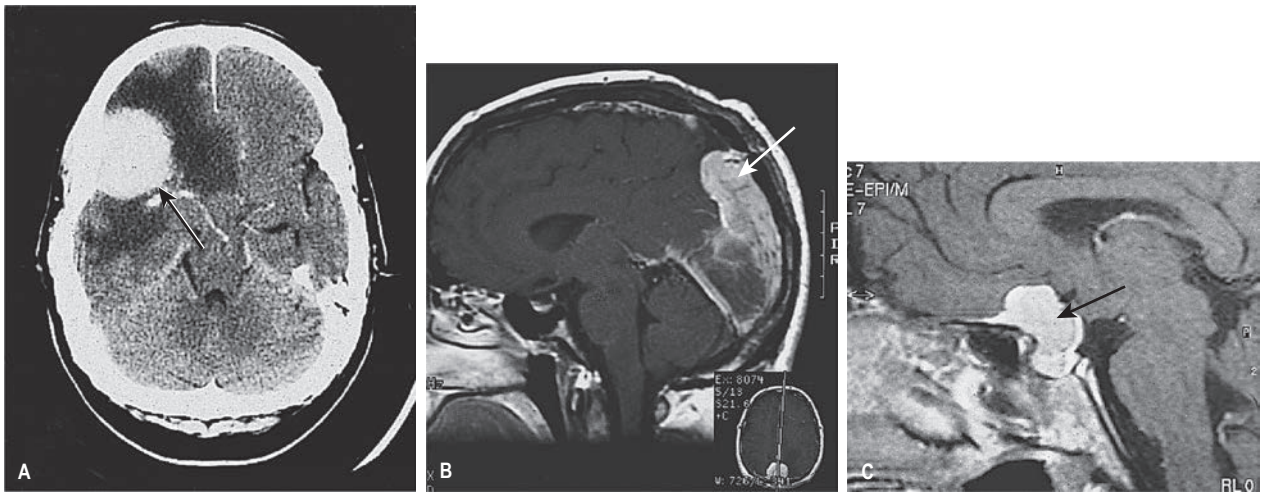


Fig. 28.30 **Meningiomas *(arrowed)*.** **(A)** Frontal: CT. **(B)** Falx meningioma (occipital): T1-weighted MRI. **(C)** Within sella and suprasellar extension: T1-weighted MRI.

Direct effects of mass lesions

The hallmark of a direct effect of a mass is local progressive deterioration of function. Tumours can occur anywhere within the brain. Three examples are given:

- ***A left frontal meningioma*** causing a developing frontal lobe syndrome over several years with vague disturbance of personality, apathy, and impaired intellect. This was followed by expressive aphasia, progressive right hemiparesis as the corticospinal pathways became involved, and then pressure headaches and papilloedema reflecting raised ICP.
- ***A right parietal lobe glioma*** causing a left homonymous field defect through involvement of the optic radiation. Cortical sensory loss in the left limbs and left hemiparesis followed over 3 months. Partial seizures (episodes of tingling of the left limbs) then developed.
- ***A left VIIIth nerve sheath neurofibroma (acoustic neuroma, schwannoma)*** in the cerebellopontine angle causing, over 3 years, progressive deafness (VIII), left facial numbness (V) and weakness (VII), followed by cerebellar ataxia on the same side.

With a hemisphere tumour, epilepsy and the direct effects commonly draw attention to the problem. The rate of progression varies greatly, from a few days or weeks in a highly malignant glioma, to several years with a slowly enlarging mass such as a meningioma. Cerebral oedema surrounds mass lesions; its effect is difficult to distinguish from that of the tumour itself.

Raised intracranial pressure

Raised ICP causing headache, vomiting, and papilloedema is a relatively unusual presentation of a mass lesion in the brain. These symptoms often imply hydrocephalus: obstruction to CSF pathways. Typically, this is produced early by posterior fossa masses that obstruct the aqueduct and IVth ventricle, but only later by lesions above the tentorium.

Shift of the intracranial contents produces features that coexist with the direct effects of any expanding mass:

- ***Distortion of the upper brainstem*** as midline structures are displaced either caudally or laterally by a hemisphere mass (see **Fig. 28.29**) causes impairment of consciousness, progressing to coning and death as the medulla and cerebellar tonsils are forced into the foramen magnum.
- ***False localizing signs***, termed false only because they do not point *directly* to the site of the mass, include:
  - ***A VIth nerve lesion***, first on the side of a mass and later bilaterally as the VIth nerve is compressed or stretched during its long intracranial course.
  - ***A IIIrd nerve lesion*** developing as the temporal lobe uncus herniates caudally, compressing the IIIrd nerve against the petroclinoid ligament. The first sign is ipsilateral pupil dilatation as parasympathetic fibres are compressed.
  - ***Hemiparesis*** on the same side as a hemisphere tumour—that is, the side that might not be expected—from compression of the contralateral cerebral peduncle on the edge of the tentorium (the Kernohan notch phenomenon).

Seizures

Seizures are a common presenting feature of malignant brain tumours. Focal seizures, with or without altered awareness, that may evolve into bilateral tonic–clonic seizures, are characteristic of many hemisphere masses, whether malignant or benign. The pattern of focal seizure is of localizing value (see 'Focal seizure types' section).

Investigations

Imaging

Both CT and MRI are useful in detecting brain tumours; MRI is superior for posterior fossa lesions. Benign and malignant tumours, brain abscess, tuberculosis, neurocysticercosis, and infarction have characteristic, but not entirely reliable, radiological appearances, and refined imaging techniques and biopsy are often necessary. MR angiography is used occasionally to define blood supply and MR spectroscopy to identify patterns typical of certain gliomas. PET is sometimes helpful to locate an occult primary tumour with brain metastases.

Routine tests

Since metastases are more common than primary tumours, routine tests, such as chest X-ray, and cross-sectional body imaging, should be performed.

Lumbar puncture

This is contraindicated when there is a space-occupying lesion with mass effect, or a mass in the posterior fossa, as withdrawing CSF may provoke immediate coning. CSF examination is rarely helpful unless there is meningeal infiltration (malignant meningitis).

Biopsy and tumour removal

Stereotactic biopsy via a skull burr hole is carried out to ascertain the histology of most suspected malignancies. With a benign tumour, such as a symptomatic, accessible meningioma, craniotomy and removal are usual.

Management

This is discussed in Chapter 41.

Further reading

Louis DN, Perry A, Wesseling P, et al. The 2021 World Health Organization classification of tumors of the central nervous system: a summary. *Neuro Oncol* 2021; 23:1231–51.

HYDROCEPHALUS

Hydrocephalus is an excessive accumulation of CSF within the head, caused by a disturbance of formation, flow or absorption. High pressure and ventricular dilation result (**Fig. 28.31**).

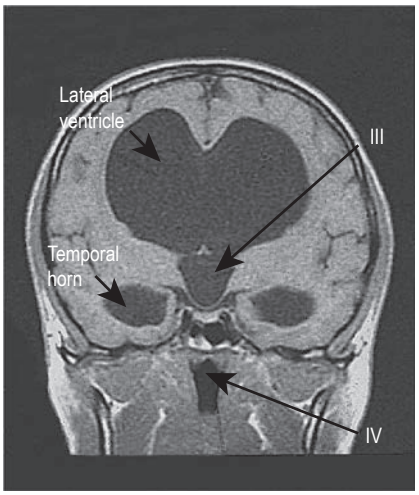


Fig. 28.31 Hydrocephalus with gross IIIrd, IVth, and lateral ventricle enlargement *(arrowed)*: T1-weighted MRI.

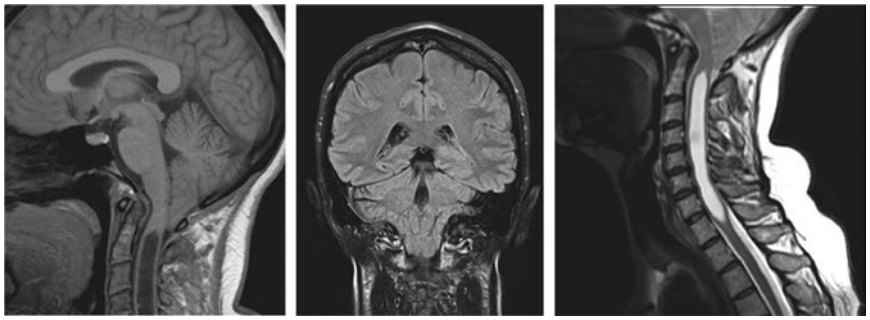


Fig. 28.32 **Chiari I malformation with spinal cord syrinx.**

Infantile hydrocephalus

Head enlargement in infancy occurs in 1 in 2000 live births. There are several causes:

- ***Chiari malformations.*** The cerebellar tonsils descend into the cervical canal (Fig. 28.32). Associated spina bifida is common. Syringomyelia may develop (see 'Syringomyelia and syringobulbia' section).
- ***Stenosis of the aqueduct of Sylvius.*** Aqueduct stenosis is either congenital (genetic) or acquired following neonatal meningitis/haemorrhage.
- ***Dandy–Walker syndrome.*** There is cerebellar hypoplasia and obstruction to IVth ventricle outflow foramina.

Hydrocephalus in adults

Hydrocephalus is sometimes an unsuspected finding on imaging. Stable childhood hydrocephalus can become apparent in adult life ('arrested hydrocephalus') but can suddenly decompensate. Combinations of headache, cognitive impairment, features of raised ICP, and ataxia develop, depending on how high the CSF pressure rises and how rapid the onset is. Elderly patients with more compliant brains may present with gradual-onset gait apraxia and subtle cognitive slowing.

Hydrocephalus may be caused by:

- ***posterior fossa and brainstem tumours*** obstructing the aqueduct or IVth ventricular outflow
- ***subarachnoid haemorrhage***, head injury, or meningitis (particularly tuberculous), causing obstruction of CSF flow and impaired reabsorption
- ***a IIIrd ventricle colloid cyst*** causing intermittent hydrocephalus—recurrent prostrating headaches with episodes of lower limb weakness, or, ominously, loss of consciousness
- ***choroid plexus papilloma*** (rare) secreting CSF.

Frequently, the underlying cause of hydrocephalus remains obscure.

Management

Ventriculoperitoneal shunting is necessary when progressive hydrocephalus causes symptoms. Removal of tumours is carried out when appropriate. Endoscopic IIIrd ventriculostomy may be performed.

Idiopathic normal pressure hydrocephalus

This describes a syndrome of enlarged lateral ventricles in elderly patients with the clinical triad of:

- a gait disorder—gait apraxia
- dementia
- urinary incontinence.

Ventriculoperitoneal shunting may be required, but the diagnosis of iNPH and selection of patients for shunting is a matter of some controversy. A trial of prolonged drainage of lumbar CSF over several days may predict response to shunt insertion. A sizeable number of patients identified as potentially having iNPH, often raised for the first time as a possibility on brain imaging based on non-specific imaging features, will instead turn out to have a neurodegenerative disorder.

Secondary normal pressure hydrocephalus is a form of communicating hydrocephalus secondary to reduced resorption of CSF, for instance following subarachnoid haemorrhage or prior meningitis.

BRAIN INJURY

Traumatic brain injury

In most Western countries, head injury accounts for about 250 hospital admissions per 100 000 population annually. Traumatic brain injury (TBI) describes injuries with potentially permanent consequences. For each 100 000 people, 10 die annually following TBI; 10 to 15 are transferred to a neurosurgical unit, the majority of these requiring rehabilitation for a prolonged period of 1 to 9 months. The prevalence of survivors with a major persisting handicap is around 100/100 000. Road traffic accidents and excessive alcohol use are the principal aetiological factors in this major cause of morbidity and mortality, in many countries.

Skull fractures

Linear skull fracture of the vault or base is one indication of the severity of a blow, but in itself is not necessarily associated with any brain injury. Healing of linear fractures takes place spontaneously. ***Depressed*** skull fracture is followed by a high incidence of post-traumatic epilepsy. Surgical elevation and debridement are usually necessary.

Principal local complications of skull fracture are:

- ***meningeal artery rupture*** causing extradural haematoma (see Chapter 11, 'Subdural and extradural bleeding' section)
- ***dural vein tears*** causing subdural haematoma (see Chapter 11, 'Subdural and extradural bleeding' section)
- ***CSF rhinorrhoea/otorrhoea*** and consequent meningitis.

Mechanisms of brain damage

Older classifications attempted to separate ***concussion***, transient coma for hours followed by apparent complete clinical recovery, from brain ***contusion***—that is, bruising—with prolonged coma,

focal signs, and lasting damage. Pathological support for this division is poor. Mechanisms of TBI are complex and interrelated:

- neuronal and axonal damage from direct trauma
- diffuse axonal injury—shearing and rotational stresses on decelerating brain, sometimes at the site opposite impact (the *contrecoup* effect)
- brain oedema and raised ICP
- brain hypoxia
- brain ischaemia.

Clinical course

In a mild TBI, a patient is stunned or dazed for a few seconds or minutes. Following this, the patient remains alert without post-traumatic amnesia (PTA). Headache can follow; complete recovery is usual.

In more serious injuries, duration of unconsciousness, and particularly of PTA, helps grade severity. PTA lasting more than 24 hours defines severe TBI. The Glasgow Coma Scale (GCS) is used to record the degree of coma; this has prognostic value. A GCS below 5/15 at 24 hours implies a serious injury; 50% of such patients die or remain in a vegetative or minimal conscious state (see 'Disorders of consciousness' section). However, prolonged coma of up to some weeks is occasionally followed by good recovery.

Recovery after severe TBI takes many weeks or months. During the first few weeks, patients are often intermittently restless or lethargic, and have focal deficits such as hemiparesis or aphasia. Gradually, they become more aware, though they may remain in PTA. PTA denotes a diffuse (but transient) confusional state, during which mechanisms of learning are particularly disrupted. The patient is unable to lay down any continuous memory despite being awake; it is not a pure amnestic syndrome, however, and behavioural disturbance is very common. After emergence from PTA, the patient will have little or no memory of that period. The state may last some weeks or more, and may not be obvious clinically: it is important not to miss it as it has crucial management and prognostic implications. PTA is one of the most reliable predictors of outcome: the longer the duration of PTA, the lower the likelihood of return to paid work.

Late sequelae

Sequelae of TBI are major causes of morbidity and can have serious social and medicolegal consequences. They include:

- ***Incomplete recovery***, for example cognitive impairment, hemiparesis.
- ***Post-traumatic epilepsy***.
- ***The post-traumatic (post-concussion) syndrome.*** This describes the vague complaints of headache, dizziness, and malaise that follow even minor head injuries. Litigation is frequently an issue. Depression is prominent. Symptoms may be prolonged.
- ***Benign paroxysmal positional vertigo*** (see Chapter 21).
- ***Chronic subdural haematoma*** (see Chapter 11, 'Subdural haematoma' section).
- ***Hydrocephalus*** (see 'Hydrocephalus' section).
- ***Chronic traumatic encephalopathy.*** This follows repeated and often minor injuries. It consists of cognitive impairment and extrapyramidal and pyramidal signs, seen typically in professional boxers, in whom it is often termed ***dementia pugilistica.*** Repeated concussions in athletes playing rugby and American football have also been implicated, and specific neuropathological findings have been identified in patients with CTE.

Management

Immediate management

Attention to the airway is vital. If there is coma, depressed fracture, or suspicion of intracranial haematoma, CT imaging and discussion with a neurosurgical unit are essential. Indications for CT vary from imaging all minor head injuries in some centres to more stringent criteria elsewhere.

In many severe TBI cases, assisted ventilation will be needed. ICP monitoring is valuable in comatose or sedated patients, and raised ICP requires careful treatment. Care of the unconscious patient is described in Chapter 13 ('Neurocritical care' section). Prophylactic antiepileptic drugs have been shown to be of no value in prevention of late post-traumatic epilepsy. Trials of various putative neuroprotective treatments have shown no benefits. For instance, induced hypothermia lowers ICP but has not been shown to improve outcomes and should not be used routinely.

Rehabilitation

TBI cases require skilled, prolonged, and energetic support. Survivors with severe physical and cognitive deficits require rehabilitation in specialized units. Rehabilitation includes care from a multidisciplinary team, including physiotherapists, occupational therapists, speech and language therapists, and clinical psychologists. Many survivors are left with cognitive problems (amnesia, neglect, disordered attention, and motivation) and behavioural/emotional problems (temper dysregulation, depression, and grief reactions). Long-term support for both patients and families is necessary.

Hypoxic–ischaemic brain injury

The incidence of hypoxic–ischaemic brain injury (HIBI) or hypoxic–ischaemic encephalopathy (HIE) is rising as the numbers of cardiac arrest survivors rise. Most patients with HIBI have had a period of both reduced perfusion of the brain and reduced oxygenation. HIBI can result from reduced perfusion during a cardiac arrest, from catastrophic haemorrhagic or multifocal ischaemic stroke, or from systemic vascular collapse. A purer anoxic insult is generated by carbon monoxide poisoning, drowning or asphyxiation, and prolonged hypothermia. The degree of brain injury suffered by these patients is highly variable, and recovery depends on multiple factors including the underlying aetiology, the age of the patient, the presence of co-morbidities, and the availability of specialist rehabilitation. The critical care management of acute brain injury is discussed in Chapter 13.

Disorders of consciousness

After severe TBI, diffuse HIBI, or other diffuse brain insults that disrupt widespread networks, some patients will exhibit disorders of consciousness. Consciousness consists of two components, alertness or wakefulness (the *level* of consciousness) and awareness (the *content* of consciousness). These components may be affected in tandem or differentially. The care of patients with long-term disorders of consciousness throws up ethical and legal dilemmas, and requires careful assessment and multidisciplinary decision-making.

Disorders of consciousness may be of short duration or last indefinitely. Prolonged disorders of consciousness (PDOC) last longer than 4 weeks. A proportion of patients in PDOC are young and otherwise fit, and so may survive for many years. Some have the potential to recover awareness and hence a degree of functional independence or communication. Identifying who these are requires an experienced MDT. Serial assessment is vital as fatigue and fluctuation can otherwise seriously skew the impression. The Coma Recovery Scale—Revised (CRS-R) is one example of a scoring system used for the documentation of responses in these patients.

Disorders of consciousness extend along a spectrum (Figs. 28.33 and 28.34) and comprise:

- ***Coma***, in which both the content and level of consciousness are severely reduced. The acute management of comatose patients is described in Chapter 11.
- ***Vegetative state*** (VS), which is usually a consequence of extensive cortical damage. Lower brainstem function is intact and so breathing is normal without the need for mechanical ventilation, and the patient appears *alert* with eye-opening and sleep–wake cycles. The 'vegetative' functions (breathing, sleep-wake cycle) are preserved, hence the name. However, there is no sign of *awareness* or response to environmental stimuli, except reflex movements. Feeding is supported by artificial means.
- ***Minimally conscious state*** (MCS) describes patients with limited, often fluctuating, awareness, with inconsistent but reproducible responses, for example movements in response to voice; crying or laughing in response to emotional stimuli; and vocalization in response to questions. A patient may emerge from VS into MCS, and the trajectory of any improvement is the most useful piece of information when trying to assess prognosis. In practice the border between VS and the lower reaches of MCS are blurred, and the term 'low awareness state' can be used to avoid erroneous precision.
- ***Continuing, Chronic, and Permanent MCS and VS:*** MCS and VS are defined as 'continuing' when they have been present for 4 weeks. VS can be considered 'chronic' if it has persisted for >3 months following a non-TBI, or >1 year following a TBI. 'Permanent' VS or MCS denotes that there is no chance of regaining awareness, and should only be diagnosed by a suitably qualified clinician. In general, in the UK, a disorder of consciousness will be considered permanent if there has been no change in trajectory for 6 months or longer, as measured by serial expert assessment. In the UK, the diagnosis of permanent VS no longer has legal relevance, but it can still be useful information for care teams. These definitions have changed over time and are different in different countries.
- ***Emergence from PDOC*** is characterized by demonstrating functional interactive communication (by any means) or functional object use. Anybody who has been in PDOC for more than a few months is likely to have severe permanent physical and cognitive disability, even if they emerge.

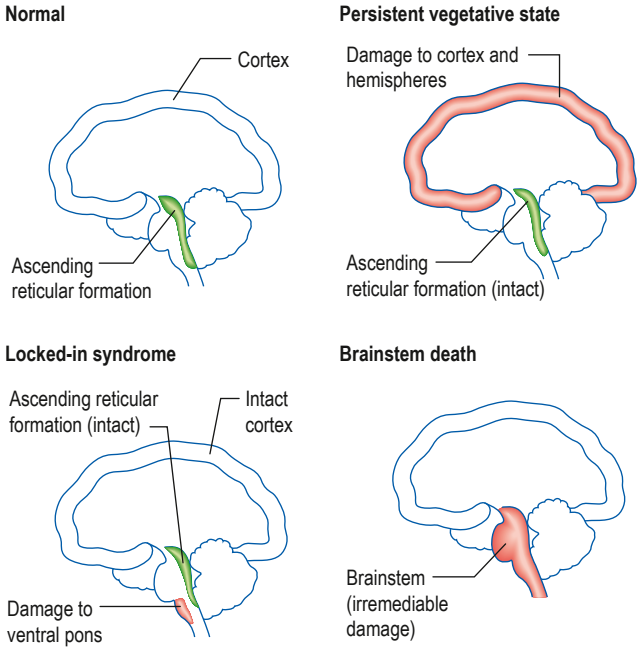


Fig. 28.33 Anatomy of vegetative state, locked-in syndrome, and brainstem death. *(Adapted from Bates D. Coma and brainstem death.* Medicine *2004; 32.)*

Locked-in syndrome is not a disorder of consciousness. Rather, consciousness is normal but voluntary movement is impossible, typically as a result of discrete brainstem pathology (Fig. 28.34). These patients can usually communicate using movements of the eyes or eyelids.

Brainstem death is discussed in Chapter 13 ('Death by neurological criteria' section).

Prognosis in brain injury and disorders of consciousness

Prognosis depends on the cause of PDOC or coma and the extent of brain damage sustained. Metabolic and toxic causes of coma have the best prognosis when the underlying problem can be corrected.

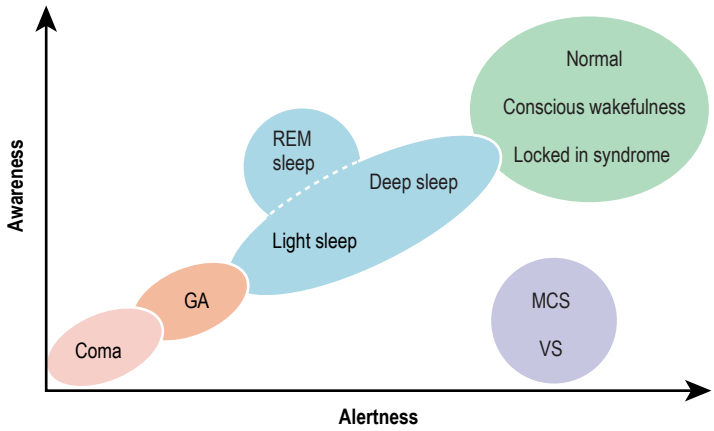


Fig. 28.34 Graphical representation of the disorders of consciousness, illustrating the dissociation between *alertness* and *awareness*. *GA,* General anaesthesia; *MCS,* minimally conscious state; *REM,* rapid eye movement; *VS,* vegetative state. *(Adapted from Laureys S. The neural correlate of (un)awareness: Lessons from the vegetative state.* Trends in Cognitive Science *2005; 556–559.)*

Prognosis after TBI is highly variable, owing to the heterogeneous nature of the condition, and many patients with TBI are young and have greater rehabilitation potential. Following HIBI, such as after cardiac arrest, only 11% make a good recovery. Of those patients who survive but do not regain consciousness soon after the injury, a substantial proportion will remain in a vegetative or minimally conscious state.

It is often useful when considering prognosis after brain injury to consider the best- and worst-case scenario in terms of recovery. The provision of ongoing supportive care (such as artificial hydration and nutrition) is governed in UK law not by the diagnosis of VS or MCS, or its judged permanence, but by considering whether the patient could ever regain a quality of life that they would themselves find acceptable. These decisions require a holistic consideration of *best interests* and discussions should not be delayed until a diagnosis of 'chronic' or 'permanent' VS is made.

Further reading

Royal College of Physicians. *Prolonged Disorders of Consciousness Following Sudden Onset Brain Injury: National Clinical Guidelines*. London: RCP, 2020.
Stocchetti N, Maas AI. Traumatic intracranial hypertension. *N Engl J Med* 2014; 370:2121–2130.

SPINAL CORD DISEASE

The cord extends from C1, the junction with the medulla, to the lower vertebral body of L1, where it becomes the conus medullaris. Blood supply is from the anterior spinal artery and a plexus on the posterior cord. This network is supplied by the vertebral arteries and by several branches from lumbar and intercostal vessels, including the artery of Adamkiewicz (the largest anterior segmental medullary artery, arising in most people between T8 and L1 and supplying the lower spinal cord via its anastomosis with the anterior spinal artery).

Spinal cord compression

The principal features of chronic and subacute cord compression are spastic paraparesis or tetraparesis, radicular pain at the level of compression, and sensory loss below the compression (**Box 28.40**).

For example, in compression at T4, a band of pain radiates around the thorax, characteristically worse on coughing or straining. Spastic paraparesis develops over months or days, depending on underlying pathology; acute-onset processes will however cause an initial flaccid paraparesis before spasticity sets in. Numbness commencing in the feet rises to the level of compression. This is called the ***sensory level*** and is usually 2 to 3 dermatome levels below the level of anatomical compression. Retention of urine and constipation develop.

? Box 28.40 Causes of spinal cord compression

- Spinal cord tumours
  - Extramedullary, e.g., meningioma or neurofibroma
  - Intramedullary, e.g., ependymoma or glioma
- Vertebral body destruction by bone metastases, e.g., prostate primary
- Disc and vertebral lesions:
  - Chronic degenerative and acute central disc prolapse
  - Trauma
- Inflammatory:
  - Epidural abscess
  - Tuberculosis
  - Granulomatous
- Epidural haemorrhage/haematoma

Aetiology

Disc and vertebral lesions. Central cervical disc and thoracic disc protrusion cause cord compression (see Chapter 32, 'Pain in the lower back' section). Chronic compression due to cervical spondylotic myelopathy is frequently seen in clinical practice and is the most common cause of a spastic paraparesis in an elderly person.

Trauma. Any trauma to the back is potentially serious and the patient should be immobilized until the extent of the injury can be determined.

Spinal cord tumours. Extramedullary tumours, such as meningiomas and neurofibromas, cause cord compression (**Fig. 28.35** and **Box 28.41**) gradually over weeks to months, often with root pain and a sensory level. Vertebral body destruction by bony metastases, such as in prostate or breast cancer, is a common cause of spinal cord compression.

Intramedullary tumours (e.g., ependymomas) are less common and typically progress slowly, sometimes over many years. Sensory disturbances similar to syringomyelia may develop (see 'Syringomyelia and syringobulbia' section).

Tuberculosis. Spinal tuberculosis is the most common cause of cord compression in countries where the disease is common. There is destruction of vertebral bodies and disc spaces, with local spread of infection. Cord compression and paraparesis follow, culminating in paralysis: Pott's paraplegia.

Spinal epidural abscess.

Epidural haemorrhage and haematoma. These are rare sequelae of anticoagulant therapy, bleeding disorders, or trauma, and can follow LP when clotting is abnormal. A rapidly progressive cord or cauda equina lesion develops.

Fig. 28.35 Thoracic meningioma *(arrowed)* compressing the spinal cord: T2-weighted MRI.

i Box 28.41 Principal spinal cord neoplasms

Extradural
- Metastases:
  - Bronchus
  - Breast
  - Prostate
  - Lymphoma
  - Thyroid
  - Multiple myeloma

Intradural extramedullary
- Meningioma
- Nerve sheath tumours (schwannoma, neurofibroma)
- Lipoma

Intramedullary
- Ependymoma
- Astrocytoma
- Haemangioblastoma
- Intramedullary metastases

Management

Acute spinal cord compression is a medical emergency. Early diagnosis and treatment are vital. MRI is the imaging technique of choice. A trap for the unwary is imaging the wrong part of the spine: a lumbar spine MRI only covers the conus and the cauda equina, and if spinal cord compression is possible then the whole spine must be imaged, particularly in TB or cancer where multiple sites of compression are common.

Routine tests (e.g., chest X-ray) may indicate a primary neoplasm or infection. Surgical exploration is frequently necessary; if decompression is not performed promptly, irreversible cord damage ensues. Results are excellent if benign tumours and haematomas are removed early. Radiotherapy is used to treat cord malignancies, or compression due to inoperable malignant vertebral body disease causing cord compression.

Other spinal cord disorders

Inflammatory cord lesions (transverse myelitis)

See 'Transverse myelitis' section.

Anterior spinal artery occlusion

There is acute paraplegia and loss of spinothalamic (pain and temperature) sensation, with infarction of the anterior two-thirds of the spinal cord. This may result from aortic atherosclerosis, dissection, trauma, or cross-clamping in surgery. Vasculitis, emboli, haematological disorders causing thrombosis, and severe hypotension are other causes. Occlusion of the artery of Adamkiewicz, which supplies the thoracic anterior spinal artery, causes watershed infarction of the cord, typically at the T8 level where perfusion is relatively poor.

Arteriovenous malformations of the cord

Although rare and difficult to diagnose, spinal arteriovenous malformations (AVMs) are potentially curable. The two main types seen are dural arteriovenous fistulas (acquired) and true intramedullary AVMs (probably congenital but gradually enlarging). Dural arteriovenous fistulas occur mainly in middle-aged men due to formation of a direct connection between an artery and vein in a dural nerve root sleeve. This causes arterialization of veins with venous hypertension, and thus oedema and congestion of the spinal cord at and below the affected level. Presentation is with a gradually progressive myelopathy over months or a few years, often with thoracic back pain. MRI usually shows cord swelling (Fig. 28.36) and may show the enlarged arterialized veins over the surface of the cord.

Fig. 28.36 Dural arteriovenous fistulas.

Spinal angiography demonstrates the fistula and allows endovascular ablation with glue, often with complete resolution of symptoms if permanent neuronal damage has not already occurred.

Genetic disorders—hereditary spastic paraparesis

Several genetic disorders may present with a gradually evolving UMN syndrome resembling a myelopathy. Typically, spasticity and stiffness dominate the clinical picture rather than weakness, especially in hereditary spastic paraparesis (HSP). Muscle relaxants, such as baclofen, improve gait. There are more than 80 known genes associated with HSP, some causing 'pure' spasticity and others with associated neurological features, such as thinning of the corpus callosum.

Other genetic disorders such as adrenoleukodystrophy may cause a slowly progressive spastic paraparesis (including in manifesting female carriers), as can the spinocerebellar ataxias (SCAs) or presenilin-1 mutations.

Vitamin B_{12} deficiency

Subacute combined degeneration of the cord resulting from vitamin B_{12} deficiency (see Chapter 25, 'B vitamins' section) is the most common example of metabolic disease causing spinal cord damage. Abuse of nitrous oxide may precipitate functional B_{12} deficiency with normal serum B_{12} levels, diagnosed by demonstrating high serum homocysteine and/or methylmalonic acid concentrations. The dorsal columns are preferentially affected causing a marked sensory ataxia, but the term 'combined' refers to frequent additional involvement of the lateral columns (corticospinal tracts). Mixed upper and lower motor neuron signs are classically observed owing to peripheral nerve involvement. Emergency B_{12} replacement is mandatory (and must be initiated before replacement of folate, which can otherwise precipitate SACDC).

Other causes of a spastic paraparesis

MND may present initially with a spastic paraparesis before LMN features develop (see 'Motor neuron disease' section). Paraneoplastic disorders, radiotherapy, copper deficiency, liver failure, and rare toxins (e.g., lathyrism) may cause spinal cord damage. Not all causes of paraparesis relate to spinal cord pathology; beware a parasagittal cerebral meningioma presenting with a paraparesis due to bilateral compression of the leg area of the motor cortex.

Care of the patient with paraplegia

Where patients are left with a severe paraplegia, there are several issues to address in long-term care, and specialist nursing is vital.

Bladder management. The bladder does not empty and urinary retention results. Patients self-catheterize or develop reflex bladder emptying, helped by abdominal pressure. Early treatment of urine infections is essential. Chronic kidney disease may develop from chronic obstruction or repeated urinary tract infection.

Bowel function. Constipation and impaction must be avoided. Following acute paraplegia, manual evacuation is necessary; reflex emptying develops later.

Skin care. Risks of pressure ulcers and their sequelae are serious. Meticulous attention must be paid to cleanliness and regular turning. The sacrum, iliac crests, greater trochanters, heels, and malleoli should be inspected frequently. Pressure-relieving mattresses are useful initially until patients can turn themselves. If pressure ulcers develop, plastic surgery may be required. Pressure palsies, such as those of ulnar nerves, can occur.

Lower limbs. Passive physiotherapy helps to prevent contractures. Severe spasticity, with flexor or extensor spasms, may be helped by muscle relaxants such as baclofen or by botulinum toxin injections.

Rehabilitation. Many patients with traumatic paraplegia or tetraplegia return to self-sufficiency (especially if the level is at C7 or below). A specialist spinal rehabilitation unit is necessary. Lightweight, specially adapted wheelchairs provide independence. Tendon transfer operations may allow functional grip if hands are weak. Autonomic dysreflexia may be a problem. Patients with paraplegia have substantial practical, psychological, and sexual needs.

Syringomyelia and syringobulbia

A syrinx is a fluid-filled cavity within the spinal cord. Syringobulbia means a cavity in the brainstem. Syringomyelia is frequently associated with a Chiari malformation (see Fig. 28.32). The abnormality at the foramen magnum probably allows normal pulsatile CSF pressure waves to be transmitted to fragile tissues of the cervical cord and brainstem, causing secondary cavity formation. The syrinx is in continuity with the central canal of the cord. Syrinx formation may also follow spinal cord trauma or infection and lead to secondary damage years later, and can also be caused by intrinsic cord tumours.

Pathophysiology

The expanding cavity in the cord gradually destroys the neurons closest to the central canal: spinothalamic neurons, anterior horn cells, and lateral corticospinal tracts. In the medulla (syringobulbia), lower cranial nerve nuclei are affected.

Clinical features

Cases associated with the Chiari I malformation usually develop symptoms around the age of 20 to 30. Neck, upper back, and shoulder pain exacerbated by exertion or coughing is typical. Spinothalamic sensory loss—pain and temperature—leads to painless upper limb burns and trophic changes. Paraparesis develops. The following are typical signs of a substantial cervical syrinx (Fig. 28.37, and see Fig. 28.32):

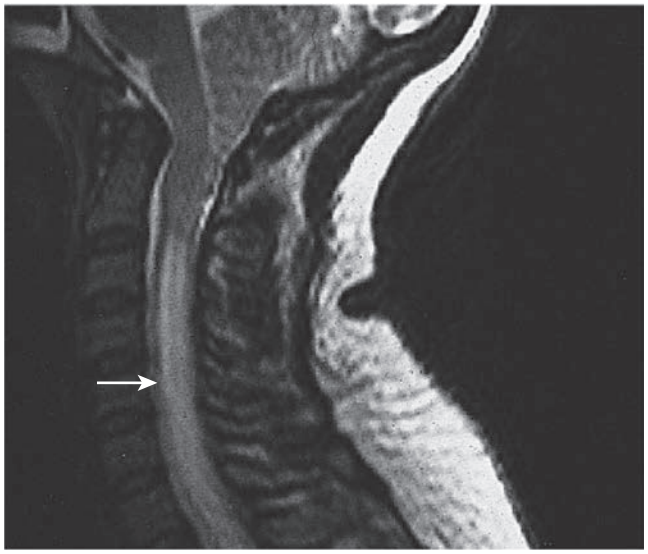


Fig. 28.37 Cervical syrinx—cavity within cord *(arrowed)*: T2-weighted MRI.

- ***a 'suspended' area of dissociated sensory loss***—that is spinothalamic loss in a cape-like distribution without loss of light touch
- ***loss of upper limb reflexes***
- ***muscle wasting*** in the hand and forearm
- ***spastic paraparesis***—initially mild
- ***brainstem signs***—as the syrinx extends into the brainstem (syringobulbia), there may be tongue atrophy and fasciculation, bulbar palsy, a Horner's syndrome, and impairment of facial sensation.

Investigations and management

MRI demonstrates the cavity and may shed light on the underlying cause. Syringomyelia is gradually progressive over several decades. Sudden deterioration sometimes follows minor trauma or occurs spontaneously. Surgical decompression of the foramen magnum in the presence of a Chiari often causes the syrinx to collapse.

Further reading

Ginsberg L. Disorders of the spinal cord and roots. *Pract Neurol* 2011; 11:259–267.

DEMENTIA

Dementia is broadly a clinical syndrome of decline in neurocognitive function, of sufficient severity to cause significant functional impairment. One or more domains of cognition will be affected, including often episodic memory or language function, but also potentially frontal executive function, visuospatial function, praxis, social cognition, and complex attention. Unlike delirium, dementia is chronic and in most cases progressive (rather than acute and fluctuating).

There are multiple causes (Box 28.44), but the key ones are ***neurodegenerative***. The DSM-V diagnostic criteria for major neurocognitive disorder (intended to be synonymous with dementia) allow for non-progressive and even reversible dementia diagnoses, but in the majority of cases, these are more usefully described as dementia mimics, such as cognitive impairment from autoimmune encephalitis or hydrocephalus.

Dementia robs patients of their independence, and is a serious burden on carers and a major socioeconomic challenge for society as a whole.

Epidemiology

Dementia is common, and increasing in prevalence as a result of an ageing population and better case ascertainment. Age is the main risk factor, followed by family history. Over the age of 65, there is a 7% prevalence; over the age of 85, the prevalence rises to over 30%.

Clinical assessment

There are two main considerations:

- Does the patient have dementia?
- Is the pattern of cognitive deficits, tempo of progression, or associated features suggestive of a distinct cause? Cognitive problems need to be interpreted in the context of estimated premorbid abilities (e.g., based on educational attainments or occupation).

A well-taken history is the key to an accurate dementia diagnosis. Taking a collateral account from a relative or friend is essential. Patients may tend to downplay or deny symptoms (anosognosia) or

constantly look to the relative for answers (the 'head-turning sign'). See Box 28.42 for the key elements in history-taking.

Examination

Conversation with the patient during history-taking may be as revealing as formal cognitive assessment, but many patients hide deficits well behind an intact social façade.

Bedside cognitive assessment

The mini-mental state examination (MMSE) is commonly used as a quick screen of cognitive function but has its limitations, such as poor sensitivity for milder cognitive impairment and frontal lobe dysfunction, especially in those with high pre-morbid abilities. The Montreal Cognitive Assessment (MoCA) has greater sensitivity for mild cognitive impairment (MCI) and also incorporates some testing of executive function (the trail-making test) but is as short as the MMSE. Licensing and cost limit the use of both tools. The Addenbrooke's Cognitive Examination–III (ACE-III) is a tool developed to address the deficiencies of the MMSE, and is short enough to use in clinical practice. It has some value in diagnosing frontotemporal dementia as well as Alzheimer's disease, although none of these tools can be used in isolation to differentiate between dementia syndromes.

It is useful to ask patients to give an account of recent news events to assess episodic memory, and to ask about route-finding to assess topographical memory ('Can you tell me how to get from your home to the clinic?').

Individual cognitive domains can be tested separately in detail: for example, clock drawing for visuospatial (parietal lobe) function; naming and reading tasks for language function; verbal fluency, conceptual similarity (to test abstract thinking) and stop–go tasks, which are components of the Frontal Assessment Battery (FAB).

Check for primitive reflexes (frontal release signs), such as grasp, palmo-mental and pout reflexes, and perseveration or utilization behaviour with frontal lobe involvement.

Test:

- ideomotor limb praxis—copying hand gestures and miming tasks, e.g., 'Show me how you brush your teeth'.
- orofacial praxis, e.g., 'Show me how you would blow out a candle'.

A complete neurological examination (to look for evidence of papilloedema, parkinsonism, myoclonus, and gait disorders, for example), is also necessary, in addition to general examination and assessment of mental state.

Investigations

Investigations (Box 28.43) are aimed at identifying treatable causes and helping support a clinical diagnosis of dementia type. For most patients, this should include the elements listed below.

Blood tests. These are primarily to rule out reversible causes of cognitive impairment and should include a full blood count and measurement of vitamin B_{12}, thyroid function, urea and electrolytes, liver function, glucose, ESR, HIV, and syphilis serology.

Brain imaging. CT is adequate to exclude structural lesions (Fig. 28.38), such as tumours or hydrocephalus. The superior anatomical resolution of MRI can identify patterns of regional brain atrophy and so help to distinguish between different types of degenerative dementia (e.g., hippocampal atrophy in Alzheimer's disease (AD) versus temporal lobe and frontal atrophy in frontotemporal dementia). Imaging also allows assessment of brain 'vascular load'. MRI is always preferable to CT in patients with cognitive disorders. Serial volumetric imaging can be helpful to view the trajectory of atrophy over time.

Detailed neuropsychometric assessment. Appropriate in most patients, this allows quantification of the relative involvement of different cognitive domains and may be helpful if performed serially over time to assess progression.

CSF examination. Measurement of CSF protein biomarkers has been shown to be useful in distinguishing between different types of dementia; for example, in AD, CSF tau is raised and Aβ42 reduced (or the Aβ42:Aβ40 ratio is reduced). In Creutzfeldt–Jakob disease (CJD) and other rapidly progressive dementias, protein 14–3–3 is

Box 28.42 Taking a dementia history

- Memory:
  - Is the patient repetitive, e.g., with questions?
  - Is there a temporal gradient of amnesia—preservation of more distant memories with amnesia for recent events?
  - Is there difficulty learning to use new devices, e.g., computer, mobile phone?
- Functional ability:
  - Has work performance or ability to cook and do domestic tasks declined?
  - Has responsibility for finances and administration shifted to the partner?
  - Does the patient get easily muddled?
- Personality and frontal lobe function:
  - Has personality changed?
  - More aggressive/apathetic/lacking initiative
  - Lacking empathy
  - Disinhibition
  - Change in food preference or religiosity
- Language:
  - Difficulty with word finding or remembering names
- Visuospatial ability:
  - Does the patient get lost in familiar places?
  - Difficulty dressing, e.g., putting jacket on the wrong way round
- Psychiatric features:
  - Features of depression
  - Hallucinations (visual or auditory?)
- Tempo of progression
- Family history of dementia
- Alcohol and drug use
- Medication
- Any other neurological problems, e.g., parkinsonism, gait disorder, strokes

Box 28.43 Tests in dementia

Blood tests

- Full blood count, erythrocyte sedimentation rate, vitamin B_{12}
- Urea and electrolytes
- Glucose
- Liver biochemistry
- Serum calcium
- Thyroid stimulating hormone, T_3, T_4
- HIV serology
- Syphilis serology

Imaging

- CT or MRI brain scan

Other—selected patients only

- Cerebrospinal fluid—including tau and Aβ42 measurement
- Genetic studies, e.g., for AD and FTD genes, HD, prion mutations
- Electroencephalography
- Advanced brain imaging (e.g., FDG-PET, SPECT, amyloid PET, or DaTScan)
- Brain biopsy

Fig. 28.38 CT scan showing a large frontal meningioma presenting with apathy and frontal dementia.

increased. A new CSF assay for abnormal prion protein, ***real-time quaking-induced conversion (rt-QUIC)***, has proved to be highly sensitive and specific for CJD.

Specialist tests—in selected patients (e.g., those <65 years old, those with a positive family history or those with atypical features including rapid onset), more intensive investigation may be necessary. This may include EEG, genetic tests (e.g., for Huntington's, familial AD, and FTD genes), voltage-gated potassium channel antibodies, anti-neuronal antibodies (for paraneoplastic limbic encephalitis); or on rare occasions, brain biopsy may be appropriate.

New imaging modalities. Functional imaging (FDG-PET or SPECT) can sometimes be useful to identify areas of brain hypometabolism/hypoperfusion. Radionuclide scans, using radioactively labelled ligands that bind directly to amyloid, allow amyloid deposition in the brain to be directly visualized, and are used widely as a research tool for clinical trial enrolment. Amyloid imaging is now beginning to enter clinical practice in some specialist centres, but appropriate patient selection and interpretation is vital. A positive scan lends support to a diagnosis of Alzheimer's disease in the right clinical context, but amyloid deposition can arise as a co-pathology in patients with another cause of dementia, or in asymptomatic individuals.

Mild cognitive impairment

Mild cognitive impairment (MCI) is an intermediate state between normal cognition and dementia. Often mild memory impairment, greater than expected for age but not sufficient to be classified as dementia, is the only symptom ('amnestic MCI'). MCI may be a pre-dementia state, with 10% to 15% of patients per year developing overt dementia.

Causes and risk factors

There are many causes of dementia (Box 28.44), by far the most common being AD. Cause varies according to age (Fig. 28.39). Most commonly, multiple pathologies converge to cause dementia in older people although there will generally be a dominant aetiology.

Aetiologies are distinct from risk factors—and many risk factors for dementia have been identified, some of which are modifiable (Box 28.45). Many of these increase the risk for multiple dementia types (several vascular risk factors increase the risk of Alzheimer's disease as well as vascular dementia).

? Box 28.44 Causes of dementia

Degenerative
- Alzheimer's disease
- Dementia with Lewy bodies
- Frontotemporal dementia
- Huntington's disease
- Parkinson's disease
- Progressive supranuclear palsy
- Prion diseases, e.g., Creutzfeldt–Jakob

Vascular
- Vascular dementia
- CADASIL (Cerebral Autosomal Dominant Arteriopathy with Subcortical Infarcts and Leukoencephalopathy)

Metabolic[a]
- Liver failure

Toxic[a]
- Alcohol
- Heavy metals
- Recreational drugs (e.g., cocaine, solvents)
- Medications (e.g., chemotherapy)

Traumatic
- Severe or recurrent traumatic brain injury (including chronic traumatic encephalopathy)

Structural lesions[a]
- Hydrocephalus (including idiopathic normal pressure hydrocephalus)
- Tumours

Infections/Inflammatory[a]
- HIV
- Neurosyphilis
- Whipple's disease
- Autoimmune/paraneoplastic encephalitis
- Multiple sclerosis

Other
- Mixed dementia/multiple aetiologies

[a]Potentially reversible aetiologies, in which case the presentation is best classified as a dementia mimic.

Dementia syndromes

Alzheimer's disease

Although technically a definitive diagnosis can only be made by histopathology, in practice the clinical features are sufficiently characteristic that a diagnosis can usually be made with considerable accuracy in life, supported by diagnostic investigations. The key clinical features are:

- ***Memory impairment*.** Memory-led cognitive decline is the hallmark of Alzheimer's disease. Episodic (day-to-day) memory is affected. There is progressive loss of ability to learn, retain, and process new information. There is a characteristic temporal gradient, with relative preservation of remote memory and amnesia for more recent events. This type of recent memory relies on the hippocampus and related structures in the mesial temporal lobe, which is the area of the brain most severely affected in classical AD. This type of anterograde long-term memory deficit is often incorrectly termed 'short-term memory loss', which actually refers to loss of working memory: for example, digit span, which is preserved in the early stages of AD. Procedural (how to do things) and semantic (knowledge of general facts and concepts) memory is also spared until the later stages.

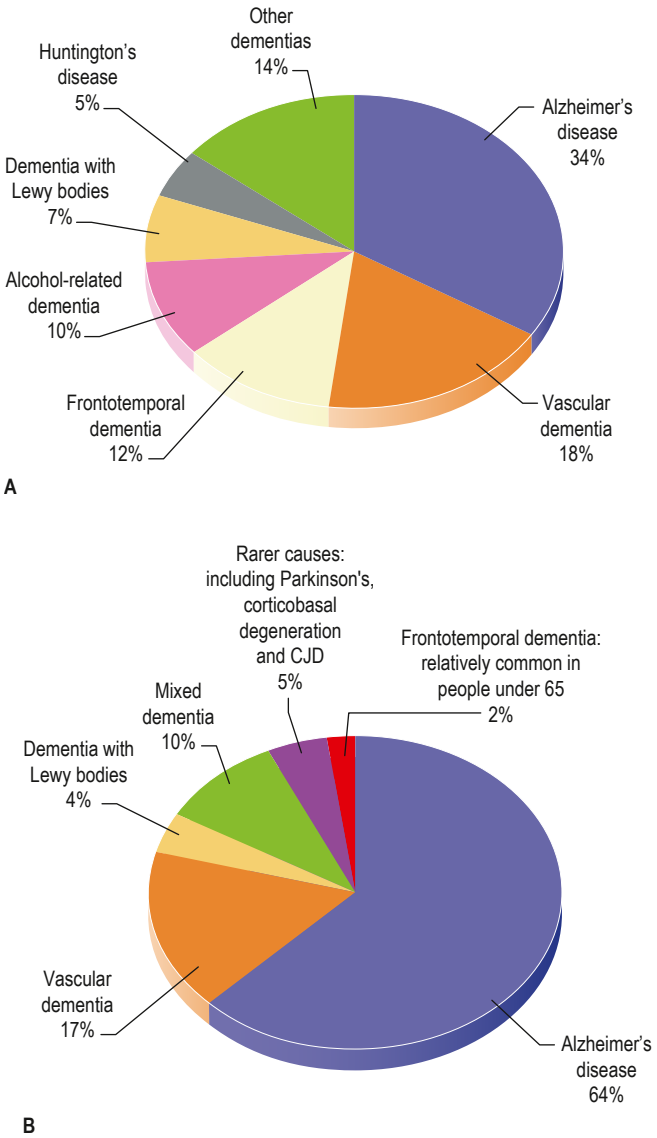


Fig. 28.39 Causes of dementia at different ages. **(A)** Under 65 years. **(B)** At 65 years or more. *CJD,* Creutzfeldt–Jakob disease. *(Data from Prince M, Knapp M, Guerchet M et al.* Alzheimer's Society Dementia UK Update. *London: Alzheimer's Society; 2014.)*

Box 28.45 Risk factors for dementia

- Age
- Genetic factors
- Hypertension
- Smoking
- Diabetes
- Chronic kidney disease
- Obesity
- Lack of physical activity
- Poor diet
- Alcohol excess
- Low levels of cognitive engagement
- Depression
- Traumatic brain injury
- Hearing loss
- Social isolation
- Air pollution

- ***Topographical disorientation***, or difficulty navigating, is frequently an early sign, as expected from the role of the hippocampus in topographical memory. There may be a history of getting lost.
- ***Aphasia.*** Language usually becomes impaired as the disease advances. Difficulty with word finding (anomia) is characteristic.
- ***Apraxia.*** Ability to carry out skilled motor activities is impaired (see 'Apraxia' section).
- ***Visuospatial dysfunction*** including difficulties with parking or filling out forms (parietal).
- ***Altered auditory processing.*** Inability to pick out a conversation in a noisy environment can be a very early sign of AD.
- ***Frontal executive dysfunction.*** Organizing, planning, and sequencing are impaired.
- ***Anosognosia.*** Lack of insight by the patient into their difficulties is common; they may be reluctant to seek medical attention but be brought to clinic by a family member. Where insight is preserved, depression is common.

- ***Personality change.*** In contrast to other dementias such as frontotemporal dementia, basic personality, and social behaviour remain intact until late AD, but more subtle behavioural and emotional changes are common.
- ***Tempo.*** Disease onset is insidious and often not noticed by family members initially. Progression is gradual but inexorable over a decade or longer, with eventual severe deficits in multiple cognitive domains.
- ***Atypical variants:***
  - ***Posterior cortical atrophy (PCA).*** The visual variant of AD is caused by initial involvement of the occipital lobes and occipitoparietal regions. Patients have complex visual symptoms that may be difficult to describe; they often say that it is easier to see distant than close-up objects. Visuoperceptive deficits manifest with difficulties recognizing faces or objects in spite of intact acuity; they will be unable to read fragmented letters. Visuospatial skills are also affected, as are nearby parietal functions (reading, praxis, numeracy), while memory is initially well preserved. The ***biparietal variant*** is a form of PCA where the dorsal ('where?') occipitoparietal stream is affected more than the ventral ('what?') occipitotemporal network. PCA is occasionally caused by other neurodegenerative pathologies (such as dementia with Lewy bodies, corticobasal degeneration, or prion disease).
  - ***Logopenic-variant primary progressive aphasia* (lvPPA).** Occasionally, AD will present with early language impairment and preserved episodic memory. There will be word-finding pauses, length-dependent difficulties with sentence repetition, and phonological errors in speech with intact comprehension and grammar.
  - ***Frontal variant*** AD causes prominent behavioural and/or dysexecutive features.
- ***Late non-cognitive features.*** Myoclonus may develop, sometimes followed by seizures (the cortex is the main site of pathology). Sleep–wake cycle reversal and incontinence may place a great strain on carers. Gegenhalten or paratonic rigidity (resistance to passive movement that increases with the velocity of the movement) may be seen, and seizures can occur. Overall motor function may be strikingly preserved, so patients are capable of wandering and getting lost. Swallowing may become impaired, leading to aspiration pneumonia—often a terminal event.

Investigations

MRI typically shows characteristic atrophy of mesial temporal lobe structures, including hippocampi, progressing eventually to generalized cerebral atrophy. Imaging may be normal in the early stages and selective regional atrophy is seen in AD variants, such as occipital lobe volume loss in PCA. Characteristic MRI and psychometric testing abnormalities are sufficient to make a diagnosis if the clinical picture is suggestive (a progressive amnestic cognitive disorder in an older person). CSF tau and β-amyloid measurement is helpful in cases of diagnostic difficulty, and there may be a role for amyloid PET (see above).

Molecular pathology

Although the cause of AD is still not known, a great deal is now understood about the molecular pathology. The pathological and biochemical hallmarks are:

- The deposition of β-amyloid (Aβ) in amyloid plaques in the cortex. The amyloid precursor protein (APP) is processed by secretase enzymes to form pathogenic $A\beta_{1–42}$ monomers, which polymerize into extracellular amyloid plaques (Fig. 28.40). Amyloid may also be laid down in cerebral blood vessels, leading to amyloid angiopathy, which frequently co-occurs with AD. Amyloid plaques are frequently associated with dystrophic neurites (swollen axons and dendrites) and an inflammatory response.
- Structural and conformational changes in tau protein, that is hyperphosphorylation and aggregation into paired helical filaments, lead to the formation of neurofibrillary tangles. These protein aggregates damage synapses and ultimately lead to neuronal death. Early pathological changes in the brain, including Aβ deposition, predate clinical symptoms, and diagnosis by up to 25 years.
- A basal forebrain cholinergic deficit occurs and may explain the therapeutic response to cholinesterase inhibitor drugs.

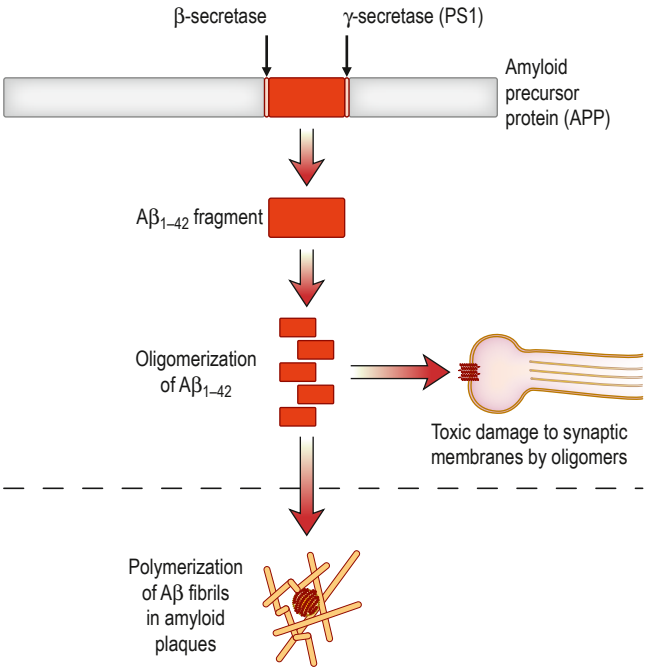


Fig. 28.40 Molecular pathogenesis of Alzheimer's disease. Processing of the amyloid precursor protein (APP) molecule by secretase enzymes releases $A\beta_{1–42}$ monomers, which are the building blocks of amyloid plaques. Aβ oligomers may damage synaptic membranes. Mutations in APP or γ-secretase (presenilin-1) cause genetic forms of Alzheimer's. Disruption of the balance between formation and degradation of $A\beta_{1–42}$ is thought to be a major problem.

Genetics of AD

A first-degree relative with AD confers a doubled lifetime risk of AD. There are rare autosomal dominant monogenic early-onset forms of familial AD with high penetrance, caused by mutations in specific genes; taken together, these account for only 1% of cases of AD.

Amyloid precursor protein (APP). Point mutations in the APP gene can cause AD, and the presence of three copies of the APP gene on chromosome 21 in Down's syndrome patients is responsible for the high incidence of AD in that condition (a majority of patients with Down's syndrome will develop AD by the sixth decade of life).

Presenilin (PS)-1 and 2. Mutations in these genes affect the γ-secretase enzyme function (see Fig. 28.40). PS1 mutations account for 50% of monogenic forms of AD. The PS1/2 and APP genes may be sequenced for mutations in selected early-onset cases (<65) with a family history.

Other genes. The E4 allele of the apolipoprotein E gene confers an increased risk of AD (2–3 times lifetime risk), especially if two copies of the E4 allele are inherited (8–12 times risk). Several other candidate genes have been identified as risk factors for AD, with modest effect size, in large genome-wide association studies.

Acquired risk factors

Age is the main risk factor for AD, as incidence increases exponentially with age. Vascular risk factors, including hypertension, diabetes, and dyslipidaemia also increase AD risk (and may be modifiable); head trauma, smoking, and air pollution also increase risk. AD shares risk factors with other causes of dementia (**Box 28.45**).

Dementia with Lewy bodies and Parkinson's disease dementia

Dementia with Lewy bodies (DLB) is characterized by the early onset of visual hallucinations, fluctuating cognition with variation in attention and alertness, REM sleep behaviour disorder, and parkinsonism. Delusions, dysautonomia, and transient loss of consciousness occur. The visual hallucinations often take the form of people or animals, or the sense of a presence ('extracampine hallucinations'). In REM sleep behaviour disorder, the normal atonia associated with REM sleep is lost and so dream enactment takes place. Memory loss may be absent in the early stages, and instead the cognitive profile is typified by deficits in attention and visuospatial function; unlike in other types of dementia, extreme cognitive fluctuations can occur, which can be mistakenly attributed to delirium. Lewy bodies, inclusions containing aggregates of the protein α-synuclein first described in Parkinson's disease, are found in the cortex.

In DLB, the cognitive features dominate; parkinsonism may evolve later and is typically mild. In Parkinson's disease dementia (PDD), cognitive problems are a late feature, occurring at least 1 year after onset and usually after the age of 75. Both conditions may respond to cholinesterase inhibitors. Patients with DLB/PDD may be very sensitive to neuroleptic drugs with dramatic worsening.

Vascular dementia

This common cause of dementia is due to several mechanisms, including multi-infarct dementia, cerebral small vessel disease, and post-stroke dementia. Most vascular dementias are of mixed cause. Vascular dementia is distinguished from AD by its clinical features and imaging (but the two frequently coexist). Cognition classically declines in a stepwise fashion, with the accumulation of new infarcts, but it can also be progressive. Cognitive decline may follow a clinically diagnosed stroke or strokes (post-stroke dementia), or may arise without a recent stroke. Apraxic gait disorder, extrapyramidal rigidity worse in the legs, and urinary incontinence are common additional features. Widespread small-vessel disease seen on MRI is the typical finding and may produce a variety of cognitive deficits, reflecting the site of ischaemic damage, but the most common is a frontal-subcortical profile, with slowed processing, reduced verbal fluency, and dysexecutive features.

Cerebral amyloid angiopathy can also cause significant small vessel disease and a vascular dementia, with or without coexistent Alzheimer's disease. Genetic cerebral angiopathies, such as cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL), may cause a progressive subcortical dementia.

However, incidental cerebrovascular disease is a common finding in the elderly. Cognitive decline should not be ascribed to vascular dementia based purely on an imaging report.

Frontotemporal dementia

Frontotemporal dementia (FTD) is the term used to describe a group of neurodegenerative disorders characterized by asymmetric frontal lobe and temporal lobe atrophy, and changes in behaviour, personality, and/or language function. Onset is usually below the age of 65 and there is often a family history. FTD is considerably less common than AD, prevalence being approximately 10 per 100 000 before the age of 65. Some patients with FTD will also develop a motor syndrome (parkinsonism or MND), because of overlapping neuropathology.

There are three distinct clinical syndromes, depending on which anatomical region is affected first.

Behavioural variant FTD is the most common subtype, and characterized by personality change, emotional blunting, apathy, disinhibition, carelessness, and behavioural change with striking preservation of episodic memory.

The other two clinical subtypes are two variants of ***primary progressive aphasia***, and are characterized by progressive impairment of language function. (The classification here is confusing, with *logopenic variant PPA* being typically caused by Alzheimer's disease [see above], but the other two major subtypes, *non-fluent variant* and *semantic variant PPA*, being types of FTD.)

Semantic variant primary progressive aphasia (svPPA), with asymmetric atrophy of the dominant anteroinferior temporal lobe, produces progressive difficulty with comprehension of words (e.g., the patient will not be able to tell you what an owl is, or point to a picture of an owl). Speech will be fluent and well-structured but lacking in meaningful content. Eventually there is a profound loss of object knowledge (semantic knowledge), beyond just the words themselves (the patient will not know that daffodils are yellow). Insight is impaired.

Non-fluent variant primary progressive aphasia (nfvPPA) is associated with peri-Sylvian atrophy, with involvement of Broca's area (inferior frontal gyrus) and the insula. Speech is effortful and distorted, with loss of verbal fluency, and grammar is lost resulting in increasingly telegraphic speech ('expressive agrammatism'). Insight is preserved and the patient may be very frustrated or depressed.

The frontal and temporal presentations eventually merge as cognitive decline becomes more widespread.

Pathology

About 30% of cases are familial, with the majority of the heritability accounted for by autosomal dominant mutations in the microtubule-associated protein tau *(MAPT)*, progranulin (GRN), or chromosome 9 open reading frame 72 *(C9orf72)* genes. The pathology is heterogeneous, grouped together under the umbrella of 'frontotemporal lobar degeneration'. There are three major types defined by the deposition of an abnormally aggregated protein: phosphorylated tau (FTLD-tau), transactive response DNA-binding protein 43 (FTLD-TDP), or fused in sarcoma (FTLD-FUS).

Many patients with FTD will develop motor symptoms, either of motor neuron degeneration or parkinsonism. Fifteen percent of patients with FTD will develop overt MND, and others have features of parkinsonian disorders (PSP features are particularly common in patients with nfvPPA). These linked disorders are all caused by FTLD pathology.

There is no cure or specific treatment at present for FTD.

Prion diseases, including Creutzfeldt–Jakob disease

Prion diseases (see Chapter 42, 'Transmissible spongiform encephalopathies [prion diseases]' section) are transmissible neurodegenerative disorders with a long incubation period, caused by

accumulation of misfolded native prion protein (PrP^c). Misfolding and conformational change in PrP^c is caused either by exposure to the abnormal misfolded isoform of the protein (PrP^{sc}) or by mutations in the PrP gene *(PRNP)*, leading to toxic accumulation of PrP^{sc} as amyloid in beta-pleated sheets. Neuronal cell damage and 'spongiform' change in the brain result, the clinical correlate being a rapidly progressive dementia in most cases.

Creutzfeldt–Jakob disease

Creutzfeldt–Jakob disease (CJD) is the most common prion disease in humans, the animal equivalents being bovine spongiform encephalopathy (BSE) in cattle and scrapie in sheep. CJD may be sporadic, iatrogenic, or familial.

Sporadic CJD (sCJD) is the most common form, occurring over the age of 50, with an incidence of approximately 1 per million. It is thought to be due to spontaneous stochastic conformational change in PrP^c to PrP^{sc} with a subsequent 'domino effect': PrP^{sc} replicates by auto-catalytic templating, inducing misfolding in other PrP molecules. A rapidly progressive dementia leads to death within 6 months of onset. Rapidly progressive cognitive decline should always lead to suspicion of CJD. The presence of myoclonus is also a clinical clue (present in 90%). Monoclonal antibodies and antisense oligonucleotide treatments are being studied.

Iatrogenic CJD is transmitted from neurosurgical instruments (prions are resistant to sterilization), transplant material (e.g., corneal grafts), and cadaveric pituitary–derived growth hormone taken from patients with CJD or presymptomatic CJD. Iatrogenic CJD has a long incubation period of several years.

Familial CJD (rare) is associated with PRNP gene mutations. Other genetic prion diseases, such as ***fatal familial insomnia***, also occur.

Variant CJD (vCJD) was first seen in the UK in 1995. vCJD patients are younger than sporadic cases, with a mean age of 29. Early symptoms are neuropsychiatric, followed by ataxia and dementia with myoclonus or chorea. vCJD has a longer course than sporadic CJD—up to several years. vCJD and BSE are caused by the same prion strain, and there is epidemiological and experimental evidence that vCJD is transmitted by consumption of BSE-contaminated beef. Transmission via blood transfusion may also occur. Most patients with vCJD and sporadic CJD have a specific polymorphism at codon 129 of the PRNP gene that leads to susceptibility.

Diagnostic evaluation

Neuropathological confirmation is required for definitive diagnosis of sporadic and acquired prion diseases, but brain biopsy should be avoided unless a treatable alternative is being considered. In sCJD, MRI typically shows high signal on DWI, FLAIR, and T2 in the caudate head, putamen, and along the cortex (cortical ribboning). A majority of patients will have periodic sharp wave complexes on EEG. However, the most useful diagnostic test is real-time quaking-induced conversion (RT-QuIC) on CSF samples: this is a 'seeding assay' that detects the presence of PrP^{sc} through an induced conformational change in recombinant prion protein. RT-QuIC is highly sensitive and specific for sporadic CJD. CSF tau and protein 14-3-3 are also used as adjunctive biomarkers but are less accurate than RT-QuIC.

Variant CJD cannot be confirmed with RT-QuIC, but tonsillar biopsy (confirming the presence of tissue PrP^{sc}) is diagnostic. MRI demonstrates FLAIR hyperintensity in the pulvinar of the thalamus bilaterally, or in both pulvinar and dorsomedial thalamus ('hockey stick sign').

Other dementias

Other neurodegenerative disorders may include dementia as one of their clinical manifestations. For example, CBD and PSP are intimately linked to frontotemporal dementia. Dementia may be a feature of a number of genetic and metabolic disorders, such as Huntington's disease (see 'Huntington's disease' section) and the leukodystrophies.

Management of dementia

It is rare for a reversible cause for a dementia presentation to be found, such as hydrocephalus or a frontal meningioma. Other disorders, such as limbic encephalitis or severe depression, may present as dementia mimics and every effort should be made to distinguish these treatable conditions from degenerative dementias.

Management is supportive, to preserve dignity and to provide care for as long as possible in the familiar home environment. The burden of illness frequently falls on relatives. Dementia clinical nurse specialists form a central part of the multidisciplinary team.

General measures. Some evidence suggests that participation in cognitively demanding activities in later life may protect against or delay the onset of dementia. Managing vascular risk factors may slow the progression of cognitive decline. There may be modest benefit from vitamin E in slowing progression of mild to moderate AD.

Cognitive enhancing drugs. These have a modest symptomatic benefit in AD, equivalent to an increase of 1 to 2 points on the MMSE. They are not disease-modifying, so do not slow or prevent progression. While there has been some dispute about the place for these drugs, they contribute to patients being able to prolong independence and remain at home for longer than might otherwise be the case.

- ***Acetylcholinesterase inhibitors*** (donepezil, rivastigmine, and galantamine) increase brain acetylcholine levels by inhibiting CNS acetylcholinesterase. Patients with AD have a central cholinergic deficit. There is usually a small but significant improvement in memory, cognition, and function. Cholinesterase inhibitors are also effective in DLB and Parkinson's dementia but not in FTD or vascular dementia. Side-effects, particularly cholinergic gastrointestinal symptoms, may be a problem. The presence of cardiac conduction deficits is a contraindication.
- ***Memantine*** is an NMDA receptor antagonist. It is used in moderate or severe AD or where cholinesterase inhibitors are not tolerated. There is some evidence that a combination of memantine and cholinesterase inhibitors is better than either used alone. It is generally well tolerated.

Disease-modifying treatments. It is possible that we are on the cusp of a revolution in dementia treatment, with the development of disease-modifying treatments that have shown evidence of target engagement. The most promising treatments so far are monoclonal antibodies directed against amyloid (Aβ) peptide. One of the barriers is the need to administer these therapies very early in the course of the disease for maximal efficacy, before significant irreversible neuronal damage has accumulated—in many cases, this will be before dementia can be clinically appreciated. Adverse effects caused by these agents include amyloid-related imaging abnormalities (ARIA), which are frequently asymptomatic but may occasionally cause life-threatening brain swelling.

- ***Aducanumab*** is a recombinant monoclonal antibody directed against amyloid beta, which has been shown to substantially

reduce amyloid plaques in the brain, but evidence of clinical efficacy has been inconsistent. It was the first agent to gain accelerated approval from the FDA in the USA, but has now been withdrawn by the manufacturer.

- ***Lecanemab*** is another mAb that binds to and clears Aβ. A phase 3 clinical trial in patients with early AD showed that lecanemab slowed cognitive decline by 27%. It has been approved in the USA.
- Other potential avenues being explored include agents that reduce the downstream effects of pathological Aβ signalling, inhibitors of secretase enzymes that process APP into Aβ fragments, and tau aggregation blockers.

Psychiatric and behavioural problems. Depression is common in dementia and may be difficult to distinguish from dementia symptoms, such as apathy and worsening cognitive function. A trial of an antidepressant is appropriate where depression is suspected. Distinguishing depressive pseudodementia from organic dementia can be difficult but is crucial. Behavioural disturbance (e.g., due to agitation or delusions) and hallucinations may occur in late-stage disease. Use of antipsychotic medications is associated with significantly increased stroke risk in patients with dementia and should be used only as a last resort.

Financial and legal issues. Patients may want to set up a Lasting Power of Attorney (if they retain mental capacity to do so in legal terms) to allow a spouse or relative to deal with their financial affairs on their behalf when they lose the capacity to do so. Patients and carers may be entitled to state financial benefits.

Driving. After a diagnosis of dementia, patients in the UK have a duty to inform the Driver and Vehicle Licensing Authority (DVLA), who may request a driving safety assessment.

Further reading

Bang J, Spina S, Miller BL. Frontotemporal dementia. *Lancet* 2015; 386:1672–1682.
Howard R, McShane R, Lindesay J et al. Donepezil and memantine for moderate-to-severe Alzheimer's disease. *N Engl J Med* 2012; 366:893–903.
Knopman DS, Amieva H, Petersen RC et al. Alzheimer disease. *Nat Rev Dis Primers* 2021 7:33.
Mitchell SL. Advanced dementia. *N Engl J Med* 2015; 372:2533–2540.
National Institute for Health and Care Excellence. *Dementia: Assessment, Management and Support for People Living with Dementia and Their Carers*. NG97; NICE 2018; http://www.nice.org.uk.
Walker Z, Possin KL, Boeve BF et al. Lewy body dementias. *Lancet* 2015; 386:1683–1697.

MOTOR NEURON DISEASE

Motor neuron disease (MND) is a devastating condition causing progressive weakness and eventually death, usually as a result of respiratory failure or aspiration. It is relatively uncommon, with an annual incidence of 2/100000 (but a lifetime risk of 1/300). Presentation is usually between the ages of 50 and 75. Below age 70, men are affected more often than women. Amyotrophic lateral sclerosis (ALS) is the term more commonly used for MND in some countries.

Pathogenesis

MND predominantly affects lower and upper motor neurons in the ventral horn of the spinal cord, cranial nerve motor nuclei, and frontal cortex. However, other neuronal systems may also be affected; 15% of patients also develop FTD (see 'Frontotemporal dementia' section) and up to 50% have some measurable frontal lobe cognitive impairment. MND is usually sporadic and of unknown cause, with good epidemiological evidence for a multistep process underlying the pathogenesis (the nature of the individual steps remains unknown, although genetic factors may account for a few steps). Age, family history, and cigarette smoking are the main known risk factors.

Ubiquitinated cytoplasmic inclusions containing the RNA processing proteins TDP-43 and FUS are the pathological hallmarks found in axons, indicating that protein aggregation may be involved in pathogenesis, as with other neurodegenerative disorders. Altered RNA processing, oxidative neuronal damage, and glutamate-mediated excitotoxicity have also been implicated in pathogenesis.

Between 5% and 10% of cases of MND are familial. A hexanucleotide GGGGCC repeat expansion in the *C9ORF72* gene on chromosome 9 is the most common cause of familial MND in European populations, and causes MND, FTD, and MND-FTD overlap syndromes. Mutations in the free radical scavenging enzyme superoxide dismutase *(SOD-1)* account for 15% of familial MND cases in European populations but 30% in selected Asian populations; *SOD-1* MND is pathologically distinct and not associated with FTD. Mutations in a number of other genes, including some that cause both FTD and MND such as *TDP-43* and *FUS*, have been identified.

Clinical features

Three main clinical subtypes are seen. The sensory system is not involved but a proportion of patients do report some sensory symptoms (but should not have major sensory signs).

- ***Amyotrophic lateral sclerosis (ALS).*** This is the classical presentation with simultaneous involvement of upper and lower motor neurons, usually in one limb, spreading gradually to other muscles in adjacent regions of the body. The typical picture is one of progressive focal muscle weakness and wasting (e.g., in one hand), with muscle fasciculations due to spontaneous firing of abnormally large motor units formed by surviving axons branching to innervate muscle fibres that have lost their nerve supply. The 'split hand' sign (disproportionate wasting of the thenar eminence and first dorsal interosseous, with sparing of the hypothenar eminence) may be present. Approximately 25% of patients have bulbar-onset disease, with a worse prognosis than limb-onset ALS, in which case dysarthria, dysphagia, nasal regurgitation of fluids, and choking are the presenting symptoms. Examination often reveals UMN signs, such as brisk reflexes (a brisk reflex in a wasted muscle is a classic sign), extensor plantar responses, and spasticity. A fasciculating tongue with slow, stiff tongue movements and brisk jaw jerk is the classic finding in bulbar-onset disease. Sometimes, an asymmetric spastic paraparesis is the presenting feature, with LMN features developing months later. Relentless progression of signs and symptoms over months allows confirmation of a diagnosis that may initially be suspected.
- ***Progressive muscular atrophy (PMA).*** This is a pure LMN presentation with weakness, muscle wasting, and fasciculations, usually starting in one limb and gradually spreading to involve other adjacent spinal segments. Survival is longer than in ALS but most patients diagnosed with PMA in life have evidence of UMN involvement at postmortem, implying that these are different faces of the same disease.
- ***Primary lateral sclerosis (PLS)*** (rare, 1%–2%). This is the least common form of MND and is confined to UMNs, causing a slowly progressive tetraparesis and pseudobulbar palsy. Recent criteria for PLS require isolated UMN signs for longer than 4 years. Survival is longer than in ALS.

Many authors consider that PMA and PLS are all part of the clinical spectrum of ALS. Overt evidence of mixed UMN and LMN

Box 28.46 Gold Coast criteria for diagnosing ALS (2019)

Progressive motor impairment documented by history or repeated clinical assessment, preceded by normal function.
AND
Upper and lower motor neuron dysfunction in ≥1 body region, or lower motor neuron dysfunction in ≥2 body regions. If only one body region is affected, UMN and LMN dysfunction must both be present.
AND
Investigations have excluded other disease processes.

involvement will often emerge, meaning that the majority of patients who do not have both lower and upper motor neuron signs initially are best classified as having ***upper motor neuron predominant*** or ***lower motor neuron predominant ALS***.

Diagnosis

Diagnosis is largely clinical. Gold Coast Criteria for the diagnosis of ALS focus on the progression of motor impairment, having excluded other causes (**Box 28.46**). There are no diagnostic tests but investigations (such as MRI and NCS) allow exclusion of other disorders. Denervation of muscles due to degeneration of LMNs is confirmed by EMG, and may confirm subclinical involvement of muscle groups, such as paraspinal muscles, allowing an earlier diagnosis in some patients.

Cervical spondylosis causing radiculopathy with myelopathy (resulting in upper and lower motor neuron signs) can cause diagnostic difficulty. Motor neuropathies, such as multifocal motor neuropathy, can also appear like MND (see 'Multifocal motor neuropathy with conduction block' section).

Prognosis and management

Death due to MND usually occurs within 3 to 5 years, although there are rare cases who survive for a decade or longer.

No treatment has been shown to influence outcome substantially. Patients should be supported by a specialist multidisciplinary team with access to palliative care and a clinical nurse specialist. Non-invasive ventilatory support and feeding via a gastrostomy help prolong survival and improve quality of life. Medication options are limited:

Riluzole, a sodium-channel blocker that inhibits glutamate release, modestly slows progression and increases life expectancy. Real-world studies suggest it may increase median survival by 6 to 19 months.

Edaravone has been approved in only a few countries, including the USA. It is a free radical scavenger that has shown some beneficial effects in a subgroup of patients with early disease.

Sodium phenylbutyrate-taurursodiol (PB-TURSO) has been shown to modestly slow clinical deterioration in ALS. Availability is very limited and the cost is high.

Tofersen is an antisense oligonucleotide that silences *SOD-1* and has been shown to slow functional decline by a small amount in patients with ALS caused by *SOD-1* variants. It is given intrathecally and is only available in a few countries.

Further reading

Brown RH, Al-Chalabi A. Amyotrophic lateral sclerosis. *N Engl J Med* 2017; 377:162–172.
Renton AE, Majounie E, Waite A et al. A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron* 2011; 72:257–268.

CONGENITAL DISORDERS

Cerebral palsy (CP) is an umbrella term encompassing disparate disorders that are caused by non-progressive abnormalities in fetal or infant brain development, and are characterized by permanent motor deficits observed in early life. It is the most common form of physical disability in childhood and most affected children survive into adulthood. A variety of intrauterine and neonatal cerebral insults may cause CP, including prematurity and its complications, hypoxia, intrauterine infections, and kernicterus. In many cases, no specific cause can be identified.

Clinical features

Failure to achieve normal milestones is usually the earliest feature. Specific motor syndromes become apparent later in childhood or, rarely, in adult life.

- ***Spastic*** subtypes, including spastic diplegia (lower limb spasticity, with scissoring of gait), hemiplegia (one hemisphere is hypotrophic and the contralateral, hemiparetic limbs small), or quadriplegia (all limbs are affected).
- ***Dyskinetic*** presentations include choreoathetotic CP (with chorea—rapid dancing movements, and athetosis—slow writhing distal movements) and dystonic CP.
- ***Ataxic CP*** can occasionally be linked to a genetic cause or congenital hypoplasia of the cerebellum.
- ***Co-morbidities*** such as epilepsy and learning difficulty are common and at least as disabling as the motor deficit.

Dysraphism

Failure of normal fusion of the fetal neural tube leads to a group of congenital anomalies. Folate deficiency during pregnancy is contributory and supplements should always be given (see Chapter 25). Antiepileptic drugs, such as valproate, are also implicated (see 'Management' section in the 'Epilepsy' section). Neural tube defects may be open or closed (covered by skin); if there is a communication with the skin, bacterial meningitis may follow.

- ***Anencephaly*** is a large defect that leaves the cranial neural tube exposed and is not compatible with survival.
- ***Encephalocele*** involves extrusion of brain and/or meninges through a midline skull defect; protrusion can be minor or massive.
- ***Spina bifida*** is failure of lumbosacral neural tube fusion.
  - ***Spina bifida occulta*** or closed spinal dysraphism is isolated failure of vertebral arch fusion (usually lumbar), often seen incidentally on X-rays (3% of the population). A dimple or a tuft of hair may overlie the anomaly; clinical abnormalities are unusual.
  - ***Myelomeningocele*** consists of elements of spinal cord and lumbosacral roots within a meningeal sac. This herniates through a vertebral defect. In severe cases, both lower limbs and sphincters are paralyzed. Many babies will have a Chiari II malformation (herniation of the cerebellar tonsils and brainstem in association with the myelomeningocele) and associated hydrocephalus. ***Meningocele*** is a meningeal defect alone. The defect should be closed in the first 72 hours after birth; fetal surgery may be considered.

Further reading

Colver A, Fairhurst C, Pharoah POD. Cerebral palsy. *Lancet* 2014; 383:1240–1249.

NEUROGENETIC DISORDERS

Approximately 80% of the 20 000 to 30 000 human genes are expressed in the brain. It is therefore not surprising that of the 5000 or so Mendelian disorders of humans, a high proportion are neurological disorders. Several hundred neurological disease genes have been identified and molecular genetic testing is now part of the neurological diagnostic process. Many of these disorders are covered in individual disease sections of this chapter.

Neurocutaneous syndromes

Neurofibromatosis type 1 (von Recklinghausen's disease)

One of the most common neurogenetic disorders, neurofibromatosis type 1 (NF-1) has a prevalence of 1 in 3000. Inheritance is autosomal dominant, but 50% of cases are due to new mutations with no family history. The protein is called neurofibromin 1. NF-1 is characterized by multiple skin neurofibromas and pigmentation (*café-au-lait* patches—see Chapter 45, 'Café-au-lait macules' section, axillary freckling and Lisch nodules of the iris). The neurofibromas arise from the neurilemmal sheath.

Skin neurofibromas present as soft subcutaneous, sometimes pedunculated, lumps (see Chapter 45, 'Neurofibromatosis type 1' section). They increase in number throughout life. Plexiform neurofibromas (**Fig. 28.41**) may develop on major nerves and proximal nerve roots, sometimes involving the spinal cord. Treatment is surgical removal if pressure symptoms develop. Associated features include learning difficulties, malignant transformation of neurofibromas, and bone abnormalities including scoliosis and fibrous dysplasia.

Neurofibromatosis type 2

Neurofibromatosis type 2 (NF-2) is much less common than NF-1. It is also autosomal dominant; the gene product, merlin or schwannomin, is a cytoskeletal protein. Many neural tumours occur:

- vestibular schwannomas (usually bilateral) in 90%
- meningiomas
- gliomas (including optic nerve glioma)
- cutaneous neurofibromas (30%).

Tuberous sclerosis (epiloia)

Features of this rare multisystem, autosomal dominant condition include adenoma sebaceum, renal tumours, and glial overgrowth in the brain (cortical tubers and subependymal nodules). Epilepsy (70%) and learning difficulties (50%) are common complications.

Von Hippel–Lindau disease

This rare condition is dominantly inherited. Cerebellar, spinal, and retinal haemangioblastomas develop and can be surgically removed. Tumours—renal cell carcinoma and phaeochromocytomas—may also occur. Polycythaemia sometimes develops.

Hereditary ataxias

A wide variety of genetic disorders cause cerebellar ataxia as the sole or predominant clinical feature. In children, consideration needs to be given to potentially reversible inborn errors of metabolism, typically autosomal recessive; the rest are progressive heredodegenerative disorders that can be autosomal recessive, X-linked, mitochondrial, or autosomal dominant. Several of the common disorders are due to trinucleotide repeat insertions (see Chapter 5, 'Triplet repeat mutations and genetic anticipation' section).

Fig. 28.41 **Plexiform neurofibroma.**

Increasingly, investigation of hereditary ataxias is undertaken through whole genome sequencing.

Early-onset ataxia

Most early (<20 years of age) onset inherited ataxias are autosomal recessive. Friedreich's ataxia is by far the most common, caused by a GAA trinucleotide repeat expansion in the frataxin gene (involved in mitochondrial iron metabolism). Onset is in the early teens with progressive difficulty in walking due to cerebellar ataxia and sensory neuropathy. Associated features include scoliosis, cardiomyopathy, optic atrophy, areflexia, and diabetes.

Ataxia telangiectasia is caused by homozygous variants in a DNA repair gene (ataxia-telangiectasia mutated gene, *ATM*), presents earlier than Friedreich's and affected individuals have telangiectasias, immunodeficiency, and an increased risk of malignancy.

Metabolic ataxias should be ruled out in childhood-onset presentations, as they are potentially treatable. The presentation may be intermittent (e.g., urea cycle disorders). Ataxia with vitamin E deficiency is a rare recessive ataxia that can mimic Friedreich's and may respond to high-dose vitamin E supplementation. Wilson's disease can also cause ataxia and may present in childhood or later in life.

Late-onset ataxia

Adult-onset (>20 years of age) inherited ataxias are usually dominantly inherited SCAs and there are over 40 different genetic forms, several of which are caused by CAG trinucleotide repeats. Some are 'pure' cerebellar ataxias and others are associated with additional features, such as peripheral neuropathy, pyramidal and extrapyramidal signs, movement disorders, retinal disorders, and cognitive impairment. Cerebellar atrophy is a common but non-specific imaging finding.

- ***SCA3*** (Machado-Joseph disease), with a CAG repeat in the ataxin 3 *(ATXN3)* gene is the most common autosomal dominant SCA. Patients have progressive ataxia, with oculomotor

abnormalities, and may have other features including spasticity, neuropathy, extrapyramidal features, and cognitive impairment. In common with other polyglutamine repeat disorders, the disease demonstrates anticipation.

- ***SCA2*** is caused by a CAG repeat expansion in the ataxin 2 *(ATXN2)* gene. Typical age of onset is in the fourth decade of life with a large range. Some patients develop features of motor neuron degeneration.
- ***SCA6***, with a CAG repeat expansion in the voltage-gated calcium channel subunit alpha 1A *(CACNA1A)* gene, is a 'pure' ataxia. Onset can be between the ages of 20 and 65, and the disease accounts for a significant proportion of apparently sporadic late-adult-onset progressive cerebellar ataxia.
- The recessively inherited cerebellar ataxia, neuropathy, and vestibular areflexia syndrome ***(CANVAS)*** has recently been shown to be caused by homozygous intronic repeat expansions in the replication factor C subunit 1 *(RFC1)* gene. Patients often have a chronic cough with varying combinations of cerebellar ataxia, sensory ataxia, and vestibulopathy. It is likely under-recognized.

PARANEOPLASTIC SYNDROMES

Neurological disease may accompany malignancy in the absence of metastases. These paraneoplastic syndromes are associated in many cases with anti-neuronal antibodies, believed to be involved in the generation of signs and symptoms. Numerous anti-neuronal antibodies have been described.

Clinical presentations with a high risk of underlying tumour include:

- subacute cerebellar degeneration
- opsoclonus-myoclonus syndrome
- limbic encephalitis (see 'Autoimmune encephalitis' section)
- encephalomyelitis
- subacute sensory neuronopathy (affecting the DRG)
- Lambert–Eaton myasthenic syndrome (LEMS; see 'Lambert–Eaton myasthenic syndrome' section) and MG with thymoma
- chronic gastrointestinal pseudo-obstruction (enteric neuropathy)
- dermatomyositis.

In addition to these 'classical' paraneoplastic neurological syndromes, polyradiculoneuropathy, stiff person syndrome (SPS), myelopathy, neuromyotonia, and anti-NMDA receptor encephalitis can all be paraneoplastic.

The neurological syndrome usually precedes evidence of the neoplasm: often a small-cell lung cancer, thymoma, or breast or ovarian cancer. Diagnosis is based on the clinical pattern and antibody profile. A classical syndrome or positive onconeuronal antibody should prompt a search for cancer; the neoplasm may be occult as it is held at bay by the immune response that also causes the neurological manifestations. Sometimes, when a cancer is not found initially, serial testing is required. Treatment is directed at both the underlying cancer and the immune system; except in selected syndromes treatment is often unsatisfactory.

PERIPHERAL NERVE DISEASE

Mechanisms of damage to peripheral nerves

Peripheral nerves consist of two principal cellular structures: the nerve nucleus with its axon, and the myelin sheath, which is produced by Schwann cells between each node of Ranvier (see Fig. 28.1). Blood supply is via vasa nervorum. Several mechanisms, some coexisting, cause nerve damage.

Demyelination

Schwann cell damage leads to myelin sheath disruption. This causes a marked slowing of conduction, seen, for example, in Guillain–Barré syndrome (GBS) and many genetic neuropathies.

Axonal degeneration

Axon damage causes the nerve fibre to die back from the periphery. Conduction velocity initially remains normal (compare demyelination) because axonal continuity is maintained in surviving fibres. A wide range of toxic and metabolic disorders damage peripheral nerves, as their long axons (requiring cellular transport of proteins from cell body to nerve terminals) make them uniquely vulnerable. This explains the concept of length-dependent neuropathy with the longest, most vulnerable axons (to the toes) being affected first.

Compression

Focal demyelination at the point of compression causes disruption of conduction. This typically occurs in entrapment neuropathies, such as CTS.

Infarction

Microinfarction of vasa nervorum occurs in diabetes (along with other mechanisms of peripheral nerve damage) and arteritis, such as polyarteritis nodosa and eosinophilic granulomatosis with polyangiitis (see Chapter 32, 'Small-vessel vasculitis' section). Wallerian degeneration occurs distal to the infarct.

Infiltration

Infiltration of peripheral nerves occurs by inflammatory cells in leprosy, by granulomas, such as in sarcoid, and by neoplastic cells.

Nerve regeneration

Regeneration occurs either by remyelination—Schwann cells produce new myelin sheaths around an axon—or by axonal growth down the nerve sheath with sprouting from the axonal stump. Axonal growth takes place at up to 1 mm/day.

Types of peripheral nerve disease

See Fig. 28.42.

- ***Neuropathy*** simply means a pathological process affecting a peripheral nerve or nerves.
- ***Mononeuropathy*** means a process affecting a single nerve.
- ***Mononeuritis multiplex*** means that several individual nerves are affected.

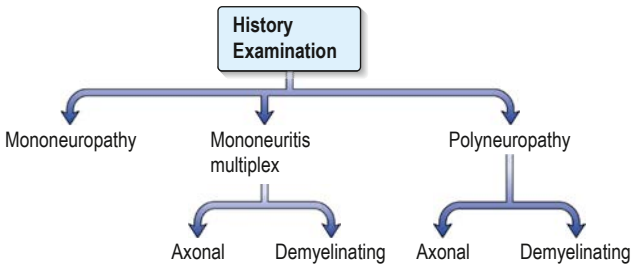


Fig. 28.42 Peripheral neuropathies. The type of neuropathy (axonal or demyelinating) can be assessed by electrical nerve studies (see 'Electromyography and nerve conduction studies' section).

- ***Polyneuropathy*** describes diffuse, typically symmetrical disease, usually commencing peripherally. The course may be acute, chronic, static, progressive, relapsing, or towards recovery. Polyneuropathies are motor, sensory, sensorimotor, and autonomic. They are classified broadly into demyelinating and axonal types, depending on which principal pathological process predominates. It is often impossible to separate these clinically. Many systemic diseases cause neuropathies. Widespread loss of tendon reflexes is typical, with distal weakness and distal sensory loss. ***Length-dependent*** distal dying-back axonopathies start at the toes and progress upwards; the fingertips are affected when the process reaches the knees.
- ***Radiculopathy*** means disease affecting nerve roots; ***plexopathy*** affects the brachial or lumbosacral plexus.
- ***Small fibre involvement*** may be seen with many neuropathies, particularly diabetic neuropathy, or may form an isolated small fibre neuropathy with prominent painful dysaesthesiae and burning. NCS only record from large fibres so will not detect a pure small fibre neuropathy.
- ***Neuronopathy*** affects the cell body (as in MND, or sensory ganglionopathy) rather than the peripheral axon or myelin sheath. The pattern is not length dependent, there may be global involvement (including the face) in sensory neuronopathy, and the distribution can be quite variable.

Diagnosis is made by clinical pattern, NCS/EMG, identification of systemic or genetic disease, and occasionally nerve biopsy (usually sural or radial). Diagnosis of small fibre neuropathy is often clinical but when confirmation is needed skin biopsy for intraepidermal nerve fibre density is useful. Identifying treatable causes and types of neuropathy is the main focus of work-up. Where there is an obvious cause with a chronic length-dependent neuropathy, extensive work-up is unnecessary, but if there is acute onset, rapid progression, or unclear cause, then more tests will be needed.

Clinical approach to neuropathy

The clinical approach to a patient with suspected neuropathy is summarized in **Box 28.47**.

i Box 28.47 Clinical approach to peripheral neuropathy

Is it a neuropathy?
- LMN weakness, worse distally or in a single nerve distribution
- Reduced or absent reflexes
- Sensory loss in the distribution of a single nerve, or in stocking/glove distribution
- Large fibre sensory involvement causes sensory ataxia; small fibre neuropathy presents with painful dysaesthesiae
- Differentials for flaccid weakness may include an acute cord lesion, muscle disease (more often proximal weakness but plexopathy/radiculopathy can also be proximal and muscle disorders occasionally cause distal weakness), neuromuscular junction disorder

What is the time course?
- Acute: GBS, some toxins (heavy metals), metabolic (B_1 deficiency, porphyria)
- Subacute: B_{12} deficiency, paraneoplastic, vasculitis, diabetic amyotrophy, vasculitis (often sequential involvement of individual nerves)
- Chronic: everything else (CIDP is occasionally relapsing)

What is the pattern?
- Mononeuropathy (often compressive)
- Mononeuritis multiplex (think vasculitis if painful, cryoglobulinaemia, leprosy)
- Polyneuropathy:
  - Symmetrical length-dependent neuropathy—usually axonal dying back but could be demyelinating, especially genetic
  - Proximal and distal weakness, symmetrical but non-length-dependent—consider demyelinating inflammatory neuropathy (GBS, CIDP)
  - Asymmetrical/patchy polyneuropathy—could be a confluent mononeuritis multiplex, or think about plexopathy/neuronopathy
- Plexopathy (diabetic amyotrophy, brachial neuritis, infiltrative, radiation)
- Radiculopathy (many potential causes but often compressive)
- Neuronopathy (sensory neuronopathy—think paraneoplastic, drugs, autoimmune including Sjögren's)
- Presence of cranial nerve involvement—rare in most neuropathies. Think GBS, occasionally CIDP, neoplastic/granulomatous causes, and MND. Reconsider muscle/neuromuscular junction process

Motor/sensory/autonomic involvement
- For pure motor syndromes, look very carefully for UMN signs to indicate ALS.
- Acute pure motor neuropathy can be GBS (acute motor axonal neuropathy variant), infection (botulism, diphtheria), porphyria (motor-predominant).
- More subacute or chronic motor neuropathies include lead toxicity, multifocal motor neuropathy with conduction block (MMN), distal hereditary motor neuropathy.
- Sensory involvement: is there prominent sensory ataxia (subacute combined degeneration of the cord, sensory neuronopathy)? Slowly progressive distal symmetrical sensory loss with sensory ataxia and tremor in an older person may indicate DADS (distal acquired demyelinating sensory neuropathy, often associated with an IgM paraprotein and antibodies to myelin-associated glycoprotein).
- Burning pain: implies a small fibre neuropathy (diabetes, Sjögren's, idiopathic, HSAN).
- The commonest cause of chronic late-onset pure sensory neuropathy/neuronopathy is probably *RFC1* gene mutation (CANVAS, see 'Late-onset ataxia' section).
- Marked autonomic involvement in a chronic neuropathy may suggest amyloidosis or diabetes. Acutely, autonomic involvement is seen in GBS.
- Most neuropathies are mixed sensorimotor but motor predominance (e.g., AIDP) vs sensory predominance (e.g., chemotherapy-induced polyneuropathy) should be noted.

Is it axonal or demyelinating?
- Demyelinating—immune mediated (GBS, CIDP, MMN, IgM paraproteinaemic neuropathy) or genetic (CMT1)
- Axonal—most other causes of neuropathy
- Mixed axonal and demyelinating features often seen in diabetic polyneuropathy

Associated features
- Pain—pain of nerve ischaemia in vasculitic neuropathy; burning/tingling pain with small fibre involvement
- Features of underlying systemic disease e.g., rash
- Nerve thickening (leprosy, genetic neuropathies)
- Joint/foot deformities (Charcot joints suggest chronicity; pes cavus implies neuropathy started during foot growth)
- Chronic cough—often predates the symptomatic neuropathy by decades in CANVAS
- UMN signs: for example, upgoing plantars and absent knee jerks. Seen in: motor neuron disease; subacute combined degeneration of the cord, Friedreich's ataxia, adrenomyeloneuropathy, HIV (or dual pathology)

ALS, Amyotrophic lateral sclerosis; *AIDP,* acute inflammatory demyelinating polyradiculoneuropathy; *CANVAS,* cerebellar ataxia with neuropathy and vestibular areflexia syndrome; *CIDP,* chronic inflammatory demyelinating polyradiculoneuropathy; *CMT,* Charcot–Marie–Tooth disease; *GBS,* Guillain-Barré syndrome; *HIV,* human immunodeficiency virus; *HSAN,* hereditary sensory and autonomic neuropathy; *MMN,* multifocal motor neuropathy with conduction block; *UMN,* upper motor neuron.

Box 28.48 Nerve compression and entrapment

Nerve	Entrapment/compression site
Median	Carpal tunnel (wrist)
Ulnar	Cubital tunnel (elbow)
Radial	Spiral groove (of humerus)
Posterior interosseous	Supinator muscle (forearm)
Lateral cutaneous of thigh	Inguinal ligament
Common peroneal	Neck of fibula
Posterior tibial	Tarsal tunnel (flexor retinaculum—foot)

Mononeuropathies

Peripheral nerve compression and entrapment

Nerves are vulnerable to mechanical compression at a few key sites (Box 28.48), such as the common peroneal nerve at the head of the fibula, or the ulnar nerve at the elbow. Entrapment develops in relatively tight anatomical passages, such as the carpal tunnel. Focal demyelination predominates at the compression site, and some distal axonal degeneration occurs.

These neuropathies are recognized largely by their clinical features. Diagnosis is confirmed by NCS.

The most common are mentioned here. All are seen more frequently in people with diabetes.

Carpal tunnel syndrome

This common mononeuropathy, median nerve entrapment at the wrist, is usually known as carpal tunnel syndrome (CTS). CTS is typically not associated with any underlying disease. It is, however, seen in:

- hypothyroidism
- pregnancy (third trimester)
- rheumatoid disease
- acromegaly
- amyloidosis.

Symptoms, signs, and management are discussed in Chapter 32 ('Carpal tunnel syndrome' section).

Ulnar nerve compression

The nerve is compressed in the cubital tunnel at the elbow. This follows prolonged or recurrent pressure and elbow fracture ('tardy ulnar palsy', as onset is very delayed).

There is clawing of the hand, wasting of interossei and hypothenar muscles, and weakness of interossei and medial two lumbricals with sensory loss in the little finger and splitting of the ring finger. Decompression and transposition of the nerve at the elbow are sometimes helpful but often disappointing.

Radial nerve compression

The radial nerve is compressed acutely against the humerus: for example, when the arm is draped over a hard chair for several hours, known as Saturday night palsy. Wrist drop and weakness of brachioradialis and finger extension follow. Recovery is usual, though not invariable, within 1 to 3 months.

Lateral cutaneous nerve of the thigh compression

This is also known as meralgia paraesthetica and is described in Chapter 32 ('Meralgia paraesthetica' section).

Common peroneal nerve palsy

The common peroneal nerve is compressed against the head of the fibula following prolonged squatting, yoga, pressure from a cast, prolonged bed rest or coma, or for no apparent reason. There is foot drop and weakness of ankle eversion. The ankle jerk (S1) is preserved. A patch of numbness develops on the anterolateral border of the dorsum of the foot and/or lateral calf. Confusion with an L5 motor radiculopathy may occur. Recovery is usual, though not invariable, within several months.

Hereditary neuropathy with pressure palsies (HNPP)

The genetic converse of Charcot–Marie–Tooth disease 1A (see 'Charcot–Marie–Tooth disease' section), this dominantly inherited disorder is due to deletion (as opposed to duplication) of a region of the *PMP-22* gene. Patients are susceptible to pressure palsies after minor compression episodes; even the brachial plexus may be involved. There is also a mild background neuropathy that develops gradually. Genetic testing can be performed.

Mononeuritis multiplex

Owing to the potential for serious and/or treatable causes, this pattern is a red flag, especially when neuropathies occur away from compression sites. Causes include:

- vasculitis, including eosinophilic granulomatosis with polyangiitis
- amyloidosis
- leprosy
- malignancy
- neurofibromatosis
- HIV and hepatitis C infection (cryoglobulinaemia)
- multifocal motor neuropathy with conduction block (MMN)/multifocal acquired demyelinating sensory and motor neuropathy (MADSAM)
- HNPP (see above)
- multiple compression neuropathies.

Several nerves, such as the ulnar, median, radial, and lateral popliteal, become affected sequentially or simultaneously. When multifocal neuropathy becomes confluent, there is difficulty distinguishing it from polyneuropathy. A carefully taken history will often reveal a mononeuritic onset, or asymmetrical polyneuropathy should raise a suspicion.

Polyneuropathies (peripheral neuropathy)

Many diseases cause polyneuropathy, by far the most common being diabetes. The diagnosis should not stop with identification of the polyneuropathy, but should involve a full diagnostic work-up to identify the underlying cause (Box 28.49). However, despite thorough investigation, the aetiology remains unknown in up to 50% of cases. Selected types of neuropathy are discussed below.

Clinical features. Duration, time course, distribution, and pattern of the different types of polyneuropathy vary considerably (Box 28.47).

Neurophysiological features. NCS allow separation into axonal and demyelinating forms. Demyelinating neuropathies are much more likely to be immune-mediated in aetiology and responsive to immunomodulation. With axonal neuropathies, the focus should be on finding a reversible metabolic or toxic cause.

Diagnostic investigations (in addition to neurophysiology). A stepped approach can be taken (Box 28.50).

Box 28.49 Causes of polyneuropathy

Demyelinating

Immune mediated
- Guillain–Barré syndrome
  - Acute inflammatory demyelinating polyradiculoneuropathy (AIDP)
  - Miller Fisher syndrome
- Chronic inflammatory demyelinating polyradiculoneuropathy (CIDP)
- Multifocal motor neuropathy with conduction block

Paraproteinaemic
- IgM paraproteinaemic neuropathies
  - Anti-MAG neuropathy
- POEMS (polyneuropathy, organomegaly, endocrinopathy, monoclonal gammopathy, and skin changes) syndrome—can be mixed axonal/demyelinating

Hereditary
- Charcot–Marie–Tooth Disease (hereditary motor sensory neuropathy, HMSN) types 1 (AD), 4 (AR), and CMTX1
- Hereditary neuropathy with liability to pressure palsies (HNPP)
- Refsum disease

Toxic
- Selected drugs including amiodarone, tacrolimus, interferon-alpha, TNF-alpha blockers
- Immune checkpoint inhibitors

Axonal

Metabolic
- Diabetes mellitus (often have mixed axonal and demyelinating features)
- Uraemia
- Hepatic disease
- Thyroid disease
- Porphyria
- Vitamin/mineral deficiency:
  - B_1 (thiamine)
  - B_6 (pyridoxine; pyridoxine excess can also cause a sensory ganglionopathy)
  - B_{12}
  - Copper

Toxic
- Drugs (see **Box 28.52**)
- Alcohol
- Industrial toxins, e.g., lead, acrylamide

Infiltrative
- Amyloidosis (familial, senile, and AL)
- Malignancy, especially lymphoma

Hereditary
- Charcot–Marie–Tooth Type 2 (AD and AR)
- Hereditary motor neuropathy (HMN)
- Hereditary sensory and autonomic neuropathy (HSAN)
- CANVAS (cerebellar ataxia with neuropathy and vestibular areflexia syndrome)
- Friedreich's ataxia, ataxia-telangiectasia

Infective
- HIV
- Lyme disease
- Hepatitis B/C (commonly vasculitic neuropathy causing a mononeuritis multiplex)

Inflammatory/autoimmune
- Primary or secondary systemic vasculitis
- Non-systemic vasculitic neuropathy
- Neuropathies associated with systemic inflammatory diseases, e.g., Sjögren's syndrome (may be associated with a sensory neuropathy or neuronopathy)
- Paraneoplastic (may cause neuronopathy)
- Axonal GBS variants
  - Acute motor axonal neuropathy (AMAN)
  - Acute motor and sensory axonal neuropathy (AMSAN)

Other
- Idiopathic sensorimotor neuropathy
- Critical illness neuropathy
- Neuropathies associated with IgA or IgG paraproteins

Box 28.50 Investigations in peripheral neuropathy

Initial investigations
- Blood tests:
  - Full blood count, erythrocyte sedimentation rate, vitamin B_{12}, folate
  - Renal, liver, thyroid function, bone profile
  - Glucose, HbA_{1c}
  - Protein electrophoresis/immunofixation, immunoglobulins
  - Antinuclear antibody
  - HIV, hepatitis B + C
- Chest X-ray
- Urine: Bence Jones proteins

Selected patients
- Nerve conduction studies: to differentiate axonal from demyelinating forms
- Blood tests:
  - Anti-ganglioside antibodies
  - Anti-myelin-associated glycoprotein (anti-MAG) antibodies
  - Lyme disease
  - Cryoglobulins, vasculitis screen
  - Anti-Ro and anti-La
  - Porphyrins
  - Genetic tests (e.g., Friedreich's ataxia, CMT)
  - Serum-free light chains
  - Anti-neuronal antibodies
- Search for malignancy
- CSF analysis for protein (with demyelinating neuropathy)
- Nerve biopsy
- Labial salivary gland biopsy for Sjögren's
- Slit-skin smear for leprosy
- Skin biopsy for intraepidermal nerve fibre density (to diagnose small fibre neuropathy)

CMT, Charcot–Marie–Tooth disease.

Immune-mediated neuropathies

Guillain–Barré Syndrome

Guillain–Barré syndrome (GBS) is the most common cause of acute flaccid paralysis, and the most common acute polyneuropathy (3/100 000 per year). GBS describes any type of acute immune-mediated polyneuropathy, usually arising after an infection when the immune system reacts against epitopes on peripheral nerves that are similar to those on infectious agents (molecular mimicry).

Clinical features

The canonical syndrome is ***acute inflammatory demyelinating polyradiculoneuropathy*** (AIDP), in which symmetrical proximal and distal weakness progress rapidly, often associated with back pain and varying degrees of sensory loss, reaching nadir typically over 2 weeks (but a maximum of 4 weeks). While the legs are typically affected before the arms, this is not always the case, and many patterns of progression can be seen (including descending from the face). Descending paralysis should raise the differential of botulism but does not exclude GBS. Absent or reduced reflexes are a vital clue. Cranial nerve involvement is common. Bulbar and respiratory failure are much-feared complications, with about one in five patients requiring mechanical ventilation. Autonomic involvement is very common and can be fatal through inducing arrhythmias and blood pressure instability.

Several GBS variants are seen:

AIDP is the most common type of GBS in Western countries. The typical clinical picture is described above. Demyelinating features (slowed nerve conduction velocity, conduction block, temporal dispersion) will be seen on NCS.

Acute motor axonal neuropathy (AMAN) is frequently preceded by *Campylobacter jejuni* infection, and carries a worse prognosis than AIDP. Antibodies to various gangliosides (e.g., GM1) may be found. The damage is principally axonal and CSF protein is more likely to be normal than in AIDP.

Miller–Fisher syndrome consists of the triad of ataxia, areflexia, and ophthalmoplegia, and 90% of patients will have GQ1b antibodies. Pure MFS does not require immunomodulatory treatment and recovery is excellent. One-quarter of patients will develop clinical weakness as in GBS.

Diagnosis

The differential diagnosis includes other acute paralytic illnesses, such as botulism, cord compression, muscle disease, and myasthenia.

Diagnostic work-up includes a lumbar puncture, done mainly to exclude other causes (e.g., infective polyradiculitis). The classical finding is albuminocytological dissociation (elevated protein, often >1 g/L, with a normal white cell count). CSF protein may be normal early in the course of the illness in about one-third of patients.

Neurophysiological studies have a higher yield if done after at least one week of onset, and are not essential for diagnosis, but will delineate the type of GBS (e.g., axonal or demyelinating) and may be prognostically helpful.

Anti-ganglioside antibodies can shed light on the aetiopathogenesis but take too long to be useful diagnostically in the acute setting; an HIV test is mandatory as GBS can arise as a seroconversion illness. MR imaging may be needed to rule out other processes (e.g., cauda equina syndrome or acute myelitis); nerve root enhancement is a sensitive but non-specific sign of GBS.

Management and outcome

Management priorities in GBS include serial monitoring of weakness, bulbar function, respiratory function, and timely escalation to intensive care. Breathlessness or low oxygen saturations are a very late sign of respiratory muscle weakness and should not be relied upon: bedside spirometry is crucial. A forced vital capacity of less than 20 mL/kg or that has fallen by more than 30% from baseline warrants escalation to ITU. Cardiac monitoring is essential in patients with suspected dysautonomia. Thromboprophylaxis and avoidance of pressure ulcers are important.

Immunomodulatory treatment is warranted for patients who cannot walk unaided for more than 10 m, or who have rapidly progressive weakness, severe bulbar or autonomic dysfunction, or respiratory insufficiency. Treatment options include intravenous immunoglobulin (IVIG, 0.4 g/kg daily for 5 days), or plasma exchange (five sessions). The disease is almost always monophasic. Functional recovery is aided by expert rehabilitation, and is good in most cases but the course can be prolonged (recovery can continue for up to 3 years). Ten percent of patients will have ongoing severe motor impairments. The mortality of GBS is 5% even with modern treatment.

Chronic inflammatory demyelinating polyradiculoneuropathy

Chronic inflammatory demyelinating polyradiculoneuropathy (CIDP) develops over months, causing progressive or relapsing proximal and distal limb weakness with sensory loss. Several variants, such as multifocal acquired demyelinating sensory and motor neuropathy (MADSAM), pure sensory and pure motor forms, occur. In some cases, cranial nerves may be involved.

There is no single diagnostic test but CSF protein is raised and patchy demyelination is usually seen on NCS. Some cases are associated with a serum paraprotein. Nerve biopsy is sometimes required. CIDP usually responds to long-term immunosuppression with steroids or to intravenous immunoglobulin.

Multifocal motor neuropathy with conduction block

A distal immune-mediated focal demyelinating motor neuropathy (often asymmetrical and predominantly in the hands) develops gradually over months with profuse fasciculation; hence there is confusion with MND (see 'Motor neuron disease' section). Conduction block and denervation are seen electrically. IgM antibodies to the ganglioside GM_1 are found in over 50% of cases.

Treatment is usually with regular intravenous immunoglobulin infusions that produce rapid but transient improvement. Steroids may cause the condition to worsen and should be avoided.

Paraproteinaemic neuropathies

Up to 70% of patients with a serum paraprotein have a neuropathy and some 10% of patients with no other identifiable cause for their neuropathy have a paraprotein. Most are associated with monoclonal gammopathy of unknown significance (MGUS; see Chapter 39, 'Monoclonal gammopathy of undetermined significance' section) but they are also seen in myeloma (see Chapter 39, 'Myeloma and other plasma cell disorders' section). The paraprotein may be pathogenic for the neuropathy (e.g., anti-MAG) or coincidental (particularly with non-IgM paraproteins associated with an axonal neuropathy).

IgM paraproteins: usually a demyelinating neuropathy. The paraproteins are often directed against myelin-associated glycoprotein (anti-MAG). The anti-MAG phenotype is a slowly progressive distal neuropathy with ataxia and prominent tremor (DADS—distal acquired demyelinating symmetric neuropathy). Anti-MAG neuropathy does not respond well to IVIG but may respond to rituximab, cyclophosphamide, or fludarabine. DADS can also be seen without a paraprotein in which case it is regarded as a CIDP variant.

POEMS syndrome (***P***olyneuropathy [demyelinating], ***O***rganomegaly [hepatomegaly 50%], ***E***ndocrinopathy [reduced testosterone usually], an ***M*** [para]protein band, and ***S***kin changes): probably caused by vascular endothelial growth factor (VEGF) release from a plasmacytoma. A mixed demyelinating and axonal neuropathy occurs. Treatment is of the plasmacytoma/plasma cell dyscrasia.

Primary AL amyloidosis causes an axonal painful neuropathy with autonomic dysfunction. Treatment is directed at the underlying plasma cell dyscrasia, without which prognosis is poor.

Cryoglobulinaemia causes a vasculitic neuropathy, presenting as painful mononeuritis multiplex or asymmetric polyneuropathy.

Metabolic, toxic, and vitamin deficiency neuropathies

Causes of the most common neuropathies are shown in **Box 28.49**.

Metabolic neuropathies

Diabetes mellitus

This is the most common cause of neuropathy in high income countries; 50% of patients with diabetes have neuropathy after 25 years. Good glycaemic control is protective against this microvascular complication of diabetes. Several varieties of neuropathy occur (see Chapter 35, 'Diabetic nephropathy' section):

- ***distal symmetrical sensorimotor neuropathy:*** slowly progressive, sensory loss predominates in a length-dependent stocking and glove distribution. Late cases are accompanied by weakness. The neuropathy may be asymptomatic and should be screened for regularly; it is related to diabetes duration and

glycaemic control. Mixed axonal and demyelinating features are frequently seen.

- ***mononeuropathy and multiple mononeuropathy (mononeuritis multiplex):***
  - cranial nerve lesions (e.g., medical third nerve palsy)
  - individual mononeuropathies (e.g., CTS) or mononeuritis multiplex
- ***diabetic amyotrophy*** or lumbosacral radiculoplexus neuropathy: a reversible vasculitic plexoneuropathy. There is pain followed by acute wasting and weakness of one leg, often associated with weight loss. Usually the weakness will plateau and then there will be some partial recovery; the contralateral leg or even the arms can become involved in some cases.
- ***autonomic neuropathy:*** postural hypotension, gastroparesis, and enteric neuropathy with constipation or diarrhoea can be intractable.

Diabetic neuropathy is typically not reversible with glycaemic control, but progression may be slowed with good management of the diabetes. Good footcare and control of neuropathic pain are the mainstays of treatment.

Uraemia

Progressive sensorimotor neuropathy develops in chronic uraemia. Response to dialysis is variable; the neuropathy may improve after transplantation.

Thyroid disease

A mild chronic sensorimotor neuropathy is sometimes seen in both hyperthyroidism and hypothyroidism. Myopathy also occurs in hyperthyroidism (see Chapter 34, 'Hyperthyroidism' section).

Porphyria

In acute intermittent porphyria (see Chapter 36, 'Porphyrias' section), there are episodes of a severe, mainly proximal neuropathy in the limbs, sometimes with abdominal pain, confusion, and coma. Alcohol, barbiturates, and intercurrent infection can precipitate attacks.

Amyloidosis

Polyneuropathy or multifocal neuropathy develops (see Chapter 39, 'Monoclonal gammopathies of clinical significance' section).

Toxic neuropathies

Alcohol

Polyneuropathy, mainly in the lower limbs, occurs with chronic alcohol use. It is a common cause of neuropathy. A myopathy may accompany it. For other neurological consequences of alcohol, see **Box 28.51**.

Drugs and industrial toxins

Many drugs (**Box 28.52**) and a wide variety of industrial toxins cause polyneuropathy. Toxins include:

- lead—motor neuropathy
- acrylamide (plastics industry), trichlorethylene, hexane, fat-soluble hydrocarbons, such as solvent abuse
- arsenic, thallium, and heavy metals.

Vitamin deficiency neuropathies

Vitamin deficiencies cause nervous system damage that is potentially reversible if treated early, and progressive if not. Deficiencies, often of multiple vitamins, develop in malnutrition.

Box 28.51 Neurological effects of ethyl alcohol

- Acute intoxication:
  - Disturbance of balance, gait, and speech
  - Coma
  - Head injury and sequelae
- Alcohol withdrawal:
  - Morning shakes
  - Tremor
  - Delirium tremens
- Thiamine deficiency:
  - Polyneuropathy
  - Wernicke–Korsakoff syndrome
- Epilepsy
- Acute intoxication
- Alcohol withdrawal
- Hypoglycaemia
- Cerebellar degeneration
- Cerebral infarction
- Cerebral atrophy, dementia
- Osmotic demyelination syndrome (ODS)
- Marchiafava–Bignami syndrome (corpus callosum degeneration, rare)

Box 28.52 Drug-related neuropathies

Drug	Neuropathy	Mode/site of action
Phenytoin	M	A
Chloramphenicol	S, M	A
Metronidazole	S, S/M	A
Isoniazid	S, S/M	A
Dapsone	M	A
Anti-retroviral drugs	S > M	A
Nitrofurantoin	S/M	A
Vincristine	S > M	A
Paclitaxel	S > M	A
Disulfiram	S, M	A
Cisplatin	S	A
Amiodarone	S, M	D, A
Chloroquine	S, M	A, D
Suramin	M > S	D, A
Immune checkpoint inhibitors	M > S	D

A, Axonal; *D*, demyelinating; *M*, motor; *S*, sensory.

Thiamine (vitamin B_1)

Dietary deficiency causes ***beriberi*** (see Chapter 25, 'Beriberi' section). Its principal features are polyneuropathy and cardiac failure. Thiamine deficiency also leads to Wernicke's encephalopathy and Korsakoff psychosis (see Chapter 30). Alcohol is the most common cause in Western countries and, rarely, anorexia nervosa or vomiting of pregnancy.

Pyridoxine (vitamin B_6)

Deficiency causes a mainly sensory neuropathy. In practical terms, this is seen as limb numbness developing during anti-tuberculosis therapy in slow isoniazid acetylators (see Chapter 20, 'Management' section in the "Tuberculosis' section). Prophylactic pyridoxine 10 mg daily is given with isoniazid.

Vitamin B_{12} (cobalamin)

Deficiency causes damage to the spinal cord, peripheral nerves, and brain.

Subacute combined degeneration of the cord (SACD). Combined cord and peripheral nerve damage is a sequela of pernicious anaemia and other causes of vitamin B_{12} deficiency (see Chapter 25, 'B vitamins' section). Nitrous oxide abuse causes functional B_{12} deficiency and can lead to SACD. Initially, there is numbness and tingling of fingers and toes; distal sensory loss, particularly of the posterior column; absent ankle jerks; and, with cord involvement, exaggerated knee jerks, and extensor plantars. Optic atrophy and retinal haemorrhage may occur. Rarely, neuropsychiatric changes are prominent. In later stages, sphincter disturbance, severe generalized weakness, and dementia develop.

Vitamin B_{12} levels should be checked; activated B_{12}, methylmalonic acid, and homocysteine levels may be particularly useful in the case of functional B_{12} deficiency, where the absolute levels may be normal. Parenteral B_{12} reverses nerve and cord damage although recovery may be incomplete; effect on the brain may be less impressive. Copper deficiency is a very rare cause of a similar picture.

Genetic neuropathies

Inherited neuropathy may occur as a 'pure' neuropathy disorder (e.g., Charcot–Marie–Tooth disease) or as part of a neurological multisystem disorder (e.g., spinocerebellar ataxia).

Charcot–Marie–Tooth disease

Charcot–Marie–Tooth (CMT) disease is a complex group of heterogeneous hereditary motor and sensory neuropathies (HMSNs) with multiple causative genes. Distal limb wasting and weakness typically progress slowly over many years, mostly in the legs, with development of pes cavus and variable loss of sensation and reflexes. In advanced disease, severe foot drop results but patients usually remain ambulant. Mild cases have pes cavus and toe clawing that can pass unnoticed.

- ***CMT1*** (HMSN1) is a group of autosomal dominant demyelinating neuropathies.
  - ***CMT 1A*** is the most common (70% of CMT; 1:2500 births), caused by duplication (or point mutation) of a 1.5 megabase portion of chromosome 17 encompassing the peripheral myelin protein 22 gene (*PMP-22*, 17p11.2). Deletions of the same portion of the gene cause HNPP, hereditary neuropathy with liability to pressure palsies (see 'Hereditary neuropathy with pressure palsies [HNPP]' section).
  - ***CMT 1B***—the second most common autosomal dominant demyelinating neuropathy due to mutations in the myelin protein zero gene *(MPZ)* on chromosome 1 (1q22).
- ***CMT 2*** *(HMSN 2)*—rare axonal polyneuropathies caused by a large number of different mutations, including *MFN2* or *KIFIB*; there is pes cavus with distal weakness and variable sensory loss. CMT2 can be AD or AR.
- ***CMTX1***—an X-linked dominant HMSN on chromosome Xq13.1; the gene product is a gap junction B1 protein (GJB1), also known as connexin 32. Nerve conduction velocity is in the intermediate range. Patients can also experience relapsing encephalopathy.
- ***CMT4*** is autosomal recessive with demyelinating pathology (very rare).

Hereditary motor neuropathy (HMN), also known as distal spinal muscular atrophy or distal hereditary motor neuropathy, is a rare cause of a pure motor syndrome, often affecting distal muscles and progressing slowly.

Hereditary sensory and autonomic neuropathies cause variable sensory loss, autonomic dysfunction, and neuropathic pain syndromes.

Other polyneuropathies

Neuropathy in cancer

Polyneuropathy or sensory neuronopathy is seen as a paraneoplastic syndrome (non-metastatic manifestation of malignancy). Many chemotherapy agents cause polyneuropathy. Polyneuropathy occurs in myeloma and other plasma cell dyscrasias via several mechanisms (see above). Individual nerves may be infiltrated with malignant cells, such as in lymphoma.

Neuropathies in systemic diseases

Vasculitic neuropathy occurs in SLE (see Chapter 32, 'Systemic lupus erythematosus' section), polyarteritis nodosa (Chapter 32, 'Polyarteritis nodosa' section), granulomatosis with eosinophilia (Chapter 32, 'Small-vessel vasculitis' section), and rheumatoid disease (Chapter 32, 'Rheumatoid arthritis' section). Non-systemic vasculitic neuropathy (localized only to the peripheral nerves) is much less common. Both multifocal neuropathy and a more confluent sensorimotor polyneuropathy occur. Vasculitic neuropathy responds to treatment with powerful immunosuppression including cyclophosphamide and rituximab; a nerve biopsy is often warranted prior to treatment.

Symmetrical axonal sensorimotor neuropathy is also more common in people with connective tissue disorders and other chronic inflammatory states; the exact pathogenesis is often opaque. Sjögren's syndrome can be associated with a sensory neuronopathy, or a more length-dependent sensory-predominant neuropathy.

Autonomic neuropathy

Autonomic neuropathy causes postural hypotension, urinary retention, erectile dysfunction, nocturnal diarrhoea, diminished sweating, impaired pupillary responses, and cardiac arrhythmias. This can develop in diabetes and amyloidosis (AL or TTR), and may complicate GBS. Many varieties of neuropathy cause autonomic problems in a mild form. Occasionally, such as with amyloidosis, a severe autonomic neuropathy may occur.

Neuromuscular weakness complicating critical illness

Some 50% of critically ill intensive care unit patients with multiple organ failure and/or sepsis develop an axonal polyneuropathy. This can delay ventilatory wean. Typically, distal weakness and absent reflexes are seen during recovery from critical illness. Resolution is usual.

Chronic sensorimotor neuropathy: no cause found

This situation is common. Progressive symmetrical numbness and tingling occur in hands and feet, spreading proximally in a glove and stocking distribution. Distal weakness also ascends. Tendon reflexes are lost. Symptoms may progress, remain static, or occasionally remit. Autonomic features are sometimes seen.

NCS usually show axonal degeneration. Nerve biopsy helps to classify selected cases: for example, diagnosing CIDP or unsuspected vasculitis.

Box 28.53 Common root and plexus problems

Nerve root
- Cervical and lumbar spondylosis
- Trauma
- Herpes zoster
- Tumours, e.g., neurofibroma, metastases
- Meningeal inflammation, e.g., syphilis, Lyme disease, arachnoiditis

Plexus
- Trauma
- Malignant infiltration
- Radiation damage
- Neuralgic amyotrophy (brachial neuritis)
- Diabetic amyotrophy (diabetic lumbosacral radiculoplexus neuropathy)
- Thoracic outlet syndrome (cervical rib)

Plexus and nerve root lesions

The common conditions that cause these are summarized in Box 28.53.

Cervical and lumbar degeneration

Spondylosis describes vertebral and ligamentous degenerative changes occurring during ageing or following trauma. Several factors produce neurological signs and symptoms:
- osteophytes—local overgrowth of bony spurs or bars
- thickening of spinal ligaments
- congenital narrowing of the spinal canal
- disc degeneration and protrusion (posterior and lateral protrusion: cord and root compression)
- vertebral collapse (osteoporosis, infection)
- rheumatoid synovitis (see Chapter 32, 'Rheumatoid arthritis' section)
- ischaemic changes within cord and nerve roots.

Narrowing of disc spaces, osteophytes, narrowing of exit foramina, and narrowing of the spinal canal are also seen on X-rays and MRI in the *symptomless* population, commonly in the mid- and lower cervical and lower lumbar region, and so imaging must not be over-interpreted.

Lateral cervical disc protrusion

The patient complains of pain in the arm. A C6-7 protrusion is the most common problem (Fig. 28.43). There is root pain that radiates to the C7 myotome (triceps, deep to scapula, and extensor aspect of forearm), with sensory disturbance, tingling, and numbness in the C7 dermatome.

In an established C7 root lesion, there is:
- weakness/wasting—triceps, wrist, and finger extensors
- loss of the triceps jerk (C7 reflex arc)
- C7 dermatome sensory loss.

Although the initial pain can be very severe, most cases recover with rest and analgesics. It is usual to immobilize the neck for a short time; gentle mobilization and physiotherapy should follow. Disc protrusion with root compression is seen on MRI. Root decompression is sometimes helpful.

Lateral lumbar disc protrusion

The L5 and S1 roots are commonly compressed by lateral prolapse of L4–L5 and L5–S1 discs; the root number below a disc interspace is compressed. There is low back pain and ***sciatica:*** pain radiating from the back to buttock and leg. Onset is typically acute. This can follow lifting, bending or minor injury. When pain follows such an event, it is tempting to ascribe the disc protrusion to it. However, commonly, lateral lumbar disc protrusion is spontaneous; lifting or injury is usually only brings forward an inevitable disc prolapse.

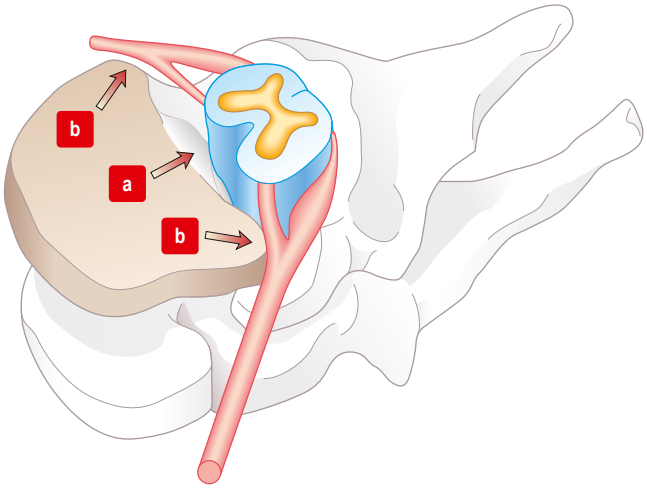


Fig. 28.43 Central and lateral disc protrusions. (a) Central disc protrusion compressing the cord. **(b)** Lateral disc protrusion compressing nerve roots.

Straight-leg raising is limited. There is reflex loss, such as ankle jerk in an S1 root lesion, and weakness of plantar flexion (S1) or great toe extension (L5); footdrop occurs in severe cases (L5). Sensory loss is found in the affected dermatome.

Most sciatica resolves with initial rest and analgesia, followed by early mobilization. MRI is sometimes appropriate. Surgery is indicated when a substantial persistent symptomatic disc lesion is shown.

Acute low back pain

Acute low back pain is extremely common. Often, pain is of disc or facet joint origin. Significant nerve root compression is unusual. Maintenance of activity and a trial of gentle manipulation are recommended (see also Chapter 32, 'Pain in the lower back' section).

Cervical spondylotic myelopathy

This is a relatively common disorder of older adults. Posterior disc protrusion (Fig. 28.44), common at C4–5, C5–6, and C6–7 levels, causes spinal cord compression. Congenital spinal canal narrowing, osteophytic bars, ligamentous thickening, and ischaemia are contributory. Usually, there are no or few neck symptoms. The patient complains of slowly progressive difficulty walking as a spastic paraparesis develops. There may be signs of myeloradiculopathy: LMN radicular signs at the level of the lesion (e.g., absent biceps and supinator jerks with a C5–6 disc protrusion), and brisk reflexes below the lesion (brisk triceps).

MRI demonstrates the level and extent of cord compression and T2 signal change is usually evident in the cervical cord at the point of maximal compression. Neck manipulation should be avoided.

Decompression by anterior cervical discectomy may be necessary when cord compression is severe or progressive. Improvement of the pyramidal signs may occur but is not guaranteed; progression is generally halted.

Spinal stenosis

A narrow spinal canal is developmental and frequently symptomless, but a congenital narrowing of the cervical canal predisposes the cervical cord to damage from minor disc protrusion later.

In the lumbosacral region, further narrowing of the canal by disc protrusion causes root pain and/or buttock and lower limb claudication. As the patient walks, nerve roots become hyperaemic and swell, producing buttock and lower limb pain with numbness (claudication). Surgical decompression is required.

Fig. 28.44 **C5/6 disc *(arrowed)* compressing the cord: T2-weighted MRI.**

Cauda equina syndrome

In severe cases of lumbar stenosis, or with large central disc herniations below the level of the conus, cauda equina syndrome may result, with bilateral leg weakness affecting multiple myotomes (L3–S1) and sphincter disturbance due to involvement of sacral nerve roots. This is a neurosurgical emergency.

Neuralgic amyotrophy (brachial neuritis)

Severe pain in the muscles around one shoulder is followed by wasting and weakness, usually of infraspinatus, supraspinatus, deltoid, and serratus anterior, evolving over several days. The cause is unknown but believed to be inflammatory; an allergic or viral basis is postulated. Rarely, a similar condition develops in distal upper limb muscles or in a lower limb. Recovery of wasted muscles is usual, but not invariable, over several months or years. There is no specific proven treatment.

Thoracic outlet syndrome

A fibrous band or cervical rib extending from the tip of the C7 transverse process towards the first rib compresses the lower brachial plexus roots, C8 and T1. There is forearm pain (ulnar border), T1 sensory loss, and thenar muscle wasting, principally of abductor pollicis brevis. A Horner's syndrome may develop. The rib or band can be excised. Frequency of this diagnosis varies widely; thoracic outlet problems are sometimes invoked to explain ill-defined arm symptoms, typically on poor evidence.

A rib or band can also produce subclavian artery or venous occlusion. Neurological and vascular problems rarely occur together.

Malignant infiltration and radiation plexopathy

Metastatic disease of nerve roots or the brachial or lumbosacral plexus causes a painful radiculopathy and/or plexopathy. An example is apical bronchial carcinoma (Pancoast's tumour) causing a T1 and sympathetic outflow lesion: wasting of small hand muscles, pain, and T1 sensory loss with an ipsilateral Horner's syndrome. This also occurs in apical tuberculosis.

In the upper limb, radiotherapy following breast cancer can produce a plexopathy. This is classically painless (as opposed to breast cancer metastases, which will produce a painful plexopathy). EMG in post-radiation plexopathy shows myokymia.

Further reading

Overell JR. Peripheral neuropathy: pattern recognition for the pragmatist. *Pract Neurol* 2011;11:62–70.
Rossor AM, Evans MRB, Reilly MM. A practical approach to the genetic neuropathies. *Pract Neurol* 2015;15:187–198.
https://www.gbs-cidp.org. *GBS/CIDP Foundation International.*

MUSCLE DISEASES

Definitions

- ***Myopathy*** means a disease of voluntary muscle.
- ***Myositis*** indicates inflammation.
- ***Muscular dystrophies*** are inherited disorders of muscle cells.
- ***Myasthenia*** means fatiguable (worse on exercise) weakness—seen in NMJ diseases.
- ***Myotonia*** is sustained contraction/slow relaxation.
- ***Skeletal muscle channelopathies*** are ion channel disorders of muscle cells.

Weakness is the predominant feature of muscle disease. A selection of these conditions is given in **Box 28.54**.

Pathophysiology

Muscle fibres are affected by:

- acute inflammation and fibre necrosis (e.g., polymyositis, infection)
- immunological damage (e.g., immune-mediated necrotizing myopathy)
- genetically determined metabolic failure (e.g., Duchenne muscular dystrophy)
- infiltration by inflammatory tissue (e.g., sarcoidosis)
- fibre hypertrophy and regeneration
- mitochondrial diseases
- ion channel disorders (e.g., chloride channel mutations in hereditary myotonias).

The NMJ is subject to immune-mediated damage (MG, LEMS), toxins (botulism, organophosphate poisoning), and genetic defects (congenital myasthenic syndrome).

Diagnosis

Clinical features, including the distribution of weakness, wasting or hypertrophy, and the tempo of progression and presence of family history contribute to a clinical diagnosis. Several investigations help make a definitive diagnosis.

Serum muscle enzymes

Serum creatine kinase (CK) is a marker of muscle fibre damage and is greatly elevated in many dystrophies, such as Duchenne, and in inflammatory muscle disorders, such as dermatomyositis. Elevated CK is not specific for muscle disease and can go up (usually to about 1000 IU/L) in MND, other neurogenic disorders, or after exercise; ethnicity has an effect on baseline CK.

Box 28.54 Muscle and neuromuscular junction disease: classification

Acquired
- Idiopathic inflammatory myopathy
  - Polymyositis (including anti-synthetase syndrome and overlap myositis)
  - Dermatomyositis
  - Immune-mediated necrotizing myopathy
  - Inclusion body myositis
- Other inflammatory
  - Viral, bacterial, and parasitic infection
  - Sarcoidosis
- Endocrine/toxic
  - Corticosteroids/Cushing's
  - Thyroid disease
  - Calcium disorders
  - Alcohol misuse
  - Drugs, e.g., statins, anti-retrovirals

Genetic
- Muscular dystrophies
  - Duchenne
  - Facioscapulohumeral
  - Limb girdle and others
- Myotonic dystrophy
  - Myotonic dystrophy type 1
  - Myotonic dystrophy type 2
- Skeletal muscle channelopathies
  - Myotonia congenita/paramyotonia congenita
  - Hypokalaemic/hyperkalaemic periodic paralysis
- Metabolic
  - Myophosphorylase deficiency (McArdle's syndrome)
  - Other defects of glycogen and fatty acid metabolism
  - Pompe disease
- Mitochondrial disease
- Congenital myopathies

Neuromuscular junction disorders
- Myasthenia gravis
- Lambert–Eaton myasthenic syndrome (LEMS)
- Botulism
- Congenital myasthenic syndrome

Neurogenetic tests

These are essential in muscular dystrophies and mitochondrial disease.

Electromyography

Characteristic EMG patterns are as follows:
- ***Myopathy***. Short-duration spiky polyphasic muscle action potentials are seen. Spontaneous fibrillation is occasionally recorded.
- ***Myotonic discharges***. A characteristic high-frequency whine is heard.
- ***Decrement and increment***. In MG, a characteristic decrement in evoked muscle action potential follows repetitive motor nerve stimulation. The reverse occurs, that is increment, following repetitive stimulation in LEMS (see 'Lambert–Eaton myasthenic syndrome' section).
- In ***denervation***, profuse fibrillation potentials are seen.

Muscle biopsy

Unlike neural tissue, skeletal muscle can be easily biopsied to provide a definitive diagnosis using powerful molecular immunohistochemical techniques. Histology and muscle histochemistry of fibre types demonstrate denervation, inflammation, and dystrophic changes. Electron microscopy is often valuable. In dystrophies, immunohistochemistry in specialist laboratories allows identification of the abnormal muscle protein and a precise molecular diagnosis.

Imaging

MRI shows ***signal changes*** within muscles in some cases of myositis, and fatty replacement of muscle in chronically damaged muscles.

Clinical approach to muscle disease

The clinical approach to patients with muscle disease is described in **Box 28.55**.

Inflammatory myopathies

Inflammatory myopathies, including polymyositis, dermatomyositis, immune-mediated necrotizing myopathy, and inclusion body myositis (IBM), are described in Chapter 32 ('Polymyositis and dermatomyositis' and 'Inclusion body myositis' sections). While IBM is often classified as an idiopathic inflammatory myopathy, it does not respond to immunosuppression and is a combined inflammatory and degenerative disease of muscle. Granulomatous muscle infiltration and inflammation may occur in sarcoidosis and other disorders such as rheumatoid arthritis, causing a mild myopathy. Viral myositis may also occur, and muscles may be involved in other infections such as neurocysticercosis (see Chapter 42, '*Taenia solium* and cysticercosis' section).

Metabolic and endocrine myopathies

Corticosteroids and Cushing's syndrome

Proximal weakness occurs with prolonged high-dose steroid therapy and in Cushing's syndrome (see Chapter 34, 'Cushing's syndrome' section). Selective type-2 fibre atrophy is seen on biopsy.

Thyroid disease

Several myopathies occur (see also Chapter 34, 'Hyperthyroidism' section). Thyrotoxicosis can be accompanied by severe proximal myopathy. There is also an association between thyrotoxicosis and MG, and thyrotoxicosis can cause an acquired hypokalaemic periodic paralysis. Both associations are seen more frequently in South-east Asia. In ophthalmic Graves' disease, there is swelling and lymphocytic infiltration of extraocular muscles (see Chapter 34, 'Graves' orbitopathy [ophthalmopathy]' section).

Hypothyroidism is sometimes associated with muscle pain and stiffness, resembling myotonia. A proximal myopathy also occurs.

Disorders of calcium and vitamin D metabolism

Proximal myopathy develops in hypocalcaemia, rickets, and osteomalacia (see Chapter 33, 'Rickets and osteomalacia' section).

Hypokalaemia

Acute hypokalaemia (e.g., with diuretics) causes flaccid paralysis reversed by potassium, given slowly (see Chapter 27, 'Hypokalaemia' section). Chronic hypokalaemia leads to mild, mainly proximal, weakness. (See also periodic paralysis; 'Hypokalaemic periodic paralysis' section).

Alcohol and drugs

Severe myopathy or rhabdomyolysis with muscle pain, necrosis, and myoglobinuria occurs in acute excess. A subacute proximal

Box 28.55 Clinical approach to muscle disease

Is it myopathy?

- Muscle weakness, typically proximal, and often symmetrical (with exceptions). Difficulty climbing stairs, rising from a seated or supine position; difficulty brushing hair or lifting arms above the head
- Distal weakness is less common. This may result in difficulty turning a key in a lock, or foot drop
- Reflexes will be reduced commensurate with the degree of weakness; if they are absent, that may be a sign of neuropathy. Muscle atrophy can ensue if the condition is chronic; rapid wasting is more common with a neurogenic process
- Myotonia (delayed relaxation after a forceful contraction; look for the slow-to-release handshake) may indicate dystrophic or non-dystrophic myotonias
- Cramps, contractures, stiffness, myoglobinuria may suggest a muscle disease
- NMJ disorders are accompanied by fatiguability and variability
- Dysphagia and dysarthria can be features of both nerve and muscle pathology; dysarthria always comes before dysphagia in motor neuron disease, so isolated oropharyngeal dysphagia is more likely to occur in myopathy or myasthenia (or from a non-neurological cause)
- Pain is not a feature of most muscle diseases. It can occur with metabolic myopathies, and fascial inflammation in idiopathic inflammatory myopathies will cause pain; otherwise, frequently the cause of pain will be rheumatological or secondary to altered forces in the presence of muscle weakness
- Sensory changes are absent

What is the time course?

- Acute/subacute: inflammatory myopathy, viral myositis, toxic insults, botulism
- Chronic and slowly progressive: muscular dystrophies, myotonic dystrophy, mitochondrial diseases
- Relapsing or variable: myasthenia gravis, metabolic myopathies, channelopathies

Age of onset

- At birth (floppy baby)—congenital myopathies including nemaline rod myopathy and central core disease; congenital muscular dystrophies (unlike the congenital myopathies, which tend to be static, congenital muscular dystrophies cause increased muscle breakdown with age); congenital myasthenic syndrome
- Childhood—congenital myopathies, endocrine/metabolic disorders, Pompe disease, juvenile dermatomyositis, Duchenne, Becker, FSHD, LGMD, Emery-Dreifuss muscular dystrophy, mitochondrial disorders
- Adulthood—a large number of disorders (including some that usually present in childhood but have a late-onset form)

Family history

- Autosomal dominant: myotonic dystrophy, FSHD, LGMD1, central core disease, oculopharyngeal muscular dystrophy
- Autosomal recessive: LGMD2, metabolic myopathies
- X-linked: Becker, Duchenne, Emery-Dreifuss muscular dystrophy
- Maternal transmission: mitochondrial

What is the pattern?

- Assess for involvement of the following muscle groups: ocular, facial, bulbar, neck, trunk, respiratory, shoulder girdle, hands/forearm, hip girdle, leg/foot
- Proximal weakness involving shoulder and hip girdles is the usual pattern
- Distal weakness may point to a specific disorder (weakness of long finger flexors in IBM, distal weakness in myotonic dystrophy, rare distal myopathies)
- Scapuloperoneal weakness (involvement of proximal arms and peroneal muscles) seen in FSHD (may be very asymmetrical), Emery-Dreifuss muscular dystrophy, Pompe disease
- Facial involvement in FSHD or myotonic dystrophy (myopathic facies). Can they whistle?
- Respiratory muscle involvement in Duchenne, Becker, Pompe disease, rapidly progressive inflammatory myopathies, and congenital myopathies
- Axial weakness in ryanodine receptor 1 myopathy; prominent neck extensor weakness in FSHD, idiopathic inflammatory myopathies, myotonic dystrophy, myasthenia gravis (beware: head drop can also be seen in MND)
- Ophthalmoplegia and/or ptosis may point to mitochondrial myopathy, myasthenia gravis, oculopharyngeal muscular dystrophy, or myotonic dystrophy (ptosis without ophthalmoplegia)

Provoking features

- Exercise-induced myalgia/rhabdomyolysis in metabolic myopathies
- Cold temperatures worsen myotonia congenita
- Drugs (e.g., aminoglycosides) and intercurrent illness can provoke a myasthenic crisis
- Large carbohydrate load can precipitate periodic paralysis

Associated features

- Recurrent rhabdomyolysis and myoglobinuria in metabolic myopathies
- Classical rash of dermatomyositis (see Chapter 32)
- Interstitial lung disease in antisynthetase syndrome (see Chapter 32) or features of systemic connective tissue disorder in overlap myositis (see Chapter 32)
- Contractures (inability to fully stretch a muscle because of fibrosis) are a feature of some types of inherited muscle disease (e.g., Emery-Dreifuss muscular dystrophy, calpainopathy—LGMD 2A/R1)
- Contractures can also refer to sustained painful muscle contractions, as seen in McArdle's disease
- Cardiac involvement can be a diagnostic pointer and also a worrying feature (arrhythmias in myotonic dystrophy, cardiomyopathy in Pompe disease, several muscular dystrophies; cardiac myositis can be a feature of anti-SRP necrotizing myopathy and rarely other types of inflammatory myopathy)
- Presence of frontal balding and cataracts in myotonic dystrophy; other multisystem features, e.g., diabetes, sensorineural hearing loss, epilepsy, in mitochondrial myopathies

FSHD, Facioscapulohumeral muscular dystrophy; *IBM,* inclusion body myositis; *MND,* motor neuron disease; *LGMD,* limb-girdle muscular dystrophy; *NMJ,* neuromuscular junction; *SRP,* signal recognition peptide.

myopathy occurs with chronic alcohol use. A similar acute myopathy occurs with cocaine, diamorphine, and amfetamine use.

Medications

Drug-induced muscle disorders include proximal myopathy (steroids), muscle weakness (lithium), painful muscles (fibrates), rhabdomyolysis (a fibrate combined with a statin, or interaction between statins and other drugs such as certain antibiotics), mitochondrial myopathy (zidovudine), and malignant hyperthermia (volatile anaesthetics, succinylcholine). Most respond to drug withdrawal. Immune checkpoint inhibitors, D-penicillamine, and TNF-alpha blockers can trigger an immune-mediated myopathy. Statins are also associated with immune-mediated necrotizing myopathy, in which case antibodies to 3-hydroxy-3-methyl-glutaryl-coenzyme A reductase (HMGCR) are found. The precise role of statins in the pathogenesis of this condition is unclear (and it can occur without prior statin exposure). Immunomodulatory therapies are necessary.

Myophosphorylase deficiency (McArdle's syndrome)

McArdle's syndrome is a muscle-only glycogenosis where there is a deficiency of muscle glycogen phosphorylase (PYGM). It presents with exercise intolerance, muscle cramps, fatigue, anaesthetic

problems, and myoglobinuria after exercise in adults; a 'second wind' phenomenon (improvement after 10 mins of exercise), as the switch to fatty acid metabolism is made, is classical. Diagnosis is made by genetic testing (or on a muscle biopsy, especially in atypical cases). Patients have a normal lifespan. Sucrose should be given before exercise, and regular mild-to-moderate aerobic exercise improves symptoms.

Acid maltase/acid alpha glucosidase deficiency (Pompe disease)

Pompe disease is an autosomal recessive lysosomal storage disease that leads to accumulation of glycogen within the lysosome (see Chapter 36). The classic form presents in infancy with hypotonia and hypertrophic cardiomyopathy, but a late-onset form with progressive limb-girdle weakness, respiratory failure, and bent spine also occurs. Acid maltase activity can be readily measured in a dry blood spot, prior to proceeding to genetic testing, and enzyme replacement therapies are available. A newborn screening programme is in place in many countries.

Malignant hyperthermia

Widespread skeletal muscle rigidity with hyperpyrexia as a sequel of volatile anaesthetics or succinylcholine (depolarizing neuromuscular blocking agent) is due in most cases to genetic defects in the pathways regulating sarcoplasmic reticulum calcium release. It is often inherited as a dominant trait, with variants in the muscle ryanodine receptor 1 *(RYR1)*, or the dihydropyridine receptor *(CACNA1S)*. Death during or following anaesthesia can occur in this rarity. Dantrolene is of some help for rigidity.

Neuromuscular junction disorders

Myasthenia gravis

Myasthenia gravis (MG) is an autoimmune disorder of NMJ transmission, characterized by weakness and fatiguability of proximal limb, bulbar, and ocular muscles, the latter sometimes in isolation. The prevalence is about 4 in 100 000. MG is twice as common in women as in men, with a peak age incidence around 30 and a second smaller peak in incidence in older men.

Antibodies to acetylcholine receptor protein (anti-AChR antibodies) are pathogenic. Immune complexes of anti-AChR IgG and complement are deposited at the postsynaptic membranes, causing destruction of AChRs. Antibodies against muscle-specific receptor tyrosine kinase (anti-MuSK antibodies) have been identified in anti-AChR antibody-negative cases.

Thymic hyperplasia is found in 70% of MG patients below the age of 40. In some 10%, a thymic tumour is found (paraneoplastic myasthenia), the incidence increasing with age. Thymoma-associated MG is more treatment-resistant.

There is an association between MG and other autoimmune disorders: thyroid disease, rheumatoid disease, pernicious anaemia, and SLE. Transient MG is sometimes caused by D-penicillamine treatment.

Clinical features

Weakness and fatiguability are typical. Limb muscles (proximal), extraocular muscles, speech, facial expression, and mastication muscles are commonly affected. Symptoms are worse towards the end of the day. Fluctuating diplopia and ptosis (often asymmetric) are frequent early symptoms, and symptoms may be confined to the eyes: ocular myasthenia. Respiratory and bulbar difficulties can occur in generalized myasthenia. Early complaints of fatigue are often dismissed.

Complex extraocular palsies, ptosis, and fatiguable proximal weakness are found on examination (prolonged upgaze should be checked and limb power tested after repetitive contractions). The reflexes are initially preserved but may be fatiguable: that is, they disappear following repetitive activity. Wasting is sometimes seen after many years.

Investigations

- ***Serum anti-AChR and anti-MuSK antibodies.*** Anti-AChR antibodies are present in some 80% to 90% of cases of generalized MG, and about 50% of cases of ocular MG. These antibodies are highly specific for MG and confirm the diagnosis. Anti-MuSK antibodies define a subgroup of MG patients (up to 10% in generalized disease) characterized by weakness predominantly in bulbar, facial, and neck muscles.
- ***Repetitive nerve stimulation and single-fibre EMG.*** A characteristic decrement occurs in the evoked muscle action potential during repetitive stimulation. Single-fibre EMG of orbicularis oculi is more sensitive but less specific than repetitive stimulation and shows block and jitter.
- ***Imaging and other tests.*** Mediastinal MRI or CT is needed to look for thymoma in all cases. Antibodies to striated muscle suggest a thymoma.

Course and management

MG fluctuates in severity; most cases have a protracted, life-long course. Respiratory impairment, nasal regurgitation, and dysphagia occur; emergency ventilation may be required. Simple monitoring tests, such as the duration for which an arm can be held outstretched, and the vital capacity, are useful.

Exacerbations are usually unpredictable and unprovoked but may be brought on by infections and by aminoglycosides and other drugs.

Drug treatment

Oral anticholinesterases

Pyridostigmine (60 mg tablet) is widely used. The duration of action is 3 to 4 hours, and the dose (usually 3–6 tablets daily) determined by response. Pyridostigmine prolongs acetylcholine action by inhibiting acetylcholinesterase. Overdose of anticholinesterases causes severe weakness (cholinergic crisis). Muscarinic side-effects, such as colic and diarrhoea, are common; oral propantheline (antimuscarinic) helps to control these. Anticholinesterases help weakness but do not alter the natural history of myasthenia.

Immunosuppressant drugs

These drugs are used in patients who are not fully controlled on pyridostigmine or who have severe weakness. Steroids are usually used and are the mainstay of treatment for myasthenia. There is improvement in 70%, although steroid dose should be increased slowly as there is a risk of steroid-dipping (an initial relapse triggered by steroids). Azathioprine, mycophenolate, and other immunosuppressants are introduced as steroid-sparing agents but take a long time to work.

Thymectomy

When a thymoma is present, the potential for malignancy makes surgery necessary but the myasthenia may not improve. In patients without thymoma, thymectomy improves myasthenia in patients with thymic hyperplasia and positive AChR antibodies. Practice varies but the evidence supports offering thymectomy to patients under the age of 50 with generalized AChR antibody-positive MG, with the aim of inducing remission, reducing the need for immunosuppression, and reducing hospital admissions.

It is not a cure and the effects take time; thymectomy should only be performed in patients with stable myasthenic symptoms. The benefits are less clear in older patients who have more atrophy of the thymus, but thymectomy can be considered on an individual basis; it can also be offered to patients with double-antibody-negative disease and patients with treatment-refractory ocular myasthenia.

Plasmapheresis and intravenous immunoglobulin

These produce a rapid dramatic response and are used in exacerbations and severe myasthenic crisis.

Biologics

Rituximab is used for treatment-refractory disease although the benefits for patients with MuSK antibodies seem to be greater than those with AChR antibodies. Eculizumab is a recombinant monoclonal antibody that reduces complement activation. It has been shown to be effective in trials of treatment-refractory MG but there is a significant increased risk of meningococcal disease. Ravulizumab is a similar agent with a longer half-life. Efgartigimod (an antibody fragment) and rozanolixizumab (monoclonal antibody) both block the neonatal Fc receptor and can be used as chronic immunotherapy or as a bridge therapy to minimize steroid exposure.

Other rare myasthenic syndromes exist, such as congenital myasthenia.

Lambert–Eaton myasthenic syndrome

Lambert–Eaton myasthenic syndrome (LEMS) is a paraneoplastic manifestation of small-cell bronchial carcinoma due to defective acetylcholine release at the NMJ. A smaller proportion of cases are autoimmune without underlying malignancy. Proximal limb muscle weakness, mimicking a myopathy, sometimes with ocular/bulbar muscle involvement, develops, with some absent tendon reflexes: a cardinal sign. Weakness tends to improve after brief maximal muscle contraction, and absent reflexes return (compare myasthenia). More sustained exercise results in fatiguability. Diagnosis is confirmed by repetitive nerve stimulation (increment; see above). Antibodies to voltage-gated calcium channels are found in most cases (90%). Treatment with 3,4-diaminopyridine (DAP) is reasonably safe and effective.

Muscular dystrophies

These progressive, genetically determined disorders of skeletal and sometimes cardiac muscle have a complex clinical and neurogenetic classification.

Duchenne muscular dystrophy and Becker's muscular dystrophy

These are inherited as X-linked recessive disorders, though one-third of cases are spontaneous mutations. Duchenne muscular dystrophy (DMD) occurs in 1 in 3000 male infants. There is absence of the gene product *dystrophin*, a rod-shaped cytoskeletal muscle protein in DMD. In Becker's dystrophy, dystrophin is present but levels are low. DMD is usually obvious by the fourth year, and often causes death by the age of 20.

Dystrophin is essential for muscle cell membrane stability. Deficiency leads to reduction in three glycoproteins (α-, β-, and γ-sarcoglycans) in the dystrophin-associated protein complex (DAP complex) that links dystrophin to laminin within cell membranes.

Becker's muscular dystrophy is less severe than Duchenne and weakness only becomes apparent in young adults.

Clinical features

A boy with DMD is noticed to have difficulty running and rising to his feet; he uses his hands to climb up his legs (Gowers' sign). There is initially proximal limb weakness with calf pseudohypertrophy. The myocardium is affected. Severe disability is typical by the age of 10.

Investigations

The diagnosis is often suspected clinically. CK is grossly elevated (100–200 times normal). Biopsy shows variation in muscle fibre size, necrosis, regeneration and replacement by fat, and on immunochemical staining, absence of dystrophin.

Management

There is no curative treatment but steroids delay progression. New gene-editing therapies have been shown to increase dystrophin expression (but clinical benefit has not yet been established). Ataluren allows ribosomal read-through of a premature stop codon and so can be used for those patients with a nonsense mutation; it is approved in the UK and EU. Several other agents promote exon-skipping; access is limited. Physiotherapy helps prevent contractures in the later stages. Non-invasive respiratory support and multidisciplinary care improve life expectancy.

Carrier detection. Females with an affected brother have a 50% chance of carrying the DMD gene. In carriers, 70% have a raised CK, and usually EMG abnormalities and/or changes on biopsy. Carrier and prenatal diagnosis is available with genetic counselling.

Limb-girdle and facioscapulohumeral dystrophy

These less severe but disabling dystrophies are summarized in **Box 28.56**. There are many other varieties of dystrophy;

Box 28.56 Limb-girdle and facioscapulohumeral dystrophies

	Limb-girdle	Facioscapulohumeral
Inheritance	Recessive, dominant, or X-linked	Dominant, new mutations common
Onset	10–20 years	10–40 years
Muscles affected	Proximal: shoulders, pelvic girdle	Face, shoulders, scapular winging, upper arms, foot drop. Typically asymmetric
Progression	Severe disability <25 years	Life expectancy normal, slow progression
Associated features	Very variable. May include contractures, cardiac involvement, respiratory failure	Occasionally, deafness and retinal abnormalities
Diagnosis	Multiple genetic types Diagnosis on muscle biopsy, immunohistochemistry, and genetic testing	Genetic testing (muscle biopsy usually not required but may show inflammatory changes)

facioscapulohumeral dystrophy is one of the most common. Genes for numerous forms of limb-girdle muscular dystrophy have been identified. CK is usually moderately elevated.

Other inherited muscle disorders

- *Emery-Dreifuss muscular dystrophy* is a genetically heterogeneous group of disorders with onset in childhood, typified by early contractures, slowly progressive humeroperoneal weakness, and cardiac disease with cardiomyopathy and conduction defects.
- *Oculopharyngeal muscular dystrophy* causes progressive ptosis, ophthalmoplegia, dysphagia, and dysarthria starting in middle life; it mimics a mitochondrial myopathy.
- *Distal muscular dystrophies* such as Nonaka myopathy (hereditary IBM) start in the distal muscles and progress proximally.
- *Congenital myopathies* are caused by abnormalities of muscle development, and present with hypotonia in early life, often with breathing difficulties; they include central core disease (*RYR1* gene mutations) and nemaline rod myopathy (occasionally presents later in life).
- *Congenital muscular dystrophies* present in the first 2 years of life and cause progressive muscle breakdown over time; there may be associated central nervous system defects.

Myotonic syndromes

Myotonia denotes continued, involuntary muscle contraction after cessation of voluntary effort: that is, failure of muscle relaxation. EMG is characteristic (see 'Electromyography and nerve conduction studies' section). Myotonia can be caused by myotonic dystrophy, or by skeletal muscle channelopathies (myotonia congenita and paramyotonia congenita). Electrical myotonia can also be seen as a non-specific feature of other myopathies and with drug toxicity.

Myotonic dystrophy (DM)

This autosomal dominant condition is a genetic disorder with two different triple repeat mutations: most commonly, an expanded CTG repeat in a protein-kinase *(DMPK)* gene (DM1). There is a correlation between disease severity, age at onset, and approximate size of triplet repeat mutations. There is progressive distal muscle weakness, with ptosis, weakness, and thinning of the face and sternomastoids. Myotonia is typically present. Muscle disease is part of a syndrome comprising:

- cataracts
- frontal baldness
- cognitive impairment (mild)
- oesophageal dysfunction (and aspiration)
- cardiomyopathy and conduction defects (sudden death can occur in DM1)
- small pituitary fossa and hypogonadism
- glucose intolerance
- low serum IgG.

This gradually progressive condition usually becomes evident between 20 and 50 years. DM2 (or proximal myotonic myopathy—PROMM) is much less common. It is caused by an expanded CCTG repeat in a zinc finger protein gene. The clinical manifestations are similar to DM1 but less severe; there is more prominent proximal (hip-girdle) weakness.

Skeletal muscle channelopathies

Non-dystrophic myotonias

Myotonia congenita: clinical myotonia, usually mild, becomes evident in childhood, but may be ignored until much later in life. The gene, *CLCN1*, codes for a muscle chloride channel. The myotonia, which persists, is accentuated by rest and by cold. Warm-up phenomenon means the myotonia is relieved with repeated activity. Diffuse muscle hypertrophy occurs, meaning that although physically weak, the patient will have bulky muscles. Recessive and dominant forms occur.

Paramyotonia congenita worsens with exercise and cold and is caused by mutations in the sodium channel gene, *SCN4A*; it can be associated with hyperkalaemic periodic paralysis.

Hypokalaemic periodic paralysis

This disorder is characterized by generalized weakness, including that of bulbar muscles, which often starts after a heavy carbohydrate meal or following exertion. Attacks last for several hours. The disorder often first comes to light in the teenage years and tends to remit after the age of 35. Serum potassium is usually below 3.0 mmol/L in an attack. The weakness responds to (slow) intravenous potassium chloride. It is usually inherited as an autosomal dominant trait caused by mutation in a muscle voltage-gated calcium channel gene *(CACNA1S)*. Other mutations in the sodium channels *(SCN4A)* and potassium channels *(KCNE3)* also occur. Acetazolamide sometimes helps prevent attacks. Weakness can be caused by diuretics. A similar condition can also occur with thyrotoxicosis.

Andersen-Tawil syndrome presents with hypokalaemic periodic paralysis but is caused by mutations in a potassium channel gene, *KCNJ2*; patients have dysmorphic features and cardiac conduction defects that can lead to sudden death.

Hyperkalaemic periodic paralysis

This condition, also autosomal dominant, is characterized by attacks of weakness, sometimes with exercise. Attacks start in childhood and tend to remit after the age of 20; they last about 30 to 120 minutes. Myotonia may occur. Serum potassium is elevated. An attack can be terminated by intravenous calcium gluconate. There are point mutations in a muscle voltage-gated sodium-channel gene *(SCN4A)*. Acetazolamide or a thiazide diuretic can be helpful.

Stiff person syndrome

Stiff person syndrome (SPS) is a rare autoimmune disease, more common in females, causing axial muscle stiffness with abnormal posture, spasms, and falls. Attacks of stiffness are sometimes provoked by noise or emotion, but sometimes occur spontaneously. Between attacks, which last from hours to days or even weeks, the patient may appear normal.

Widespread muscle stiffness is typical during an attack; there are no other neurological signs. SPS has been mistaken for Parkinson's, dystonia, and functional neurological disorder. Anti-glutamic acid decarboxylase antibodies (anti-GAD) are found in very high titre in more than 50% of cases and are believed to be involved in the generation of muscle stiffness. Continuous motor activity in paraspinal muscles is seen on EMG.

Treatment with diazepam, other muscle relaxants, and intravenous immunoglobulin or plasma exchange can be helpful during attacks.

A form of SPS is also seen occasionally as a paraneoplastic condition associated with antibodies to the synaptic protein amphiphysin.

Mitochondrial diseases

These comprise a complex group of rare disorders involving muscle, peripheral nerves, and CNS, characterized by morphological

and biochemical abnormalities in mitochondria. Multisystemic involvement including diabetes, deafness, and cardiac involvement is common. Mitochondrial DNA is inherited maternally (see Chapter 5, 'Mitochondrial chromosome disorders' section), but many important mitochondrial proteins and functions are encoded in the nuclear genome (e.g., *POLG*). Therefore, inheritance of mitochondrial disorders can be autosomal dominant, autosomal recessive, X-linked, or maternal. The spectrum is wide, ranging from optic atrophy to myopathies, neuropathies, and encephalopathy.

- *MELAS* (*m*itochondrial *e*ncephalomyopathy, *l*actic *a*cidosis, *s*troke-like episodes) is one well-recognized form.
- Chronic progressive ophthalmoplegia (CPEO) is another.
- *MERRF* describes *m*yoclonic *e*pilepsy with abnormal muscle histology, the muscle appearance being described as *r*agged *r*ed *f*ibres.
- Leigh syndrome describes a subacute necrotizing encephalomyelopathy, usually in children.

Further reading

Barohn RJ, Dimachkie MM, Jackson CE. A pattern recognition approach to patients with a suspected myopathy. *Neurol Clin* 2014; 32:569–593.
Gorman G, Chinnery P, DiMauro S et al. Mitochondrial diseases. *Nat Rev Dis Primers* 2016; 2:16080.
Ryan A, Matthews E, Hanna MG. Skeletal muscle channelopathies: periodic paralysis and nondystrophic myotonias. *Curr Opin Neurol* 2007; 20:558–563.
https://www.musculardystrophyuk.org *Muscular dystrophy UK (UK national charity)*.

Bibliography

Howard R, Kullmann D, Werring DJ, Zandi M (eds). *Neurology: A Queen Square Textbook*. 3rd edition. Chichester: Wiley Blackwell; 2024.

Significant websites

http://jnnp.bmjjournals.com *Journal of Neurology, Neurosurgery and Psychiatry.*
https://pn.bmj.com *Practical Neurology.*
http://www.theabn.org *Association of British Neurologists information service*.

Questions for this chapter can be found on Elsevier eBooks+.

29 Stroke medicine

Ganesh Subramanian, Margaret O'Connor, Soma Banerjee, and Abhijit Das

Background 977
Pathogenesis and aetiology 977
TOAST classification 978
Stroke classification by vascular territory 981
Anterior circulation clinical stroke syndromes 981
Posterior circulation clinical stroke syndromes 981
Transient ischaemic attack 982
Acute ischaemic stroke 983
Imaging 983
Intravenous thrombolysis 988
Mechanical thrombectomy 988
Stroke unit care, rehabilitation, and recovery 991
Acute stroke unit care 991
Primary and secondary ischaemic stroke prevention 993
Lifestyle modifications 993
Pharmacological approaches 994
Specific interventions 997
Intracerebral haemorrhage 998

CORE SKILLS AND KNOWLEDGE

Stroke medicine has developed as a subspecialty for doctors specializing in the acute management and rehabilitation of patients suffering a stroke. In the UK, stroke training is completed via a year of subspecialist training available to specialist trainees from parent specialties including neurology, geriatrics, rehabilitation medicine, or acute medicine.

Key skills in stroke medicine include:

- Outlining the key elements of a stroke history
- Applying the principles of hyperacute stroke management to clinical practice, for ischaemic and haemorrhagic stroke
- Developing a personalized secondary stroke prevention and rehabilitation strategy for individual stroke patients
- Explaining the principles of assessment and management of TIA.

Opportunities to learn stroke medicine include shadowing the on-call stroke doctors and witnessing acute stroke management (including thrombolysis and mechanical thrombectomy), attending ward rounds on an acute stroke unit, and observing the work of the multidisciplinary team in providing longer-term stroke rehabilitation.

BACKGROUND

Stroke is the second most common cause of death and a leading cause of disability worldwide, with one in four people over the age of 25 suffering a stroke in their lifetime. There is a 30% mortality rate and 40% dependency rate at 1 year after a stroke. The global burden of stroke disproportionately impacts middle- and low-income countries. Stroke rates are higher in Asian and Black African populations than in White populations. The incidence of stroke increases with age. However, stroke is not just a disease of older people, with one-quarter of all strokes occurring before the age of 65 years.

Definitions

Stroke is *defined* as a syndrome of rapid onset neurological deficit caused by focal cerebral, spinal, or retinal infarction or haemorrhage. Tissue injury is confirmed by neuroimaging. The clinical history is focused on rapidly establishing the onset of symptoms as well as determining risk factors for cerebrovascular disease (Box 29.1).

Transient ischaemic attack (TIA) refers to a brief episode of neurological dysfunction due to temporary focal cerebral or retinal ischaemia without infarction, such as a weak limb, aphasia, or loss of vision, usually lasting seconds or minutes with complete recovery. TIAs may herald a future stroke. Stroke and TIA are traditionally defined by the (arbitrary) duration of symptoms less than 24 hours (time-based definition); however, with the availability of advanced neuroimaging, a TIA may be differentiated by the absence of infarction on imaging (tissue-based definition).

PATHOGENESIS AND AETIOLOGY

The underlying pathology responsible for stroke is either infarction (85%) or haemorrhage (15%). Haemorrhagic stroke is outlined in detail in the 'Intracerebral haemorrhage' section. For further information on subarachnoid haemorrhage, extradural and subdural bleeding, refer to Chapter 11.

Infarction of brain tissue occurs due to a critical reduction in regional cerebral blood flow. This results in intracellular biochemical disturbance in the ischaemic tissues (with protein synthesis inhibition, lactic acidosis, and failure of ATP generation), causing anoxic depolarization, intracellular potassium efflux, and cell death (infarction). Major advances in therapy focus on early restoration of blood flow for salvageable brain tissue that has not yet infarcted. Notably, grey matter is more susceptible to ischaemia than white matter.

Cerebral autoregulation endeavours to compensate for reduced perfusion with dilation of leptomeningeal collateral blood vessels (Fig. 29.1). Blood flow can bypass the occluded vessel via the circle of Willis to supply leptomeningeal collaterals. This provides time to salvage ischaemic brain with thrombolysis or thrombectomy particularly as collateral failure often occurs over time.

Box 29.1 Taking a history in acute stroke

- A focused and rapid history is needed for hyperacute stroke to assess for thrombolysis and thrombectomy
- Collateral history is required if the patient is aphasic, disorientated, or drowsy
- A more comprehensive history can be obtained later

Presenting symptoms

- Time of symptom onset or last known to be well (provides guidance for thrombolysis and thrombectomy)
- Time of onset for a wake-up stroke is assumed to be when an individual fell asleep in the absence of advanced imaging to demonstrate otherwise
- Pattern of onset—acute and maximal at onset (embolic), stuttering or progressive (lacunar or haemorrhage)
- Triggers/activity at onset—Valsalva associated with patent foramen ovale-related stroke; migraine associated with specific triggers
- Orientation and alertness
- Localization
  - Anterior circulation symptoms: motor weakness, difficulty mobilizing, sensory loss, dysarthria, aphasia, neglect, dysphagia, apathy, abulia, and cognitive impairment
  - Posterior circulation symptoms: 5 Ds diplopia, dizziness (vertigo), dysphagia, dysarthria, drowsiness; ataxia, incoordination, sensory disturbance, nausea, vomiting, and visual loss

Screening questions for other relevant features

- Handedness—dominant or non-dominant
- Symptoms of raised intracranial pressure—vomiting or early morning headaches worse lying down, diplopia
- Systemic symptoms—malignancy (weight loss, nausea, etc.); inflammatory conditions (mouth ulcers, joint pains, alopecia, rash, muscle pain, jaw claudication, etc.) and fever (infections as triggers for stroke or vasculopathy due to meningitis or post-stroke complications)
- Urine and bowel incontinence or difficulty passing urine (urinary retention)
- Loss of consciousness, aura, muscle jerking, tongue biting, incontinence, and a drowsy post-ictal phase (features to suggest seizure as a stroke mimic)
- Disturbance of higher functions—personality change, cognitive decline, loss of executive control, disinhibition, psychiatric symptoms
- Positive symptoms such as visual flashing, spots, or tingling are more suggestive of stroke mimic such as migraine
- Pregnancy/postpartum (alternative aetiology and caution with secondary prevention therapy)
- Other symptoms (palpitations—atrial fibrillation; neck pain—dissection)

Past medical history

- History of previous TIA and/or stroke (ischaemic or haemorrhagic) and any residual deficit (contraindication to thrombolysis with recent stroke)
- Head injury (haemorrhage or dissection)
- Atherosclerotic risk factors—hypertension, diabetes, other cardiovascular disease; hypercholesterolemia
- Embolic risk factors—atrial fibrillation, patent foramen ovale, recent myocardial infarction, prosthetic heart valve
- Pre-existent cognitive impairment
- Diabetes—time since diagnosis, degree of glycaemic control, presence of other complications
- Malignancies, systemic inflammatory disorders, immunosuppressive conditions, e.g., HIV
- Recent COVID (SARS-CoV-2)
- Recent surgery/epidural (contraindication to thrombolysis)

Medication and allergies

- Anticoagulants and time last taken—contraindication to thrombolysis
- Antiplatelets, antihypertensives
- Hypoglycaemic agents
- Oral contraceptive or hormone replacement therapy (thrombosis risk)

Family history

- Cardiovascular disease
- Young onset dementia and small vessel disease (genetic disease)
- Young onset systemic disease including renal, eye, and hearing disease (genetic link)

Activities of daily living

- Modified Rankin Scale for measuring dependence

Lifestyle factors

- Smoking, alcohol or illicit drug consumption and obesity

Social history

- Driving
- Employment
- Type of dwelling (e.g., bungalow, care home)
- Stress or bereavement
- Family support and social network

Address concerns and expectations

Focused history for hyperacute stroke management (aim <10 minutes as time is brain)

TOAST classification

The TOAST trial ('**T**rial of **O**RG10172 in **A**cute **S**troke **T**reatment') was one of the first multicentre trials of acute stroke management carried out in the USA, and the classification system used in the trial (the 'TOAST classification') is the most widely used aetiological classification system for ischaemic stroke. The different subtypes of stroke are summarized in Box 29.2 below.

Large artery atherosclerosis

Atherosclerosis is the most common process contributing to stroke with increasing age, with large artery atherosclerosis associated with nearly 20% of ischaemic stroke. It involves a chronic inflammatory process with platelet activation and aggregation after atherosclerotic plaque disruption leading to thrombus formation, with distal embolization to cerebral vessels causing an acute thromboembolic stroke or TIA. Less commonly, gradual atherothrombotic vessel occlusion can occur leading to cerebral hypoperfusion and haemodynamic/hypoperfusion-related stroke. Alternatively, if this process occurs in the presence of adequate collateral circulation, patients may remain asymptomatic despite almost-complete vessel occlusion.

Arterial branch points, such as the origin of the great vessels arising from the aorta, the proximal internal carotid artery, and its distal intracranial branches, are particularly affected (Fig. 29.2). Populations from the Black or Asian ethnic groups tend to have more hypertension, diabetes, and a higher incidence of intracranial atherosclerosis compared with White populations (where extracranial atherosclerotic disease is more common). Carotid extracranial atherosclerosis is correlated with coronary artery and peripheral vascular atherosclerosis.

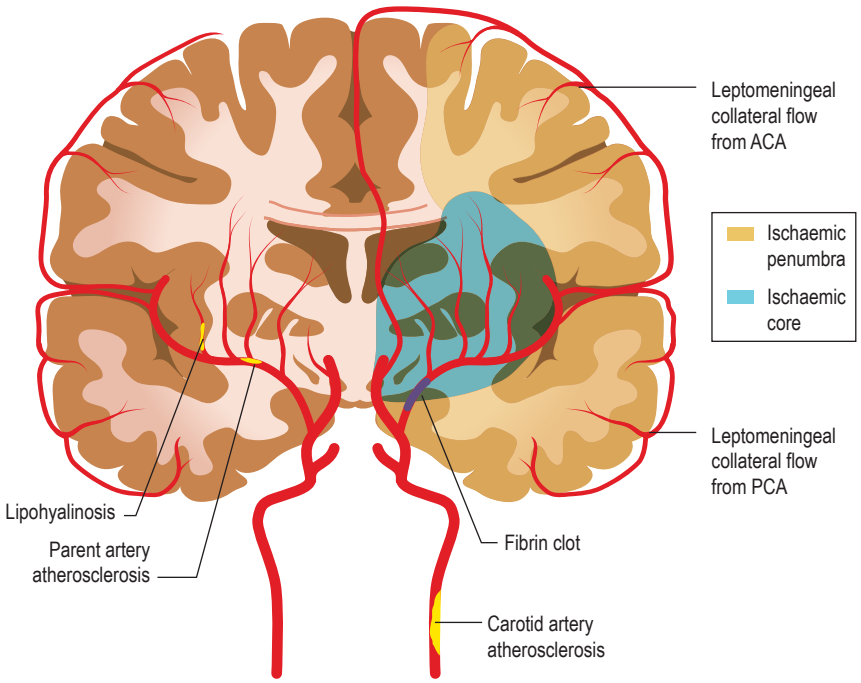


Fig. 29.1 **Collateral circulation.**

Box 29.2 Aetiology outlined based on TOAST classification

Large artery atherosclerosis
Carotid atherosclerotic disease >50% Intracranial atherosclerosis Aortic arch atheroma
Cardioembolism
High risk, e.g., atrial fibrillation or mechanical heart valve Medium risk, e.g., patent foramen ovale, bioprosthetic (non-mechanical) heart valve
Small vessel occlusion (lacune)
Other determined causes
Nonatherosclerotic vasculopathies (aortic or carotid dissection, CNS vasculitis primary or secondary such as giant cell vasculitis, infectious vasculopathy such as varicella zoster virus, illicit drug abuse, fibromuscular dysplasia, dolichoectasia) Hypercoagulable disorders (malignancy, antiphospholipid syndrome, thrombophilia, systemic lupus erythematosus, disseminated intravascular coagulopathy, hyperhomocystinaemia) Haematological disorders (essential thrombocytosis, polycythaemia rubra vera, sickle cell) Infections (HIV, neurosyphilis) Genetic causes (cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy CADASIL, mitochondrial diseases, Fabry's disease) Sympathomimetic drugs (e.g., cocaine)
Undetermined
Two or more causes Negative evaluation Incomplete evaluation

The principal risk factors for atherosclerosis include older age, male sex, smoking, hypertension, hyperlipidaemia, diabetes, obesity, physical inactivity, a diet high in saturated fat, alcohol excess, chronic inflammatory disease, metabolic syndrome, and genetic/ethnic factors. (See Fig. 29.14, Risk factor attributable burden of stroke in the 'Secondary prevention' section.)

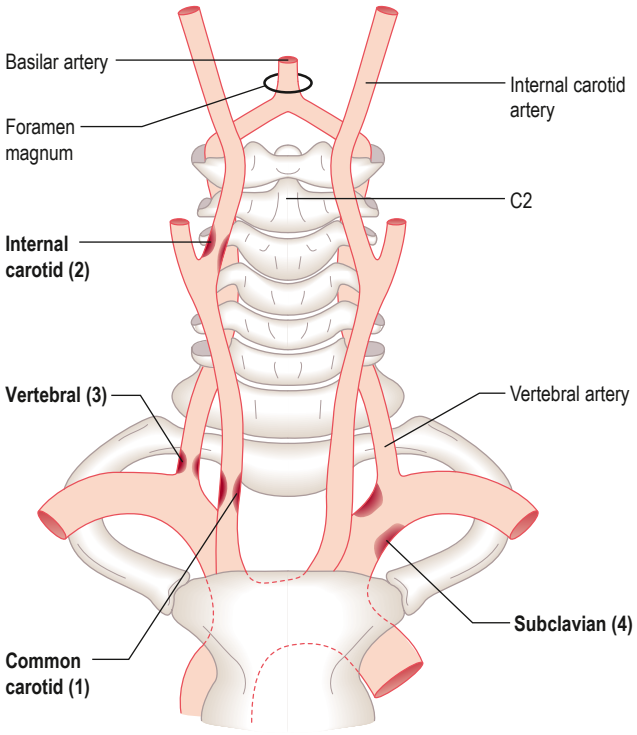


Fig. 29.2 **Principal sites of stenoses in extracerebral arteries.**

Cardioembolism

The heart is a common source of brain emboli, which travel from the heart to intracranial vessels causing occlusion and ischaemia (Fig. 29.3). Cardioembolism is responsible for a quarter of all ischaemic strokes. Infarcts in multiple vascular territories are very suggestive of a proximal source of emboli in the heart or aorta. Other features of cardioembolic stroke include wedge infarcts involving cortical and

subcortical tissue. Cardioembolic strokes often cause severe disability (Fig. 29.4).

Atrial fibrillation is the most common source with the left atrial appendage being the main source of thrombus. Other high-risk causes of cardioembolism include valvular heart disease, including rheumatic mitral valve stenosis and infective endocarditis. Mural thrombosis may occur in dilated cardiomyopathy or associated with akinetic left ventricle segments, although risk has reduced with coronary reperfusion therapy after myocardial infarction. Similarly, advances in therapy have led to a reduction in the thromboembolism risk associated with prosthetic heart valves (careful anticoagulation) and aortic arch atheroma (effective secondary prevention in high-risk patients).

Lower-risk causes include patent foramen ovale (PFO), a persistent remnant of the foramen ovale in the intra-atrial septum that is found in a quarter of the general population (see Chapter 17). While the majority of PFOs are asymptomatic, paradoxical embolism can occur via right-to-left shunting between the venous and arterial system. Likewise, pulmonary arteriovenous fistulas may also act as a conduit for paradoxical embolization. Rarer causes include atrial myxomas and iatrogenic causes, such as cardiac bypass and air embolism.

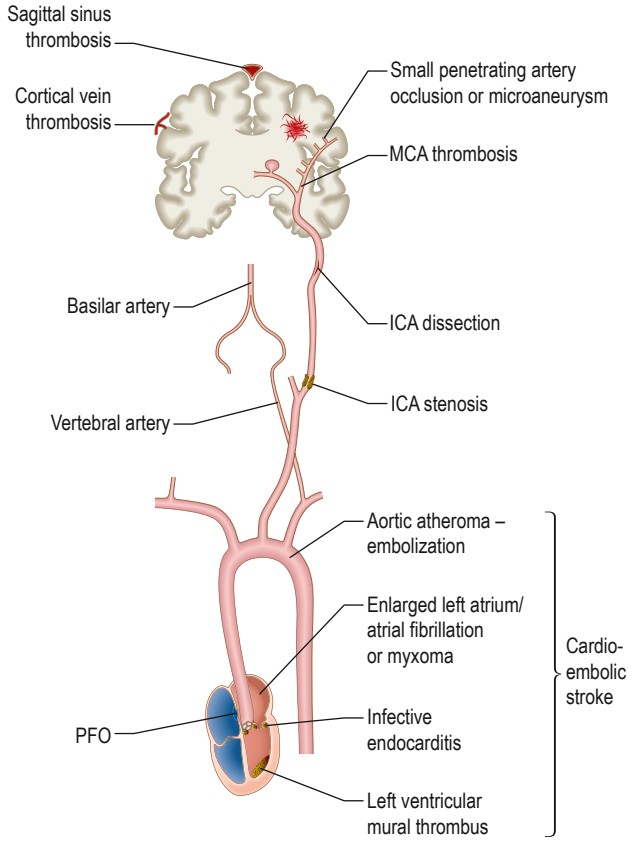


Fig. 29.3 Pathophysiology of ischaemic stroke.

Small vessel disease

Small vessel disease is the most prevalent brain disease, contributing to a quarter of all ischaemic strokes, 45% of dementia, and 80% of intracerebral haemorrhage (ICH) cases. Small penetrating arterial branches supply the deep brain parenchyma and are affected by a different pathological process: a diffuse vasculopathy—traditionally identified as lipohyalinosis but now recognized to include endothelial dysfunction and subtle increases in blood–brain barrier permeability. Other putative mechanisms include hypertensive arteriosclerosis, branch atheromatous disease, microembolism or cerebral hypoperfusion. The hallmark of this disease process is formation of lacunes (2–15 mm diameter fluid-filled areas of infarcted brain tissue in the deep cerebral white matter, basal ganglia, or pons). Neuroimaging may reveal gradual accumulation of white matter hyperintensities, perivascular spaces and cerebral microbleeds. Clinically, there may be stepwise deterioration associated with lacunar strokes, gradual cognitive impairment and gait impairment, or the disease may remain silent for many years. The related phenomenon of vascular dementia is discussed in more detail in Chapter 28. The most common risk factors include hypertension and diabetes. Smoking and excess alcohol contribute to a lesser extent. While lacunar infarcts are traditionally considered non-embolic, 10% to 15% may be embolic, and so the aetiological work-up must incorporate a search for common cardioembolic sources such as atrial fibrillation. More recently, cerebral amyloid angiopathy (CAA) is a recognized cause of small vessel disease with radiological features extending from

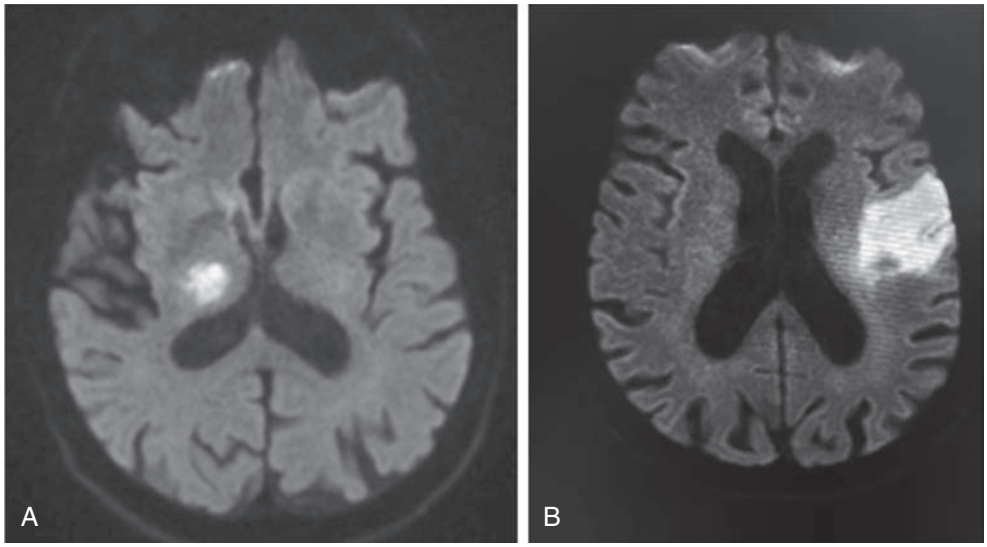


Fig. 29.4 Lacunar and embolic stroke pattern on neuroimaging. The left hand image **(A)** is the lacunar infarct in the right basal ganglia; the right hand image **(B)** is the embolic infarct of the left middle cerebral artery.

multispot white matter hyperintensities to centrum semi-ovale perivascular spaces as well as the more frequently recognized haemorrhagic manifestations (see 'Intracerebral haemorrhage' section).

Stroke of other determined aetiology

Carotid and vertebral artery dissection

Cervical artery dissection can occur at any age, but accounts for 5% to 25% of stroke cases in younger people, often with few established stroke risk factors. Dissection results from separation of arterial wall layers with an intramural haematoma, and this can lead to arterial stenosis, occlusion, or aneurysm formation. Thrombosis within the true lumen due to tissue factor release and platelet activation leads to embolization of the brain from the site of dissection, sometimes 2 to 3 weeks after the initial event.

It may be a sequelae of neck trauma (often trivial) or hyperextension; for example, after whiplash, osteopathic manipulation, hair washing in a salon, or exercise. Vascular abnormality (e.g., fibromuscular dysplasia) and collagen diseases (e.g., Marfan's) are well known to predispose to dissection.

Most dissections are in large extracranial neck vessels, with carotid stenosis typically occurring 2 to 3 cm beyond the carotid bulb, producing a tapering or flame-shaped stenosis and extending to the petrous canal on vascular imaging. Vertebral artery dissection typically occurs distally near the skull base and may extend intracranially. Pain in the neck or face is often the clue leading to diagnosis. Horner's syndrome may occur with carotid dissection, as the sympathetic plexus is intimately associated with the carotid artery in the neck. Neurological deficits vary or may be absent.

Venous stroke

Thrombosis within intracranial venous sinuses, such as the superior sagittal sinus, or in cortical veins is a rare form of stroke (see Chapter 11). This may occur in association with hypercoagulable states including pregnancy, thrombotic disorders, malignancy, obesity, oral contraceptive use, dehydration, and infections (COVID-19). Venous sinus thrombosis should be suspected in a stroke which does not follow the usual arterial territories. The most frequent presentations are headaches, seizures, focal neurological deficits, and encephalopathy.

Hypoperfusion

Systemic hypoperfusion states like reduced cardiac output and/or profound hypoxaemia may lead to border-zone infarction in the watershed areas between vascular territories, particularly if there is severe stenosis of proximal carotid vessels. It can affect either the superficial (fronto-parietal or parieto-occipital areas) or deep watershed territories—the latter causing 'rosary bead' appearance in neuroimaging.

There are many other stroke aetiologies, some requiring specific management for example infectious or inflammatory disease (see **Box 29.2**).

Stroke of undetermined aetiology

Approximately one-quarter of all strokes are cryptogenic without an identified source (not attributed to cardioembolism, small vessel disease, or large artery atherosclerosis). These can be further divided based on neuroimaging features into lacunar or embolic stroke patterns (**Fig. 29.4**). Embolic source of undetermined source (ESUS) may occur because of undiagnosed atrial fibrillation and complex aortic plaque among other causes.

STROKE CLASSIFICATION BY VASCULAR TERRITORY

Vascular anatomy

The circle of Willis is supplied by the two internal carotid arteries (the anterior circulation) and by the vertebrobasilar (posterior) circulation. The distribution of the anterior, middle, and posterior cerebral arteries that supply the cerebrum is shown in **Figs. 29.5 and 29.6**. Occasionally, posterior cerebral arteries arise from internal carotid arteries rather than basilar system—these are termed fetal posterior cerebral arteries. This is of relevance as an occipital infarction may arise from internal carotid artery atherosclerosis.

Knowledge of normal arterial anatomy and likely sites of atheromatous plaques and stenoses helps understanding of the main stroke syndromes. **Box 29.3** outlines the major vascular territories and their corresponding stroke syndromes. Principle motor and sensory pathways are shown in **Figs. 28.11 and 28.15**. Brainstem tracts are shown in **Fig. 29.7**.

Anterior circulation clinical stroke syndromes

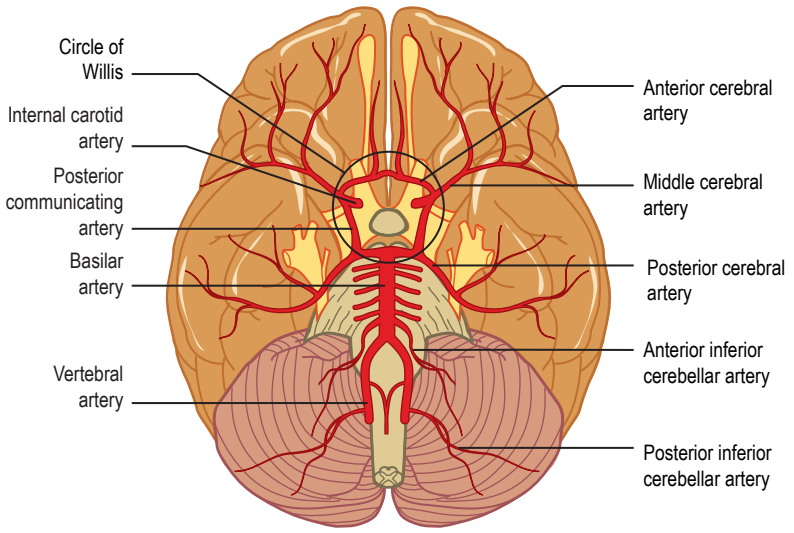
This includes infarcts in the territory of the internal carotid, middle cerebral (MCA), anterior cerebral (ACA), and ophthalmic arteries. Complete MCA occlusion results in a devastating stroke as multiple cortical areas are infarcted, with contralateral hemiplegia and facial weakness (primary motor cortex), hemisensory loss (primary sensory cortex), neglect syndromes (parietal lobe), eye deviation towards the affected side (frontal eye fields), aphasia (dominant hemisphere lesions) and hemianopia (optic radiation). Brain swelling of infarcted tissue leads to a high risk of death due to brain herniation—so-called 'malignant MCA infarction'. A similar picture to MCA stroke is caused by internal carotid occlusion, although collateral circulation may reduce the infarct size (see Fig. 26.31).

MCA branch occlusions produce fragments of the picture described above, such as hemiparesis, monoparesis, or aphasia. Occlusion of lenticulostriate perforating arteries (or MCA occlusion with collateral circulation protecting the cortex) causes infarction of deep subcortical structures such as the internal capsule, resulting in hemiplegia and hemisensory deficits. ACA infarcts are significantly less common than MCA infarcts and typically produce weakness affecting the leg more than arm, and frontal lobe deficits such as apathy or apraxia (see **Box 29.3** and Chapter 28).

Posterior circulation clinical stroke syndromes

Posterior cerebral artery (PCA) infarcts are typically embolic. Brainstem infarction causes complex signs, depending on the relationship of the infarct to cranial nerve nuclei, long tracts, and brainstem connections (Boxes 29.3 and 29.4; **Fig. 29.7**):

- The lateral medullary syndrome (Wallenberg's syndrome) is a common brainstem vascular syndrome presenting as acute vertigo with cerebellar and other signs, including Horner's syndrome (**Box 29.4**) (**Fig. 29.7**). It follows thromboembolism in the posterior



The circle of Willis is comprised of the following vessels:
- Anterior cerebral artery
- Middle cerebral artery
- Anterior communicating artery

→ Anterior circulation

- Posterior cerebral artery
- Vertebral artery
- Basilar artery

→ Posterior circulation

Anterior circulation - areas supplied by branches of internal carotid artery (anterior cerebral artery and middle cerebral artery)
Posterior circulation - areas supplied by branches of vertebral and basilar arteries (anterior inferior cerebellar artery, posterior inferior cerebellar artery, posterior cerebral artery).
Posterior communicating artery usually does not give rise to branches - it completes the circle of Willis

Fig. 29.5 Arteries supplying the brain and cord.

inferior cerebellar artery (PICA) or its branches, vertebral artery thromboembolism or dissection (see **Box 29.3**).

- Cerebellar infarcts occur in isolation or as part of a more extensive brainstem syndrome. Swelling of the cerebellum may cause brainstem compression and coma or obstructive hydrocephalus necessitating decompressive surgery.
- Basilar artery thrombosis is more common than embolism. The clinical picture depends on the level of the occlusion and the branch vessels affected. High lesions cause midbrain infarction, including coma and locked-in syndrome (see Chapter 28) or top of the basilar syndrome, with the PCA also affected, causing devastating midbrain, occipital lobe, and thalamic infarction (**Box 29.3**).
  - Contralateral homonymous hemianopia results from unilateral lesions, while bilateral lesions cause cortical blindness (Anton's syndrome) (**Box 29.3**). Neglect syndromes and visual agnosias are associated with involvement of more anterior visual association areas.
  - The PCA supplies the thalamus, and posteromedial temporal lobe and infarction of these structures causes confusion or memory impairment.

TRANSIENT ISCHAEMIC ATTACK

At least 15% of ischaemic strokes are preceded by a TIA. Early diagnosis and treatment of a TIA is imperative to prevent a stroke from occurring.

Clinical features

TIAs cause sudden loss of function (negative symptoms), usually lasting for minutes only, with complete recovery and no evidence of infarction on imaging (tissue-based definition). The symptom onset is abrupt and is often at maximal intensity at the time of onset. Focal neurological symptoms might include hemiparesis or hemisensory loss; speech disturbance (e.g., dysarthria or dysphasia); visual impairment (e.g., amaurosis fugax, diplopia, or hemianopia); and vertigo. See **Box 29.1**.

Amaurosis fugax

This is a sudden transient loss of vision in one eye. This is due to the passage of emboli through the retinal arteries, and the embolus is sometimes visible through an ophthalmoscope during an attack (Hollenhorst plaque). A TIA causing an episode of amaurosis fugax may be the first clinical evidence of internal carotid artery (ICA) stenosis—a warning sign of incipient ICA territory stroke. Carotid duplex or CT/MRI imaging of the carotid vessels should be part of the assessment. Evidence of a carotid stenosis ≥50% should prompt discussion with the vascular surgery team about consideration of carotid surgery (see 'Secondary prevention' section). Auscultation of the carotid artery (for detection of bruit suggesting carotid stenosis) is not recommended as it lacks both sensitivity and specificity.

Giant cell arteritis should also be considered as a potential cause of amaurosis fugax, and assessment should include asking about unilateral headache, scalp tenderness, and jaw claudication; and CRP/ESR must be checked. Suspicion of this serious condition should prompt referral for rheumatology review, consideration of a temporal artery biopsy, and early commencement of steroids.

TIA diagnosis

Diagnosis of TIA is often based solely on its description, and good history taking is therefore key. Collateral history from a bystander can be extremely helpful if they had witnessed the event. Transient loss of consciousness points towards an alternate diagnosis (not a TIA). MRI brain imaging is recommended, looking for evidence of acute infarction.

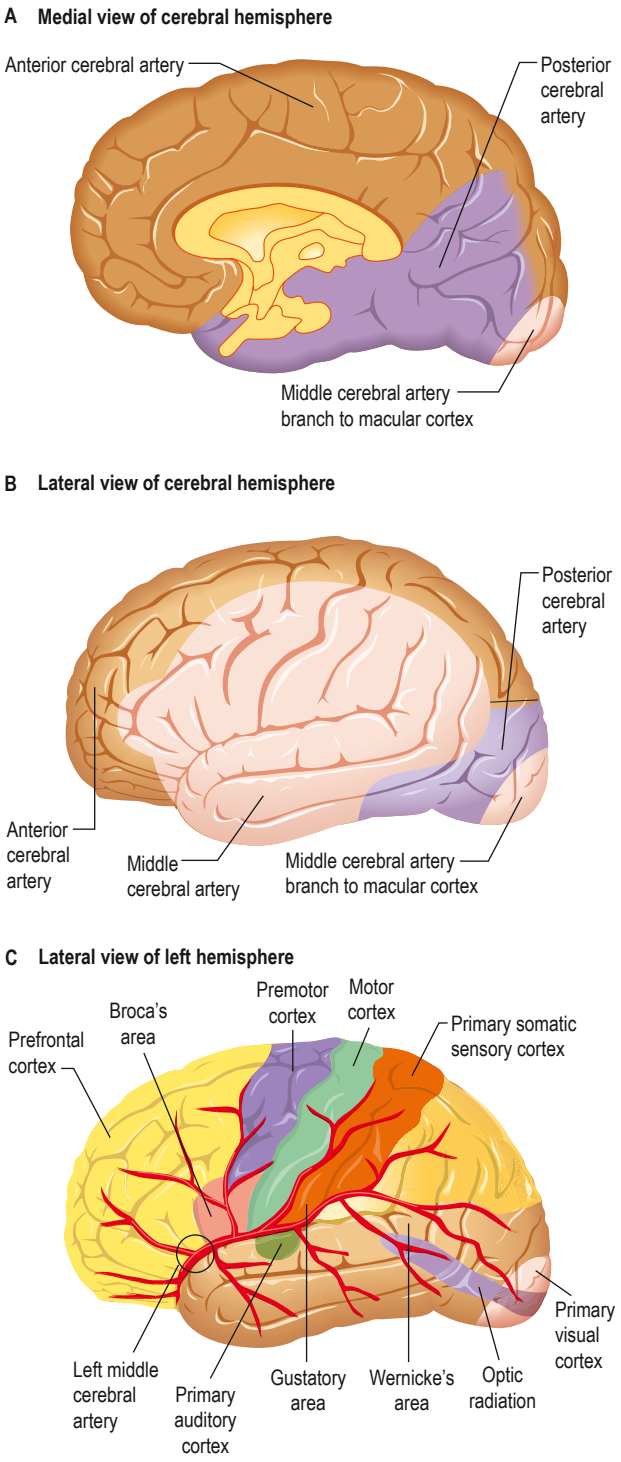


Fig. 29.6 The territories supplied by the three major cerebral arteries (A and B). (C) Cortical areas supplied by left middle cerebral artery branches (the dominant hemisphere in 95% of right-handed individuals.)

Differential diagnosis and stroke mimics

Approximately 50% of presentations to a TIA clinic are diagnosed as TIA mimics.

Migraine is the commonest mimic presenting to TIA clinics. Migraine often produces positive symptoms (e.g., headache, paraesthesia, visual flashes/fortification), but TIA often produces negative symptoms (weakness, numbness, loss of vision, etc.). The onset and evolution of symptoms is usually slower with migraine aura than TIA (minutes rather than seconds). Limb weakness is rarely due to migraine. It is important to remember that migraine is a diagnosis of exclusion.

Focal epilepsy is usually recognized by its positive features (e.g., limb jerking) and progression over minutes. In a TIA, involuntary limb movements can occur occasionally—***limb-shaking TIA***—and are pathognomonic of severe carotid stenosis causing transient focal cerebral hypoperfusion.

Other causes of TIA mimics are listed in **Box 29.5**. Careful history, examination, and focused investigations are essential to distinguish between these and a genuine TIA.

Prognosis

TIAs are associated with a cumulative early risk of stroke of 5.2% at 7 days. The risk is the highest in the first 24 hours after TIA. All patients with a suspected TIA event should be referred to a TIA clinic, with assessment, investigation, and management as an emergency, within 24 hours of symptoms. Historical scores such as the ABCD^2 score are no longer recommended for risk stratification. Early treatment of TIA reduces the risk of stroke by 80%, and includes early initiation of antiplatelet medication, statins, and blood pressure management (see 'Secondary prevention' section).

ACUTE ISCHAEMIC STROKE

Acute ischaemic stroke is a medical emergency. Hyperacute treatments for this condition are time sensitive, with significant advances in management options available over the last three decades.

Initial assessment should start with an ABCD approach in the emergency department. A rapid history and examination when assessing for hyperacute stroke therapy focuses on the following:

- time of onset (if known, or alternatively 'last known well' time)
- the clinical syndrome
- medications including anticoagulants
- past medical history including previous stroke or TIA
- vascular risk factors.

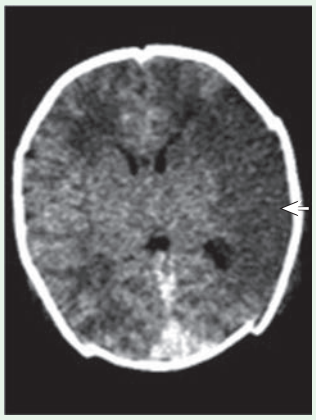
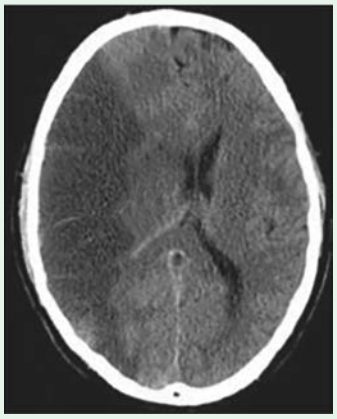
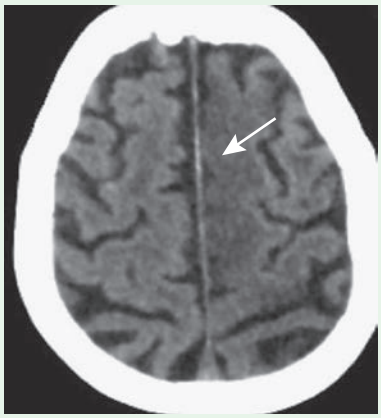
This is followed by a focused examination using the National Institute of Health Stroke Scale (NIHSS) (see **Box 29.6**) and establishing the pre-stroke level of function. The assessment should include focus on indications and contraindications for reperfusion therapies such as thrombolysis and/or thrombectomy. During this rapid assessment, bloods should be taken in parallel, and an intravenous cannula inserted (preferably avoiding the affected limb).

Once the initial assessment is complete (aim <10 mins), the patient undergoes rapid brain imaging. The choice of imaging modality will depend on the patient's presentation, time of onset, and eligibility for reperfusion therapies.

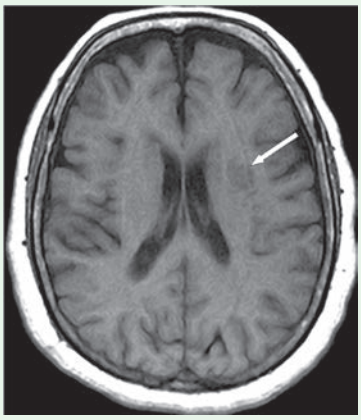
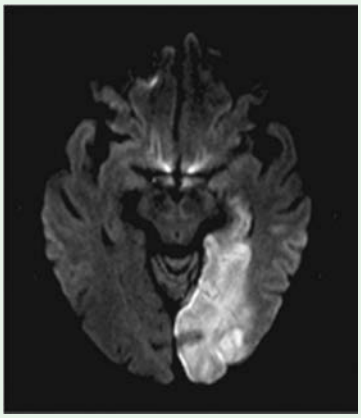
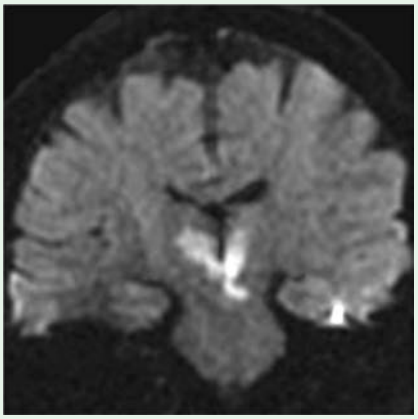
Imaging

The first aim of imaging with acute ischaemic stroke is to exclude haemorrhage. This usually involves first-line non-contrast CT head

Box 29.3 Major vascular territories with stroke syndromes

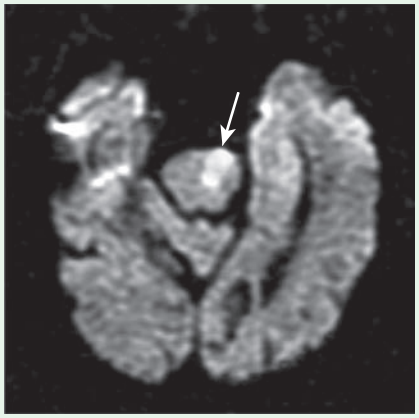
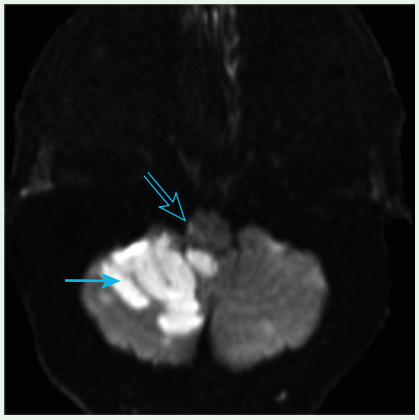
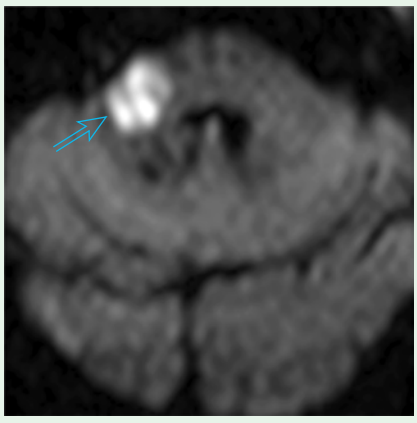
Some stroke clinical syndromes and associated vascular territories		
Vascular supply/region affected	**Neurological deficits**	**Diagram**
Anterior circulation		
Left MCA (dominant hemisphere)	Right-sided weakness involving face and arm > leg Right sensory loss Dysphasia Right visual field defect Gaze palsy (eye looks towards the affected hemisphere)	
Right MCA (non-dominant)	Contralateral left-sided weakness involving face and arm > leg Visual and/or sensory neglect Left sensory loss Denial of disability Left visual field defect	
ACA (medial and superior parts of the frontal lobe, and anterior parietal lobe)	Contralateral motor weakness—leg predominant (medial motor cortex leg area) Incontinence Abulia Amnesia (cingulate cortex)	

Box 29.3 Major vascular territories with stroke syndromes—cont'd

Lateral lenticulostriate branches of MCA (internal capsule, basal ganglia except caudate)	Contralateral motor, sensory or sensorimotor loss face = arm = leg Severe dysarthria (corticobulbar fibres) Lack of cortical deficits (e.g., aphasia)	
Posterior circulation		
(refer to Fig. 29.7 for brainstem tracts and nuclei)		
PCA (occipital cortex, medial temporal lobe, thalamus, midbrain)	Homonymous hemianopia (occipital) Cortical blindness (Anton's syndrome) if both occipital lobes are affected May be macular sparing of vision (MCA branches) Visual neglect (non-dominant occipital) Memory impairment, hemisensory loss or paraesthesia, transcortical aphasia if dominant, behavioural change (thalamus) Contralateral hemiparesis and ipsilateral III nerve palsy (Weber syndrome midbrain)	
Artery of Percheron or branches of PCA (paramedian thalamus)	Coma-like state Cognitive impairment Thalamic pain/dysaesthesiae Aphasia (transcortical sensory)	

Continued

Box 29.3 Major vascular territories with stroke syndromes—cont'd

Basilar artery and pontine perforator arteries (pons) **See Fig. 29.7 for midbrain nuclei and tracts**	Crossed motor or sensory findings (e.g., one side of the face and the opposite side of the body) Contralateral hemiparesis or tetraparesis (corticospinal tracts) Ipsilateral complete facial weakness and dysarthria (VII nucleus lower motor neuron) Contralateral hemisensory loss (spinothalamic tract) Ipsilateral facial sensory loss (V CN palsy) Contralateral ataxia (cerebellar peduncle) therefore contralateral ataxic hemiparesis Diplopia due to contralateral internuclear ophthalmoplegia (medial longitudinal fasciculus) Gaze paresis (away from the lesion)	
Posterior inferior cerebellar artery or vertebral artery (lateral medullary syndrome)	Vertigo, nausea, vomiting, nystagmus, truncal ipsipulsion (vestibular nucleus) Ipsilateral limb ataxia and incoordination (inferior cerebellar peduncle) Crossed sensory signs Facial sensory loss (V nucleus and tract) Contralateral sensory loss of body (spinothalamic tract) Dysphagia, dysphonia, hiccup (nucleus ambiguous IX and X) Ipsilateral Horner's syndrome (descending sympathetic fibres)	
AICA (lateral inferior pons and cerebellum)	Ipsilateral nystagmus, vertigo, nausea, vomiting (VIII nucleus or nerve) Ipsilateral facial weakness (VII palsy) Ipsilateral deafness, tinnitus (labyrinthine artery and cochlear nerve) Ipsilateral ataxia and incoordination (cerebellar peduncle and cerebellar hemisphere) Crossed sensory signs Ipsilateral impaired facial sensation (V nucleus) Contralateral impaired pain and temperature on body (spinothalamic tract)	

Note: 95% of right-handed individuals have left-sided language dominance. Approximately 50% of left-handed individuals have left-handed language dominance.

(NCCT), although MRI head is also acceptable (but less accessible for many centres). The second aim of imaging in acute ischaemic stroke is to review the extent of ischaemic change if present. This can be done using a scoring system such as the ASPECTS (Alberta Stroke Programme Early CT Score) (**Fig. 29.8**) derived from the CT or MRI images. Typically, reperfusion therapies may not be offered if the extent of the ischemic change is already too extensive (e.g., very low ASPECTS score).

If the patient is a candidate for mechanical thrombectomy (MT), additional CT angiogram imaging (CTA) should be completed, looking for a thrombus occluding a large intracranial vessel (middle cerebral artery, internal carotid artery, or basilar artery). Some centres complete NCCT and CTA in all stroke admissions to assess all these factors in one sitting. This may be preferable, as it saves time during the initial assessment phase of management.

Newer techniques such as CT or MR perfusion (CTP or MRP), allow assessment of the area of ischaemia in more detail; for delineation of the ischaemic core and penumbra (**Fig. 29.9**).

The ischemic core refers to the part of the brain in acute ischaemic stroke with severely reduced blood flow, that has already infarcted and thus is thought to have irreversible neuronal damage. The surrounding brain tissue served by collateral vessels may have only moderately reduced perfusion despite a large vessel occlusion. This area is known as the penumbra where tissue damage is still

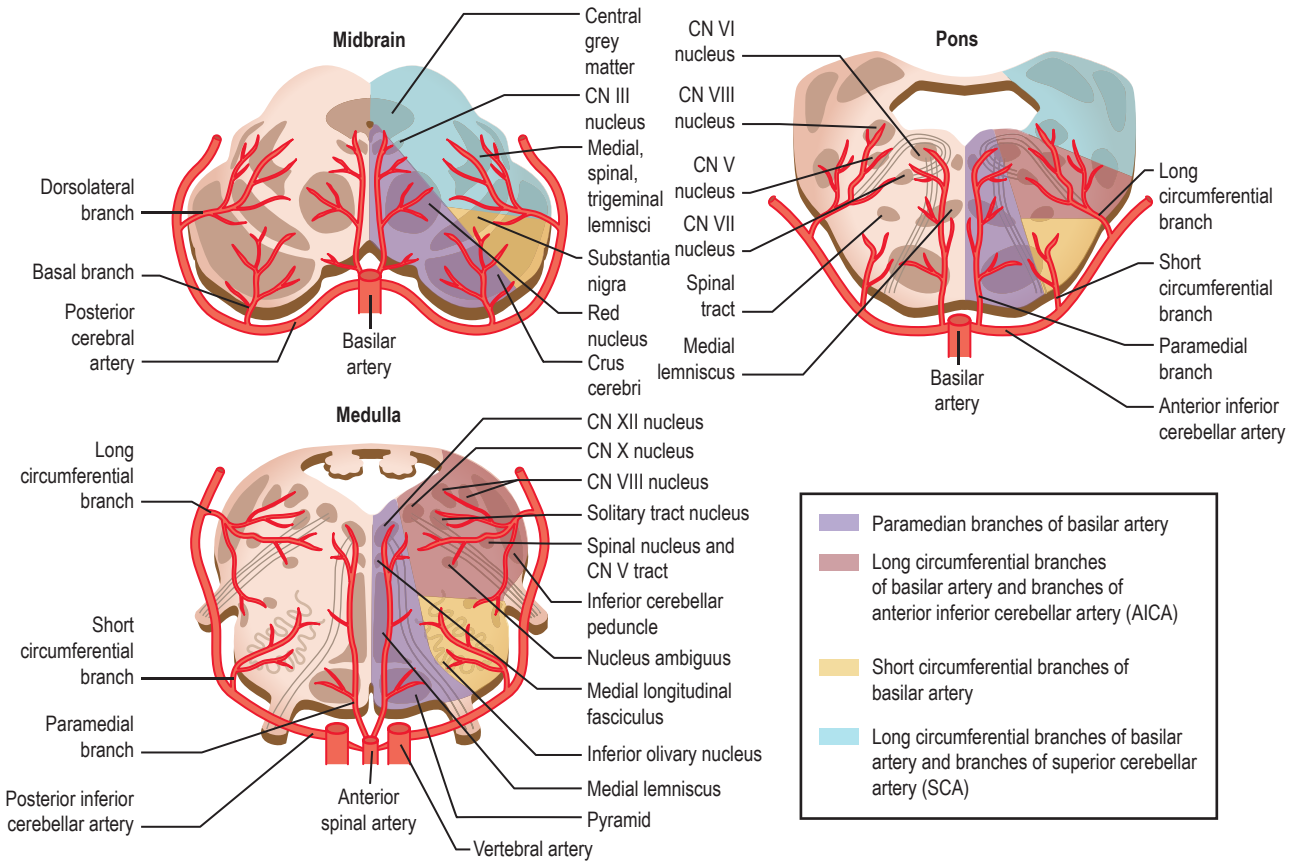


Fig. 29.7 Brainstem anatomy—tracts and nuclei. *(https://n.neurology.org/content/97/20_Supplement_2/S158.)*

Box 29.4 Features of brainstem infarction

Clinical feature	Structure involved
Hemiparesis or tetraparesis	Corticospinal tracts
Sensory loss	Medial lemniscus and spinothalamic tracts
Diplopia	Oculomotor system
Facial numbness	Vth nerve nuclei
Facial weakness	VIIth nerve nucleus
Nystagmus, vertigo	Vestibular connections
Dysphagia, dysarthria	IXth and Xth nerve nuclei
Dysarthria, ataxia, hiccups, vomiting	Brainstem and cerebellar connections
Horner's syndrome	Sympathetic fibres
Coma, altered consciousness	Reticular formation

potentially reversible if the occlusion can be targeted with reperfusion therapies.

CT or MR perfusion is typically performed in cases where patients have presented in a late window (beyond 4.5 hours; but within 24 hours of onset or last known well). This allows assessment of 'salvageable penumbral tissue' and allows careful use of reperfusion therapies if a favourable profile is seen (equating to a small core and large penumbra) (**Box 29.7**).

Box 29.5 TIA and stroke mimics

- Migraine (commonest)
- Syncope
- Focal seizure or Todd's paresis (after a seizure)
- Vestibular dysfunction/BPPV
- Metabolic (e.g., hypoglycaemia)
- Transient global amnesia
- Peripheral nerve disorder
- Amyloid spells
- Space occupying lesion (primary or metastases)
- Bell's palsy
- Postural hypotension
- Cardiac arrhythmia
- Subdural haematoma
- Anxiety/functional neurological disorder
- Acute episode of demyelination
- Intracranial infection (e.g., cerebral abscess, meningoencephalitis)
- Spinal cord lesions
- Paraneoplastic disorder
- Degenerative brain disease (e.g., primary progressive aphasia, spinocerebellar degeneration)
- Trigeminal neuralgia
- Medication (side-effects)

Box 29.6 NIHSS

1 a—Level of consciousness	0—Alert, keenly responsive 1—Not alert, but arousable by minor stimulation 2—Not alert; requires repeated stimulation 3—Unresponsive or responds only with reflex
1 b—Level of consciousness questions	0—Answers two questions correctly 1—Answers one question correctly 2—Answers neither question correctly
1 c—Level of consciousness commands	0—Performs both tasks correctly 1—Performs one task correctly 2—Performs neither task correctly
2—Best gaze	0—Normal 1—Partial gaze palsy 2—Forced gaze deviation
3—Visual	0—No visual loss 1—Partial hemianopia 2—Complete hemianopia 3—Bilateral hemianopia
4—Facial palsy	0—Normal symmetric movements 1—Minor paralysis 2—Partial paralysis 3—Complete paralysis
5—Motor arm—Left arm	0—No drift 1—Drift 2—Some effort against gravity 3—No effort against gravity 4—No movement
Motor arm—right arm	0—No drift 1—Drift 2—Some effort against gravity 3—No effort against gravity 4—No movement
6—Motor leg—left leg	0—No drift 1—Drift 2—Some effort against gravity 3—No effort against gravity 4—No movement
Motor leg—right leg	0—No drift 1—Drift 2—Some effort against gravity 3—No effort against gravity 4—No movement
7—Limb ataxia	0—Absent 1—Present in one limb 2—Present in two limbs
8—Sensory	0—Normal 1—Mild to moderate sensory loss 2—Severe or total sensory loss
9—Best language	0—No aphasia 1—Mild to moderate aphasia 2—Severe aphasia 3—Mute, global aphasia
9—Dysarthria	0—Normal 1—Mild to moderate dysarthria 2—Severe dysarthria
Extinction	0—No abnormality 1—Visual/tactile/auditory/spatial or personal inattention 2—Profound hemi-inattention or extinction
Score	0—42

If any of the above cannot be tested (e.g., amputation or intubation), score 9 for that domain and exclude that from the NIHSS score.

Artificial imaging (AI) solutions have been successfully developed to calculate an automated ASPECTS score, while also being able to detect an occlusion in a large vessel. These AI solutions are also able to calculate core and penumbra maps from CTP imaging (Fig. 29.9). External validation of this software is ongoing.

Intravenous thrombolysis

Intravenous thrombolysis (IVT) involves administering fibrinolytic treatment (tissue plasminogen activator) to dissolve the thrombus or embolus, thereby restoring perfusion to an affected area of brain. IVT has been shown to increase the odds of an excellent outcome after acute ischaemic stroke. The benefits are greater the earlier the treatment is given (Fig. 29.10). Of note, there is an increase in the risk of intracranial haemorrhage with this treatment, but the absolute risk of haemorrhage is less than the benefit of treatment with IVT (see schematic). The contraindications for IVT are listed in Box 29.8; these should be reviewed prior to proceeding with treatment. When the patient is being considered for IVT, the aim is to deliver this treatment within 30 minutes of the patient arriving in the emergency department. Benefits of thrombolysis are shown in Fig. 29.11.

IV thrombolysis windows

Thrombolytic drug options include i.v. alteplase (i.v. bolus and 1-hour infusion), or i.v. tenecteplase (i.v. bolus only), with the choice of agent left to the discretion of the treating team.

IVT is licensed for up to 4.5 hours after the onset of stroke. However, stroke guidelines advise an extension of the time window owing to advances in imaging (Box 29.8). For example, in wake-up strokes that happen during sleep, one method involves the use of perfusion imaging with evidence of small core and large penumbra (with significant mismatch) up to 9 hours from symptom onset.

Hyperacute MRI has been used in wake-up strokes to look for mismatch between diffusion-weighted imaging (DWI) and fluid-attenuated inversion recovery (FLAIR) sequences. MRI DWI sequences will pick up the core infarct within minutes. On the other hand, typically the MRI FLAIR sequences will not clearly show the infarct until approximately 4.5 hours. Therefore, an infarct that is present on DWI but is absent on FLAIR can be assumed to be in its early stages (as a surrogate for knowing the exact time of onset); and such patients can be offered IVT, if no contraindications.

Mechanical thrombectomy

MT is a procedure that involves directly removing the clot causing an occlusion of a large vessel (ICA, MCA, or basilar), using an arterial catheter inserted through the groin (or wrist) that directly accesses the intracranial occlusion. The clot is then removed using a device that can either aspirate or trap the clot using a retrievable stent system (Fig. 29.12). This treatment has revolutionized management of patients with large vessel occlusion who were previously more likely to have a poor outcome. Landmark trials published in 2015 assessed the efficacy of MT, in addition to standard treatment with IVT, in patients with anterior circulation large vessel occlusion. These trials demonstrated a significant increase in the proportion of patients who were alive and independent at 3 months after stroke, over and above IVT alone. Recent

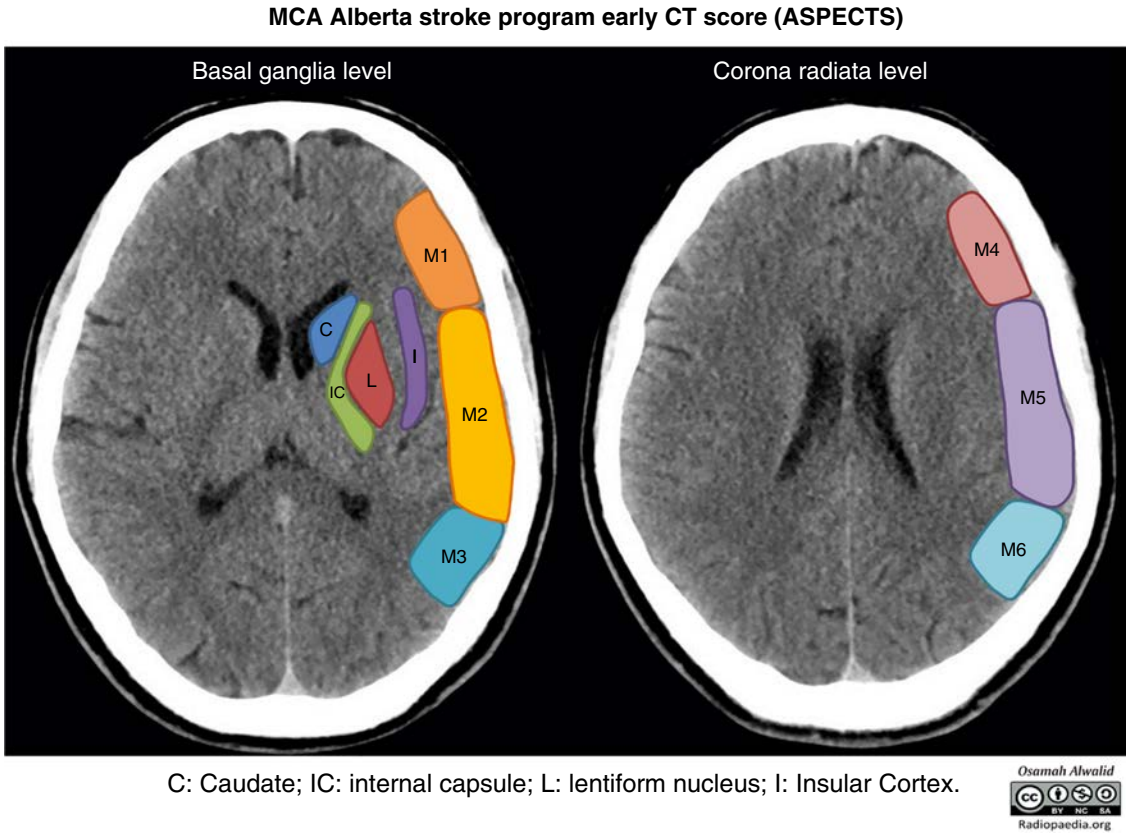


Fig. 29.8 Alberta Stroke Programme Early CT Score (ASPECTS). Calculation of the extent of early ischaemic changes using the ASPECTS score; 10 pre-specified (highlighted) areas are assessed, and 1 point subtracted for each area involved; with a maximum score of 10 (no ischaemia) and minimum score of 0 (extensive ischaemia). *(Credit: https://radiopaedia.org/articles/alberta-stroke-programme-early-ct-score-aspects?lang=gb.)*

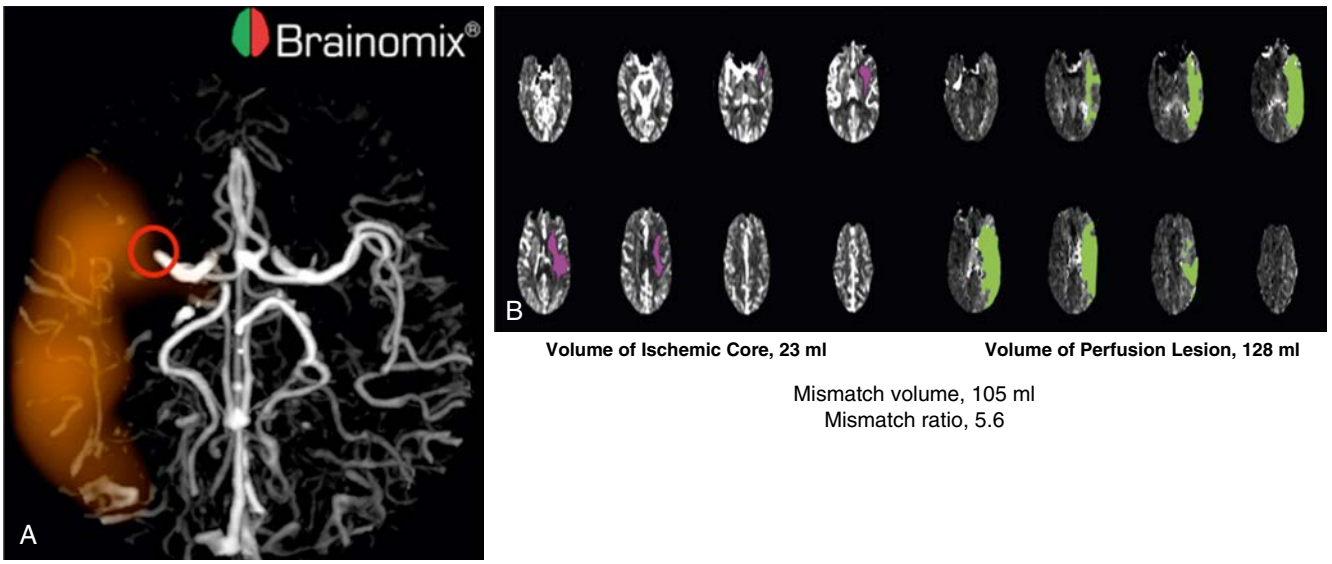


Fig. 29.9 Automated software: detection of large vessel occlusion and CT perfusion maps of core/penumbra. (A) Right middle cerebral artery large vessel occlusion using automated detection software. **(B)** Core-perfusion maps of left middle cerebral artery large vessel occlusion with small core 23 mL and large perfusion defect of 128 mL and mismatch volume of 105 mL (please note that these two scans are from different patients).

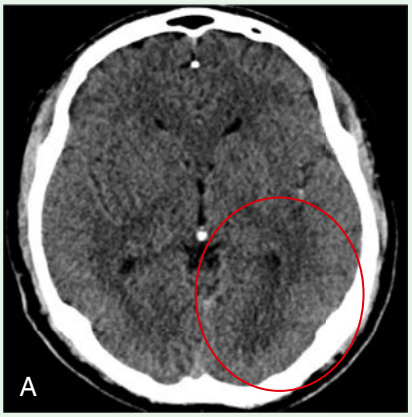
Box 29.7 Stroke imaging table

Brain parenchymal imaging
Assess for infarcted brain tissue, intracranial bleeding, tumour
Stroke is a clinical diagnosis and a normal non-contrast CT brain is common in the first few hours after acute stroke ***(see scan A below)***
Imaging may differentiate embolic and lacunar features especially MRI
(see over acute infarct on non-contrast CT and MRI diffusion weighted)
	• Non-contrast computed tomography (CT) brain ***(scan A below)***
	  • Consent needed for minimal risk radiation exposure in childbearing females
	• Magnetic resonance (MR) brain
	  • Principle imaging modality for TIA
	  • Issue with compatibility of pacemakers and metallic implants
	  • No exposure to radiation
	  • No known adverse effects to the fetus but use with caution

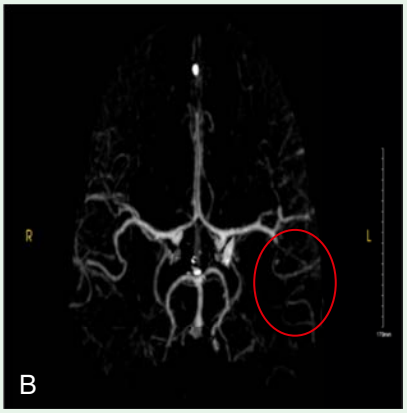
Vascular imaging—intracranial and extracranial
Assess for acute large or medium vessel occlusion or stenosis to determine aetiology such as atherosclerosis, dissection, thrombus, vasculitis, vasospasm, or moyamoya
(See over CT angiogram with left MCA large vessel occlusion)
	• CT angiography from aortic arch to brain ***(scan B below)***
	  • Caution with advanced kidney disease
	• MR angiography with or without contrast (gadolinium)
	  • Caution with advanced kidney disease (nephrogenic sclerosis)
	• Ultrasound Doppler to assess for carotid and vertebral disease
	  • No radiation exposure
	  • Limited access to upper neck near skull base
	  • User dependent

Perfusion imaging
To determine core to perfusion mismatch pre thrombectomy, thrombolysis, or for diagnostic purposes with mimics
	• CT or MR perfusion (requires injection of contrast, see ***scans C and D below***)
	• Automated software for CT perfusion
	  • Provides automated maps of perfusion estimating ischaemic core, penumbra, and mismatch
	  • Not externally validated

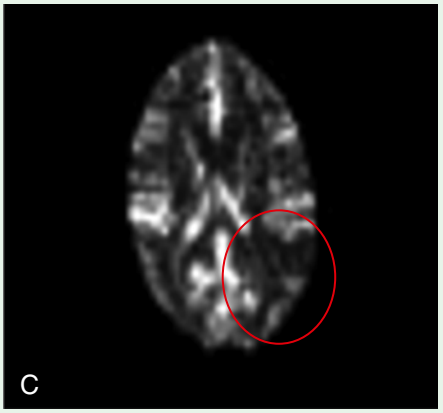
Contrast-enhanced imaging Assesses for tumour and vasculitis	• CT or MRI contrast-enhanced scan
Catheter angiogram or digital subtraction angiogram Completed during thrombectomy, after intracerebral bleed to assess for aneurysm or vascular malformation and the assess for cerebral vasculitis	• Catheter angiogram is associated with a 0–5 to 1% risk of major vascular and neurological complications as it involves direct catheterization of the large arteries of the neck and head



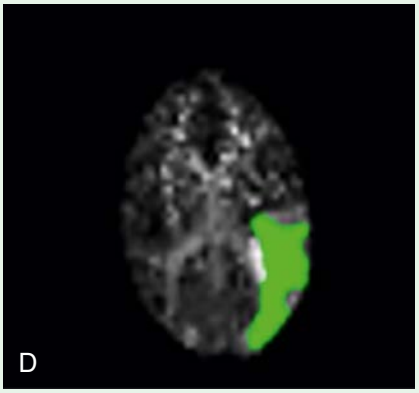
Non contrast CT scan showing a dense MCA in the Sylvian fissure (M2 segment)



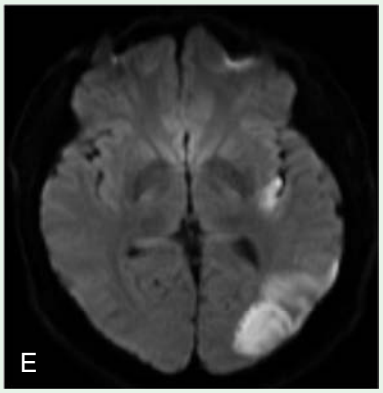
CT angiogram showing abrupt cut off of the M2 branch (arrow) with very poor visualisation of the vessel distal to the occlusion



CT Perfusion showing no area of core damage (i.e., no irreversible damage)



CT perfusion showing 'penumbra' (i.e., salvageable brain with prompt treatment e.g., i.v. thrombolysis +/- mechanical thrombectomy)



rt-PA was administered 2 hours after the onset of symptoms with recanalisation of M2. Diffusion weighted MRI done on day 6 showing left parieto-occipital lobe and left posterior insular cortex acute infarction, a much smaller area of infarction than the initial penumbra at risk

A and B, Reproduced from Ralston SH, Penman ID, Strachan MWJ et al. Davidson's Principles and Practice of Medicine, *23rd edn. Edinburgh: Elsevier; 2018. C–E Courtesy of Professor Adrian Dixon, Cambridge Radiology Department, UK. Adapted from Fig. 26.4 and Fig. 26.38.*

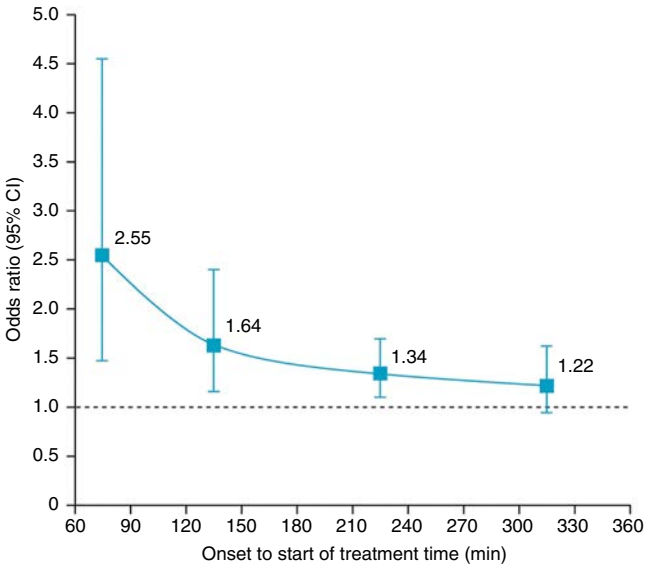


Fig. 29.10 Time is brain: Likelihood of an excellent outcome (no or minimum disability) after thrombolysis, depending on onset to treatment time.

evidence shows that this treatment can be delivered safely up to 24 hours after symptom onset in anterior and posterior circulation large vessel occlusion and is associated with excellent outcomes in selected patients. The number needed to treat (NNT) to reduce disability is 3, a number unmatched by most other medical interventions in medicine.

Most recent guidelines recommend that stroke patients with anterior circulation large vessel occlusion (ICA or MCA) on CTA and presenting with an NIHSS 6 or more, should be considered for MT. Beyond 12 hours, advanced imaging with CTP or MRI is recommended to demonstrate salvageable penumbra on perfusion imaging.

Posterior circulation MT (basilar) may be considered within 24 hours of onset for those with an NIHSS of 10 or more, with favourable imaging characteristics (e.g., good p-ASPECT—posterior circulation ASPECT).

MT is a game-changing treatment; however, resource and infrastructure deficits mean that availability is often limited. This results in significant inequity in access and a postcode lottery for patients with a large vessel occlusion causing stroke. This therefore remains a priority area for improvement for stroke services across the world.

STROKE UNIT CARE, REHABILITATION, AND RECOVERY

Acute stroke unit care

Acute stroke unit care is delivered in a geographically defined hospital-based unit, staffed by a specialized, multidisciplinary team including nurses, physicians, physiotherapists, occupational therapists, speech therapists, psychologists, dieticians, social workers, and others. Acute stroke unit care is associated with reduced mortality, disability, and institutionalization by approximately 25% compared to non-specialist hospital care. Early supported discharge to the community for those with mild to moderate disability is associated with improved outcomes, including reduced length of hospital stay and increased independence in daily activities compared to traditional in-hospital care (Box 29.9).

Post-stroke complications

A biopsychosocial approach to stroke care is essential. Assessment is routine for stroke complications (urinary tract infection, pneumonia, venous thromboembolism, pressure ulceration, and seizures) with targeted screening for delirium, cognitive impairment, depression, anxiety, and fatigue. Measures to reduce the risk of venous thromboembolism in immobile patients require further research but include early mobilization and intermittent pneumatic compression. Graduated compression stockings are not recommended. Prophylactic low-molecular-weight heparin reduces thromboembolism risk at the expense of increased symptomatic ICH.

Dysphagia

Dysphagia is associated with pneumonia and death. Early dysphagia screening (<4 hours) provides the opportunity for intervention including modified fluids, diet, and positioning. Enteral feeding for severe dysphagia is preferably within 1 to 2 days via nasogastric tube. Dysphagia often improves, as neurological pathways have bilateral cortical representation, so percutaneous gastrostomy is often deferred to the subacute phase (such as approximately 4 weeks).

Space occupying brain infarction/malignant hemispheric cerebral infarction

Occlusion of the internal carotid artery or proximal middle cerebral artery can lead to a large (malignant) hemispheric infarction with cerebral oedema. Raised intracranial pressure compresses leptomeningeal collateral circulation and leads to herniation of the temporal lobe onto the brainstem. It is more commonly seen in younger individuals and is associated with a high (>75%) mortality. Decompressive craniectomy involves removing part of the skull to allow the brain to swell (Fig. 29.13) and provided within 48 hours reduces the risk of death and poor outcomes, although most individuals who survive have substantial disability. Most benefit is seen in those under age 60 years.

Impairment of mobility

Gradual mobilization within 48 hours is recommended (short and frequent episodes) but very early intensive mobilization within 24 hours may be harmful especially for those with more severe strokes. Aerobic training is advised where feasible, in addition to resistance training and repetitive, task-specific rehabilitation activities such as repetitive walking practice. The multidisciplinary team should have expertise in multiple modalities which may provide benefits such as constraint-induced movement therapy, functional electrical stimulation, balance training, mirror therapy, mental practice, repetitive transcranial magnetic stimulation, vocational rehabilitation, and others.

Spasticity and pain

Botulinum injection will address severe spasticity, supported by dedicated physiotherapy techniques. These have better tolerance and efficacy than oral agents such as baclofen (gamma-aminobutyric acid B receptor agonist) or tizanidine (alpha-2 adrenergic receptor agonist), which are associated with sedation and abrupt withdrawal should be avoided. Neuropathic pain may be treated with gabapentin, pregabalin, amitriptyline, or duloxetine. Education on optimal shoulder position may reduce subluxation.

Box 29.8 Indications, contraindications, and dosing for thrombolytic therapy

Indications for TPA

Diagnosis of acute ischemic stroke causing new **disabling symptoms** to patient regardless of NIHSS score
Treat even if improving symptoms provided the deficit remains disabling
- Disabling symptoms definition
  - A deficit that, if unchanged, would prevent the patient from performing basic activities of daily living (i.e., bathing, ambulating, toileting, hygiene, and eating) or returning to work
- Non-disabling strokes are *excluded* unless there is large vessel occlusion
  - Isolated mild aphasia (patient still able to communicate meaningfully), isolated facial droop, mild cortical hand (especially non-dominant), mild hemi-motor loss or mild hemisensory loss, and mild hemiataxia (patient still able to ambulate)

Treatment can be started **within 4.5 hours of symptom onset**
OR
in the **Extended time window** (4.5–9 hours or wake-up strokes) **using advanced imaging** (CT perfusion/MRI)

Initial glucose levels >2.8 mmol/L
If glucose is ≤2.8 mmol/L, treat hypoglycaemia first then reassess

Absolute contraindications

Intracerebral haemorrhage on non-contrast CT brain

Severe hypertension (>185/110 mmHg) refractory to aggressive hyperacute antihypertensive treatment. Blood pressure must be <185/110 to commence TPA

Recent (<3 months) stroke, traumatic brain injury, intracerebral or intraspinal surgery

Known coagulopathy, bleeding disorder, severe cirrhosis, or platelet count <100 (Do not wait for a platelet count unless history is suggestive of a platelet disorder)

Anticoagulation treatment with warfarin with INR > 1.7

Direct oral anticoagulants (DOACs) within 48 hours for whom there is no specific coagulation test available
Longer time may be needed if abnormal renal function
- Specific coagulation tests include: calibrated anti–Xa-activity for factor Xa inhibitors, thrombin time for dabigatran, or the DOAC blood concentrations as available in your laboratory
- For patients with acute ischemic stroke of <4.5 hours who used dabigatran within the last 48 hours, reversal with Idarucizumab and subsequent thrombolysis maybe considered

Active internal bleeding

Recent (<10 days) cardiopulmonary resuscitation or obstetric delivery

Bacterial endocarditis or pericarditis

Symptoms suggestive of subarachnoid haemorrhage even if normal CT brain

Aortic dissection

Relative contraindications

Seizure at onset unless residual neurology is considered to be stroke related

Lumbar puncture within previous 7 days

Intracerebral neoplasm, arteriovenous malformation (AVM) or larger cerebral aneurysms

Blood glucose ≤2.8 mmol/L. Consider alteplase if deficits are not resolving with intravenous dextrose. Hyperglycaemia is not a contraindication

Recent major trauma (<14 days) with internal injuries or visceral biopsy

Recent (<14 days) gastrointestinal or genitourinary tract haemorrhage, or major surgery at a non-compressible site

History of intracranial haemorrhage. Consider if treatable cause (e.g., trauma, subarachnoid haemorrhage with aneurysm repair)

Acute pancreatitis

Recent (<7 days) puncture of a non-compressible blood vessel (e.g., subclavian vein)

Systemic malignancy with increased haemorrhage risk

Caution if recent (<3 months) myocardial infarction or pericarditis

Pregnancy

Caution is advised if there is a very large area of hypodensity on the non-contrast CT brain signifying an established recent infarct (>1/3 MCA)

Thrombolysis dosing

Dose of intravenous alteplase (tissue plasminogen activator)
- Total dose 0.9 mg/kg (max. 90 mg)
- 10% of the total dose by initial bolus over 1 minute
- Remainder is infused over 1 hour

Dose of intravenous tenecteplase
- The dose of tenecteplase is 0.25 mg/kg (maximum total dose 25 mg)
- Give in a single intravenous bolus over 5 seconds (saline flush as incompatible with dextrose)

Bowel and bladder dysfunction

Constipation is common and needs polypharmacy review (to stop aggravating medications), laxative treatment, and a structured bowel management programme. Urine incontinence, which occurs frequently, is addressed via behavioural interventions and adaptations such as bladder retraining prior to pharmacological intervention. Prolonged use of indwelling urinary catheters should be avoided.

Ongoing disability

Fatigue may be negotiated via self-management programmes. If there are ongoing swallow difficulties, comfort feeding or palliative care may be appropriate goals of care in certain circumstances. Significant social support may be needed on discharge. Reference must be made to national driving guidelines with onward referral of individuals with persistent disability to driving assessment and support services.

Goals of therapy

Rehabilitation can be delivered within the acute hospital setting, rehabilitation unit, and via early supported discharge facilitating transition to the community and integration of care. Goal-directed care with shared decision-making is central to recovery from stroke and directed by an individual patient's preferences and values. Rehabilitation provides activities to promote neuroplasticity, neural reorganization, and recovery, and facilitates compensation to overcome deficits in function. The International Classification of Functioning, Disability and Health (WHO 2001) provides a useful conceptual framework. This identifies stroke as a pathology causing impairment of body function, but furthermore recognizes limitations to activity and impact on participation in life providing targeted goals for rehabilitation. From the outset, communication and engagement with the patient, family, and carers is important to support effective, patient-centred rehabilitation and discharge planning.

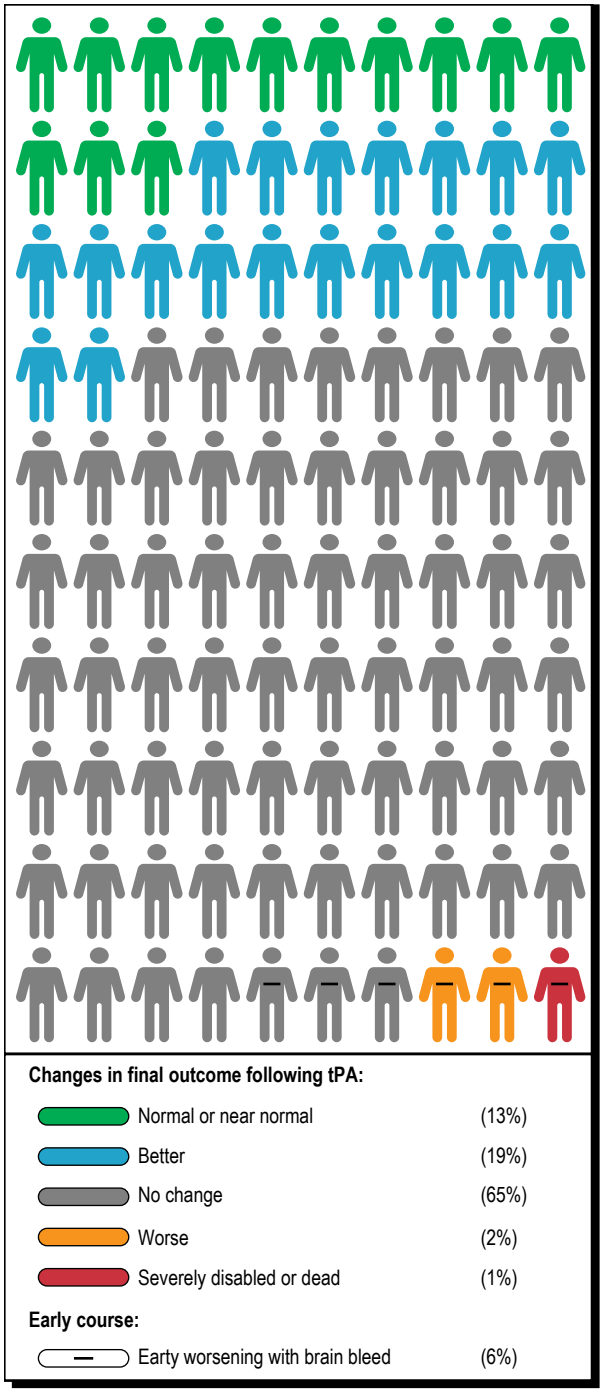


Fig. 29.11 Outcomes for thrombolysis therapy for acute ischaemic stroke within 3 hours of symptom onset. *(Credits: UCLA Stroke—centre figure.)*

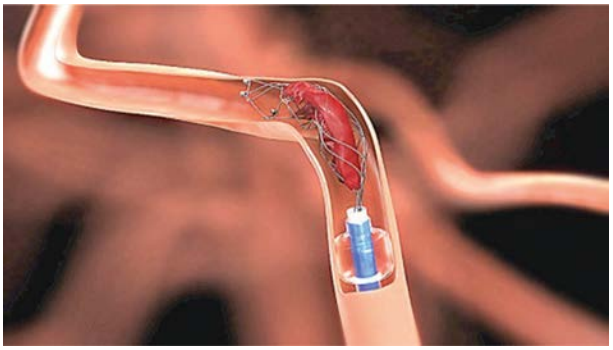


Fig. 29.12 Solitaire stentriever device for removing clot from intracranial vessel.

PRIMARY AND SECONDARY ISCHAEMIC STROKE PREVENTION

Secondary prevention of stroke

Ischaemic stroke is a complex and heterogeneous disease. A comprehensive approach to identify the underlying causal mechanism(s) is needed, to urgently inform targeted secondary stroke prevention. The highest risk of a further stroke is in the first 1 to 2 weeks after the index event, particularly for mechanisms such as high-grade carotid disease, intracranial stenosis, and atrial fibrillation. Atherosclerosis is the leading cause of stroke and therefore management of vascular risk factors is important across all age ranges and stroke subtypes. Treatment to reduce recurrence is called secondary prevention with comparable assessment and treatment for TIA and minor ischaemic stroke due to shared aetiology.

Primary prevention of stroke and cardiovascular disease

Primary prevention describes individual-level preventative strategies that target those at higher risk of stroke but without established disease. Population-based strategies are essential to address modifiable vascular risk factors, as the majority of stroke occurs in those at low to moderate cardiovascular disease risk. Risk factors, such as hypertension, smoking, abdominal obesity, high body mass index, poor diet, low physical activity, diabetes, excess alcohol consumption, high LDL cholesterol concentrations, kidney dysfunction, air pollution, and psychosocial factors may account for nearly 90% of the overall burden of stroke, with hypertension being the single most important risk factor (Fig. 29.14). On an individual level, anticoagulation for atrial fibrillation using risk stratification models (e.g., CHADS2 VASC score) is key for reducing cardioembolic strokes. With better population-based risk factor control, stroke incidence is reducing in older adults but despite this continues to increase in those under 50 years.

Personalized and goal-directed management of risk factors

Personalized medicine will ensure that treatment is targeted to each patient's unique needs, stroke aetiology, and co-morbidity profile (Box 29.2).

Lifestyle modifications

A healthy Mediterranean and low salt diet along with regular physical activity are cornerstones of treatment. Changing lifestyle requires behavioural change models instituted by the multidisciplinary team in partnership with the patient. Alcohol consumption should be minimized with smoking cessation essential. Specific physical activity targets, where feasible, entail low/moderate-intensity aerobic activity for 10 minutes on 4 days per week, or more vigorous aerobic activity for 20 minutes twice per week. Obesity is associated with cardiovascular disease and mortality. Targeting a healthy body mass index is advised on this basis, despite a lack of high-quality data for impact on stroke recurrence.

Box 29.9 Stroke management outline

Immediate general medical measures
- ABCDE (airway, breathing, circulation, disability, exposure) approach
- Assess Glasgow Coma Scale
- Apply monitors, check vital signs, obtain intravenous access, and send urgent laboratory tests
- Complete the standardized neurological examination using the NIHSS (Box 29.6)
- Consider intensive care and intubation for unconscious patients who are unable to protect their airway (e.g., GCS < 8)
- Proceed with urgent CT brain and CT angiogram
- Further testing such as electrocardiogram and chest radiograph are completed after acute stroke assessment, imaging, and thrombolysis unless clinically urgent (e.g., coexistent chest pain)

When haemorrhage is excluded, is thrombolysis appropriate (Box 29.7)?
- Immediate non-contrast CT brain imaging is needed (some centres use MRI)
- If no evidence of intracerebral haemorrhage, assess if patient fulfils indications for thrombolysis (disabling deficit within 4.5 hours)
- Check for contraindications to thrombolysis
- Do not wait for urgent laboratory test results before administration of thrombolysis unless coagulopathy suspected or other reason for clinical concern based on contraindication checklist
- Further urgent imaging with CT perfusion is required in the extended/unknown time window beyond 4.5 hours (known time of stroke onset 4.5–9 hours using perfusion imaging; wake-up strokes within 9 hours of the midpoint of sleep)
- MRI diffusion flair mismatch can be used to guide thrombolysis for strokes occurring during sleep with no clear time of onset and presenting to hospital within 4.5 hours of waking

When haemorrhage is excluded, is thrombectomy appropriate?
- CT angiogram is needed to confirm a large vessel occlusion and collateral status
- Is the patient presenting within 24 hours of symptom onset?
- Is there a significant deficit, usually NIHSS 6 or more?
- What is the pre-morbid modified Rankin score or dependency level of the patient? Consider excluding if severe dependency or short life expectancy indicating that the benefit of intervention may be limited.
- Is the non-contrast CT brain ASPECTS score very low identifying a very large established infarct with less potential for benefit?
- Refer rapidly. If in doubt, discuss with neuro-interventionist colleagues

Cerebral haemorrhage
- If the non-contrast CT brain confirms an intracerebral haemorrhage, implement a bundle of care including urgent blood pressure management, immediate reversal of anticoagulants, neurosurgical consultation, and avoidance of early do not resuscitate orders, where appropriate

Admit to the multidisciplinary stroke unit
- Urgent swallow assessment (<4 hours)
- Prophylaxis for venous thromboembolism
- Targeted blood pressure management (varies with thrombolysis administration and reperfusion status)
- Monitor and treat medical complications
- Early multidisciplinary review (physiotherapy, occupational therapy, speech and language specialist review etc.)
- Investigate aetiology (Box 29.2)
- Early initiation of secondary prevention measures (personalized, specific for aetiology, and goal directed)
- Lifestyle advise

Rehabilitation
- Commence in the acute stroke unit
- Referral onwards to community-based rehabilitation with early supported discharge or inpatient rehabilitation facilities
- Driving advice
- Refer to patient and community support groups

NIHSS, National Institute of Health Stroke Scale.

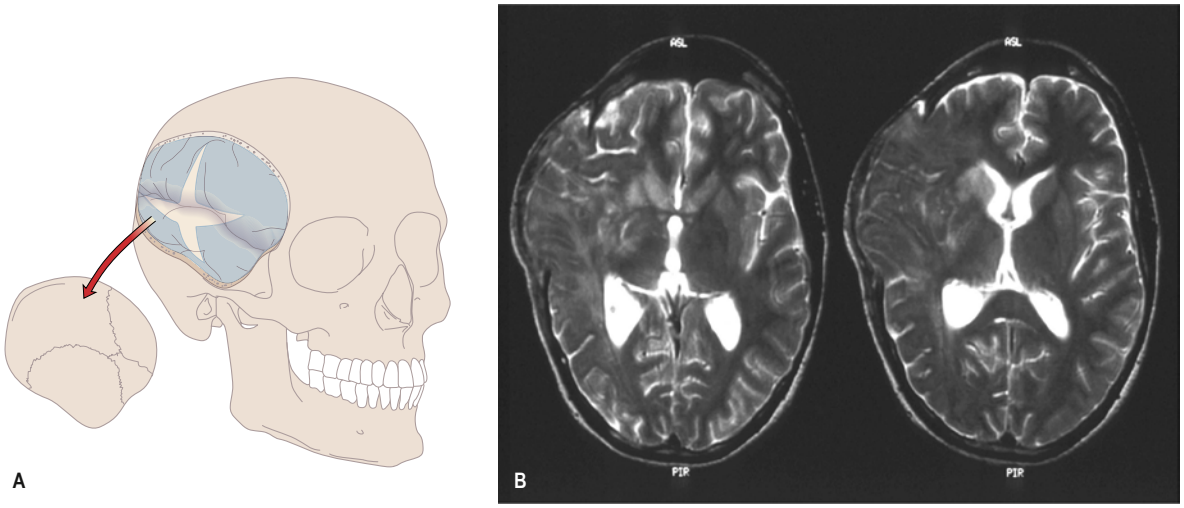


Fig. 29.13 Decompressive craniectomy for stroke. **(A)** Bone flap removed. **(B)** MRI scan showing brain swelling after bone flap removal.

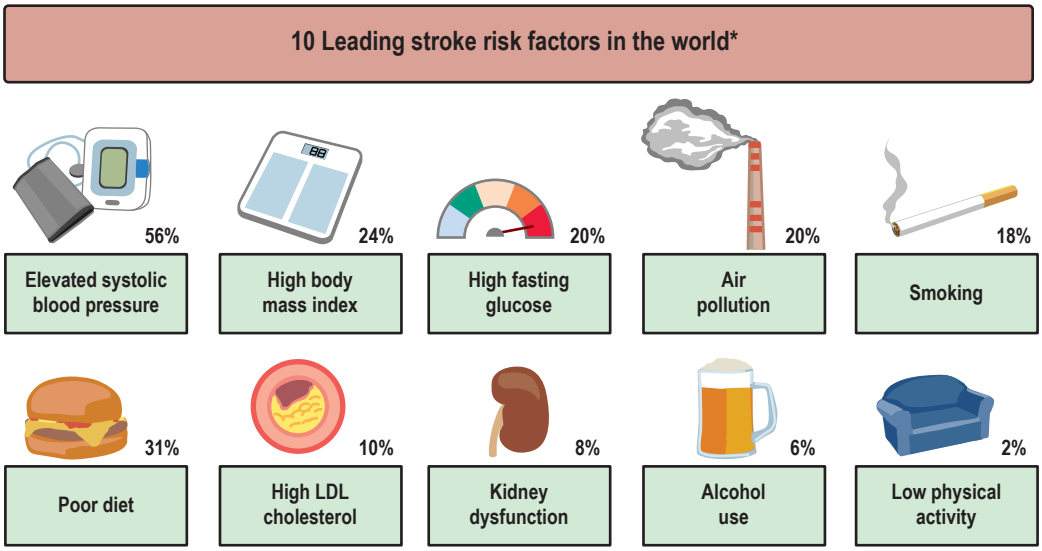
Smoking cessation

A strong association exists between smoking and stroke, with cessation associated with a 30% reduction in stroke recurrence. Behavioural intervention models increase smoking cessation and in combination with pharmacotherapy (e.g., varenicline, bupropion, nicotine replacement therapy) should be offered to all smokers (see Chapter 20 for further information on smoking cessation).

Pharmacological approaches

Blood pressure management

There is a linear relationship between blood pressure and stroke. Secondary prevention trials confirmed that lowering blood pressure is associated with reduced stroke and cardiovascular events proportionate to the magnitude of BP lowering.



*The sum of stroke burden attributable to the risk factors exceeds 100% because the effect of many of these risk factors overlap and are mediated partly or wholly through another risk factors.
Percentages show stroke-related disability-adjusted life years attributable to each risk factor.

Fig. 29.14 Burden of stroke attributable to risk factors. *(Credit: https://www.world-stroke.org/assets/downloads/WSO_Global_Stroke_Fact_Sheet.pdf.)*

A BP lowering target of <130/80 (or equivalent home BP reading <125/75) is recommended and well-tolerated, with a less intensive target such as 140–150 mmHg considered for those with orthostatic hypotension, bilateral severe carotid stenosis, or for frail older people.

Antihypertensive medication options should be personalized to the co-morbidity profile. Individualized treatment might mean choosing ACE inhibitors first line with coexistent systolic heart failure. Beta-blockers provide the least protection but may be beneficial for patients with ischemic heart disease or rate control of atrial fibrillation.

Mechanisms to improve compliance such as home or ambulatory BP monitoring can be instituted. Hypertension should be routinely treated with lifestyle modifications including the 'dietary approaches to stop hypertension' (DASH) diet, exercise, weight, and alcohol reduction.

Antiplatelet therapy

Single antiplatelet treatment long-term with clopidogrel 75 mg daily has replaced aspirin as first line in the absence of an indication for anticoagulation. Recent evidence supports the early use of dual antiplatelet therapy in patients with high-risk TIA or minor ischaemic stroke to reduce early stroke recurrence provided there is a low risk of bleeding. One such regimen combines aspirin 75 mg once daily with clopidogrel 75 mg once daily for NIHSS 0–3, after initial loading doses of 300 mg for aspirin and clopidogrel. Ticagrelor can be considered in this setting instead of clopidogrel (NIHSS 0–5). Single antiplatelet treatment is used from the outset in moderate to severe strokes.

The risk of haemorrhage is higher with dual antiplatelet treatment so treatment duration must be minimized, usually 21 days extending up to 3 months in limited circumstances such as stroke associated with intracranial stenosis. A proton pump inhibitor may reduce the risk of gastrointestinal haemorrhage with concurrent dual antiplatelet therapy.

Anticoagulation: atrial fibrillation and other conditions

Atrial fibrillation is associated with nearly a quarter of all strokes, therefore initial screening with an electrocardiogram and 24 hours of cardiac monitoring is essential. Prolonged monitoring for subclinical atrial fibrillation is recommended in embolic stroke, where the mechanism has not been determined after initial investigations. There is debate on the appropriate duration of prolonged monitoring in this setting with many suggesting at least 14 days. This can be carried out using an external monitoring device (>48 hours to 2 weeks or beyond) or an implantable loop monitor (ILR) positioned subcutaneously for up to 3 years. Atrial fibrillation may be confirmed on a single 12-lead electrocardiogram, based on more than 30 seconds of atrial fibrillation on intermittent (24-hour) cardiac monitoring or more than 5 minutes on implanted monitoring devices such as a pacemaker or ILR, according to consensus criteria.

Anticoagulation is recommended for persistent or paroxysmal atrial fibrillation and atrial flutter associated with stroke. Risk factors for atrial fibrillation include age over 50 years, frequent premature atrial contractions, dilated left atrium, high brain natriuretic peptide, hypertension, obesity, diabetes, and obstructive sleep apnoea with many factors serving as triggers (e.g., hyperthyroidism, cardiac surgery, and pneumonia).

Direct oral anticoagulants (DOACs), including direct thrombin or factor Xa inhibitors, are the standard of care after stroke (in preference to vitamin K antagonists) for non-valvular atrial fibrillation. DOACs are associated with better efficacy and safety having lower rates of intracranial haemorrhage, stroke, and thromboembolic events. DOACs are commenced once intracranial bleeding is excluded; immediately for TIA and as early as 48 hours for minor–moderate stroke. Anticoagulation for major stroke can be considered from 6 days after onset. Contraindications to DOACs must be considered along with drug interactions via cytochrome p450 or p-glycoprotein (e.g., anti-tuberculous and anti-epileptic medications). Modifiable

Box 29.10 Investigative work-up for ischaemic stroke

Laboratory	Routine labs; full blood count, full biochemical profile, thyroid function tests, lipids and glucose, HbA_{1c}, C-reactive protein, erythrocyte sedimentation rate, coagulation tests (prothrombin time, partial thromboplastin time)
Urine testing (as appropriate)	• Urine dipstick (infection, vasculitis) • Urine or serum toxicology (cocaine, amfetamines, may lead to vasospastic or vasculitic stroke) • Urine pregnancy testing (HCG)
Bladder ultrasound	• To assess the post void residual volume of urine
Hemodynamic monitoring	BP monitoring in hospital • Adhere to thrombolysis targets initially and secondary prevention targets on discharge • With differences in right and left arm blood pressure consider subclavian stenosis or aortic dissection
Electrocardiographic (ECG) monitoring to assess for atrial fibrillation	12-lead ECG; initial 24-hour cardiac monitor

Cardiac imaging
To assess for structural cardiac disease and aortic atherosclerosis particularly to elucidate an embolic stroke mechanism
	• Echocardiography
	  • Completed commonly for clinical or ECG evidence of cardiac disease or evidence of embolic stroke
	• Bubble echocardiogram with agitated saline bubble study in stroke patients 60 years and under to look for patent foramen ovale (PFO)
	  • One should consider appropriateness for PFO closure before screening for PFO
	• Transoesophageal echocardiography in selected cases
	  • Detection of PFO, atrial septal aneurysm, aortic arch atheroma, endocarditis, native and prosthetic valve abnormalities; for embolic stroke of unknown aetiology with multiple territory infarction or recurrent ischaemia where aetiology is not determined

Further tests to identify aetiology/manage risk factors	
Blood cultures	Consider if suspicion for endocarditis, infection, or with fever or raised CRP
Cardiac monitoring for embolic stroke	Prolonged external monitor—minimum 14 days or implantable loop monitor for atrial fibrillation detection
Haemodynamic monitoring to assess for BP control	Home BP monitoring or 24-hour ambulatory monitor
Other laboratory tests	• In cryptogenic stroke, testing for infections may be appropriate such as HIV, syphilis antibody, hepatitis B and C, VZV, or lyme, cryptococcus, and quantiferon (for TB) testing. The mechanism of stroke varies with each infection. Cryoglobulins may be considered • With anaemia and hypercalcaemia, check serum electrophoresis and immunoglobulins • Antiphospholipid screen: anti-cardiolipin antibodies (IgG and IgM), B-2 glycoprotein1, lupus anticoagulant (repeated after 3 months if elevated): consider these under age 50 years in absence of cardiovascular risk factors or if a history of miscarriages • An abnormal full blood count could prompt consideration for testing with a myeloproliferative neoplasm panel (such as thrombocytosis) • Autoimmune (ANA, ANCA, RF, anti-CCP, complement C3, C4) LDH, ferritin, fibrinogen, immunoglobulins, creatine kinase: consider if raised inflammatory markers, symptoms suggestive of vasculitis or systemic inflammatory disorder • Testing for heritable thrombophilia is not recommended in patients with stroke, regardless of age; except in selected cases such as neonatal stroke or prior history of venous thrombosis • In cerebral vein thrombosis, consider thrombophilia screening, testing for JAK2 mutation in the absence of clear provoking factors and a normal blood count, and testing for antiphospholipid syndrome
Lumbar puncture	• Consider lumbar puncture in an appropriate clinical context such as possible intracranial infectious cause of stroke, primary or secondary cerebral vasculitis • Clues include headache, cognitive impairment, seizures, or systemic symptoms • Cerebral vasculitis may cause elevated cerebrospinal fluid white cell count and protein
Sleep apnoea assessment	• Consider polysomnography which is the gold standard to calculate apnoea–hypopnoea index • Overnight oximetry is less sensitive for screening and will not differentiate central sleep apnoea
Cancer screening	• Appropriate body imaging or endoscopy where cancer is suggested by symptoms, laboratory findings, or multiterritory infarcts (Trousseau's sign) • D-dimer levels are a non-specific marker of cancer
Genetic and other testing	Genetic testing for monogenic causes of stroke in exceptional circumstances—for example CADASIL: mid-adult-onset cognitive decline, migraine with aura, neuropsychiatric disturbance, and diffuse white matter disease • Requires consent and counselling on implications

risk factors for bleeding should be addressed including uncontrolled hypertension, before DOACs are initiated. Falls are almost never a contraindication.

Exceptions where vitamin K antagonists are first line rather than DOAC include mechanical heart valve replacement, rheumatic moderate to severe mitral stenosis associated with atrial fibrillation or antiphospholipid syndrome with high-risk features and recurrent events. Vitamin K antagonists were the mainstay of treatment for renal failure with an estimated glomerular filtration of <15 mL/kg/1.73 m^2 (chronic kidney disease stage 5) and for individuals on dialysis, though certain DOACs are increasingly chosen for off-label use based on observational evidence. Recent evidence on atrial cardiopathy (a possible precursor of atrial fibrillation) and embolic stroke of undetermined aetiology (ESUS) identified no benefit

from anticoagulation therefore antiplatelets remain the mainstay of treatment.

Lipid management

Elevated LDL is a causal factor for atherosclerosis and aggressive LDL reduction in secondary prevention is associated with reduced mortality, cardiovascular events, and stroke. Statins prevent cholesterol synthesis via hydroxymethylglutaryl co-A reductase inhibition and are recommended for most stroke patients if not contraindicated (e.g., pregnancy). Statins provide particular benefit where atherosclerosis is present and for those with high cardiovascular risk and such patients should be started immediately on high-intensity statin therapy, such as atorvastatin 80 mg daily. Lower statin doses may be appropriate with potential drug interactions. For statin intolerance, use the maximally tolerated dose or trial the alternative high-intensity statin, rosuvastatin.

A target LDL <1.8 mmol/L should be reached within 1 to 3 months either with maximally tolerated statin therapy or addition of ezetimibe 10 mg daily. Ezetimibe inhibits dietary and biliary cholesterol absorption. Other LDL-lowering drugs may be considered via specialist services where LDL targets are not achieved; see Chapter 19 for discussion of these newer agents.

Diabetes

In patients with ischemic stroke and diabetes, a long-term target of HbA1c < 53 mmol/mol (<7%) is recommended for most people under age 65 without life-limiting comorbid illness to reduce the risk of microvascular complications (see Chapter 35). Antihyperglycemic agents associated with improved cardiovascular outcomes are preferred such as SGLT-2 inhibitors or glucagon-like peptide-1 receptor agonists (GLP-1RAs). GLP-1RAs (dulaglutide and semaglutide subcutaneous injection) reduce stroke risk in individuals with type 2 diabetes at high risk for cardiovascular disease. To date, SGLT-2 inhibitors have been shown to reduce cardiovascular disease and heart failure risk but not stroke.

Contraception and hormone replacement

The combined oral contraceptive pill should be avoided after a stroke. Similarly postmenopausal oestrogen therapy should be discouraged after stroke. Evidence is conflicting on the association between menopausal hormone replacement therapy and stroke; however, there may be a small increased risk particularly for higher oral oestrogen doses. See Chapter 47 for a fuller discussion.

Specific interventions

Carotid intervention

Multiple international guidelines recommend surgical carotid endarterectomy for individuals with recent anterior circulation non-disabling stroke who have symptomatic, ipsilateral, carotid stenosis of 50% to 99% using the North American Symptomatic Carotid Endarterectomy Trial (NASCET) measurement technique. Many factors impact on risk-benefit including co-morbidity, 5-year life-expectancy and plaque morphology including intraplaque haemorrhage on MRI. The optimal timing is between 2 and 14 days after the initial stroke. Endarterectomy is not recommended for complete carotid occlusion. Early surgical risks include stroke, myocardial infarction, haematoma, and cranial nerve injury due to traction, which usually recovers over months. Cerebral hyperperfusion syndrome can present postoperatively with headache and seizure after revascularization of a severe carotid stenosis if high blood pressure in the postoperative period is not controlled.

Carotid stenting is associated with a higher risk of stroke, especially in older patients, but a lower risk of myocardial infarction and cranial nerve injury. Stenting may be preferred in specific situations such as inaccessible carotid bifurcation or post-radiotherapy.

Optimizing vascular risk factors is the mainstay of treatment for carotid stenosis, along with other atherosclerotic stroke aetiologies such as aortic arch atheroma, vertebral or intracranial stenosis. Surgical intervention is not routine beyond carotid stenosis.

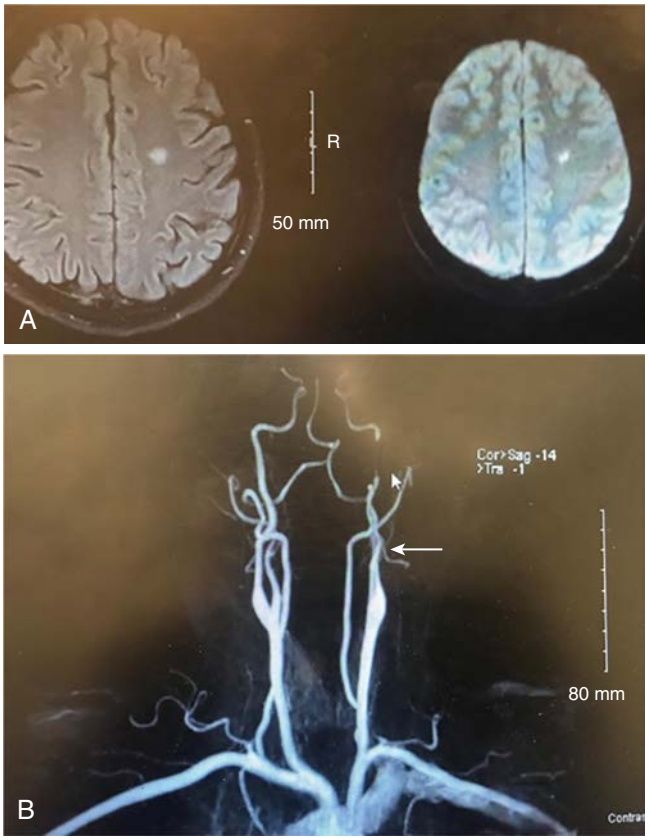


Fig. 29.15 MRI imaging of acute cerebral infarction (diffusion positive) and subsequent magnetic resonance angiography showing left extracranial carotid dissection. **(A)** Left subcortical acute infarct on MRI brain, hyperintense on flair and diffusion-weighted imaging. **(B)** Subsequent MRA confirms a left distal extracranial internal carotid dissection. This spares the carotid bulb which is the usual location of atherosclerotic occlusion.

Dissection of vessels

For extracranial cervical artery dissection (**Fig. 29.15**), antiplatelet or anticoagulant treatment is recommended for at least 3 months with assessment of recanalization at a later interval. Dual antiplatelet therapy is an option in the peri-stroke stroke period (first 21 days) for high-risk TIA and minor stroke. For intracranial arterial dissection, the potential risk of subarachnoid haemorrhage and aneurysm must be considered, with a preference therefore for antiplatelet treatment.

Embolic stroke of undetermined source

Antiplatelet therapy is recommended rather than anticoagulation.

Patent foramen ovale

Evidence confirms the benefit of PFO closure in stroke patients with embolic features under 60 years in the absence of another stroke mechanism (cryptogenic). The PASCAL classification is recommended for patient selection. Those with high-risk anatomical features such as atrial septal aneurysm and a large right-to-left shunt gain the most benefit. A multidisciplinary team discussion including cardiologists is recommended to inform decisions regarding PFO closure, the timing of which should be within 6 months post stroke. PFO closure is discussed further in Chapter 17.

Left atrial appendage occlusion

Thrombus in the left atrial appendage (a muscular, trabeculated pouch attached to the left atrium) is a primary cause of stroke associated with atrial fibrillation. A left atrial appendage occlusion (LAAO) device plugs off the left atrial appendage removing connection to the circulation. LAAO can be considered after cardioembolic stroke associated with non-valvular atrial fibrillation where anticoagulation is contraindicated due to a high bleeding risk (e.g., severe thrombocytopaenia or bleeding). Bleeding risk must be balanced with risk of device-related thrombosis. The risk of LAAO procedural complications is not trivial (4%–7%) despite the advancement in techniques.

Infective endocarditis

This is an uncommon stroke aetiology, but diagnosis requires a high index of suspicion. Over a third of patients with endocarditis have neurological manifestations, including stroke, and many more have covert ischaemic brain lesions. Infective endocarditis carries a high risk for haemorrhagic conversion of ischemic lesions, with rupture of mycotic aneurysms or septic necrotic arteritis. Targeted pathogen-specific antimicrobial therapy, as well as surgical intervention where appropriate, are the mainstay of treatment. Thrombolysis is contraindicated. Antithrombotics after endocarditis-related strokes are generally avoided due to the high haemorrhagic risk. Individualized decision-making is needed for those at high thrombotic risk such as mechanical prosthetic valves.

Cerebral venous sinus thrombosis

Anticoagulation with therapeutic heparin either low-molecular-weight heparin or unfractionated heparin are the treatment of choice in the acute phase even in the presence of secondary ICH. Despite limited evidence, endovascular thrombectomy may provide rescue treatment in highly selected cases. Anticoagulation with vitamin K antagonists is recommended to be continued for 3 to 12 months or indefinitely depending on aetiology. Evidence is also emerging for selected DOACs for this indication.

Vaccine-induced immune thrombotic thrombocytopaenia after certain adenoviral vector COVID (SARS-CoV-2) vaccinations is associated with unusual thrombosis including CVST. Therapy is different in this scenario. Non-heparin–based anticoagulation, DOAC, fondaparinux, or argatroban (parenteral direct thrombin inhibitor) are recommended, in conjunction with steroids, intravenous immunoglobulin, or plasmapheresis.

Other aetiologies

Inflammatory conditions (e.g., primary angiitis of the central nervous system and other secondary vasculitis) will require high-dose steroids and/or glucocorticoid-sparing agents. Infectious causes (e.g., varicella zoster, human immunodeficiency virus) will warrant antiviral treatment. Cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL) is the most common monogenic disorder causing stroke particularly in the young. Genetic testing after appropriate counselling, alongside intensive cardiovascular risk factor modification including antiplatelet therapy, are the mainstay of treatment, although with limited evidence. Identification of alternative genetic causes such as Fabry's disease may warrant enzyme replacement therapy.

Recurrent ischaemic stroke

For patients presenting with recurrent ischaemic stroke, a detailed reassessment of aetiology, compliance with treatment and secondary prevention is sensible. For cardioembolic stroke on DOAC, there is currently no evidence that switching to an alternative anticoagulant reduces stroke recurrence. Assess for a competing aetiology (e.g., atherosclerosis), review for drug–drug interactions and medication compliance, and ensure appropriate DOAC dosing. LAAO may be a useful adjunct in selected cases. Clopidogrel resistance may be considered in patients with stroke while taking clopidogrel, which is linked to a genetic polymorphism of cytochrome CYP2C19.

INTRACEREBRAL HAEMORRHAGE

Introduction

Spontaneous ICH is defined as non-traumatic bleeding in the brain parenchyma and is responsible for approximately 10% to 15% of strokes. It is associated with a higher mortality than ischaemic stroke (up to 50%) and most survivors face severe long-term disability. A large haematoma may act as a space-occupying lesion, causing raised intracranial pressure with brain displacement and herniation. The incidence is higher in Black African and Asian ethnic groups compared to White populations, potentially attributable to the higher occurrence of hypertension. Although the overall mortality from ICH is decreasing globally, the incidence in low-/middle-income countries is almost twice that of high-income countries and attributed to lack of awareness about primary preventive strategies and inadequate access to medical care.

Other types of intracranial bleeding (including subarachnoid and subdural bleeds) are covered in Chapter 12.

Aetiology (Fig. 29.16)

Hypertension

Hypertension is the most common risk factor for ICH accounting for approximately 70% to 80% of cases. Common locations for bleeding include the basal ganglia, thalamus, pons, cerebellum, and subcortical white matter. Deep perforator arteriopathy (degeneration of small, deep, penetrating arteries) relating to chronic hypertension and arteriosclerosis are the principal pathologies.

Cerebral amyloid angiopathy

Deposition of beta-amyloid in leptomeningeal and cortical small arteries, arterioles, and capillaries, in those over 50 years, can cause recurrent lobar ICH. CAA may also present with transient focal neurological episodes (amyloid spells) which are atypical for TIA. CAA is associated with apolipoprotein E genotypes E2 and Alzheimer's disease. The Boston criteria 2.0 provides a framework for the diagnosis of CAA. MRI with blood-sensitive sequences identify cerebral microbleeds, cortical superficial siderosis, and convexity subarachnoid haemorrhage with the latter two imaging characteristics particularly associated with a high risk of ICH reoccurrence.

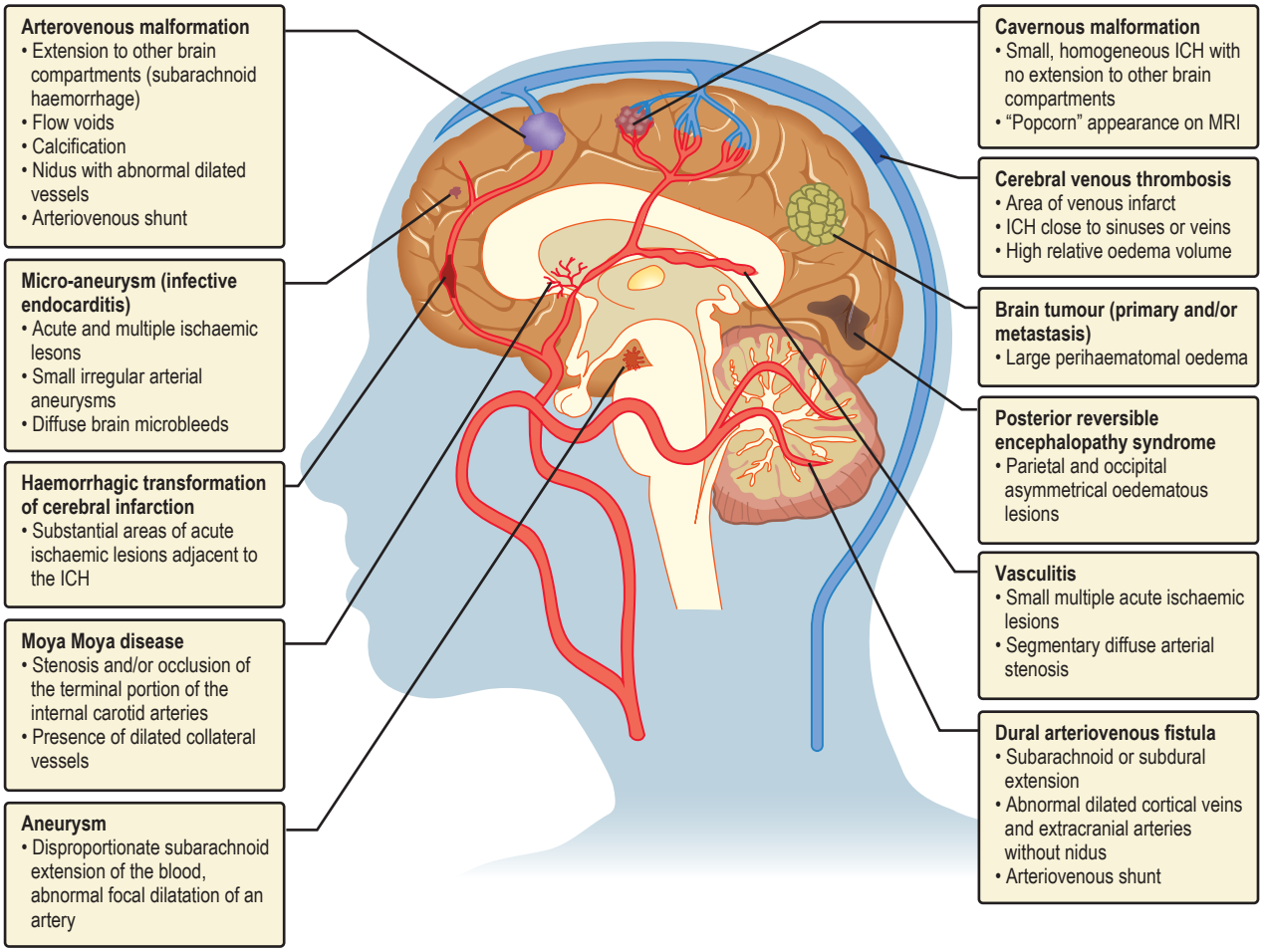


Fig. 29.16 Clues to treatable causes of ICH. *(https://www.nature.com/articles/s41572-023-00424-7.)*

Structural/vascular malformations

Brain neoplasms, arteriovenous malformations, cavernomas, and aneurysms cause approximately 20% of ICHs.

Haemostatic causes

Several studies have shown that anticoagulants increase the risk of ICH 8- to 11-fold. In comparison to vitamin K antagonists, DOACs are associated with a 50% reduction in ICH. Risk factors for anticoagulant-related ICH include older age, uncontrolled hypertension, diabetes, concomitant use of antiplatelet treatment, history of stroke or ICH as well as markers of cerebral small vessel disease (white matter hyperintensities, microbleeds, and lacunes). The use of dual antiplatelet treatment doubles the risk of ICH compared to monotherapy. Reperfusion therapies such as thrombolysis have been linked with an ICH risk between 1.7% and 8.8%.

Other aetiologies

- Sympathomimetic drugs (cocaine, amfetamines)
- Infective endocarditis
- Moyamoya
- Haemorrhagic transformation of a large ischaemic infarct may sometimes mimic a haemorrhage
- Venous sinus thrombosis or cortical vein thrombosis
- Chronic kidney disease is a risk factor.

Clinical features

At the bedside, there is no reliable way of distinguishing between cerebral haemorrhage and infarct. ICH is frequently associated with severe headache or decreased conscious level. Patients on oral anticoagulants should be assumed to have had a haemorrhage unless it is proven otherwise.

Investigations

Brain haemorrhage is seen on CT imaging immediately as intraparenchymal, intraventricular, or subarachnoid hyperdensity. This contrasts with CT brain imaging in the acute phase of an ischaemic stroke which may be normal or only show subtle signs of infarction. The CT scan is regarded as the reference standard for diagnosing ICH due to its rapidity, high sensitivity, and specificity for all forms of ICH, and widespread availability (Fig. 29.17).

Several non-contrast CT markers of expanding bleeds include the blend (blood-fluid level), black hole, island, and satellite signs. The spot sign (CT angiography) is an independent predictor of haematoma expansion, clinical deterioration, and mortality.

Structural causes must be suspected and investigated using CT/MRI angiography in those individuals without an explanation for the ICH, for example, those under age 45 years or those aged 45 to 70 years without a history of hypertension, malignancy, and anticoagulant use. Digital subtraction angiography is warranted in patients at high risk of macrovascular causes following negative CT/

Fig. 29.17 Non-contrast CT scan demonstrating a left basal ganglia bleed.

MRI angiography (**Box 29.7**). Additional imaging modalities including CT/MRI venography should be considered if venous sinus thrombosis is suspected. An interval MRI/MRA can be considered at 3 months where the aetiology has not been determined.

Baseline blood tests should include urgent coagulation studies, platelets, glucose, and toxicology where appropriate. For warfarin, point-of-care INR testing should be employed to minimize delays in reversal.

Medical management

ICH is a neurological emergency, and the treatment should be delivered in either a stroke unit or intensive care unit. Treatment is often provided as a bundle of care. Frequent monitoring of GCS and neurological signs is essential, and neurosurgery may be required. Anticoagulation should be rapidly reversed where possible (see Chapter 40).

Acute blood pressure management in ICH is recommended for people with ICH symptom onset within the last 6 hours, aimed to achieve a target of SBP 130–140 mmHg in those with SBP ranging between 150 and 220 mmHg. BP lowering should be achieved within 1 hour if possible and maintained for 7 days minimizing variability. For SBP > 220 mmHg, or GCS < 6 (elevated ICP suspected), caution should be exercised.

The third Intensive Care Bundle with Blood Pressure Reduction in Acute Cerebral Haemorrhage Trial (INTERACT3) demonstrated that implementing the care bundle—targeting SBP 140 to 130 mmHg within 1 hour of treatment initiation, intensively controlling elevated blood glucose (6.1–7.8 mmol/L for non-diabetic patients, 7.8–10.0 mmol/L for diabetic patients) as soon as possible, treating pyrexia to maintain a body temperature below 37.5°C within 1 hour, and reversing abnormal anticoagulation—resulted in better functional outcomes (measured on the modified Rankin Scale) at 6 months post-treatment compared to usual care. This included favourable effects on survival and health-related quality of life.

Mechanical ventilation may be required for individuals with a low GCS. Corticosteroids provide no benefit and should be avoided. Normoglycemia should be maintained. Prophylactic epileptic drugs are not recommended.

Surgical management

Posterior fossa haematoma may cause obstructive hydrocephalus or coma due to brainstem compression; urgent neurosurgical clot evacuation is lifesaving (and is required where the haematoma is >3 cm, shows mass effect, or the patient has a low GCS). The role of surgical decompression for non-cerebellar bleeds is less clear-cut and liaison with neurosurgeons is vital. Use of minimally invasive surgical techniques to evacuate haematomas is an area of active research. The Trial of Early Minimally Invasive Removal of Intracerebral Hemorrhage (ENRICH), predominantly recruiting patients with lobar haemorrhage with a hematoma volume of 30 to 80 mL, demonstrated that minimally invasive hematoma evacuation within 24 hours resulted in better functional outcomes at 180 days compared to best medical management alone in lobar haemorrhage. The SWITCH trial provided weak evidence that decompressive craniectomy plus best medical treatment might be superior to best medical treatment alone in people with severe deep ICH (volume 30–80 mL), but survival was associated with severe disability in both groups (mRS 4–6). An external ventricular drain may be required for obstructive hydrocephalus as seen with intraventricular haemorrhage or cerebellar mass effect.

Secondary prevention

Identifying the underlying aetiology is key to appropriate secondary prevention. Control of hypertension is central for many individuals with a target blood pressure goal of <130/80. It may be appropriate to resume anticoagulation if reversible risk factors for bleeding are identified however the optimal timing is not well established.

Further reading

Dawson J, Béjot Y, Christensen LM, et al. European Stroke Organisation (ESO) guideline on pharmacological interventions for long-term secondary prevention after ischemic stroke or transient ischemic attack. *Eur Stroke J* 2022; 7:I-II.
Mead GE, Sposato LA, Sampaio Silva G, et al. A systematic review and synthesis of global stroke guidelines on behalf of the World Stroke Organization. *Int J Stroke*. 2023;18(5):499–531.
National Clinical Guideline for Stroke for the UK and Ireland. London: Intercollegiate Stroke Working Party; 2023. Available at: www.strokeguideline.org.

Significant websites

https://biasp.org/clinical-guidelines-standards/
https://eso-stroke.org/guidelines/eso-guideline-directory/#acute-stroke
https://www.world-stroke.org/
https://www.stroke.org.uk/
https://www.nihr.ac.uk/explore-nihr/specialties/stroke.htm

Questions for this chapter can be found on Elsevier eBooks+.

30 General hospital psychiatry

Peter Byrne, Sotirios Posporelis, and Sridevi Siramahalingappa

Clinical skills 1002
Approach to the patient 1002
Formulation 1003
Risk assessment 1003
Treatment approaches in psychiatry 1006
Prescribing in psychiatry 1007
Background 1009
Liaison psychiatry 1009
The liaison psychiatry team 1010
Ethical and legal issues in psychiatry 1010
Consent and confidentiality 1010
Patients wanting to leave hospital 1011
The legal framework for mental health treatment 1011
Aetiology of mental disorders 1012
Common psychiatric emergencies 1013
Self-harm and suicide 1013
Acute disturbance 1014
Alcohol withdrawal 1014
Neuroleptic malignant syndrome and serotonin syndrome 1016
Catatonia 1017
Common psychiatric disorders 1017
Depression 1017
Bipolar affective disorder 1019
Anxiety disorders 1021
Panic disorders 1022
Anxiety syndromes 1022
Disorders specifically associated with stress 1023
Obsessive–compulsive disorder 1024
Psychosis 1025
Psychotic disorders 1026
Psychotic disorders due to other medical conditions 1027
Medically unexplained symptoms and functional neurological disorders 1029
Addictions 1030
Alcohol dependence 1030
Opioid addiction 1032
Novel psychoactive substances 1033
Personality disorder 1033
Eating and feeding disorders 1036
Anorexia nervosa 1036
Bulimia nervosa 1039
Avoidant/restrictive food intake disorder 1039
Additional considerations in selected patient groups 1039
Neurodiversity and learning disability 1039
Attention deficit hyperactivity disorder 1040
Mental health problems in older people 1040
Patients in the criminal justice system 1041
Diagnoses of exclusion: factitious disorder and factitious disorder induced by another 1041
Prevention of mental illness: public mental health 1042

CORE SKILLS AND KNOWLEDGE

Psychiatry is a branch of medicine, and liaison psychiatry (LP) is a subspecialty of psychiatry that is based in the general hospital. LP sees patients on a consultation basis to bridge between general hospital colleagues and mental health clinicians, most of whom are community-based. The LP team assesses and treats emergencies (self-harm and suicide attempts, acute confusion, psychosis, etc.), as well as provides advice on patients with physical conditions who have related or unrelated psychological morbidity.

Key skills include:

- Taking a psychologically informed history
- Recording a mental state examination
- Assessing decision-making capacity and how this informs consent
- Making a formulation (complex understanding of interacting factors)
- Assessing risk to self or others
- Assessing the effects of alcohol, cannabis, other misused substances and psychiatric medication.

These skills can be developed by shadowing members of a liaison psychiatry team (including psychiatrists, psychiatric liaison nurses, and psychologists) as they review new cases. This can be supplemented by spending time in inpatient psychiatric units and shadowing community mental health teams—where mental health can be assessed in the broader context of the patient's home situation and community relationships.

CLINICAL SKILLS

Approach to the patient

There are several steps to follow when assessing patients in psychiatry:

- the psychiatric history (Boxes 30.1–30.3)
- the mental state examination (MSE; e.g., following the 'ASEPTIC' structure set out in Box 30.4); detailed assessment of cognition is set out in Box 30.5
- obtaining a collateral history, which is always worth the effort if possible (Box 30.6)
- physical examination and investigations for signs of physical illness that might be affecting mental state (Boxes 30.7 and 30.8).

Information from these different sources are combined to produce a 'formulation' which places a patient's likely mental health diagnosis in the wider context of their physical, psychological, and social context, and has far greater utility than a simpler model of 'diagnosis'.

Box 30.1 sets out the circumstances of a psychiatric interview. Taking a psychiatric history is quite different from history-taking in other areas of medicine, and developing these skills is a steep learning curve. The experience may be exhausting for both patient and interviewer.

Box 30.1 Preparation for psychiatric interviewing

Decisions about the circumstances of the interview

How much time have you got? Do you have time to read old notes?

Will you interview on your own? Might the patient bring a relative?

Which location is best, providing adequate privacy for patient and adequate safety for you?

Do you need an interpreter? Are there other communication difficulties (e.g., autism, deafness)?

How will you record key information (you may forget dates, medication doses)?

Good beginnings

After introductions, explain what your role is: student, doctor, psychiatrist

Explain that you will try not to trigger/retraumatize, but you will ask personal questions

Define the limits of confidentiality and with whom you will share information

Record if your patient is detained under legislation, or any special circumstances

Additional diagnostic information

Even with a thorough history from the patient and MSE, a collateral history will clarify many aspects of the diagnoses (Box 30.6). The key findings of the MSE determine the formulation and management plan.

Box 30.2 Key sections of psychiatric history

History of presenting complaint

- What is the problem? When did it start? How did it develop: onset/progress/severity/social impairment (e.g., unable to work)?
- What makes it better or worse?
- Past psychiatric history
- Age of onset of symptoms and first contact with services (there is always a time gap); nature and progression of difficulties; diagnoses
- Hospital admissions: when, approximate length; voluntary or under legal section
- Past treatments—medication, psychological therapy, electroconvulsive therapy (ECT): what has helped in the past, what has not, medication type, doses prescribed, and actual doses taken (i.e., concordance with prescription); list side-effects

Medical history

- Past and current physical illness and treatments
- Current medication, including any over-the-counter drugs taken regularly; any drug can be purchased online or acquired.
- See Box 30.3
- Drug allergy: list what happened (rash, anaphylaxis, other symptoms) with what substance

Family history and living conditions

- Who's at home with you now?
- Housing type: adequacy for needs of patient/others. Income, work details, debts
- Parents: biological/adoptive/step-parents and siblings: age, state of health or cause and age of death, quality of relationships, then and now? Currently, who supports the patient and who exacerbates their problems?
- Family history of mental disorder includes substance misuse and suicide

Premorbid personality

- How would they describe themselves, how would others describe them? (use quotation marks)
- Prevailing mood and stability of mood; impulse control
- Nature of relationships with others: partners, friends, colleagues (e.g., close and confiding, casual only)
- Leisure interests (hobbies)
- Tolerance of stress and coping style, use of substances to manage stress, modify mood or social interaction

Personal history

- Obstetric and birth: planned/unplanned conception, maternal physical and mental health during pregnancy; birth full-term/premature, obstetric events and complications, low birth weight, congenital abnormalities, neonatal illness, maternal separation, and bonding
- Development and milestones: delays in interaction with others, speech; motor control, walking, toilet training; sleep difficulties; emotional or behavioural difficulties, hyperactivity; physical illness
- Family atmosphere and stability: abuse?
- Were there periods of separation from caregivers (e.g., in hospital due to childhood illness; in foster care due to parental difficulties)?
- Social development: key friendships, experience of bullying, any juvenile delinquency
- Educational attainment: specific learning difficulties, school refusal, age left education and qualifications
- Occupation: periods of employment, nature of work/skills
- Psychosexual: age of first sexual experience, sexual orientation, number, length and quality of significant relationships, marriage(s), children from previous relationships

Risk history

- Risk episodes: previous self-harm and suicide attempts; self-neglect and exploitation by others (financial, sexual), thoughts of and actual harm to others
- Forensic history (always try for collateral): arrests, charges, and convictions: nature of offences, outcome (custodial sentence, community service, probation); always ask about violent actions
- Include criminal activities where patient was not arrested, crime not detected

Substance misuse history

- Alcohol use, amounts (in units) (Fig. 30.7)
- Illicit substance use: type, pattern of use including frequency, dependency; associated problems—occupational, social, relationship, health, and criminal activity

Collateral history from informants (e.g., friends, family, GP, work colleagues). Any contradictions from patient's account? See Box 30.6

Part of every first encounter is an additional, bespoke physical examination. This may reveal evidence of substance misuse (Box 30.9) or of physical co-morbidities which are common in patients with psychiatric disease (Box 30.10). All hospitals have protocols for routine physical examinations. This process is guided by current symptoms, likely disease processes (given this presentation, patient's age, sex, risk factors, etc.) and the speciality where that patient is being admitted. Psychiatry has its own priorities when a patient is admitted to hospital—even if a recent physical exam by others looks complete: Box 30.7.

The history (physical symptoms, medical history, medications, etc.), MSE, and physical exam determine routine then bespoke investigations (Box 30.8). Abnormal findings need discussion that day (sometimes urgently) with other members of the mental health team. For example, patients may present with what is thought to be a behavioural or 'psychiatric' presentation who have actually suffered and intracranial bleed (picked up on neuroimaging, Fig. 8.39), or may have acute encephalitis or an unusual epilepsy syndrome (evidenced by abnormal electroencephalography; Fig. 28.19). Acute disease processes that will likely deteriorate on a psychiatric ward merit urgent discussions with medical colleagues.

Further reading

Byrne P, Byrne N. *Clinical Cases Uncovered: Psychiatry*. Wiley: 2008.
Oyebode F. *Sims' Symptoms in the Mind: An Introduction to Descriptive Psychopathology*, 7th edn. Elsevier 2022.

Formulation

Given the multifactorial nature of psychiatric disorders, many clinicians synthesize findings from their assessments using the four categories below as a 'formulation':

- **Predisposing factors** render someone vulnerable to developing a psychiatric disorder across their life course. These may include genetics, *in utero* environment, early development, adverse childhood experiences (ACEs), and personality, among others.
- **Precipitating factors** predate the onset of the disorder and are usually considered triggers that again can be divided into the categories of the biopsychosocial model. These interact with the predisposing factors and may at times lead to mental disorder. Biological examples include dysphoric medications (Box 30.3) and the strongest legal depressant, alcohol. Common psychological precipitants are losses (bereavement, relationship breakup), and social precipitants include 'exit events' such as loss of employment, moving home, or moving country (migrants).
- **Perpetuating factors**, also called maintaining factors, are highly relevant to clinical practice as identifying and managing these is key to successful treatment of a disorder. In general hospitals, biological examples include pain from physical disease processes, complications from medical procedures, and anxiety arising from uncertainty about medical outcomes. Psychological factors that prolong distress might be personality types that adjust badly to change. Social factors include unsuitable and insecure housing, other financial concerns, and life-changing injuries/symptoms that interfere with life goals.
- **Protective factors** predict the duration and severity of episodes. These include mature coping skills (gentle exercise, mindfulness, avoiding excess alcohol to reduce stress), a good support network, strong identity, and a high sense of belonging in a community, spirituality/religion, among others.

Risk assessment

Risk is defined as the likelihood of a negative event occurring, and its assessment is one of the most important outcomes of a clinical interview. While every healthcare professional should be proficient in determining risk for their patients, it is especially important for mental health professionals, as patients with mental disorders have a higher risk of self-harm. Risk assessments are dynamic,

Box 30.3 Common drugs and dedications associated with mood, confusion, and/or psychosis

Category/speciality	Drug example	Depression	Mania	Confusion	Psychosis
Widely available	Alcohol	+++	++	+++	+
Anti-inflammatory drugs	Corticosteroids	+++	++	+++	+++
	Interferon	+++	++	+	++
Analgesics	Opioids	++	+	++	+
Cardiology	Beta-blockers	+		+	
Respiratory medicine	Theophyllines	+		++	+
Gastroenterology	Protein pump inhibitors Anticholinergics	+ +		+ +++	+
Renal medicine	Nil specific	++ (renal failure)		'dialysis dementia'	
Endocrine	Replace/block hormones	Rare			
Infectious diseases	Anti-tuberculosis and antimalarials Antiretrovirals	++ ++	++ ++	++ ++	++ +
Neurology	Anti-Parkinson's Antiepileptics	++ ++	++ +	+++ ++	++ +
Psychiatry	Antidepressants Antipsychotics Benzodiazepines	 ++ ++	++ +	+ ++ +++	
Oncology	Chemotherapy	++		+	
Dermatology	Isotretinoin Antihistamines	+++ +	+	+ ++	+

Box 30.4 Mental State Exam details (ASEPTIC)

A—Appearance and behaviour

General appearance, physical state, abnormal movements, and behaviour

- Style and manner of dress, hygiene: self-neglect
- Physical state: any signs of distress or physical illness, drug/alcohol withdrawal, self-harm scars
- Quality of eye contact (e.g., fixed stare, avoidant)
- Manner of engagement during interview (e.g., suspicious/guarded/relaxed): comment on rapport
- Distractibility and preoccupation with internal world (e.g., appearing to respond to auditory hallucinations)
- Hyper-/hypoactivity: relevant in delirium and mood states
- Abnormalities of movement: mannerisms are idiosyncratic goal-directed behaviours (e.g., style of walking); motor stereotypes are regular repetitive non-goal directed movements (e.g., rocking)

S—Speech

Rate, amount, volume, form, and coherence

- Increased (mania)/decreased (depression); fast (pressure of speech—mania)/slow; loud (agitation, anxiety)/soft (depression)
- Verbal stereotypy: repetition of irrelevant words or phrases
- Formal thought disorder disruption to the continuity of thought

Answers may be initially appropriate but quote speech in full (inverted commas).

Common patterns of abnormality affecting the content of speech include:

- circumstantial speech, straying far from the topic before returning to the point of the question;
- tangential speech, where they do not return to answer the question or the conversation in hand, the latter represents mild derailment.

- Loosening of associations, the juxtaposition of completely irrelevant ideas. The most extreme form of thought disorder is known as word salad where meaning is indecipherable.
- Flight of ideas in hypomanic/manic states is the rapid transition between topics via internal links: connected words, themes, rhyming, puns, alliteration, e.g., 'Black bears scare me, I've a black bag'), or the inclusion of external distractions into the train of thought (e.g., subsequent comments on interviewer's black shoes)
  - Poverty of thought describes insubstantial speech that conveys little meaning
  - Neologisms: words or phrases invented or used idiosyncratically to denote new meaning ('I don't like my boss: he's a bosstard')

E—Emotion, meaning mood and affect

Mood, subjectively—according to the patient

Affect, objectively—as assessed by an external observer. Try to record a score

Suicidality: subjective + objective assessment (this is essential to every mental state examination)

- Subjectively patient's description of their current mood: rated out of 10, with 0 lowest they can imagine; it is useful to rate 'usual' mood as 8/10 (tell them 10/10 is the 'happiest day of their lives') and do NOT accept a negative number for worst mood.
- Objectively interviewers appraise their affect (external manifestation of emotional state) and emotional range during interview. The term euthymic describes an emotional state within the normal range and with normal reactivity. Abnormal states show incongruous affect. Emotional responses may be perplexed, blunted, or flat (lacking emotional reactions) in depression and some psychosis states. Affect can also be labile (extremely changeable range of emotions).
- Presence or absence of thoughts of self-harm: state any plans and degree of intent.
- Further enquiry about suicidal ideas. Record the spectrum from passive death wish (e.g., going to sleep and wishing s/he would not wake up), through suicidal ideation and suicidal intent, to suicide plan, and actions **(if you have not asked about suicide, you have not completed an MSE).**

P—Perception

Sensory distortions, sensory deceptions (illusions, hallucinations). Auditory hallucinations include noises and voices. Other hallucinations include visual, somatic, olfactory, taste, sexual, and touch. Some hallucinations (e.g., hypnopompic hallucinations) are part of normal experience.

- Distortions are changes in intensity or quality of real sensory phenomena (e.g., micropsia in a temporal lobe seizure).
- Deceptions are either illusions (misinterpretations of real stimuli), often in altered mood states or consciousness (e.g., hearing an innocuous noise as a sinister footstep when anxious) or hallucinations, which are internally generated perceptions in the absence of an external stimulus.
- Pseudohallucinations are experienced as arising from within the patient (e.g., 'voices in my head') rather than the external world, but they are beyond conscious control. In contrast to hallucinations, they are not experienced as having a material reality. They occur in normal grief (seeing or hearing the deceased) as well as a range of disorders including post-traumatic stress ('flashbacks') and emotionally unstable personality disorder.

T—Thought content

Morbid preoccupations (i.e., ruminations), obsessions, overvalued ideas, and delusions

- Obsessions: repetitive, intrusive, unwanted, stereotyped thoughts or images.
- Overvalued ideas: those held with a morbid intensity, but without fulfilling the criteria for a delusion. They are not argued beyond the bounds of reason (e.g., patients with anorexia nervosa are not deluded but have overvalued ideas about their weight).
- Delusions are fixed (usually false) beliefs held without evidence that is out of keeping with an individual's sociocultural background. Delusions may be primary or secondary. There are many subtypes of delusion, and these are key to differentiating the type of psychosis, for example, schizophrenia versus a depressive psychosis.
- Delusions of passivity describe the experience that one's mind (thought passivity), emotions, actions, will, or body (somatic passivity) is not under one's control. Passivity is usually linked with a delusional explanation (e.g., thoughts removed by the government), including: thought insertion—someone is inserting thoughts into one's head; thought withdrawal—someone is removing thoughts from one's head; thought broadcasting—loss of the sense of barrier between one's mind and external world: a delusion that others can hear or know what we are thinking

I—Insight

Understanding of illness and its treatment. We use the 3 'A's':

- Awareness: Do they think there is anything wrong with them?
- Attribution: If something is wrong, do they think it is a physical or psychological problem, or both? How do they describe the problem and what caused it?
- Acceptance: Do they think they need treatment, if so what? What do they think of treatment offered? Is psychiatric treatment acceptable to them?

C—Cognition

Perform detailed cognitive assessment (**Box 30.5**) if you suspect cognitive impairment.

For most patients, use the headings COMIC:

Consciousness level: describe in one word, usually alert; if impaired consider Glasgow Coma Scale or similar.

Orientation by person, place (the ward, floor level, and the hospital) day, month, and year.

Memory: a quick bedside test of short-term memory uses three unrelated words (a colour, a flower (that is not usually associated with that colour) and a place (perhaps a city). Get patient to register the three words, move on to something else, then test recall at 2 minutes. If the score is below 3/3 recall, give a hint: the word is a colour or flower or city, but record that you gave a prompt.

Intelligence quotient: only people with high IQ will give you a number; focus on levels of intelligence you observe (not language sophistication) in the light of patient's educational attainment.

Concentration: subtract serial 7s from 100, list the months of the year backwards; both can be established (time to answer, numbers of mistakes) with interpreters; in patients whose first language is not English, try days of the week backwards.

Box 30.5 Details of cognitive exam

Global	Dominant hemisphere tests	Non-dominant hemisphere tests	Frontal lobes tests
Level of consciousness (if abnormal, use the Glasgow Coma Scale), orientation in time, place, and person Attention and concentration (e.g., test naming months of the year backwards), memory Anterograde short-term ('working') memory tested by immediate recall of three given items Long-term tested by their recall 5 minutes later Retrograde memory includes public (e.g., 'Who is the prime minister?') and personal ('Where were you born?') information, semantic (e.g., how to use a fork) and episodic memory (e.g., 'What happened yesterday?') Global cognition includes IQ (usually estimated rather than formally tested; e.g., 'high', 'low normal')	Language (naming of objects, repetition of a phrase, comprehension of commands, reading and writing) Calculation (simple mathematical tests) Praxis (limb apraxia, e.g., 'Show how you wave goodbye') Finger agnosia, e.g., put a pen in their hand with eyes closed—'What's this?' Conceptual apraxia, e.g., show toothbrush—'What's this used for?') and awareness of details	Neglect (hemi spatial rather than sensory) Construction and visuospatial ability (e.g., 'draw a clock at 3:45'	Verbal fluency (e.g., 'Name as many animals as you can in a minute': tests fluency plus strategy, e.g., listing farm animals first) Similarities (e.g., 'What do a table and chair have in common?'; 'What is the difference between a mistake and a lie?') Estimates (e.g., 'How fast can a leopard run?': patients with frontal lobe lesions typically grossly overestimate) Alternating sequences (copying of alternating hand sequence, which tests sequential motor activity dependent on dorsolateral prefrontal cortex function)

Box 30.6 Collateral history: getting information from others to make/confirm the diagnosis

Index event(s)	Responses, adaptations	Possible psychiatric diagnosis (based on limited examples)
Early experiences of frequent separations, deaths in family, family breakdown	Insecure attachment styles: see Box 30.38	Personality difficulties characterized by problematic relationships; depression
Neurodevelopmental milestones: delayed or never attained; loss of milestones is unusual	Struggles with school readiness, problems learning at school	Intellectual disability (ID) is not a mental disorder but psychological co-morbidities are high in people with ID
Speech delay/difficulties (no single words by aged 2), deficits in social interactions, narrow interests and hobbies	Struggles with social imitative play, making friendships, coping with change at school/home	Autism spectrum disorder
Adverse childhood experiences (ACEs)	Effects of most ACEs can be mitigated with support from one trusted adult	Increased mental disorders—affective, anxiety, psychosis, and borderline traits (see Box 30.40)
School refusal or 'truancy'	Anxiety leading to avoidance behaviours	Anxiety disorders and/or state anxiety
School-based concentration difficulties: poor academic performance despite ability	Symptoms of conduct and emotional disorders; associated hyperactivity	Increased mental disorders—affective, anxiety, and attention deficit hyperactivity disorder (ADHD)
Failure to progress in social learning: empathy, adapting behaviour with age	Disruptive behaviours at school and home, trouble with police, violence	Personality difficulty on the dissocial domain (see Box 30.40)
Early life isolation: imposed on the child or self-imposed	Avoids new situations, favours solitary pursuits	Personality difficulty on the Detachment domain (see Box 30.40)
Late adolescence changes in social functioning: previously outgoing person withdraws over weeks then months	Prodrome symptoms: anxiety, mood changes, sleep and concentration deteriorate	Possible schizophrenia diagnosis
Adult changes in sleep and/or mood	Clarify any increase in alcohol/substances	If no alcohol/substances making things worse, likely depression
Act(s) of self-harm in all age groups	Changes observed before event that are indicative of suicidal intent; increasing isolation, completion of affairs (cancelled holidays, wrote a will), talked of being a 'burden' to others	Some people understate the extent of their difficulties and suicidal intent. Post self-harm, they are safe in hospital and may experience a 'flight into health', saying everything is just fine. Collateral history saves lives
New drug is prescribed: see Box 30.3	Changes in behaviour observed: new insomnia, new responding to voices, rising confusion or paranoia	Likely a transient mood, confusion, or psychotic state (stop the new medication) but remember the medication may have 'brought out' this change in mental state
Missing important events, new irritability due to concentration or subjective memory problems	Memory loss for recent events but intact long-term memories	Dementia—see Chapter 49

Box 30.7 Priorities in physical examination when admitting patients with mental disorders

Signs of acute medical illness: record abnormalities of vital signs (pulse and respiratory rates, blood pressure, temperature, and oximetry).

Any compromise of consciousness level—indicative of brain-based or systemic illness, or related to intoxication/withdrawal of alcohol or substances.

Signs of the effects of self-harm: cuts to limbs or body, bruises (self-harm and/or domestic violence), neck ligature marks (attempted hanging), other injuries including burns, and trichotillomania (hair pulling).

Signs of alcohol or substance misuse (**Box 30.9**).

Medical conditions that can cause/exacerbate affective, anxiety, or psychotic disorders (see **Box 30.8** for investigations).

Co-morbid medical conditions: these are the rule not the exception in people with psychosis/bipolar affective disorder over aged 30. They principally include cardiovascular and chest diseases and diabetes. Detecting these or their proximal risk factors (smoking, obesity, sedentary lifestyle) will prevent future physical morbidity and premature mortality.

Coincidental physical pathology: psychiatric patients do not get more cancers than other people (with the exception of lung cancer due to 40% smoking rates in people with psychosis) but their cancers are detected late and they are less likely to be offered treatments than the general population.

time-sensitive, and therefore require regular updating. Comprehensive risk formulations should include risk to self, risk to others (e.g., children), and risk *from* others. Each category should consider historic, present, and anticipated risk, taking into account protective factors (**Fig. 30.1**). Collateral information (**Box 30.6**) should be obtained, and neglecting this will undermine a full risk assessment. The whole process needs to be collaborative in nature and can have therapeutic potential. Risk assessments would ideally conclude with a comprehensive risk management plan and safety planning. Structured tools may be used, though these have not been shown to be useful in predicting suicide risk; they are most commonly used in forensic settings.

Treatment approaches in psychiatry

A wide range of evidence-based treatments are available for mental disorders. Most of these are delivered outside the general hospital, with many positive actions within the control of most patients. It is simplistic to consider psychiatric interventions as medications and talking therapy: for most conditions, a stepped approach to treatment is adopted of which medications are one part, having

Box 30.8 Investigations of a new patient for psychiatric admission, or at clinic

These routine tests should be performed on all or most patients being assessed in the context of a new or acute mental health problem, to screen for physical illness that might be undiagnosed or disease contributing to the disturbance of mental state

	Abnormality	Potential causes in patients with mental illness
Full blood count	Anaemia	Dietary deficiency, malabsorption or blood loss from any source e.g., gastrointestinal bleed
	Infection	Urine, chest, blood-borne–virus infection in intravenous drug users, opportunistic infection in those with HIV
	Neutropenia	Clozapine side-effect
	Thrombocytopenia	Alcohol excess
Urea and electrolytes	Acute kidney injury	Dehydration through self-neglect, nephrotoxic drug overdose
	Hypokalaemia	Vomiting or diuretic/laxative abuse in eating disorders
	Chronic kidney disease	Lithium Poorly treated medical co-morbidities, e.g., diabetes
Thyroid function	Hypothyroidism and hyperthyroidism	Thyrotoxicosis may mimic mania or psychosis Hypothyroidism is a differential diagnosis for depression
Random glucose, HbA_{1C}	Hyperglycaemia	Undiagnosed or poorly controlled diabetes
Liver profile	Raised transaminases	Alcohol excess, paracetamol overdose, medications, hepatitis
Urinalysis	Nitrites and leucocytes	Urinary tract infection causing sepsis and delirium in older adults
	Blood, protein, or glucose	Undiagnosed chronic kidney disease or diabetes
	Pregnancy test	Unknown/undisclosed pregnancy
Urine drug screen	Undisclosed substances (benzodiazepines, cannabis, methadone)	Extent of substance misuse (if minimized) may lead to reformulation of diagnosis
Electrocardiogram	Features of ischaemia	Undiagnosed ischaemic heart disease
	Prolonged QT interval	Urgently review psychiatric medications + medical referral

Additional tests may be indicated in some patients undergoing psychiatric evaluation as directed by specific features of the history and mental state examination

	Screening for	Actions
Lithium level	Low in non-adherence, high in toxicity e.g., renal failure	Review lithium dose, consider need for emergency treatment if toxicity
Drug levels for toxicity	Clozapine, other mood stabilizers, paracetamol (in overdose)	Keep within therapeutic ranges, specific treatments for toxicity

Box 30.8 Investigations of a new patient for psychiatric admission, or at clinic—cont'd

	Screening for	Actions
Cortisol	Cushing's disease	Refer to endocrinologist
Calcium	Hypercalcaemia: can disturb mental state	Consult endocrinology for parathyroid tests; rule out malignancy and myeloma
Phosphate, magnesium, calcium, vitamins, folate	Nutritional deficiencies in eating disorders, severe depression and dementia	Eating disorders emergencies are managed in general hospital then under the care of a specialist eating disorders service
Blood-borne viruses	Curable hepatitis C; treatable hepatitis B and HIV infections	Refer to hepatology, HIV, or general services
Creatinine kinase	Neuroleptic malignant syndrome: rise is CK is usually mirrored by a parallel rise in neutrophils	Refer to specialist medical/intensive care specialists for treatment
Chest X-ray	Cancer, lower respiratory tract infection, especially in smokers	Specialist respiratory referral
Brain scanning (CT and MRI)	Intracranial bleeds, trauma, CNS infection including abscess, neoplasm	Specialist neurology/neurosurgical referral
Lumbar puncture	Infection, encephalitis (see Chapter 28, 'Encephalitis' section)	Seek specialist neurological advice
Electroencephalogram	Episodes of abnormal mental state, e.g., psychosis of temporal lobe epilepsy or delirium	Specialist neurology referral. Prescribing antipsychotics will raise the seizure threshold

Box 30.9 Physical examination and signs of alcohol and substance misuse

Features of alcohol misuse
- Altered consciousness (intoxication withdrawals, hepatic encephalopathy)
- Smell of alcohol on breath
- Plethoric facies, jaundice, anaemia
- Chest diseases: tuberculosis, aspiration pneumonia
- Hand tremor, palmar erythema, spider naevi, irregular or fast pulse, high blood pressure, peripheral neuropathy
- Abdominal tenderness (gastritis, pancreatitis, hepatitis)
- Liver: tender and enlarged (alcoholic hepatitis), distended abdomen of ascites
- Lower limbs: check for balance (ataxia of cerebellar damage), muscle weakness and peripheral neuropathy

Features of substance misuse
- Altered consciousness level: depressed (opiates), increased (amfetamines and stimulants)
- Opiate/opioid withdrawal: lacrimation, rhinorrhoea, sweating
- Abnormal pupils (pinpoint in opiate intoxication, dilated in withdrawal)
- Damage to nasal passages or septum from 'snorting' cocaine
- Poor dentition
- Lymphadenopathy (opportunistic infection including HIV)
- Track marks of intravenous drug use (**Fig. 30.8**)
- Chest diseases: emphysema (tobacco/cannabis use), lung abscess (smoking heroin)
- Liver tenderness, enlarged (hepatitis)
- Groin: track marks of intravenous drug use, abscess, and sinus formation
- Proximal muscle weakness

Box 30.10 Common physical morbidities in patients with mental disorders excluding the effects of alcohol, substances, and self-harm

- Cognitive impairment: apparent (negative symptoms of psychosis, depression) or objective, e.g., dementia
- Anaemia, poor dentition, eye damage from diabetes
- Chest diseases: chronic obstructive pulmonary disease (COPD), pneumonia, tuberculosis—all three are increased in smokers
- Irregular or fast pulse (atrial fibrillation), high blood pressure, increased limb tone of high neuroleptic doses
- Obesity: compare weight with previous records, record silent effects on liver (obesity is 'diabetes in the liver') and the locomotor system (osteoarthritis); look for signs of recent weight loss (skin flaps, reduced muscle mass)
- Constipation: clozapine, opioids, anticholinergic medications
- Lower limbs: check peripheral pulses and for peripheral neuropathy and skin diseases of diabetes; signs of venous thromboembolism (hospital admission)

instituted other measures. **Box 30.11** sets out early universal interventions (steps 1–3) available to community teams, followed by more specialist treatments provided by experts (steps 4–7). Psychotherapy choice often depends on what is available locally, waiting lists, and whether the patient can commit time and money. In the UK, other than cognitive behavioural therapy (CBT), most therapies are not provided free.

It is good practice to communicate with the therapist about the modality of therapy: neither CBT nor psychodynamic approaches (involving long treatment times, examining past patterns, with less direct 'advice' than CBT) will suit all patients. Getting therapy right can be life changing for patients.

Prescribing in psychiatry

General principles of prescribing in psychiatry are covered in **Boxes 30.12 and 30.13**, with prescribing in the major psychiatric diagnoses covered in detail in the relevant sections later in the chapter.

Common classes of psychotropic medications include antidepressants, antipsychotics, and anxiolytics, although there is considerable overlap between classes and drugs formerly licensed for one indication sometimes prove effective in other contexts, for instance selective serotonin reuptake inhibitors, originally intended as antidepressants, are now commonly used to treat severe post-traumatic stress disorder and obsessive–compulsive disorder. Some medications prescribed for physical health indications (e.g., analgesics, antiepileptics, and some antiemetics) have psychotropic effects which can sometimes be used therapeutically—though they may have unwanted side-effects (**Box 30.3**).

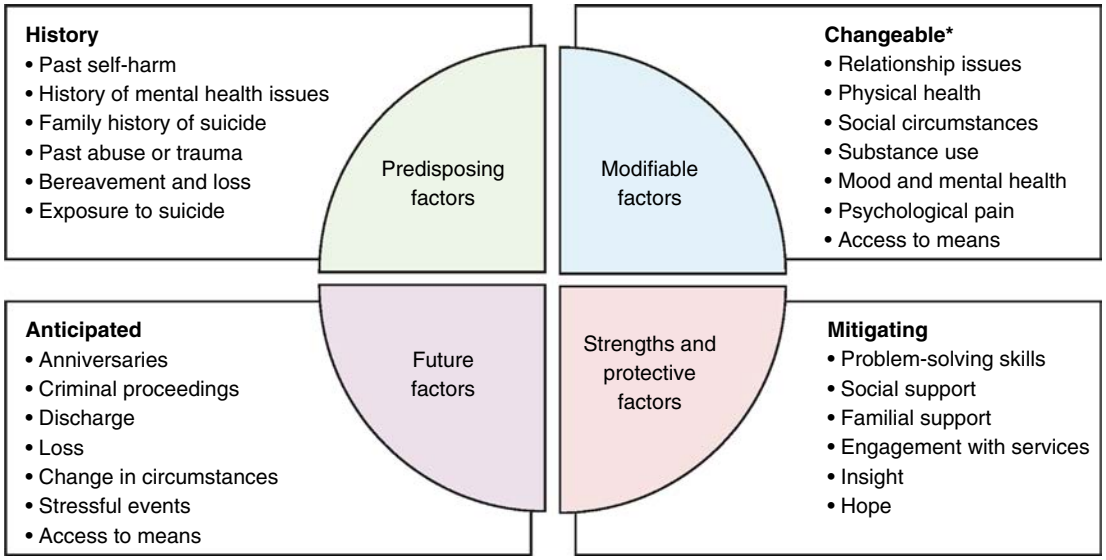


Fig. 30.1 Interactive components of risk assessment that can inform risk formulation and therapeutic management. *Acute exacerbation of any of these factors may constitute warning signs. (*From S2215036622002322/gr1.*)

Box 30.11 Treatment strategies in managing mental health conditions and other mental distress

Approach	Description	Examples within mental disorders
1	Therapeutic assessment: communication of formulation (verbally and by letter to professionals who need to know with a copy to the patient) in a useful and culturally sensitive way; plus agreed treatment plan to reduce symptoms	Diagnosing an anxiety disorder as the cause of symptoms, leading to simple advice on sleep hygiene, reducing caffeine and alcohol, gentle exercise, and renewing social networks Explaining distressing symptoms: feeling of guilt in depression; hallucinations as 'dreaming while you are awake' Diagnosing and explaining functional gastrointestinal disorders and offering diet and lifestyle advice
2	Self-management: making changes that will persist over time; engaging the patient by linking goals to their values (family life, health)	Stopping medications ('deprescribing') that are causing distressing symptoms Encouraging patients to stop behaviours that make them ill: quitting cigarettes/substances, avoiding unhealthy foods, finding alternatives to self-harm Specific advice on benefits/housing, signposting to other agencies; for some, a change of environment or meaningful work/study will drive recovery
3	Social prescribing: this requires understanding of what matters most to the patient (not 'what's the matter with that patient?') to empower them to take positive action; it requires local knowledge of what is available	Non-health interventions that will enhance the quality of life of people with mental disorders include walking groups, clubs, joining a choir, community projects, and volunteering; green social prescribing combines outdoor exercise/activity with climate action (this may be a core value in some patients) Mindfulness is a powerful anti-anxiety activity Such non-pharmaceutical interventions enhance patient experience, improve population health, reduce costs, and improve the work life of healthcare clinicians and staff
4	Psychotherapy: any intervention between patient and therapist (including non-verbal therapies such as music, art, and dance therapy) to explore the context of their distress. There are more elaborate models in better-resourced areas: hypnosis, or dialectical behavioural therapy (DBT) for borderline PD	Individual includes CBT, interpersonal therapy, psychodynamic therapy, or the third wave of therapies which combine components of these approaches, such as acceptance and commitment therapy (ACT) Modified therapies exist for specific disorders, such as focussed CBT in eating disorders (CBT-ED) Group includes family therapy, couples counselling, psychodynamic groups, or peer-led for addictions including Alcoholics and Narcotics Anonymous Less common are therapeutic communities where communal living is supported with therapies
5	Medication prescribing	These are covered within individual subsections of this chapter: prescribing is not the first step in treatment

Box 30.11 Treatment strategies in managing mental health conditions and other mental distress—cont'd

Approach	Description	Examples within mental disorders
6	Psychosocial treatments, and other treatments difficult to classify	For traumatized patients, there is eye movement desensitization and reprocessing (EMDR): patients move their eyes in a specific way while processing their traumatic memories Graded exercise therapy (GET) is delivered by physiotherapists with a proven value in chronic fatigue syndrome in the PACE trial
7	Collaborative care (a model where prescriber, therapist, and care coordinator work together to treat one patient) and integrated care (physical and mental healthcare are combined with social care)	Depression may become a long-term relapsing disorder, and collaborative care models show the best results for the reduction of symptoms, prevention of relapse and promotion of positive long-term outcomes; most community mental health teams have care coordinator models. How psychiatric services are organized varies across regions
8	Physical treatments other than medication: by nature, these are expensive, so higher up the stepped care model	Electroconvulsive therapy (ECT) has the best evidence base, and has been shown to be effective in depression, catatonia, and treatment-resistant psychoses, for example in Parkinson's disease or postpartum psychosis Transcranial magnetic stimulation (TMS) and deep brain stimulation (DBS) both have an emerging evidence base
9	Last resort: psychosurgery. This has an undistinguished history in psychosis	In severe OCD, controlled trials (randomized due to small numbers) show a modest evidence base for stereotactic ablative procedures of the subcaudate nucleus and associated areas

Box 30.12 Considerations when prescribing psychotropic medication

- Allergies: always check patient records for documented allergical reactions
- Age: adjust dose downwards, particularly in young and old, intellectual disability
- Adherence: special caution in patients with limited insight and/or cognitive impairment
- Formulation/preparation: (from preferred) liquid, orodispersible tablets, capsules, (to novel) inhalers and (least preferred) injection, depot, and suppositories: consider patient's needs/preferences
- Route of administration: oral, nasal, intramuscular, intravenous routes usually all available. Review impact on absorption and severity of side-effects
- Side-effect profile: take into consideration personal preferences and circumstances, as well as other co-morbidities (e.g., potential to lower seizure threshold in people with epilepsy or increased risk of bleeding [SSRIs] in people undergoing surgery)
- Therapeutic range: when possible, monitor drug levels to avoid under- or over-medicating and potential toxicity
- Fertility and reproduction: consider teratogenic effect of certain medications (sodium valproate, lithium)
- Pharmacodynamic and pharmacokinetic interactions: ask for a comprehensive list of prescribed, over-the-counter medication and acquired medications
- Co-morbid substance misuse: impact of recreational drugs on mental state, efficacy, and side-effect profile of medication. Consider street value and abuse potential of prescribed medication (pregabalin, quetiapine)
- Suicidality: supply only necessary amount of medication and consider supervised administration by family

Box 30.13 Potential adverse events from rapid tranquilization

- Oversedation (if severe, fatal respiratory arrest)
- Loss of consciousness
- Arrhythmias
- Cardiac arrest
- Aspiration
- Confusion
- Ataxia
- Falls
- Extrapyramidal side-effects
- Hypotension
- Seizure
- Psychological impact of restraint and forced medication

Further reading

Byrne P, Byrne N. *Clinical Cases Uncovered: Psychiatry*. Wiley: 2008.
Byrne P, Byrne N. *Psychiatric Assessment in Hutchinson's Clinical Methods*. 25th edn. 2022. **Glynn M, Drake W (eds).**
Taylor DM, Gaughran F, Pillinger T. *The Maudsley Practice Guidelines for Physical Health Conditions in Psychiatry*. John Wiley & Sons Limited. 2020.

BACKGROUND

Mental health conditions are common, affecting one in four people in the general population over a lifetime and an even higher proportion of hospital inpatients. They coexist with medical or surgical problems, and often complicate treatment. Suicide remains the leading cause of death in young people, and over a life course, poor mental health drives higher mortality and morbidity, much of which is preventable with effective care.

LIAISON PSYCHIATRY

Mental health professionals can become involved at every stage of a patient's journey within the hospital, from the emergency department to inpatient wards and beyond. Liaison psychiatry (LP) is a subspecialty within psychiatry which is dedicated to the interface between physical and mental health, addressing the psychosocial needs of patients during and around inpatient admissions. Sometimes LP services will remain involved in a person's care through

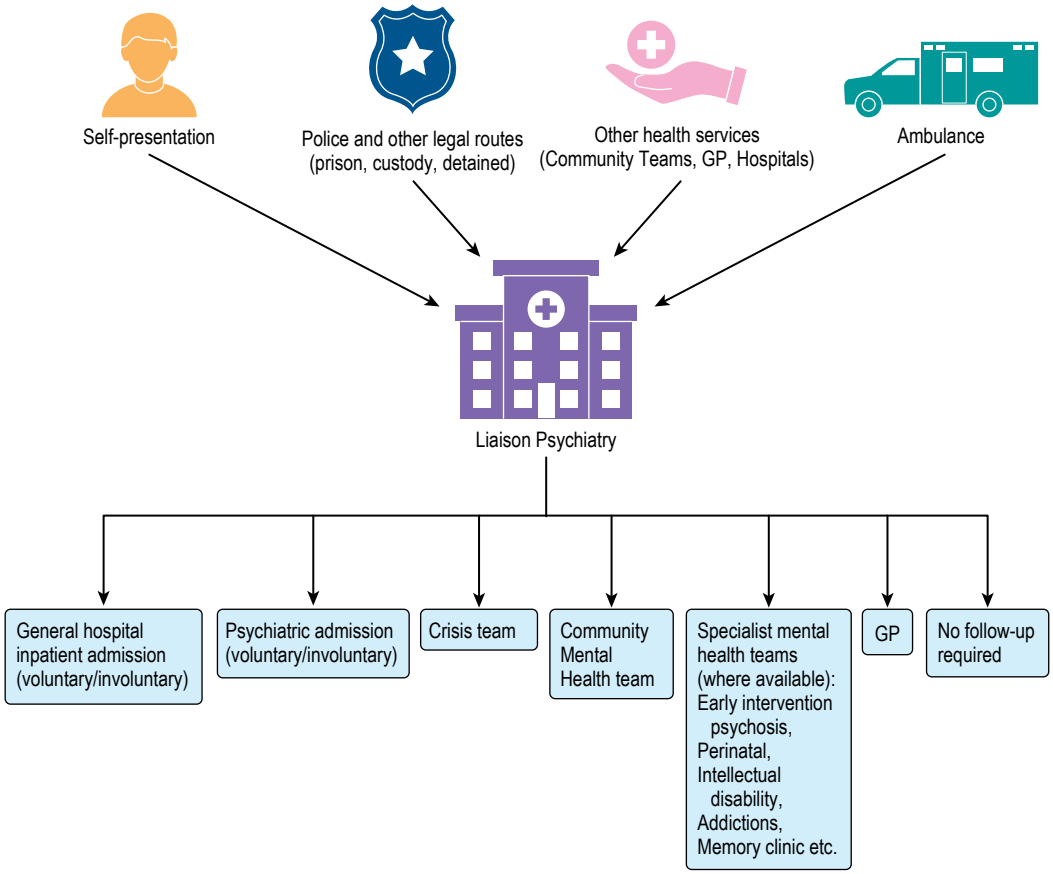


Fig. 30.2 **Liaison psychiatry and its place within other mental health services.**

outpatient clinics—perhaps run jointly with medical specialties—after discharge from hospital, or to provide alternatives to admission.

Given the time-limited nature of LP's involvement in a patient's care, it is important for these teams to establish communication channels with treating teams in the community early on in a patient's admission. This ensures all pertinent clinical information is available, so that treating clinicians can make safe decisions (Fig. 30.2). This form of collaborative approach ensures smooth transition and handover to community services when patients are ready for discharge from inpatient medical care.

The liaison psychiatry team

LP teams follow a multidisciplinary clinical model. While most teams are comprised of medical professionals specializing in psychiatry and nurses specializing in mental health, better-resourced teams have specialist psychologists, pharmacists, social workers, occupational therapists and addiction specialists working in the same team. There is a systematic under-recognition and undertreatment of the health needs of people with mental illness, and they tend to receive lower quality treatment for various other physical illnesses including cardiovascular disease, diabetes, and cancer than do people without mental illness. LP teams are therefore an essential part of a general hospital, as they ensure equal, equitable access to healthcare and provide safe, evidence-based care for all patients with psychiatric needs.

Special skillsets are required when assessing/managing patients with learning difficulties or other neurodevelopmental disorders, patients with cognitive impairment, as well as those under forensic services. LP teams might then seek the assistance/opinion of specialist mental health teams (Fig. 30.2) who might assess patients in the general hospital. The liaison component of LP teams is about learning from referrers (clinicians working in medical or surgical specialties who understand the context of their patient's physical illness), talking to community teams, perhaps their family doctor, and giving verbal feedback to referrers after interventions.

LP manages a wide range of mental disorders at any one time. People with schizophrenia are more likely to be admitted to a general hospital than a psychiatric one, with cardiorespiratory and metabolic diseases over-represented in this population. Two-thirds of people over 65 in a general hospital will have depression and/or dementia. One in five of the general population will have symptoms of depression or anxiety in any given year, and while most will never seek or receive treatment for this, they are likely to seek medical attention for related physical symptoms. Anxiety often co-occurs with low mood. Identifying how high stress levels impact on function, decision-making, and overall health is one of the core skills of every practising doctor and cannot be delegated to mental health specialists.

ETHICAL AND LEGAL ISSUES IN PSYCHIATRY

Consent and confidentiality

These key concepts underpinning the ethical and legal practice of medicine are discussed more generally in Chapter 3. In the context of mental illness and/or physical disease affecting the brain, there may be additional considerations: such illness may affect capacity to consent to treatment, and maintaining confidentiality may cause risk to the patient or others. A clinician should assure patients of confidentiality, but may tell a patient with a mental health condition, '*if, you were to tell me you have a plan to harm yourself or someone else, I could*

not keep that a secret and would have to talk to your doctors and others about how we could help prevent this'. That clinician records: 'I chose this course of action (to disclose my findings) because…'

As elsewhere in medicine, mental capacity to consent to treatment rests on the patient being able to **understand** what is said to them, recalling and **repeating back** this information, demonstrating that they are **weighing** key information, and **communicating** their decision to those around them. Capacity assessments can be carried by anyone familiar with the local code of practice, though these are usually carried out by professionals working in the health and social care services. LP teams are sometimes asked to advise in complex cases, for example, where there is uncertainty about whether a mental health condition is compromising a patient's capacity to consent to treatment for a physical health problem. In some patients with severe mental illness who are detained under legislation for the protection of themselves or others, mental capacity may be present for some interventions, but absent for others: for example, they may lack capacity to refuse a mental health assessment.

Confused patients usually understand the context of a hospital doctor who asks questions and performs a limited physical exam. A *best interest's decision* (where risks of no assessment outweigh the right to privacy in context of current situation) can be made in these circumstances. Clinicians should document *why* they chose to act, and if uncertain, talk to colleagues (and perhaps the patient's family)—always recording their reasons. Clinicians sometimes defer aspects of the physical examination in a refusing patient who lacks capacity; this is risky in an unknown patient and may lead to harm. Clinicians must document *why* they went ahead with inspection (e.g., for concealed substances or weapons) or examination (looking for injuries from self-harm, assessing vital signs) in these instances.

In most jurisdictions, mental capacity legislation exists to protect patients who lack capacity from potential abuse. For this legislation to be enacted, treating clinicians should document the disorder of brain (e.g., delirium) and/or of mind (mental disorder) that is believed to be causing impairment of capacity. For example, a depressed patient might decline lifesaving interventions, and disclose to a junior member of the team that this is because they believe they do not deserve to live. Most clinicians would judge that depression (which is treatable) has led to a temporary loss of capacity, and proceed with treatment. When assessing patients in a general hospital or clinic, it is prudent always to make clear as you begin that aspects of the assessment may be shared with other staff, starting with those who made the referral. Capacity assessments are decision specific, time-sensitive, and should take the patient's best interests into account.

Further reading

Owen GS, Spencer BWJ, Kanaan RAA. The legal and ethical framework for compulsory psychiatric treatment. *Medicine* 2016; 44(12):734–736.

Patients wanting to leave hospital

Patients sometimes change their minds about staying in hospital—choosing to leave the emergency department, declining admission to hospital, or attempting to self-discharge from the ward. It can be challenging for responsible clinicians to decide whether patients have mental capacity to decide this, especially in the context of an underlying mental disorder.

A comprehensive approach will consider the many factors potentially at play:

- Why did the patient attend this time (symptoms, diagnosis)?
- Have there been previous similar admissions?
- Might there be factors in the environment (noisy, lack of privacy, loss of autonomy) that provoked this change?
- Was it a procedure (blood test, lumbar puncture) that changed things?
- Could there be unpleasant physical symptoms (withdrawal from alcohol or substances; agitation from confusion) that escalated non-cooperation?

Patients may experience strong emotions ranging from anxiety, through agitation to outright hostility and even violence. Patients who have self-harmed may have been brought to hospital by others' decisions, or attended only for relief from a physical symptom (headache following overdose, bleeding from cuts to arms) and not for 'treatment' of self-harm: proper treatment is both to treat the physical effects of self-harm *and* address the psychosocial factors that led to it.

There are three essential areas to consider in patients who wish to self-discharge:

- **What is their mental state?** Perform a mental state exam (see later in chapter) to assess for confusion/cognitive impairment; significant anxiety; low or high mood; suicidality (thoughts, intent, and plans); psychotic symptoms; and insight into physical and mental disorders.
- **Do they have capacity to make this decision?** Can they understand what treatments are on offer, the consequences of refusing these, and the implications of self-discharge? Are they a risk to themselves (self-harm, neglect, exploitation by others) or a risk to others (children, partner, others)?
- **Who needs to know?** Which other professionals need to know about the presentation, the consequences of self-discharge, and the backup plan? It may be necessary to call the patient's relative or partner. It does not break confidentiality to seek information about a patient, even if permission to contact others is declined. The minimum output is to talk to the overall doctor in charge and write to the patient's general practitioner, usually copied to the patient.

All of the scenarios above (consent, capacity, refusing treatment, etc.) are core skills of every doctor, and cannot be delegated to LP teams given how often they arise.

The legal framework for mental health treatment

Autonomy is one of the four principles of biomedical ethics, alongside non-maleficence, beneficence, and justice (see Chapter 3). In practice, it means that patients cannot be forced or manipulated to make a treatment-relevant decision. The good intentions of clinicians are no defence in law and it is only in exceptional circumstances that treatment may be imposed—always under a strict and well-defined code of conduct.

The United Nations and the European Union have set out relevant international legal frameworks, and these helped shape and influence mental health legislation in various countries around the world, although national frameworks may have significant differences in their definitions and applications. Many countries distinguish consent to psychiatric treatment from other medical treatments. Capacity is a threshold concept: it is present or it is not. There are instances where clinicians can override decisions even in cases where capacity is maintained (e.g., in the UK a patient with depression may be deemed to have capacity to refuse admission to hospital, but still be detained under the Mental Health Act because suicidal behaviour is thought to represent a significant threat to their life).

An involuntary admission to a psychiatric hospital may be indicated, against a patient's wishes but in the interest of their health, in order to protect their safety and the safety of others (**Fig. 30.1**

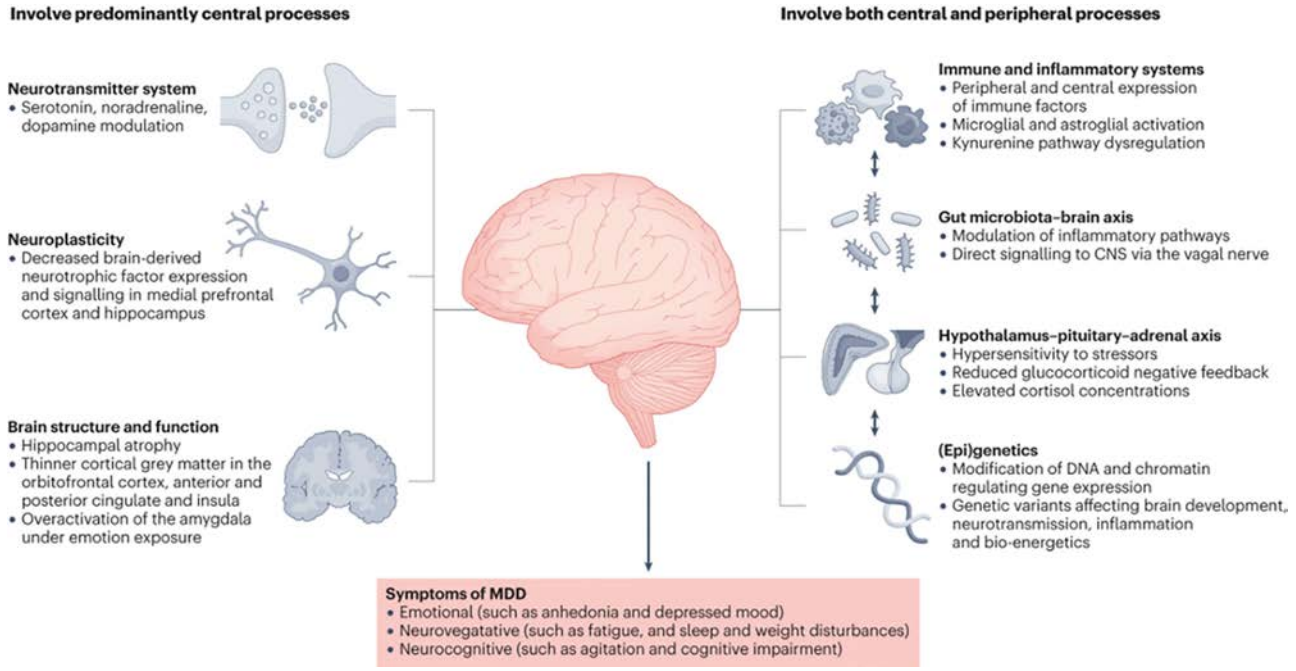


Fig. 30.3 Aetiology of depression. (*W Marx, B Penninx, M Solmi et al. (2023) Major depressive disorder.* Nature Reviews disease primers *9, 44*.)

and detailed in the 'Managing suicidal patients in healthcare settings' section). These difficult decisions are based on actual behaviours and measures (e.g., dangerous to others), and not merely on diagnosis. Legal protections incorporate the right to independent review, the right to appeal decisions, the right to an advocate, and equality before the law: that someone without this mental disorder would be treated the same.

Further reading

Henson VL, Vickery DS. Patient self-discharge from the emergency department: who is at risk? *Emerg Med J* 2005; 22:499–501.
Okai D, Owen G, McGuire H et al. Mental capacity in psychiatric patients: systematic review. *Br J Psychiatry* 2007; 191(4):291–297.
Owen GS, Spencer BWJ, Kanaan RAA. The legal and ethical framework for compulsory psychiatric treatment. *Medicine* 2016; 44(12):734–736.
Schneeberger AR, Kowalinski E, Fröhlich D et al. Aggression and violence in psychiatric hospitals with and without open door policies: a 15-year naturalistic observational study. *J Psychiatr Res* 2017; 95:189–195.

AETIOLOGY OF MENTAL DISORDERS

The aetiology of psychiatric disorders is multifactorial, transcending the now discredited mind–body divide. Understanding the underlying factors is crucial in devising a safe and effective management plan, as most commonly patients present with multiple co-morbidities. Failing to identify and treat these leads to poorer outcomes and perpetuates longstanding health inequalities with detrimental consequences for patients.

First proposed by George Engels five decades ago, the biopsychosocial model remains a useful way of understanding mental illness, though it has its challengers. According to the biopsychosocial model, mental illness can be perceived as the combined effect of biology (genetics, neurotransmitters, hormones, substances, physical diseases), psychological factors (emotions, attitudes, perceptions, personality type), and social components (income, relationships, family, culture, subculture, society). Another emerging aspect tightly linked with the social aspect of the model is environment, including social media and other influences on identity. These overlap, compound/reinforce each other, and are interlinked.

Biological mechanisms

The brain is what makes us sentient and self-aware: its processes and external influences are limitless. Depression is the most common mental health disorder encountered in primary care, and modern concepts of brain abnormalities in depression describe both central and peripheral mechanisms (**Fig. 30.3**). One of the challenges of biological psychiatry is that these changes are difficult to quantify, vary over time and vary between individuals. While some present targets for pharmacological treatment (e.g., abnormal serotonin signalling that underlies use of selective serotonin reuptake inhibitors), there is no support for the proposition that depression is a serotonin deficiency state. Such changes in biomarkers may equally be the *effects* of depression, especially if prolonged. Many researchers argue that dopamine changes are as important in depression as they are in psychosis, and that serotonin pathways are even more abnormal in obsessive–compulsive disorder.

The wider context of mental illness

Looking beyond narrow diagnostic categories, the *Perspectives of Psychiatry* approach considers patients from four perspectives: disease, dimensional, behaviour, and life story (**Box 30.14**). Many environmental factors have been identified that drive physical illnesses, and these may play an even stronger role in the development of mental disorders. Most determinants of health, both good and bad health, are outside the scope of patients' interactions with hospital services (**Box 30.15**).

Epidemiology of mental illness

Mental disorders were rising in prevalence before COVID-19, and the inequalities of the pandemic increased that rate further, disproportionately affecting the poorest in society. Preventing the onset of mental disorders across the lifespan and reducing their severity, duration, and associated harms are considered essential to a worldwide drive for sustainability (see 'Prevention of mental illness: public mental health' section).

Box 30.14 Dimensions of mental illness

The HIDE paradigm	Perspectives of	Examples
What the patient **H**as	Diseases	Medical diagnoses
Who the patient **I**s	Dimensions	Character, personality
What the patient **D**oes	Behaviours	Health + risk behaviours
What the patient has **E**ncountered in life	Life stories	Individual trauma, losses; family stories of doctors

Dimensions of illness from PR McHugh and PR Slavney (1998) The Perspectives of Psychiatry. *Baltimore MD: John Hopkins University Press.*

Box 30.15 Social determinants of health

Health determinants →	↑↑ mental illness
○ Income, social status	○ Poverty, high crime area
○ Social support networks	○ Isolation (lonely)
○ Education + literacy	○ Lower education level
○ Employment/working conds	○ Low/no employment
○ Physical environments	○ Urban birth, residence
○ Social environments	○ Lack of social support
○ Personal health practices, coping skills	○ Sedentary, substance misuse
○ Healthy child development	○ Low birth weight, parents with mental health conditions or substance misuse
○ Biological, genetic factors	○ Inherited physical conditions
○ Health services	○ Problems with access to and discrimination within health services
○ Gender; culture/subculture	○ Example: double the rates of depression in postmenarche girls and women

COPMI, Children of parents with mental illness—where environmental factors usually exceed genetic aetiology (and are therefore points of intervention to prevent mental disorders in the children).

Physical health in patients with mental illness

Despite significant health spending, many US citizens with mental disorders get no treatment for them, and people with schizophrenia there frequently end up homeless, at-best accessing emergency care only. Inequalities (differences in a society between the very rich and its poorest citizens) explain much about the clustering of diseases and the large differences in outcomes between rich and poor.

One study of the 1.14 million people in the USA diagnosed with schizophrenia (patients engaged in health services) found that a diagnosis of schizophrenia increased mortality 3.5 times compared to the general population. People with schizophrenia mostly died young, with each death reflecting significant number of *years of potential life lost* had they survived to normal life expectancy:

- 38.3 potential life years for each death from suicide (more likely in first 5 years of psychotic illness—young adults who die by suicide lose many years of life)
- 29 years in deaths from liver diseases
- 27.3 years in deaths from diabetes
- 25.6 years in deaths from ischaemic heart disease
- 24.9 years in deaths from stroke, and
- 24 years each in deaths from COPD (see Chapter 20) and lung cancer.

Health inequalities (see Chapter 9) are high in the USA but rising fastest (among Western countries) in the UK. Since the 1990s, these two countries have also described the largest increases in preventable deaths in people with severe mental illness. Within individuals, poorer physical and mental health exert synergistic effects. Children living in poor families are more likely to experience adverse childhood experiences (ACEs), and these compound pre-existing inequalities and increase the risks of future smoking, harmful alcohol use, substance misuse, unhealthy diet, sedentary lifestyle, mental disorders, suicide and self-harm, and ultimately preventable deaths.

Further reading

Byrne P. Premature mortality in people with severe mental illness: a renewed focus for our new era. *Ir J Psychol Med* 2023; 40:74–83.
Henderson C, Noblett J, Parke H et al. Mental health-related stigma in health care and mental health-care settings. *Lancet Psychiatry* 2014; 1(6):467–482.
Lund C, Brooke-Sumner C, Baingana F et al. Social determinants of mental disorders and the Sustainable Development Goals: a systematic review of reviews. *Lancet Psychiatry* 2018; 5(4):357–369.
Marmot M. *The Health Gap*. London: Bloomsbury. 2015
Olfson M, Gerhard T, Huang C et al. Premature mortality among adults with schizophrenia in the United States. *JAMA Psychiatry* 2015; 72(12):1172–1181.
Peters ME, Taylor J, Lyketsos CG, Chisolm MS. Beyond the DSM: the perspectives of psychiatry approach to patients. *Prim Care Companion CNS Disord* 2012; 14(1):PCC.11m01233.

COMMON PSYCHIATRIC EMERGENCIES

Mental disorders can precipitate various emergency presentations leading to physical illness and risk to the patient and others. A clinical approach to patients presenting with a mental health emergency involves gathering information, assessing mental state, arriving at a formulation (why is this patient presenting, right now?), and investigating physical health to clarify what risks are present, and which ones need immediate attention. Of the five emergency presentations considered in more detail below, while suicidal patients who do not require medical treatments are managed in psychiatric settings, many presentations are medical emergencies requiring immediate medical care in a general hospital alongside management by mental health professionals. What doctors do, especially prescribing (**Boxes 30.3 and 30.12**), sometimes *causes* some of these emergencies—and merits careful consideration of proximal events.

Self-harm and suicide

Self-harm refers to an intentional act of self-poisoning or self-injury and is a clear sign of emotional distress. It includes both suicidal and non-suicidal attempts with behaviours such as cutting, burning, hanging, stabbing, drowning, and jumping from heights, among others. The Office for National Statistics in England estimated the male suicide rate to be 15.8 per 100 000 in 2021, compared to a female suicide rate of 5.5 per 100 000; globally suicide rates are higher in men than women. It is thought that the overall higher rate of suicide in men—despite the higher rate of self-harm in women—may be due to the choice of more lethal methods: firearms and hanging in men versus cutting and overdose in women.

Suicide remains a major public health concern, with 817 000 deaths by suicide reported internationally in 2016. The Global Burden of Disease Study estimated it to be the 15th leading cause of years of life lost, alerting the World Health Organization to make suicide reduction one of its main aims, aspiring to reduce suicide mortality by one-third by 2030. In 2021, a nationally representative

survey of adults in the USA showed that people identifying as LGBTQ+ (lesbian, gay, bisexual, trans, and queer) are more likely to report suicide-related thoughts, plans, and attempts within the past 12 months compared with those identifying as heterosexual. While all age groups are affected, suicide is the fourth leading cause of the death in 15–29 year olds. Epidemiological studies suggest that a significant proportion of people dying of suicide have a psychiatric disorder, with affective and psychotic illnesses being the most frequent. The suicide rate for people with schizophrenia spectrum disorders is over 20 times higher than the general population. Other co-morbid medical conditions, including chronic pain, stroke, epilepsy, and head injury, also increase suicide risk.

Managing suicidal patients in healthcare settings

Emergency department (ED) clinicians are crucial in the management of patients at high suicide risk, as suicide attempts and self-harm are most often first treated in a hospital's ED. In the USA, it was reported that for every death by suicide, there were eight preceding ED visits. The risk of death is highest within 30 days of discharge from an ED or from psychiatric inpatient care.

LP teams are best placed to provide detailed mental health and risk assessments (Fig. 30.1) that determine the next steps in a patient's management. This can include crisis interventions delivered in the community (Home Treatment or Mental Health Crisis Teams; or urgent follow-ups with Community Mental Health Teams; Fig. 30.2); alternatively, psychiatric inpatient admission may be required for patients at increased risk of completing suicide. Sometimes, admission to a medical or surgical ward is required, for example, in patients requiring an acetylcysteine infusion following paracetamol overdose, and close working between LP teams and other specialties is required to support people presenting with self-harm. The decision to detain non-consenting patients under appropriate legal framework is difficult, and its timing (in relation to medical treatments) is part of this.

A good practice point in treating people from at-risk groups is to explore suicide options with open questions (be careful not to introduce a novel suicide method into the conversation) when a crisis is lifting. *What might they do in the future if overwhelmed by thoughts to end their lives?* Encourage positive strategies (phone a friend, be with people, attend a doctor or ED), and suggest other 24/7 supports: national suicide prevention lines, local resources, and bespoke services (e.g., those run by the LGBTQ+ community). Encourage them to save these numbers on their phones.

Further reading

Hawton K, Lascelles K, Pitman A et al. Assessment of suicide risk in mental health practice: shifting from prediction to therapeutic assessment, formulation, and risk management. *Lancet Psychiatry* 2022; 9(11):922–928.

Acute disturbance

Acute disturbance (AD) is an encompassing term describing escalating behaviours secondary to a change in mental state, including restlessness, agitation, aggression, and violence as well as verbal or non-verbal threats. Its aetiology is complex and includes psychiatric or other physical disorders including, yet not limited to, delirium and dementia, psychosis, stroke, and drug-induced states. A 2015 survey indicated that up to 80% of emergency department (ED) staff experienced workplace violence in the past year, with the majority of these incidents perpetrated by patients. A meta-analysis found up to 1 in 5 psychiatric inpatients (over a lifetime) may commit a violent act by causing harm or attempting to cause harm to another person. Violence and aggression in healthcare settings risk physical and psychological harm to the patient, staff, other patients, as well as to families. This applies to a range of disciplines and while emergency medicine and psychiatry are most heavily impacted, evidence suggests its effect is widespread, including renal units and geriatric and neurosurgery wards, among others.

In the first instance, if clinically appropriate, de-escalation should be used and this includes both verbal and non-verbal techniques. Building a personal connection with the patient, gently repeating their preferred name, actively listening, using non-threatening language, attempting to understand their circumstances, and offering choices may all help avert escalation to aggression or violence. Acknowledge they are upset even if you do not need to agree with their stated reasons for this. At times, involvement of others (colleagues or security staff) may be required to ensure that safety is maintained. Medical evaluation should not be side-lined even in an escalating AD patient, as several life-threatening conditions (intracerebral bleed, head injury, hypoglycaemia, adverse medication reactions, etc.) may be the root cause of AD.

Pharmacological management

Oral treatment should be offered first, and many will take up the offer of a sedative (tablet or liquid) to relieve their subjective distress; there are parenteral options in the uncooperative or violent patient or when oral route is not possible. Rapid tranquilization (RT) involves administration of medication through a parenteral route (intramuscular or intravenous), and aims to achieve a state of calmness for the patient, yet there may be instances where full sedation (induced by an anaesthetist) is required. RT is usually linked with the use of restraint; although in the UK, this is physical (by trained staff who ensure the full protecting of the patient's airway and chest), mechanical restraint is not uncommon around the world. There is a wide range of evidence-based interventions (Fig. 30.4). Although RT practice guidelines may differ in their details, benzodiazepines and second-generation (atypical) antipsychotics are generally preferred. Ketamine and dexmedetomidine are established alternatives. Regardless of the medication used, any RT requires careful monitoring of temperature, pulse, blood pressure, and respiratory rate and awareness of the medical risks involved (Box 30.13, Potential adverse events from RT).

Further reading

Patel MX, Sethi FN, Barnes TR et al. Joint BAP NAPICU evidence-based consensus guidelines for the clinical management of acute disturbance: de-escalation and rapid tranquillisation. *J Psychopharmacol* 2018; 32(6):601–640.

Alcohol withdrawal

Substance-related presentations in the ED are extremely common and relevant to a wide range of substances, including opioids, stimulants, and gamma-hydroxybutyrate (GHB), among others, but alcohol is the most prevalent by far. The number of alcohol-intoxicated patients attending the ED varies significantly between countries, approaching 15% in the UK and is associated with disruptive behaviour. In 2018/19, there were almost 1.3 million estimated hospital admissions in the UK where the primary reason for admission or a secondary diagnosis was linked to alcohol. This represents 7.4% of all hospital admissions, with two-thirds of patients being male. Although this places a significant burden on EDs, it is an opportunity for screening, detection, and intervention for alcohol-related problems, including alcohol withdrawal.

Any alcohol-dependent patient can present with symptoms indicative of alcohol withdrawal as early as three hours after their last drink. These symptoms lie on a spectrum from mild (anxiety, irritability, tremor, sweating, nausea, tachycardia) to severe (seizures,

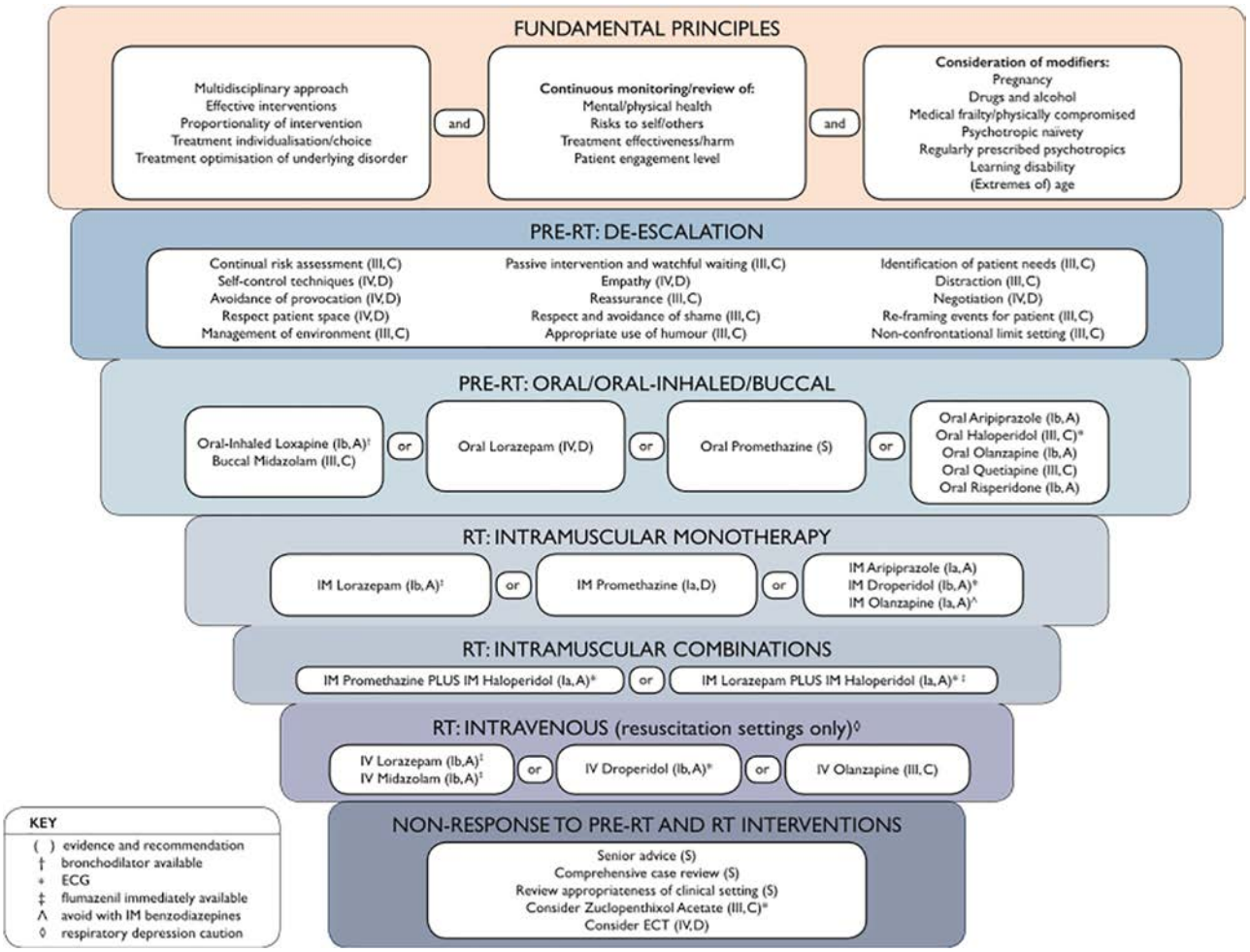


Fig. 30.4 Rapid tranquilization (RT). *(J Psychopharmacol. 2018 Jun;32(6):601-640.)*

delirium tremens). While pharmacologically assisted withdrawal (alcohol detoxification) can sometimes take place in the community, inpatient detoxification is likely to be required in all but mild withdrawals, and we should admit people with a history of severe withdrawals.

There are several detoxification regimen, all using one of the various benzodiazepines; chlordiazepoxide is most commonly used in the UK. Concurrent intravenous vitamin replacement in the form of parenteral thiamine is also recommended in at-risk patients to prevent Wernicke's encephalopathy. Patients at risk include those not eating properly for whatever reason (salience of alcohol, homelessness, crisis), those with vomiting or diarrhoea, those with other medical conditions such as diabetes, and all withdrawing patients with impaired consciousness or confusion. Early identification and treatment is paramount as severe alcohol withdrawal can be life-threatening.

Delirium tremens

This is an emergency present in about 3%–5% of those admitted to hospital for alcohol withdrawal, at times requiring ICU admission. It develops 2–4 days after the first signs of alcohol withdrawal and can be fatal in up to 4% of patients. It usually manifests as fluctuating cognitive disturbance, clouding of consciousness, and vivid visual hallucinations. Marked tremor, agitation, sleep disturbance, and autonomic dysfunction may also be present. Management requires a multidisciplinary approach that aims to safely manage severe agitation, replace B vitamins and other deficiencies, and use benzodiazepines to prevent and treat seizures. Alcohol detoxification with benzodiazepines does not cure alcohol misuse and dependence.

Wernicke-Korsakoff syndrome

This most commonly occurs in alcohol-dependent patients who stop drinking suddenly. Often, there is a physically traumatic event: surgery, road traffic collision, or myocardial infarct that increases demand on vulnerable brains. Wernicke's encephalopathy is an acute emergency with a triad of confusion, nystagmus, and ataxia. Waiting to see this triad misses the window of intervention to prevent permanent brain damage. Brain cells are continuously active and require energy generated from oxygen and glucose. Thiamine (B_1) acts as a co-factor in energy production as part of the Krebs cycle. Where dietary thiamine deficiency is followed by high glucose intake (admission to hospital and an offer of food, even sweet tea), disrupted metabolic activity results in a cellular energy deficits that can lead to neuronal loss in the mammillary body and other structures around the hypothalamus. Without B_1, transient Wernicke's encephalopathy can lead to permanent Korsakoff's psychosis (not technically a psychosis), characterized by extensive anterograde amnesia, permanent memory loss, and confabulation, where patients generate false memories to fill in gaps—potentially for the rest of their lives (Fig. 30.5). Prevention is based on rapid administration of parental vitamin preparations in at-risk individuals, before symptoms emerge.

Further reading

Schuckit MA. Recognition and management of withdrawal delirium (delirium tremens). *N Engl J Med* 2014; 371(22):2109–2113.

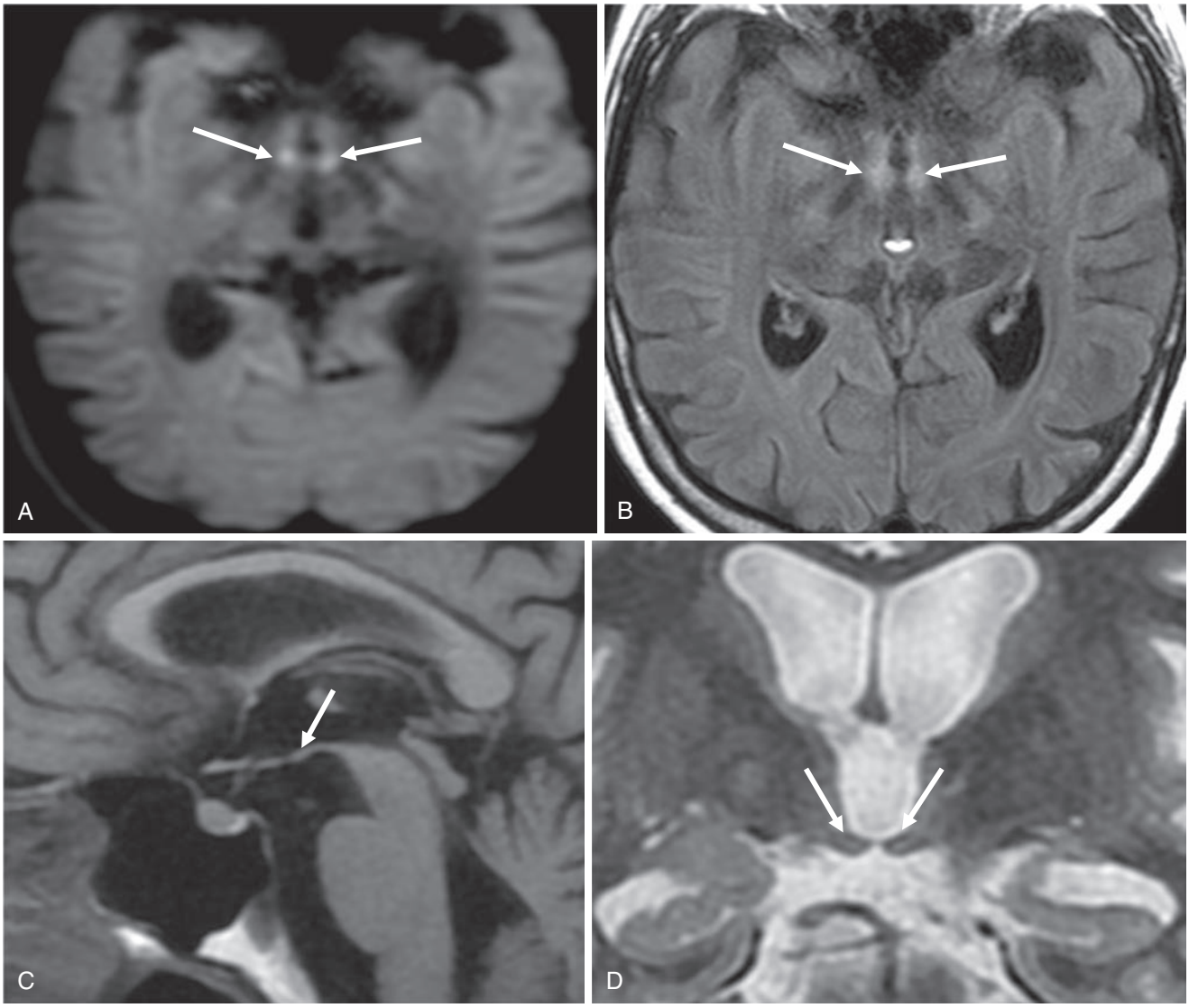


Fig. 30.5 MRI demonstrating Wernicke encephalopathy. *Arrows* indicate mamillary body hyperintensities, which evolve to atrophy in the chronic phase (Korsakoff's syndrome).

Neuroleptic malignant syndrome and serotonin syndrome

Neuroleptic malignant syndrome (NMS) is a rare, adverse reaction to psychotropic medication that has the potential to become fatal, if untreated. Although mostly linked to antipsychotic medication and dopamine blockade, it has been reported with other psychiatric medications including SSRIs, lithium, valproate, and abrupt withdrawal of dopaminergic drugs.

The clinical presentation is highly variable, including autonomic instability, hyperthermia, altered consciousness, and muscle rigidity. Elevated serum creatine kinase is common, and repeated blood investigations are needed on suspicion of the disorder, alongside liver function tests. Risk factors include exposure to high-potency first-generation antipsychotics (haloperidol), being male and being young. Mortality varies, with mortality higher in older people.

Priority should be given to discontinuing antipsychotic medication, monitoring vital signs, and providing specialist medical intensive care support. Almost all cases will need intravenous fluids to protect renal function. There is a role for benzodiazepines, bromocriptine, and dantrolene, as well as electroconvulsive therapy (ECT). Restarting antipsychotic medication will probably be required in many cases, ideally at least 5 days later, but under close supervision. First-generation antipsychotics and depot medication should then be avoided.

Serotonin syndrome

A second medication-induced challenge is the lesser recognized, but likely more prevalent, serotonin syndrome. History guides the diagnosis: patients have taken high doses of serotonergic drugs, especially in overdose or polypharmacy. An additional group may have misused (often synthetic) substances that are serotonergic. Patients are confused with high temperatures, tremors and muscle twitching, even seizures in severe cases. Muscle tone is usually normal (unlike NMS), and although creatine kinase can be elevated (as in NMS), this is unusual, and the rise is modest. Management on the intensive care unit may be indicated, all serotonergic medications should be stopped (even lithium and some antibiotics), and serotonin antagonists are administered.

Box 30.16 Catatonia signs and symptoms

- Stupor
- Cataplexy
- Waxy-flexibility
- Mutism
- Negativism
- Posturing
- Mannerism
- Stereotypy
- Psychomotor agitation
- Grimacing
- Echolalia (imitation of interviewer's speech)
- Echopraxia (imitation of interviewer's movements)

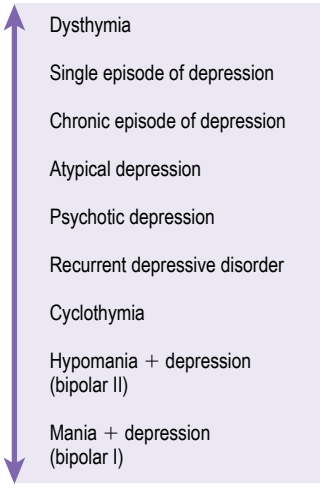


Fig. 30.6 Spectrum of affective illness.

Further reading

Prakash S, Rathore C, Rana K, Prakash A. Fatal serotonin syndrome: a systematic review of 56 cases in the literature. *Clin Toxicol* 2021; 59(2):89–100.
Schönfeldt-Lecuona C, Kuhlwilm L, Cronemeyer M et al. Treatment of the neuroleptic malignant syndrome in international therapy guidelines: a comparative analysis. *Pharmacopsychiatry* 2020; 53(02):51–59.

Catatonia

Catatonia is a neuropsychiatric syndrome first described in 1874. While originally considered part of psychotic and affective disorders, it is now known to be present in several other medical conditions including brain tumours, encephalitis, and autoimmune and metabolic disease. Observation, interview, and clinical examination are all necessary in uncovering the wide array of signs with which catatonia can present (see **Box 30.16**); two or more of these signs are considered adequate to consider catatonia in the differential diagnosis. Presence of autonomic instability is considered a complicating factor that significantly increases mortality.

There is no universal agreement on required investigations, and these are usually ordered to help exclude alternative diagnoses; neuroimaging and EEG are commonly requested. Given the similar presentation of catatonia and various encephalitides (see Chapter 28), in particular anti-*N*-methyl-D-aspartate (NMDA) receptor encephalitis, lumbar puncture may be considered. Administration of parenteral lorazepam can also be used as a catatonia challenge test to clarify diagnosis or predict future responses to treatment. Symptoms reduce with this disinhibitory challenge—even if there is a medical cause (brain tumour, CNS infection, etc.). For the challenge, the patient is given 1–2 mg lorazepam (ideally intravenously) while being monitored for response. The test is considered positive if there is a 50% reduction in catatonic signs. Benzodiazepines and ECT are considered the mainstay of treatment; followed by treatment of the underlying medical or mental disorder.

Further reading

Rogers JP, Oldham MA, Fricchione G et al. Evidence-based consensus guidelines for the management of catatonia: Recommendations from the British Association for Psychopharmacology. *J Psychopharmacol* 2023; 37(4):327–369.

COMMON PSYCHIATRIC DISORDERS

The term 'affective disorders' covers those psychiatric disorders where the primary presentation is of abnormal mood. Diagnosis is based first on recognizing mood episodes (depressive, manic, or hypomanic), and then ascribing these to different disorders based on the pattern of mood episodes over time, along with associated psychiatric features. The common mood disorders are depression, bipolar disease (types I and II), and schizoaffective disorder (see **Fig. 30.6**).

DEPRESSION

Affective (mood) disorders are very common, affecting women more than men. One in four women will have at least one depressive episode in a lifetime, with 1 in 10 of the population reaching clinical thresholds for mild depression over a 12-month period. There is a mood spectrum from normal to acute crisis/turmoil (e.g., due to sudden loss), through to a longer period of being unhappy (perhaps with sleep disturbance) or persistent low mood symptoms. Even in Western countries with universal healthcare, only 1 in 4 depressed people get any treatment.

Aetiology

Depression is unlikely a single disorder, and it has proved difficult to agree a unifying aetiological model (**Fig. 30.3**), despite a good understanding about who gets depression and factors likely to drive prolonged symptoms and relapse. **Box 30.17** sets out risk factors and profiles of people likely to experience their first and further episodes of depression.

Targeted national policy and public health interventions could prevent many depressive episodes by promoting basic income, providing safety nets, and encouraging local meaningful work. Abuse and violence can also be prevented, and evidence shows the benefit of early intervention to mitigate the effects of these. Clinicians can address several major risk factors for depression by optimizing physical health through multidisciplinary working and supporting patients to make positive changes to how they live.

Most depressive episodes do not have a single cause though many patients identify a precipitant ('last straw'). ICD11 includes seasonal depressive (affective) disorder as part of recurrent depression. Postnatal depression is covered elsewhere (see Chapter 48).

Genetic factors

Parents with a history of depression have a risk of up to 30% that their children will develop depression. Researchers describe 'activation of phenotype' with evidence that inherited abnormalities of the neuroendocrine and immune systems are activated by ACEs.

Box 30.17 Risk factors for depression

Factors increasing the risk of developing moderate to severe depression	
Sex	Female
Age	Median age of diagnosed depression varies across countries: Germany 19; France 24; UK 55
Marital status	Married < single < separated/divorced < widowed
Socioeconomic status	Poorer people and those with lower educational attainment—a proxy for lower socioeconomic group
Employment	Unemployment, shift workers (disrupted sleep), insecure employment
Abuse	Childhood abuse (emotional less than sexual), domestic violence, and other victims of violent crime
Physical illnesses	Diabetes; migraine; asthma; obesity in women
Mental disorders	A history of one or more episodes of depression predicts more episodes
Health behaviours	Smoking; inconsistent or unhealthy diet
Sleep	Daytime naps for more than 30 minutes: strongest association
Sedentary behaviour	TV, excessive mobile phone use, excessive computer use including computer gaming
Factors reducing the risk of chances of incident depression	
Physical activity	Walking for pleasure, faster walking pace, cycling, swimming, any heavy physical activity, e.g., home repairs
Sleep	Normal hours of night-time sleep
Interpersonal	Frequency of visiting or confiding in others, membership of gym/club/other social activity

Clinicians note early experiences within the context of attachment to parent or carer: attachment styles may determine childhood and adult depression.

Sex differences in the incidence of depression (from puberty, women have twice the rates of depression as men) are partly explained by evidence that female adolescents experience more adverse life events than boys and men. Later on, intimate partner violence (domestic violence) and economic inequalities (women often entering lower-paid jobs than men or earning less for the same work as men) may also drive increased depression. Bipolar disorder (and related schizoaffective disorder) has a stronger heritability than depression: 30% of one parent has bipolar, rising to 50%–60% if both parents have the diagnosis.

Clinical features

For most people, the natural history of a depression episode is around 9 months. Within this period, patients will experience lower quality of life and complications of depression. Health harm from depression might be conceptualized as *depth of low mood multiplied by time* just as adverse outcomes in diabetes are predicted by *raised glucose multiplied by time*. Lower mood is associated with 'negative outlets' meaning increased unhealthy behaviours: comfort eating, smoking, excess alcohol, substance misuse, unfocussed online time, gambling, etc., but also (as effect or cause) there will be increased isolation, loneliness, work days lost and presenteeism (attending work with low mood but with lower productivity). All of these in turn worsen depression.

Untreated depression leads to increased co-morbidity, self-harm, and suicide. For the individual, untreated depression raises lifetime risk of cardiovascular diseases and dementia, and has adverse effects on the immune system. Additional motivation to treat depressed patients includes social effects on others (children, relationships) and the wider community—they will not reach their full potential in education or work.

Diagnosis

Even if diagnostic criteria are fully met, people are excluded from a diagnosis of depression if their symptoms are 'better accounted for' by drug withdrawals, intoxication or side-effects, medical conditions with high symptom burden (pain, weight loss), or bereavement. Co-morbidity is common, especially when low mood brings out anxiety (as trait or state): mood and anxiety disorders potentiate each other. The *International Classification of Diseases 11* (ICD11) defines three clusters of symptoms which may be present in addition to low mood, for at least 2 weeks):

i. **Affective:**
   - loss of full interest and pleasure in life—as Shakespeare writes 'with what I most enjoy contented least'

ii. **Cognitive-behavioural:**
   - impaired concentration, even indecision
   - beliefs of low self-worth, including excess guilt
   - loss of hope and negative appraisals of the future
   - thoughts of death including by suicide

iii. **Neurovegetative:**
   - low energy/fatigue with minimal effort
   - appetite disturbance: this usually reduces in depressed people but can increase in a minority
   - sleep disturbance: these include difficulties getting to sleep (initial insomnia), broken sleep (middle insomnia), and/or early morning wakening (late insomnia). In severe depression, there may be the clinical signs of psychomotor retardation (like looking at a slow motion version of someone) or psychomotor agitation (extreme restlessness) in the absence of mania.

To make the diagnosis, ICD11 expects at least 5 of the 8 core symptoms above. Some depressed patients will fail to disclose suicidal thoughts, perhaps related to guilt; many will minimize their symptoms as 'too much going on' or 'everyone gets insomnia'. Suicidality must be documented in any history of depression, as well as any two additional markers of severity that may become intolerable for the patient, and thereby increase the immediate risk of suicide. **Guilt** might focus on their past acts and omissions: the depressed person's cognitive distortions make these negative, and fuel lower self-worth and loss of hope for the future.

Depressed people may also exhibit **psychotic features**—usually indicative of a severe depressive episode. Their guilt or other beliefs become delusional, and they hear (typically second person) auditory hallucinations: 'you're a bad person… you harmed people'.

After completing a history and MSE for a person with depression, it is important to document the key findings and act: to make a safety plan for the patient. If they are safe to go home now, what will they do when next they feel overwhelmed by low mood or intolerable symptoms? They will need details of local emergency mental health services and national crisis lines.

Treatment and prognosis

Because depression is very common, most patients are treated in primary care rather than specialist psychiatry services. The most common evidenced interventions are listed in Box 30.18, and many of the

Box 30.18 Treatment of affective disorders

Diagnosis/setting	Recommended treatments
New episode of less severe depression: patient declines therapy	Check capacity to refuse, give information, arrange follow-up, make sure they know they can change their mind in the future
New episode of less severe depression: patient agrees to therapy	Individual or group cognitive behavioural therapy (CBT) or behavioural activation; group exercise; interpersonal therapy or counselling; medication such as SSRIs if physical illness
New episode of more severe depression	Combination of CBT or behavioural activation plus an antidepressant: agree regular reviews to monitor for side-effects and suicidality
Prevention of relapse	Remain on medication and/or mindfulness-based CBT
Treatment-resistant depression	Check adherence to treatment, reassess co-morbidities or ongoing stress, consider second-line antidepressant (e.g., venlafaxine) and the addition of low-dose antipsychotic medication, last line: ECT
Chronic depression	Moclobemide, older agents: tricyclics, MAOIs
Depression with co-morbid personality disorder (PD)	There is no evidence that medications will improve outcome with PD: manage in multidisciplinary team, refer to specialist PD service
Psychotic depression	Combined antidepressant and antipsychotic

interventions of **Boxes 30.11 and 30.19** apply here too. In all affective disorders, the patient should be actively engaged with both the diagnosis and its treatments. Self-management is essential: recognizing changes to mood (perhaps by recording a regular mood score out of 10 in a diary or smartphone, where 0/10 is lowest possible mood, 8/10 is average good mood, and 10/10 is the happiest day of one's life). These scores can be communicated to others as necessary. This scoring system is particularly useful for patients whose mood changes with the seasons or in cycles (cyclothymia: see **Fig. 30.6**). Self-management goes beyond asking for help when needed. It includes understanding individual protective factors, innate and therapy-acquired coping skills, and positive interdependent relationships.

Primary care may need to enlist other professionals' help with more complex mood disorders in a 'stepped care' approach: see **Box 30.19**. Collaborative care requires a general physician, a care manager (who acts as a point of contact for patients, and monitors adherence, mood scores, etc.), and a mental health professional (a therapist, or in some cases a psychiatrist). The decision to prescribe is made in the context of severity, sometimes urgency, and failure to improve with other interventions (**Box 30.19**): there are additional patient issues to consider in the choice of antidepressant (**Box 30.20**).

In some patients, small or no relief of mood symptoms means chronic depression. Historical studies found that 10%–25% of patients have a bad outcome with either chronic low mood or recurrent depressive episodes despite treatments. This group may then require more complex treatments under specialist care. The majority of patients have positive outcomes—even though the risks of recurrent depression remain throughout their lives.

In recovered patients, the decision to maintain or gradually withdraw an antidepressant is made jointly with the patient, based on an informed risk-benefit analysis. Antidepressant discontinuation

Box 30.19 Stepped care model of treatment

Focus of the intervention	Nature of the intervention
STEP 4: Severe and complex depression; risk to life; severe self-neglect	Medication, high-intensity psychological interventions, electroconvulsive therapy, crisis service, combined treatments, multiprofessional and inpatient care
STEP 3: Persistent subthreshold depressive symptoms or mild to moderate depression with inadequate response to initial interventions; moderate and severe depression	Medication, high-intensity psychological interventions, combined treatments, collaborative care, and referral for further assessment and interventions
STEP 2: Persistent subthreshold depressive symptoms; mild to moderate depression	Low-intensity psychosocial interventions, psychological interventions, medication, and referral for further assessment and interventions
STEP 1: All known and suspected presentations of depression	Assessment, support, psychoeducation, active monitoring, and referral for further assessment and interventions

syndromes are well-described: typically in people on long-term treatments at higher doses using shorter-acting agents such as paroxetine. Sometimes, it is difficult to separate relapse of anxiety or mood symptoms from multiple physical symptoms, made more memorable by the FINISH mnemonic: **Box 30.21**.

Not including in **Box 30.20** are agents that have fallen out of favour for treatment of depression, including benzodiazepines and other sedatives (which lead to dependence), paroxetine (the shortest-acting SSRI, more associated with discontinuation syndromes), amitriptyline (more sedation and anti-cholinergic symptoms than duloxetine), and monoamine oxidase inhibitors (which require a restrictive tyramine-free diet, and have other significant side-effects).

Further reading

Choi KW et al. An exposure-wide and mendelian randomization approach to identifying modifiable factors for the prevention of depression. *Am J Psychiatry* 2020; 177(10):944–954.
Firth J, Solmi M, Wootton RE et al. A meta-review of "lifestyle psychiatry": the role of exercise, smoking, diet and sleep in the prevention and treatment of mental disorders. *World psychiatry* 2020; 19(3):360–380.
Gutiérrez-Rojas L, Porras-Segovia A, Dunne H et al. Prevalence and correlates of major depressive disorder: a systematic review. *Braz J Psychiatry* 2020; 42(6):657–672.
Kendall K, Van Assche E, Andlauer T et al. The genetic basis of major depression. *Psychol Med* 2021; 51(13):2217–2230.
National Research Council (US) and Institute of Medicine (US) Committee on Depression, Parenting Practices, and the Healthy Development of Children; England MJ, Sim LJ. (ed). *Depression in Parents, Parenting, and Children: Opportunities to Improve Identification, Treatment, and Prevention*. Washington (DC): National Academies Press (US); 2009. 3, The Etiology of Depression. Available from: https://www.ncbi.nlm.nih.gov/books/NBK215119/.
Taylor DM, Barnes TRE, Young AH. *The Maudsley Prescribing Guidelines in Psychiatry*, 14th edn. London: Wiley Blackwell, 2021. https://www.maudsley-prescribing-guidelines.co.uk/.
https://www.nice.org.uk/guidance/ng222.
https://www.nice.org.uk/guidance/CG91.

BIPOLAR AFFECTIVE DISORDER

Bipolar affective disorder describes an intermittent or cycling disorder of mood, with diagnosis being based on at least one episode of depression *plus* one episode of mania (bipolar type I) or hypomania (bipolar type II). Bipolar disorder is less common than depression

Box 30.20 Antidepressants, many of which are also licensed to treat anxiety disorders such as obsessive–compulsive disorder (OCD) and post-traumatic stress disorder (PTSD)

Drug name, start dose	Class	Positives	Negatives	Caution in
First line: prescribed in primary care for mild depression only if accompanied by physical symptoms in addition to CBT and social prescribing (Box 30.44); in moderate-severe depression, always prescribed in combination with CBT				
Fluoxetine 20 mg mornings up to maximum dose 80 mg	SSRI	Long acting: fewer discontinuation effects; weight neutral	Nausea for first few days; sexual dysfunction; rarely hyponatraemia and GI bleeds (especially in older people)	Liver failure, epilepsy, mania, and avoid in pregnancy and lactation; use lower doses in older patients
Sertraline 50 mg mornings up to maximum dose 200 mg	SSRI	Better tolerated, first choice if co-morbid cardiac disease	As for fluoxetine plus moderate weight gain; discontinuation problems (Box 30.21)	As fluoxetine
Citalopram 20 mg mornings to a maximum of 30 mg	SSRI	Less nausea than fluoxetine, but not as well tolerated as sertraline; available as drops for patients unable to take tablets	As for fluoxetine plus moderate weight gain; discontinuation problems	As fluoxetine; citalopram is the safest SSRI for patients with liver disease
Mirtazapine 15 mg night-time increasing to 45 mg daily—at night only	NaSSA	Sedative (promotes sleep), fast anti-anxiety effect, anti-nausea	Promotes appetite and significant weight gain often later in course; dry mouth, discontinuation effects	Liver and renal failure, epilepsy, mania; limited pregnancy data; avoid in glaucoma
Duloxetine 30 mg BD then up to 60 mg twice daily	SNRI	Good anti-anxiety effect and independent anti-pain actions	May reduce appetite and sleep, sexual dysfunction	Liver failure, epilepsy, mania, and hypertension; not safe in pregnancy
Second line: psychiatrist-led treatment; used in most severe cases (depression leading to weight loss, risks of suicide), a psychiatrist might add antipsychotic or consider ECT				
Venlafaxine 37.5 mg BD increasing to 375 mg max	SNRI	With titration, well tolerated and effective; has modified-release formulations	Blood pressure and prolactin elevation; sexual dysfunction	Liver and renal failure, epilepsy, and mania; never used in cardiac disease
Lofepramine 70 mg up to a maximum of 70 mg TDS	Tricyclic	Best tolerated tricyclic, the safest tricyclic taken in overdose	Some experience sedation by day; constipation common	Liver failure, epilepsy, mania, not used in porphyrias or cardiac disease
Moclobemide 150 mg BD up to max 300 mg BD	RIMA	Good anti-anxiety effect, even in social anxiety	Dietary consumption of tyramine-rich foods should be reduced; discontinuation more likely	Liver failure, epilepsy, mania, and avoid in pregnancy and lactation
Vortioxetine 10 mg daily, increasing to 20 mg once daily	Multimodal serotonin actions	NICE approved as second-line following two antidepressants; minimal sexual dysfunction	Skin reactions, abnormal dreams, and (as with all serotonin drugs) diarrhoea or constipation	Liver and renal failure, epilepsy, mania, and hypertension; never used in pregnancy or lactation

NaSSA, Noradrenaline and specific serotonin antidepressant; *RIMA,* reversible inhibitors of monoamine oxidase A; *SNRI,* serotonin and noradrenaline reuptake inhibitors; *SSRI,* selective serotonin reuptake inhibitors.

Box 30.21 Antidepressant discontinuation syndrome, FINISH

Flu-like symptoms (lethargy, fatigue, headache, achiness, sweating)
Insomnia (with vivid dreams or nightmares)
Nausea (sometimes vomiting)
Imbalance (dizziness, vertigo, light-headedness)
Sensory disturbances ('burning', 'tingling', 'electric-like', or 'shock-like' sensations)
Hyperarousal (anxiety, irritability, agitation, aggression, mania, jerkiness)

Source: Gabriel M, Sharma V. Antidepressant discontinuation syndrome. CMAJ. *2017 May 29;189(21):E747.*

(the lifetime prevalence of depression is 25% vs. 2.5% for mania, but while depression is commonly dealt with by primary care services, most patients with bipolar disorder require specialist psychiatric care. It is important to ask patients presenting with a depressive episode about previous episodes of elevated mood (mania or hypomania) as treatment differs considerably—mood stabilizers usually allow full restoration of social and vocational functioning.

Diagnosing elevated mood

Only some manic presentations have infectious good mood, where the patient laughs excessively or sings. To diagnose mania, it is important to identify *elevations* in **energy**, **mood**, and **interest**.

Box 30.22 Features of elevated mood: mania and hypomania

- Increased talkativeness or pressured speech (a feeling of internal pressure to be more talkative).
- Flight of ideas or experience of rapid or racing thoughts (e.g., thoughts flow rapidly and, in some cases, illogically from one idea to the next; the person reports that their thoughts are rapid or even racing and has difficulty remaining on topic).
- Increased self-esteem or grandiosity (the individual believes that he can accomplish tasks well beyond his skill level, or that they are about to become famous). In psychotic presentations of mania, this may be manifested as grandiose delusions.
- Decreased need for sleep (the person reports being able to function with only 2 or 3 hours of sleep), as distinct from insomnia, in which an individual wants to sleep but cannot.
- Distractibility (e.g., the person cannot stay on task, because attention is drawn to irrelevant or minor environmental stimuli, such as being overly distracted by outside noise during a conversation).
- Impulsive or reckless behaviour (e.g., the individual impulsively pursues pleasurable activities without regard to their potential for negative consequences, or impulsively makes major decisions in the absence of adequate planning).
- An increase in sexual drive, sociability, or goal-directed activity.

(Source: ICD11)

- The patient thinks and says they have more (even infinite) energy, with no need for rest or sleep.
- Mood may not be subjectively high (euphoric) but labile, and objectively irritable: people who know them notice a change for the worse.
- They are interested in everything, but intrusive (to other people) and ignoring all expertise but their own.

The diagnosis of bipolar is often a long time coming: on average 15 years after first presentation. So in all patients presenting with a mood disorder, we document the presence or absence of key features of mania in the clinical notes (see **Box 30.22**). Always note any significant change from the individual's usual behaviour or subjective state.

Hypomania describes elevated mood *without* psychotic features: by definition the symptoms are less severe, of shorter duration, and the patient usually has insight. Once a patient has been diagnosed with mania, they have achieved a diagnosis of **bipolar type 1** and will require long-term surveillance and medications; **Box 30.23** describes evidenced approaches. **Bipolar type 2** is an intermittent mood disorder with at least one *hypomanic* and one depressive episode. Lithium is an effective mood stabilizer but has a narrow therapeutic range and requires regular monitoring of drug concentrations to prevent toxicity (most importantly latent kidney disease).

A minority of bipolar patients can develop **rapid cycling of mood**, which presents one of the most challenging diagnostic and therapeutic scenarios in psychiatry. Rapid cyclers experience 4 or more episodes of mood disturbance (high or low) over one year—with full or partial remission to normal mood. The diagnosis cannot be made if there is substance misuse; many, however 'self-medicate' with alcohol excess, and alcohol makes rapid cycling episodes more likely. Risk factors include long-term antidepressant use and hypothyroidism; lithium remains preferred acute treatment, though it is also implicated in rapid cycling—if it is deemed unhelpful in rapid cyclers, it should be withdrawn slowly over months.

Bipolar disorder has a significantly better outcome than schizoaffective disorder (SZA) and a longer but not normal life expectancy.

Box 30.23 Treatment approaches in bipolar disorder

First episode of hypomania or mania	Consider stopping antidepressant (AD) and offer an antipsychotic (AP) regardless of whether AD had been stopped; haloperidol, quetiapine, or risperidone are first-line agents
Hypomania/mania where first AP is not tolerated	If an alternative AP is not effective, consider adding lithium or (in men with advice towards barrier contraception, and in women over 55 years old) valproate
Hypomania/mania: patient already taking lithium	Check lithium levels and adjust dose; add AP: haloperidol, olanzapine, quetiapine, or risperidone. No evidence for lamotrigine
Mania non-responsive to medication	Electroconvulsive therapy (ECT)
Bipolar in long term	Offer family therapy, individual or group therapy; lithium remains first-line medication: if ineffective, add valproate (in men with advice on barrier contraception, only in women over 55); if lithium is not tolerated, use olanzapine but watch for weight increases
Rapid cycling in long term	Carbamazepine, valproate (if safe), thyroid hormone

We diagnose SZA if there are features of schizophrenia at the same time or within a few days of the mood (depressed or manic) presentation.

Further reading

Ahmed M, Morriss R. Assessment and management of rapid-cycling bipolar affective disorder. *Adv Psychiatr Treat* 1997; 3(6):367–373.
World Health Organization *ICD11* https://icd.who.int/browse11.
https://www.nice.org.uk/guidance/cg185.

ANXIETY DISORDERS

Anxiety is a feeling of uneasiness and worry that is a normal adaptive response to a threat. However, in some individuals, this can become maladaptive, excessive, or pathological causing impairment in the socio-occupational functioning of that person. Anxiety disorders are common but often not formally diagnosed even in medical settings, and therefore unlikely to receive any treatment. Patients with panic disorder frequently present to ED when they believe that they have a medical rather than a psychiatric condition.

When patients present to physicians in any medical setting with predominant psychological symptoms, these are often recognized and managed appropriately. However, when patients present with physical symptoms related to a primary anxiety disorder (see **Box 30.24**), if unrecognized, patients undergo extensive investigations for physical disease—increasing anxiety symptoms and health services costs. Patients then receive delayed or no mental health interventions.

Despite anxiety disorders being some of the most common medical conditions, hospitals do not often confer a formal mental disorder diagnosis for anxiety unless it is overwhelming for a patient (or staff), or crosses over with other conditions. Nevertheless, pathological anxiety is an important aspect of care to consider in hospital inpatients, for example:

Box 30.24 Types of anxiety symptoms

Physical symptoms	Cognitive symptoms	Affective symptoms	Behavioural symptoms
Increased heart rate Muscle tension Sweating Feeling light-headed Fatigue Hyperventilation/choking sensation/tightness in the chest Insomnia/hypersomnia Eating and weight changes Urge to pass urine Dry mouth GI symptoms: nausea, vomiting, diarrhoea Nightmares and night terrors Tremors Syncope Numbness or tingling in hands and/or feet	Difficulty concentrating/distractibility Mind often going blank/forgetfulness Irrational fears and dread Feeling as though one is going crazy Indecisiveness Feeling in danger Flashbacks Feelings of dissociation	Feeling of fear/worry/panic/uneasiness Persistent feeling of dread/anticipating the worst outcome Fear of dying/sense of impending doom Low mood Tearfulness or anger	Avoidance Decreased performance associated with anxiety Startle response/hypervigilance

i. a primary anxiety disorder (such as generalized anxiety disorder or panic disorder) causing significant anxiety symptoms and related symptoms (e.g., poorer pain control) in a hospital inpatient
ii. medical illness mimicking anxiety disorders (e.g., endocrine: hyperthyroidism, Cushing's disease, and the rare pheochromocytoma)
iii. stressful hospital events triggering anxiety symptoms—situational anxiety when awaiting tests/results/surgery, worse when procedures are cancelled; stress when changing ward or teams, perhaps requesting a second opinion
iv. stressors triggering/worsening the symptoms of pre-existing anxiety where there are other psychiatric or neurodevelopmental conditions present, for example, agitation in a patient with autism in an unfamiliar hospital setting.

Epidemiology

Anxiety disorders are common in the general population with a point prevalence of around 6%: see **Box 30.25**. They usually start in adolescence or early adulthood usually reaching peak during middle age and then decreasing with the elderly, and are more common in women than men.

Aetiology of anxiety syndromes

The exact aetiology of anxiety disorders is unknown but it is likely to be a combination of stress, genetic and environmental factors. Anxiety disorders often run in families, suggesting that a combination of genetic and environmental stresses is responsible.

Three neurotransmitters have been implicated in anxiety disorders:

- serotonin (levels associated with happiness; exaggerated postsynaptic receptor sensitivity to serotonin has been described in anxiety disorders)
- noradrenaline (increased 'fight or flight' adrenergic activity is seen in anxiety disorders, with increased presynaptic alpha 2-receptors)
- GABA (it is thought that reduced inhibitory receptor sensitivity with low levels of GABA also results in anxiety).

In panic attacks and in PTSD, there is increased activation in the amygdala, hippocampus, locus coeruleus, thalamus, and orbitofrontal areas.

Further reading

Semple D, Smyth R. *Oxford Handbook of Psychiatry*, 4 edn. Oxford Medical Handbooks (Oxford, 2019; online edn, Oxford Academic, 1 June 2019).

In ICD11, anxiety disorders are grouped under the category of anxiety and fear-related disorders which also include specific phobias, social anxiety disorder, separation anxiety disorder, and selective mutism.

Panic disorders

In panic attacks, patients have spontaneous, unprovoked discrete intense episodes of severe anxiety that develop rapidly reaching a peak in 10 minutes and usually do not last longer than 20 minutes. Patients have a range of anxiety symptoms as described in **Box 30.24**. Patients can present frequently to medical care with a range of severity from mild (occasional, self-limited attacks) to severe attacks (patients describe unpleasant 'I'm going to die' experiences, with a fear of losing control).

Anxiety syndromes

Generalized anxiety disorder

In generalized anxiety disorder (GAD), patients have 'excessive worry': mild to moderate levels of anxiety, present for several months, at most of the time throughout the day about multiple everyday events or difficulties. Collateral history from a parent is often necessary to identify childhood/adolescent GAD symptoms. What people do (and avoid) often makes the diagnosis: school refusal, refusal to play school sports, self-isolation, and more. Patients may describe general apprehension (as free-floating anxiety), unfocussed 'worrying about worrying', or excessive worry focussed on multiple everyday events, along with the physical symptoms of anxiety (**Box 30.24**). Taken together, these cause significant impairments in socio-occupational functioning.

Agoraphobia

Agoraphobia is marked and excessive fear of public spaces or situations, for example, fear of public places (the Greek word for marketplace is *agora*) and busy shops/restaurants where escape is perceived to be difficult. Symptoms must be present for at least several months.

Box 30.25 Types of anxiety disorder

Anxiety disorder	Prevalence	Sex differences	Mean age at onset
Panic disorder	Lifetime prevalence 1.5%–3.7% (panic disorder) vs. 7%–9% for panic attacks Rates are higher in medical settings: 15% in dizziness clinics, 16%–65% in cardiac clinics, 25%–35% in HIV clinics	Women are 3 times more likely to be affected than men	Bimodal distribution Peak at 15–24 years Another peak at 45–54 years Rare after 65 years (0.1%)
Agoraphobia	6 months 2.8%–5.8%	M:F=1:3 (as panic disorder)	Bimodal distribution First peak 15–35 years Later life onset is associated with physical frailty and underlying medical problems
Simple phobia	Lifetime 12.5% 12 months 8.7% 6 months 4.5%–11.9%	M:F=1:3 Animal/situation phobias more common in women	Mean age of onset is 15 years Onset for animal phobia 7 years Situational phobias—around 20 years
Social phobia	Lifetime prevalence vary 2.4%–12.1% 12-month prevalence 6.8%	Community survey M>F Those seeking treatment F=M	Bimodal distribution with peaks at 5 years and 11–15 years Patients often do not present until they are in their 30s
Generalized anxiety disorder	6 months 2.5%–6.4% 12 months 3.1% Lifetime 5.7%	F>M	Lowest in 18–29 years—4.9% Highest 45–59 years 7.7% 60+ years 3.7% Early onset associated with childhood fears and marital/sexual difficulties Late onset is associated with stressful events, being single or being unemployed
Obsessive–compulsive disorder	0%–3%	M>F	70% onset before 25 years of age 15% after 35 years
Post-traumatic stress disorder	Lifetime 7.8%	M:F=1:2	Risk of PTSD after traumatic event 8%–13% men, 20%–30% for women Higher rates in some types of stressors, e.g., rape, torture, being prisoners of war

Phobic disorders

These are simple phobias with recurrent, intense, and excessive fear on exposure to specific objects or situations other than those covered in agoraphobia or social phobia. Symptoms persist for months/years, and patients usually avoid those situations. Examples of phobias include:

- Claustrophobia—fear of enclosed space (such as entering into an MRI scanner)
- Aerophobia—fear of flying
- Arachnophobia—fear of spiders
- Needle phobia—this is common in hospitals, and treated with reassurance, distraction, presence during venesection of a familiar family member or nurse, or occasionally, mild sedation.

Social anxiety disorder

This can be understood as a complex phobia. There is excessive anxiety restricted to social situations (meeting new people, unfamiliar situations, etc.) and the primary focus is the scrutiny by others or fear of being negatively evaluated by others. Fear of public speaking is universal (and not a disorder) but a recent change in a patient (perhaps starting University) may alert to a return of this aspect of social anxiety. Anecdotally, social anxiety increased during and after the COVID-19 pandemic, especially in younger people—who later withdrew from social activities onto their phones and online.

Disorders specifically associated with stress

In ICD11 the disorders specifically associated with stress are related to either short-term or long-term exposure to stressors or adverse events. These include acute stress reaction, adjustment disorder, post-traumatic stress disorder (PTSD), complex post-traumatic stress disorder, and prolonged grief disorder.

Core symptoms of these disorders comprise:

1. preoccupation with the stressor or its consequences, including excessive worry, recurrent and distressing thoughts about the stressor, or constant rumination about its implications; and
2. failure to adapt to the stressor with significant impairment in personal, family, social, and occupational functioning.

Acute stress reaction

The onset of symptoms of acute stress reaction are transient that can occur after a traumatic event. Symptoms usually resolve within 3 days.

Adjustment disorder

This describes a maladaptive reaction to an identifiable psychosocial stressor or multiple stressors (e.g., divorce, loss of a job, diagnosis of an illness, or recent onset of disability). The onset of symptoms is related to the adverse life event(s). Symptoms usually

emerge within a month after exposure to the stressful event but sometimes onset can occur after a longer delay, for example, 3 months after the exposure. Adjustment disorder lasts less than 6 months and diagnosis is excluded if patients meet the criteria for ICD11 criteria for (other) anxiety disorders or depression.

Post-traumatic stress disorder (PTSD)

This is a syndrome with onset following a traumatic, usually life-threatening, event. Symptoms persist for more than a month and include

i. re-experience of the traumatic events in the form of nightmares, flashbacks, or vivid intrusive memories (by day) accompanied by strong emotions such as fear
ii. avoidance of reminders of the traumatic event
iii. hypervigilance: patients describe being 'on edge', waiting for something bad to happen to them.

Complex post-traumatic stress disorder (CPTSD)

CPTSD consists of six symptom clusters: the three PTSD criteria (re-experiencing, avoidance, and hypervigilance) plus three disturbances of self-organization symptoms:

iv. emotional deregulation (heightened emotional reactivity, dissociative experiences, or emotional numbing)
v. disturbance in interpersonal relationships, and
vi. negative self-concept (feelings of guilt, shame, or worthlessness).

Normal and abnormal grief reactions

The pain of losing close family members, friends, and others is a universal human experience. A normal bereavement period is usually culturally determined but with wide variations. Anxiety (provoked by anger, denial of the loss, yearning for the deceased, sleep disturbances, etc.) is part of normal grief, and support with reassurance ('this too will pass; time heals') is the best management.

In general, if there remains continued subjective longing for the person and/or an (objective) excessive preoccupation with them six months after the bereavement (excluding the anniversary of the death), then a prolonged grief disorder, sometimes referred to as complex grief, might be present. Additional symptoms relate to cognitive distortions: patients believe they can no longer function in their post-bereavement role, with ongoing anger, even numbness, and speak of being 'unable to move on'. Assessment shows impairment of current social functions. This is a grey area of diagnosis with a significant overlap with normal experience and (sub)cultural factors, but complex grief probably affects between 4% and 10% of bereavements. A useful clinical approach involves promoting alcohol abstinence, treating co-morbid depression, offering bereavement counselling, and general supportive measures (**Box 30.11**) to prevent their loss becoming a life-changing bereavement.

Obsessive–compulsive disorder

Obsessive–compulsive disorder (OCD) is not based on a specific stress but is an episodic anxiety syndrome characterized by **obsessions** (e.g., of contamination, order or symmetry of objects, intrusive doubts) and **compulsions** (e.g., body washing, cleaning, checking, counting). Patients have recurrent intrusive, unwanted, unpleasant thoughts or images, and often a strong wish to carry out compulsive behaviours, known as rituals. Their rituals reflect their primary obsession. If their thoughts relate to hygiene or germs, there will be cleaning rituals—patients may clean their homes or hospital bedside areas 10–50 times daily, and the skin of their hands may be red or broken from handwashing. For others with OCD, their rituals are checking ('I must have left the gas cooker on'), counting (so called 'magical thinking'—'if I count to X, nothing bad will happen to my family') or other sequenced rituals that they know are unnecessary to perform, but which they feel compelled to complete. OCD patients have high co-morbidities with depression (when the thoughts and rituals recur) and other anxiety disorders. OCD diagnosis is relatively easy but there are differential diagnoses (**Box 30.26**).

Assessment of anxiety disorders

Establish onset, triggers, severity, duration, and the impact of symptoms on functioning through a careful history and MSE. Consider other co-morbid mental health diagnoses such as depression, substance misuse, or personality disorder; note coping mechanisms and previous responses to treatment. In acute hospital medical settings, consider whether the anxiety is triggered by ongoing hospital procedures or treatment. Screening anxiety scales (such as the Beck Anxiety Inventory [BAI] or Hospital Anxiety and Depression Scale [HADS]) can be used to diagnose and monitor progress.

If the age of onset of first anxiety symptoms is more than 35 years, recent onset depression is the more likely psychiatric diagnosis. However, in this over-35 group, when there is no significant past psychiatric history or family history of anxiety disorders, no significant life events for the onset of the symptoms, and no prior avoidance behaviours, then an organic cause of anxiety must be ruled out. New onset anxiety in an older person might herald dementia. Part of physical examination is to think through the possibility of a primary (treatable) medical cause for the anxiety recorded in history. Consider investigations such as full blood count, blood glucose, liver and renal profile, thyroid function, urine test for pregnancy, and drug screen (illegal drugs), then (rare) high urine catecholamines (**Box 30.8**). If there is prominent chest pain or dyspnoea, it may be necessary to rule out pulmonary embolus or acute coronary syndrome. Normal findings should reassure here.

Treatment of anxiety disorders in medical settings

Treating anxiety is part of managing all medical conditions. Identifying anxiety symptoms and behaviours (for instance, a patient repeatedly asking '*will I be alright doctor?*') and discussing them with patients can lead to direct improvement of symptoms and improves patient adherence to all treatments. It is important to intervene if a positive diagnosis of anxiety is made. Without treatment, anxiety is associated with increased disability, increased use of health service resources, poorer outcomes, and impaired quality of

Box 30.26 Differential diagnosis for OCD

Anankastic ('obsessive compulsive') personality disorder
Eating disorder—patients have overvalued ideas about eating and a morbid fear of fatness
Depression—patient may have obsessive ruminations that they are worthless or dirty: these resolve when depression is treated
Hypochondriasis—patient has excessive preoccupation of having a serious disease
Psychosis—patient may believe that someone is inserting unpleasant thoughts into their mind, or making them act in certain ways
Tourette's syndrome—characterized by motor and vocal tics, echolalia, and coprolalia (using offensive words without losing one's temper or the intent to shock)

life. Ideally, patients are best managed by involving LP in a hospital setting, but this may not always be possible as most hospitals may not have a full LP team with psychologists and occupational therapists. Therefore, all physicians must be aware of basic principles on managing anxiety in their patients.

Psychoeducation ('*your unpleasant symptoms are not a medical disease but are driven by a treatable mental health condition*') is the first step as patient feel reassured and this builds trust. Effective communication (listen, acknowledge, advise) is central to the consultation. Physicians should use open and closed questions to allow full disclosure of personal information. There may be undeclared stressors at home (marital problems, aggressive neighbours)—that cannot be easily 'fixed', but clinicians can support the person to problem solve after their discharge.

Depression, other anxiety disorders, and substance misuse need to be addressed—any improvement in these will lower the current anxiety burden. Discuss in your team how their anxiety might be provoked by (say) waiting for tests or surgery—can this be mitigated? If staff reassurance is making patients *more* anxious, then ration this.

Anxiety lowers the pain threshold. Opioids are **not** an anxiety treatment: they should be used in end-of-life care and prescribed, *short term*, for other pain if indicated. Even in a busy ward, patients can practice mindfulness as a distraction from their stresses. Visualization of unpleasant medical procedures (claustrophobic symptoms in MRI scanner) can reduce the discomfort. Benzodiazepines might be prescribed as a **once-off** to provide temporary relief from anxiety symptoms, but regular use promotes dependence and failure to develop positive coping mechanisms. While it is useful to refer to psychiatrists or psychologists for appropriate treatment options, most anxiety is treated in primary care. Consider referral to LP in hospitals or community mental health services if patients with an anxiety disorder also have:

- Thoughts/acts of self-harm or suicide
- Co-morbid severe depression or substance misuse
- Risk of self-neglect: this can be subtle, for example, anxiously avoiding a life-changing treatment if subjectively overwhelmed
- Complex medical conditions
- Treatment non-response or resistance: the anxious, stuck patient.

Psychological treatments

Cognitive-Behavioural Therapy (CBT) is the most effective treatment for anxiety disorders. Many patients avoid activities in response to anxiety, which worsens their anxiety and further worsens their functioning. The principle of treatment is that controlled exposure to the anxiety-producing stimulus will eventually lead to reduction in symptoms. Other techniques used in CBT are cognitive restructuring to change negative thoughts associated with anxiety, and behavioural modification (response prevention in OCD rituals). Single-session intensive debrief is not beneficial for patients who have experienced recent trauma. There are plentiful 'new' treatments that complement CBT: for example, guided relaxation and eye movement desensitization and reprocessing (EMDR)—described in **Box 30.11**.

Further reading

Killikelly C, Maercker A. Prolonged grief disorder for ICD-11: the primacy of clinical utility and international applicability. *Eur J Psychotraumatol* 2018; 8(Suppl 6):1476441.

Pharmacological treatments

SSRIs (**Box 30.20**) have become first-line medications for the treatment of anxiety disorders. In some patients, SSRIs can worsen anxiety in the first few days after commencing the treatment. To maximize the anxiogenic side-effect, treatment should be commenced at half or less of the usual starting dosage for depression. The dose could be titrated gradually after a week depending on the response and tolerability. SSRIs usually cause symptomatic improvement after 2 to 4 weeks of treatment.

Sertraline is often used as the first line of treatment, but if patients do not tolerate this, then another SSRI can be tried. Patients should be monitored for side-effects (GI disturbance, headache, sexual dysfunction) and drug interactions. Monitor younger patients (less than 30 years age) within a week of prescribing SSRI as there is increased risk of suicidal or self-harm ideas. Venlafaxine, mirtazapine (especially in PTSD), and pregabalin are used as alternatives.

Tricyclic antidepressants (imipramine and amitriptyline) can be used to treat anxiety disorders, and their sedation is usually welcomed by patients, but with caution in older people and vulnerable medical patients as they are associated with anticholinergic side-effects (dry mouth, constipation, urinary retention, and orthostatic hypotension) and can be dangerous if taken overdose. Beta-blockers such as propranolol can be helpful in reducing the peripheral autonomic symptoms of anxiety (tremor and tachycardia). Benzodiazepines are licenced for the temporary relief of acute anxiety but they should not be used for more than 2–4 weeks due to risk of dependence.

Further reading

Culpepper L. Use of algorithms to treat anxiety in primary care. *J Clin Psychiat* 2003; 64(suppl 2):30–33.
Marx J, Hockberger R, Walls R (eds). *Rosen's Emergency Medicine Concepts Chapter 110*. Philadelphia Elsevier; 2010:144–51.
National Institute for Health and Clinical Excellence. *Anxiety: management of anxiety (panic disorder with or without agoraphobia, and generalised anxiety disorder) in adults in primary, secondary and community care*. Last updated 15 June 2020.
Stern TA. (ed). *Massachusetts General Hospital Handbook of General Hospital Psychiatry*. 7th edn. Saunders/Elsevier. 2018.
Taylor DM, Barnes TRE, Young AH. *The Maudsley Prescribing Guidelines in Psychiatry*. 2021.
https://www.nice.org.uk/guidance/cg113.

PSYCHOSIS

Psychosis is a syndrome comprising symptoms affecting thoughts, perception, and cognition (all three are essential to a healthy mental state), ultimately affecting behaviour. Psychosis can run a chronic course and has the potential to severely impact social functioning, wellbeing, and physical health.

While historically linked with schizophrenia, there has been a gradual shift in modern concepts of psychotic disorders. This includes the introduction of the Psychosis Clinical High Risk State (CHR, see below) and the realization that 70%–80% of people defined as CHR will in fact never develop psychosis. Many physical conditions may present with psychotic features, such as Parkinson's disease, epilepsy, and a wide range of encephalitides, that shed light on the role of the immune system in psychopathogenesis. Taking into account methodological issues and significant heterogeneity between the definitions of psychosis, median global prevalence of psychotic disorders is estimated to be 4.6 per 1000 persons. However, this relatively rare disorder accounts for over 90% of psychiatric admissions.

Clinical features

The prevalence of symptoms varies between the various manifestations of psychosis. The severity and frequency with which they

appear vary within the same population. MSE findings (Box 30.4) comprise the following:

- *Appearance and Behaviour:* there is no one way someone presents/behaves during a psychotic episode. Some may be dressed appropriately and others may have signs of self-neglect/look dishevelled. Patients may be restless, agitated and occasionally engage aberrant behaviour, posing a risk to themselves and occasionally to others. Someone responding to unseen stimuli may come across as being easily distracted as if 'on drugs' when seen talking to themselves. Bizarre behaviour may also be present with patients exhibiting mannerisms or stereotypies (Box 30.4), or on the other end of the spectrum, there may be a complete absence of goal-directed movements with someone in a catatonic state.
- *Speech* can be disorganized with significant variations in rate, tone, and volume. Syntax and grammar may be affected, making comprehension difficult for the listener. Idiosyncratic use of language (e.g., neologisms: Box 30.4) is noteworthy, while in others, near poverty of speech may be present with severe 'negative symptoms'.
- *Mood* can be low or irritable, even angry. Patients may describe a delusional mood or atmosphere ('something (strange) is going on'. On other occasions, people with psychosis lack the emotional reactivity one would expect in particular settings (blunting/flattening of affect), or their emotional reactions might be inappropriate (incongruent).
- *Thoughts:* both form and content of thoughts can be affected. This may appear as loosening of associations or even tangential thoughts, and in severe cases, thoughts may be totally unintelligible. Delusional content may be present, the main theme of which varies (Box 30.27). One word can usually describe thought content: religious, persecutory (paranoid), somatic, or bizarre.
- *Perception:* many types of perceptual abnormalities are commonly present, though their prevalence tends to vary between the different subtypes of psychosis. Auditory hallucinations (Box 30.28) are colloquially referred to as 'voices' that may be commenting on someone's actions or even commanding them to do a specific action. Visual hallucinations are more frequent in patients with Parkinson's psychosis (and other organic psychoses)—these can be very vivid and detailed.
- *Cognition:* although orientation to self, time, and place appears intact in the vast majority of psychosis, other cognitive domains appear to be affected. Processing speed, working memory, executive function, and verbal learning are some of the affected domains and these are linked to poor functional outcomes.
- *Insight:* usually linked with the severity of the psychotic episode, insight may fluctuate, and is often intermediate between total absence and full presence.

Psychotic disorders

Clinical High Risk for Psychosis (CHR)

The majority of patients that go on to develop schizophrenia experience a prodromal phase (otherwise known as 'at-risk mental state'), characterized by attenuated forms of typical psychotic symptoms (such as delusions and hallucinations) that may impact on their function. They may also carry several risk factors for psychosis. The CHR group is highly heterogeneous and is identified through well-validated assessment tools, like the CAARMS (Comprehensive Assessment of At-Risk Mental States) and the SIPS (Structured Interview for Psychosis-Risk Syndromes). Being able to accurately

Box 30.27 Types of delusions

Primary delusions occur out of the blue (i.e., without prior morbid experience).	
Delusional mood	Unpleasant sense that surrounding events refer to oneself. As the mind abhors a vacuum, delusional mood is usually resolved by the formation of an explanatory sudden delusional idea (an 'autochthonous' delusion): delusional mood is unlikely in a current MSE but may be recalled retrospectively
Delusional perception	Sudden attribution of delusional meaning to a normal perception, e.g., 'the position of that cup on the table means I will be famous'
Secondary delusions (usually) evolve from pre-existing morbid psychological processes (altered mood, hallucinations, other delusions)	
Persecutory	Someone is trying to discredit, harm or kill them
Of reference	Utterances or gestures of people, such as the media, are referring to the person, e.g., 'people on this TV programme are looking at me/seem to know about me/are talking about me'
Grandiose	Being famous, having a special ability or mission
Nihilistic	Extreme negation of self or world, e.g., believing part of the body has died
Shared delusion	A false fixed belief can be acquired from a close relationship (usually a relative or partner who tends to be more dominant, perhaps more intelligent than the receiver of the secondary delusion): this is a folie à deux (2); very rarely, more than two people share the delusion: folie à trois (3) or à eux (many)

Box 30.28 Types of auditory hallucinations

Type of hallucinations	Explanation with examples
Thought echo	Hearing one's thoughts repeated aloud after one thinks them
Second person	Voices talk directly to the patient, sometime addressing using their name, including giving commands
Third person	Hallucinations discuss the patient sometimes giving a running commentary
Reflex	Normal sensory modality triggers hallucination in another modality, e.g., seeing a bus triggers somatic hallucination of electric shock
Functional	Normal sensory modality triggers hallucination in the same modality, e.g., auditory hallucinations are associated with the sound of running water
Extracampine	Hallucinations experienced as outside of sensory field, e.g., voices heard from far away, even another country
Hypnogogic	Hallucinations as the person is going to sleep; normal experience
Hypnopompic	Hallucinations after or as person wakes up; normal experience
Formication	Hallucinations of touch, e.g., small animals/insects are felt to be crawling all over the body, classically seen in organic disorders such as cocaine psychosis

predict who will develop psychosis helps develop targeted prevention strategies and treatments, but staging of psychosis remains a work in progress, reflecting the high heterogeneity of the CHR group.

The prevalence of CHR differs significantly between the general population (1.7%) and people making their first contact with mental health services (19.2%). Some CHR referrals (to services) will develop other non-psychotic disorders. Meta-analytical data suggest that only around 1 in 5 people identified as being at CHR will transition to psychosis over one year.

Schizophrenia

Schizophrenia is a chronic, heritable form of psychosis of relapsing-remitting nature. While prevalence rates may vary by geographic location, the average lifetime prevalence of schizophrenia is close to 1%. Several risk factors have been identified over the years, including sex (incidence risk ratio of 1 to 1.4 for women to men), ethnicity (higher prevalence in minority ethnic groups), urban environment, childhood and social adversity, and substance misuse (especially cannabis).

Aetiology

The pathophysiology underlying schizophrenia is considered complex, and despite advances in research methodology and relevant investigations, many questions remain unanswered. Considered a neurodevelopmental disorder borne out of the complex interplay between genetic and environmental factors, people with schizophrenia display distinct differences in multiple neurotransmitter systems (dopamine, glutamate, GABA, and more), reductions in total grey and white matter volumes, ventricular enlargement, as well as altered activation patterns as displayed using functional neuroimaging. Genome-wide association studies have identified more than 100 chromosomal loci associated with the disorder, illustrating the polygenic, complex nature of schizophrenia in which both genetic and environmental factors are relevant. Given the lack of biomarkers that can be used in a clinical setting, diagnosis remains clinical, based on operational criteria, though even these differ between the various classification and diagnostic systems.

Psychotic disorders due to other medical conditions

Autoimmune encephalitis (AE) is a debilitating brain disorder caused by aberrations of the healthy immunological and inflammatory response (see Chapter 28), caused by with a wide range of antibodies (NMDA, AMPA, GABA, LGI1, CASPR2). NMDA receptor encephalitis is the most common antibody-mediated encephalitis, and commonly presents with features of psychosis. It should be considered in any patient (especially young females) first presenting with psychotic symptoms of acute/subacute onset, particularly when accompanied by neurological signs and symptoms such as cognitive dysfunction, movement disorders, autonomic dysfunction, and seizures, among others. A significant proportion of patients present with psychiatric symptoms, even before recognition of any neurological ones. Patients may present disorientated, with extremely disorganized behaviour and bouts of aggression, at times requiring an ICU admission. Prompt treatment is associated with good outcomes, but with long hospital admissions. Protracted recovery and persistent neuropsychiatric symptoms are common, and relapse remains a possibility.

Epilepsy: the link between epilepsy and psychosis is complex and potentially bidirectional. Meta-analytical studies estimate the prevalence of psychosis in people with epilepsy to be 5.6%, with a higher risk in people with temporal lobe epilepsy. Risk factors include a family history of mental illness, young onset epilepsy, hippocampal sclerosis, structural brain lesions, and poorly controlled disease. Depending on the temporal association with ictal events, psychosis can be classified as inter-ictal, pre-ictal, ictal, and post-ictal, with the latter being by far the most common. Post-ictal psychosis may present hours to days following a seizure and presents mainly with delusional thought content, hallucinations, and potentially challenging behaviour, for example automatisms. The course of the episode is usually short with an average of 10 days. Pharmacological management needs to take into account pharmacokinetic interactions between anticonvulsants and antipsychotics, as well as the propensity of certain antipsychotics to lower seizure threshold; safer options here are amisulpride and aripiprazole.

Parkinson's disease is the second most common neurodegenerative disorder. While initially perceived as a movement disorder, it is now well-established that in addition to the classic motor symptoms (tremor, stiffness, and slowness among others), people with Parkinson's present with a wide array of non-motor symptoms, affecting cognition, mood, anxiety, sleep, bowel function, and more. Parkinson's psychosis refers to a range of psychotic symptoms commonly encountered in people with Parkinson's, ranging from low level hallucinations to a presence (a feeling that someone is near you) and passage (seeing someone crossing your peripheral field), all the way to well-formed, vivid visual hallucinations with a varying degree of insight. Delusions can be present at a later stage. Not to be conflated with delirium, the incidence of Parkinson's psychosis is estimated at around 30% in patients with Parkinson's and has been closely associated with the use of dopaminergic medication: adjustment of their dose should be considered as a first point of action. Psychosis is a poor prognostic factor and leads to increased caregiver burden. As most antipsychotics have the tendency to worsen motor function, these should be used only when other options failed. Most evidence for efficacy with minimal impact on movements lies with clozapine and pimavanserin, a novel antagonist/inverse agonist at serotonin 2A receptors.

Stroke: neuropsychiatric symptoms after a stroke are common, affecting more than 1 in 3 survivors: (for understandable reasons) anxiety is common (**Box 30.24**) and over one-third of stroke survivors become depressed. Post-stroke psychosis affects up to 5% of stroke patients, and may present a long time after a stroke has occurred (or in minor or clinically silent strokes), especially in those affecting the right (non-dominant) hemisphere. The commonest presentation is with delusions, and the disorder has been linked with poor functional outcomes and increased caregiver burden.

Traumatic brain injury (TBI): following a TBI, psychosis is 2–3 times higher than the general population, though exact prevalence estimates vary between 1% and 10%. Presentation is dominated by positive symptoms (delusions and hallucinations), whereas majority of negative/cognitive symptoms are attributed to neuronal damage from the TBI. It is important to differentiate TBI psychosis from the phenomenon of post-traumatic amnesia (PTA), common in TBI survivors, which is characterized by anterograde amnesia, agitation, and disorientation.

Drug-induced psychosis: this can occur during or soon after intoxication and/or withdrawal from a series of substances

including alcohol, cannabis, synthetic cannabinoids, opioids, sedative-hypnotics, anxiolytics, cocaine, stimulants (including amfetamines, khat, methamfetamine or methcathinone), synthetic cathinone, hallucinogens, volatile inhalants, MDMA, dissociative drugs (including ketamine or phenylcyclohexyl piperidine, PCP, often known as angel dust). Prescribed medications can also provoke psychosis (Box 30.3).

Management

Psychology

There is increased recognition of the central role psychological therapies can play in the effective management of psychosis. Psychoeducation helps patients and their carers enhance their understanding of the condition, and has been linked with reduced relapse rates in schizophrenia, with a specific place in supporting young people at high risk for psychosis. CBT, cognitive remediation therapy, and family interventions have been extensively studied, all with positive outcomes.

Social

The impact of psychosis on functioning is considerable, affecting every aspect of daily living, including life skills, personal relationships, and employability. Devising a holistic care plan that considers these aspects will improve prognosis and potentially prevent further episodes, while at the same time allowing every individual to live a fulfilling life. Social skills training, case management/care coordination, open dialogue, peer support workers, and therapeutic communities are all interventions that can boost social relationships and have been credited as beneficial in relevant research. A multidisciplinary mode of co-working with patients with psychosis improves functional outcomes; occupational therapists and social workers work alongside mental health nurses, psychologists, and psychiatrists.

Medication

First generation antipsychotics

Advances in psychopharmacology led to the manufacture of the first antipsychotic medication, chlorpromazine, in 1951. This transformed the management of people with schizophrenia, as being on an effective medication allowed patients to integrate and function within society, eventually leading to the gradual closure of most mental asylums. Other drugs of this first era of antipsychotics are collectively known as 'conventional', 'first generation', or 'typical' antipsychotics, that followed chlorpromazine (Box 30.29). Acting mainly by excessively blocking dopamine D_2 receptors, they have been linked with a series of side-effects that can become debilitating, including akathisia, dystonia, and parkinsonism; and, in chronic use, tardive dyskinesia (Box 30.29).

Second generation antipsychotics

Second-generation antipsychotics have unique additional properties that render them efficacious with a relatively more favourable side-effect profile. Examples are olanzapine, risperidone, quetiapine (all acting as serotonin 2A and dopamine 2 antagonists), aripiprazole, lurasidone, and newer agents such as cariprazine (a combined dopamine D_2, dopamine D_3, and serotonin 1A partial agonist and serotonin 2A antagonist). These 'atypical antipsychotics' are now considered to be the first-line medication option for the treatment of psychosis, and have also been used in disorders such as mania and augmenting antidepressants. Many can be given in long-acting intramuscular forms ('depot') which allows scheduled administration by community mental health teams, assuring adherence with treatment in patients with limited insight into their condition. Their side-effect profile is diverse (Box 30.29), including metabolic, endocrine and neurocognitive effects, and merits regular clinician review.

Box 30.29 Antipsychotic medications

Drug	Side-effects	Monitoring	Depot	Additional notes
Amisulpride	Hyperprolactinaemia	Renal excretion: eGFR monitoring is required when kidney function is impaired	NO	Safe for post-ictal psychosis
Aripiprazole	Akathisia described, less than first-generation APs	No specific requirements	YES	Preferred for first-episode psychosis
Cariprazine	Akathisia	No specific requirements	NO	Advantages in bipolar
Clozapine	Sedation, weight gain, hypersalivation, constipation, hypotension, neutropenia	FBC, lipids, weight, plasma glucose	NO	Best for treatment resistant-schizophrenia. Most effective in Parkinson's psychosis. Plasma levels require monitoring
Haloperidol	Akathisia, parkinsonism	Mandatory ECG	YES	Avoid in antipsychotic naïve patients. If used for rapid tranquilization, give alongside promethazine
Lurasidone	Sedation, akathisia	No specific requirements	NO	Can be effective with managing patients with a strong affective component
Olanzapine	Sedation, weight gain (worst of all APs)	Lipids, weight, plasma glucose	YES	Commonly used in rapid tranquillization
Quetiapine	Sedation; weight gain, hypotension	QTc Monitoring	NO	Preferred when opting for lighter dopamine blockade, twice daily dosing
Risperidone	Weight gain, hypotension, hyperprolactinaemia	Prolactin	YES	Can be considered for post-stroke psychosis

Clozapine

Clozapine remains a third-line medication for treatment-resistant schizophrenia (when there has been treatment failure with two other antipsychotic medications). This approach has been challenged, with research suggesting clozapine should be introduced earlier in patients in whom it is safe to prescribe. Side-effects linked to clozapine include myocarditis, constipation, and hypersalivation, among others. Agranulocytosis is a rare (0.8% of patients over a treatment lifetime), life-threatening complication. While mortality rates in schizophrenia are lower in the clozapine group, all patients on clozapine require frequent mandatory blood count and clinical monitoring.

Approach to pharmacological treatment

Predicting response to pharmacological treatment is a constant work in progress. Male sex, longer duration of untreated psychosis, severe symptoms at presentation, non-adherence, and co-morbid substance misuse disorders all predict poor response. Lower doses are preferred in first-episode patients, and these can be stopped (gradually) after 2 years under clinical supervision. All evidence points towards significant benefits for patients on antipsychotics compared to placebo, with significant reductions in the risk of relapse, inpatient admissions, and improvements in overall life expectancy. Despite this, adherence remains an issue for more than half of patients diagnosed with schizophrenia. One potential solution is the administration of the antipsychotic medication in a depot formulation; some countries use compulsory community treatment orders.

Other physical treatments: further to established pharmacological and psychosocial treatment options, there is evidence for the benefits of:

- Electroconvulsive therapy (ECT) which may be of some benefit in the medium term for psychotic symptoms resistant to medication.
- Transcranial magnetic stimulation (TMS) is a non-invasive neuromodulation technique that has been beneficial in emergent research for the negative symptoms of schizophrenia.
- Deep brain stimulation (DBS), a minimally invasive neurosurgical technique is also being considered for treatment-resistant patients in clinical research settings, though evidence is still scarce.

Major clinical trials are presently underway to test the safety and efficacy of medical cannabidiol (CBD) in treating people with psychosis. This presents a paradox: inhaled cannabis (especially high frequency, high potency in early teens) makes psychosis *more* likely, as well as being associated with other unwanted effects: lung damage, apathy, nausea and vomiting, subfertility, and cancer.

Further reading

Abboud H, Probasco JC, Irani S Autoimmune Encephalitis Alliance Clinicians Network et al. Autoimmune encephalitis: proposed best practice recommendations for diagnosis and acute management. *J Neurol Neurosurg Psychiatry* 2021; 92:757–768.
Fusar-Poli P, Borgwardt S, Bechdolf A et al. The psychosis high-risk state: a comprehensive state-of-the-art review. *JAMA Psychiatry* 2013; 70(1):107–120.
Kahn RS, Sommer IE, Murray RM et al. Schizophrenia. *Nat Rev Dis Primers* 2015; 1:15067.
Pollak TA, Lennox BR, Müller S et al. Autoimmune psychosis: an international consensus on an approach to the diagnosis and management of psychosis of suspected autoimmune origin. *Lancet Psychiatry* 2020; 7(1):93–108.

MEDICALLY UNEXPLAINED SYMPTOMS AND FUNCTIONAL NEUROLOGICAL DISORDERS

Persistent physical symptoms that are *not* clearly linked to an identifiable pathophysiological disease are common. Patients may experience debilitating symptoms of any system (back pain, tiredness, chest pain, palpitations, bowel problems, seizures, weakness, tremor) or behaviours noticed by others (frequent use of the bathroom, weak voice, difficulties with any physiological function). One in five presentations in primary care and one in two new neurology consultations are believed to be due to medically unexplained symptoms (MUS) and functional neurological disorders (FND), respectively.

The personal, social, and financial impacts of these conditions are significant, with patients sometimes experiencing negative attitudes and stigma. Despite the high use of healthcare services, correct diagnosis and appropriate treatments are often delayed, with negative consequences on prognosis and detrimental impact on patients' functioning.

Classification and terminology

Various terms have been used to describe MUS, and some classifications have proved controversial. In the latest ICD11 classification, a new term, ***bodily distress disorder***, has been introduced under 'Mental, behavioural or neurodevelopmental disorders' to describe 'the presence of bodily symptoms that are distressing to the individual and excessive attention directed towards the symptoms, which may be manifest by repeated contact with healthcare providers. This may not replace the American term 'somatiform disorders' or 'functional somatic syndromes'.

Other conditions historically considered under the category MUS include:

- **functional gastrointestinal disorders:** often called ***irritable bowel disorder***, or latterly disorders of gut-brain interaction, and included under 'Diseases of the digestive system'
- **functional neurological disorders:** now referred to as ***dissociative disorders*** under 'Mental, behavioural or neurodevelopmental disorders' (see Chapter 11, 'Functional neurological disorder' section)
- ***fibromyalgia*** (see Chapter 32): is classified as a chronic pain disorder under the broad category 'Symptoms, signs or clinical findings, not elsewhere classified'
- ***chronic fatigue syndrome/myalgic encephalopathy*** is included in ICD11 under 'post-viral fatigue syndrome' within the broader category of 'Diseases of the nervous system'.

Chronic fatigue syndrome/myalgic encephalopathy

Although several studies have demonstrated abnormalities in endocrine, neurological, and immunological parameters in patients with CFS, understanding of potential biological mechanisms underlying the condition remains poor. Historical approaches that have suggested that the disorder is 'all in the mind' are unhelpful and stigmatizing, and prevent the development of a beneficial therapeutic relationship between doctors and patients. A more helpful approach acknowledges the presence of both physical *and* psychological symptoms, with no assumptions about the underlying aetiology.

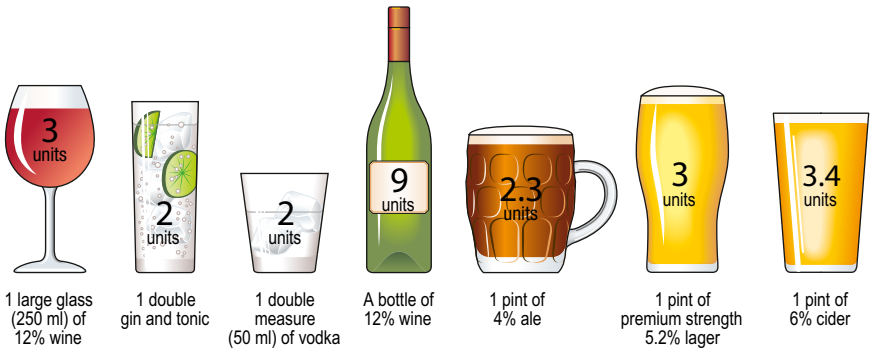


Fig. 30.7 Alcohol units.

Clinical approach to MUS/FND

While earlier theories placed disproportionate emphasis on psychological factors and life experiences, current understanding incorporates physiological processes with careful examination to uncover positive clinical features specific to MUS/FND before reaching a diagnosis. Identifying MUS/FND does not preclude other co-morbidities; for example a patient experiencing dissociative seizures (Chapter 28, 'Dissociative [non-epileptic] attacks' section) may also have epilepsy. Investigations may be required (video-EEG telemetry for seizure activity), and a rush to diagnose MUS/FND in the absence of positive investigation findings should be avoided.

Delivering and explaining a positive diagnosis of MUS/FND is the stepping stone towards a successful management plan that may include guided self-help, CBT, and input by several other members of a multidisciplinary team (e.g., physiotherapy or speech and language therapy, depending on the nature of the FND symptoms). Recovery depends on positive patient engagement with the process.

Communication of a diagnosis on MUS/FND

Patients with MUS/FND may experience symptoms which are considered 'red flags' for cancer or other serious health conditions, and anxiety about this may persist even when medical investigation has ruled these out. From a doctor's perspective, it might be regarded as good news that no more tests are indicated, but patients may continue to feel that their symptoms need an explanation. For example, a patient experiencing low back pain in the absence of physical trauma or radiological abnormality may be disappointed simply to be given medical advice on posture and gentle exercise. Although it would be wrong to decree that such back pain has a purely 'psychological' aetiology, psychological treatments (such as CBT) are proven to reduce symptom burden—likely by ameliorating the secondary effects of pain and loss of opportunities, combined with achieving behavioural changes to reduce avoidance.

Approaches to treatment

The spectrum of MUS disorders is broad, but many have the potential to become chronic and disabling and are unified by the absence of an effective biological cure. Treatment plans aim to lessen distress and disability, improve quality of life, promote effective coping strategies for symptom management, and facilitate recovery, and often require multiprofessional input organized via specialist clinics or services. Personalized integrated care plans may incorporate psychological therapies (e.g., CBT), psychosocial interventions (e.g., pragmatic active rehabilitation programmes), and drug therapies (e.g., antidepressant medications).

Further reading

Centers for Disease Control (CDC). *Myalgic Encephalomyelitis/Chronic Fatigue Syndrome Toolkit*. https://www.cdc.gov/me-cfs/.
Husain M, Chalder T. Medically unexplained symptoms: assessment and management. *Clin Med* (Lond) 2021; 21(1):13–18.
Sanal-Hayes NEM, Mclaughlin M, Hayes LD et al. A scoping review of 'Pacing' for management of Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS): lessons learned for the long COVID pandemic. *J Transl Med* 2023; 21:720.
Stone J, Burton C, Carson A. Recognising and explaining functional neurological disorder. *BMJ* 2020; 371:m3745.
White PD, Goldsmith KA, Johnson AL et al. Comparison of adaptive pacing therapy, cognitive behaviour therapy, graded exercise therapy, and specialist medical care for chronic fatigue syndrome (PACE): a randomised trial. *Lancet* 2011; 377:823–836.

ADDICTIONS

Worldwide, alcohol is one of the leading causes of disease and disability. Even at the peak of the COVID-19 pandemic, alcohol and substances had a greater combined mortality than COVID deaths. There are no safe levels of alcohol consumption and the medical recommendation is to consume less than 14 units (**Fig. 30.7**) over 7 days, spread out over 3 or more days. ICD11 classifies several disorders related to alcohol (**Box 30.30**).

Alcohol dependence

There are wide sex, age, and ethnic variations in how people metabolize alcohol. It is important to record the amounts consumed (**Fig. 30.7**) and the settings where excess is more likely. Most people underestimate and under-disclose their actual consumption; the alcohol and hospitality industries may obfuscate the total units in their products from the public. ICD11 defines alcohol misuse as a disorder of regulation of alcohol use arising from repeated or continuous use of alcohol. The characteristic feature of dependence is a strong internal drive to use alcohol (craving) plus impaired control. The diagnosis requires two or more of the three central features (see **Box 30.31**) to be evident over a period of at least 12 months but the diagnosis may be made if alcohol use is continuous for at least 1 month. Impaired control over alcohol use may describe a subjective sensation of urge or craving to use alcohol, or lack of ability to regulate the onset, level, circumstances, or termination of use. Patients admitted to hospital who meet some of these criteria should have a plan made to address potential alcohol withdrawal before it develops (generally comprising benzodiazepines and intravenous B vitamins) rather than wait for the consequences to become evident.

Box 30.30 Disorders caused by harmful use of alcohol and substances

I. Single episode of harmful use
II. Disorders due to substance use
- Harmful pattern of use (causing damage to physical/mental health or to others)
- Dependence (see definition of physiological effects, physical tolerance, withdrawal, etc. below; **Box 30.31**)
- Intoxication (this is not usually a diagnosis but a frequent part of the workload of emergency medicine departments, with acute intoxication being a temporary condition that happens due to ingestion of a large amount of the substance)
- Substance-induced delirium
- Substance-induced psychotic disorder
- Certain specified substance-induced mental or behavioural disorders
- Withdrawal states: where patients have physical and physiological symptoms after cessation of a substance; they have been using the substance for a prolonged period of time and are likely dependent on that substance

Box 30.31 Features of alcohol dependence

Note that the criteria to diagnose dependence on other substances are identical

Central features (2/3 present for 12 months is necessary to make the diagnosis)
- Compulsion (craving) to drink alcohol with stereotyped drinking pattern
- Tolerance to alcohol increases (less intoxicated on greater units)
- Stopping drinking leads to withdrawal symptoms

Other features
- Awareness of physical/psychological/other harms but continues to drink
- Neglect of other pleasurable activities, e.g., self-care, hobbies, social, occupational
- Time preoccupied with alcohol increases over time: 'salience'
- Out of control use of alcohol
- Persistent, failed attempts to cut down alcohol

For people with alcohol dependence, alcohol use becomes an increasing priority in life such that its use takes precedence over other interests or enjoyments, daily activities, responsibilities, diet, health, or personal care. Other areas of life are relegated to the periphery, a pattern that continues despite the occurrence of problems.

Withdrawal symptoms must be characteristic for the withdrawal syndrome for alcohol and must not simply reflect a hangover ('morning after' dysphoria, headache, and dehydration that follow an alcohol binge). Dependent drinkers frequently have neuropsychiatric damage from alcohol (**Box 30.32**): admission is an opportunity to identify these and discuss them with patients to enable them to make an informed decision about future drinking.

Assessment

A thorough history clarifies the age of starting to drink, drinking pattern, beverages used, quantity, duration, and the impact of alcohol on family, physical and mental health. Screening instruments like AUDIT (Alcohol Use Disorders Identification Test) are useful to detect alcohol dependence (https://auditscreen.org/). Unlike other tools, high AUDIT scores point to dependence—prompting clinicians to act to prevent withdrawal. Check for dependence and withdrawal symptoms, previous treatment, detoxification treatment, periods of abstinence, and the context of relapses.

Box 30.32 Neuropsychiatric complications of alcohol use include

- Pathological intoxication
- Dependence
- Amnesia (blackouts)
- Alcoholic hallucinosis (visual or auditory; this may occur at different times to alcohol intoxication)
- Delirium tremens
- Wernicke-Korsakoff syndrome
- Confusion
- Alcohol-related brain damage (ARBD): the most common treatable form of dementia—cognitive decline can be partially reversed with abstinence
- Cerebellar damage
- Impotence and impairment of libido
- Peripheral neuropathy (in its early stages, this is reversible with abstinence—unlike most manifestations listed here)
- Pathological or morbid jealousy—a person develops a delusion that their partner is unfaithful. This delusion is notorious for the delusional content upon which the person is most likely to take action, including violence/homicide.
- Exacerbation of existing disorder, e.g., anxiety, depression, psychosis, bulimia
- Marchiafava-Bignami disease (rare): primary degeneration of the corpus callosum associated with long-term alcohol excess

Box 30.33 FRAMES—brief intervention counselling approach for harmful alcohol consumption (and other behaviours that damage health)

Key elements of FRAMES are:

Feedback of personal risks or impairment:
Many people are usually unaware of harmful effects of the alcohol/substances that they are using. The counsellor provides feedback from the screening tools or blood results on how alcohol is affecting the person or their personal relationships, work, health, etc.

Responsibility:
The counsellor emphasizes that only the individual is responsible to change drinking pattern or behaviour change.

Advice:
Clear advice is given to reduce the harm associated with alcohol by reducing the pattern or behaviour.

Menu:
Provide information about different strategies to change drinking pattern, e.g., setting goals to limit alcohol consumption, referral to specialist teams, strategies to avoid high-risk situations, and people who can provide support.

Empathy:
Provide emotional support by engaging in a collaborative, understanding, reflective, and empathetic approach. A confrontational direct authoritative approach will not be successful in supporting patients.

Self-efficacy:
Fostering confidence and optimism in patient ability to make changes to alcohol consumption.

Identify elements of the presentation indicating risk to self or others: separation from partner, loss of job, relationships, domestic abuse, significant physical ill-health, financial problems, drunk driving, criminal charges, etc. Recording insight on MSE is a good place to check motivation for drinking, and any motivation to stop drinking. You might use the FRAMES approach within this (**Box 30.33**). Physical examination looks for signs of intoxication or withdrawal and long-term complications of alcohol (**Box 30.32**). Investigation results (**Box 30.8**) suggesting significant, harmful alcohol misuse include macrocytosis, deranged liver function, abnormal clotting

Box 30.34 The stages of change model

Stage of change	Description with worked example in relation to alcohol misuse
Precontemplation (Not ready to change)	Does not understand there is a problem with the behaviour, and not thinking about making any change
Contemplation (Getting ready)	May be ambivalent and is thinking about the change, e.g., thinking about cutting down amount of alcohol
Preparation (Ready)	Decided to make changes, and takes active steps to change
Action	Implementing smaller changes towards goals, e.g., drinking only in company, seeing a counsellor in a treatment centre
Maintenance	Avoids former unhealthy behaviours to prevent relapse
Relapse	Returns to unhealthy/harmful patterns or behaviours

times, and evidence of steatohepatitis on ultrasound or fibroscan. People who have injected drugs, even once, should be screened for (treatable) blood-borne viruses (**Box 30.8**).

Management

For heavy drinkers and those who are beginning to use health services as a result of drinking, intervention and brief advice (IBA) has demonstrated efficacy in reducing drinking, health harms, and ED attendances. Brief interventions could be of 5 to 10 minutes across 1–4 sessions based on motivational interviewing principles (**Box 30.33**). Alcohol misuse is often subculture-driven (celebrations when a sports team wins or drinks to commiserate a loss), but for many patients with alcohol dependence, their alcohol misuse loses any social connections and becomes secret, solitary drinking. Supporting people to cut down is challenging.

FRAMES (**Box 30.33**) is a model of motivational interviewing (MI) ideal for alcohol excess, substance misuse, problem gambling, and other health-harming behaviours. When working with patients with alcohol use, direct challenging may not be helpful as many are ambivalent to change. Expecting and accepting a patient's resistance to change are key concepts in MI. The clinician/counsellor identifies where the patient stands in the cycle of change and then designs sessions to support the specific issues related to the patient's goals of cutting down or abstinence. Prochaska and DiClemente proposed *readiness for change* as a dynamic process which is a vital mediator of behavioural change. This can be used for any unhealthy behaviours to drive change (**Box 30.34**).

Relapse prevention is achieved with medication, support, and contingency planning. Acamprosate and naltrexone are anti-craving medications which help in relapse prevention.

Disulfiram ('Antabuse') is less commonly used—this is an acetaldehyde dehydrogenase inhibitor, causing accumulation of acetaldehyde which causes flushing and vomiting sensations when patients drink alcohol. The intention is to link alcohol with aversive sensations, and to de-programme drinking behaviours, however, many doctors do not wish to prescribe something that they know (if their patient consumes alcohol) will make their patient very unwell, if temporarily.

Support is achieved by referral to a community additions centre or, in some cases, a residential rehabilitation unit. The mainstay of relapse prevention for 7 decades is the peer support group Alcoholic Anonymous (AA). AA, with regular meetings and its 12-step programme, has the strongest evidence for long-term abstinence. Some areas offer family therapy, but many families seek their own support in the parallel peer-run Al-Anon groups for families of alcoholics.

Box 30.35 Symptomatic treatment of opioid withdrawal symptoms

Symptom	Medication
Pain	Use non-opioid analgesics, e.g., non-steroidal anti-inflammatory drugs (NSAIDs) or paracetamol
Stomach cramps	Hyoscine butylbromide: 10–20 mg orally every 8 h
Diarrhoea	Loperamide: 4 mg orally initially, plus additional 2 mg after each loose stool up to 16 mg in 24 h
Nausea and vomiting	Metoclopramide: 10 mg 4–6 h up to three times a day
Use with caution—it can help with many of the opioid withdrawal symptoms	Clonidine, alpha-2-adrenergic agonist: as per local protocols

Further reading

Alcohol-use disorders: diagnosis, assessment and management of harmful drinking and alcohol dependence NICE clinical guideline 115 published 23/02/2011. www.nice.org.uk/guidance/CG115.

Opioid addiction

Worldwide, opioids are the second most harmful substance (after alcohol) with significant death rates from respiratory depression in accidental overdose. In 2021, the USA reported over 100 000 preventable opioid deaths, and drug-related deaths from opioids in the UK are currently the worst in Europe. Deaths are mostly from respiratory depression, when excessive doses overwhelm even highly dependent patients; by contrast with alcohol withdrawal, opioid withdrawal state is uncomfortable for patients, but not life-threatening. When patients use street opiates (heroin, morphine) and opioids (methadone, mostly illegal fentanyl), they are at a high risk of overdose, as the doses of what they buy are unpredictable. Street drugs are also riskier due to harmful contaminants, or substitution with another (unknown) agent when stocks are low. Patients may request opioids to reduce their withdrawal symptoms even when they might probably be still under the influence of opioids. A thorough assessment is required before prescribing to avoid iatrogenic respiratory depression. When in doubt, provide symptomatic treatment (**Box 30.35**) to manage objective withdrawal symptoms based on direct examination (**Box 30.9** and **Fig. 30.8**): see **Box 30.36**.

Commence opioid replacement therapy only after getting advice from specialists. The Clinical Opioid Withdrawal Scale (COWS) can be used to assess the severity of opioid withdrawal symptoms in patients. If patients are started on opioids in ICU or other medical settings for pain management, doses should be tapered slowly over several days

Acute medical management can be challenging, but the next steps to signpost patients to addiction treatment are often lifesaving (**Boxes 30.33 and 30.37**).

Opioid substitution therapy (OST) with methadone or buprenorphine is the mainstay of harm minimization in chronic opioid misuse. If someone is already on methadone substitution treatment, the correct dose should be prescribed throughout a

Fig. 30.8 Injection sites ('skin tracks') in an intravenous drug user. There is overlying hyperpigmentation and scarring of veins due to inflammation from repeated non-sterile injections as well as injections of irritating drugs and adulterants'.

Box 30.36 Opioid withdrawal—more objective features are underlined

At 8–12 h:
- Sweating, lacrimation
- Yawning, rhinorrhoea
- Dilated pupils, irritable mood
- Tremor, restless, ↑heart rate

Methadone and buprenorphine withdrawal:
- Starts after 24 hours
- Can last for 2–4 weeks

From 24 to 36 h:
- Nausea + vomiting
- ↑blood pressure and ↑heart rate
- Muscle and abdominal cramps
- Insomnia, nightmares

hospital admission. Always check—with GP or pharmacy if necessary—both the total dose and when the patient had their last dose: many have a home supply, and some may take it all if anticipating hospital admission. Most general hospitals have specialist substance misuse teams who can guide therapy. Often the total methadone dose is split into two in hospitals (given at 8 AM and 8 PM), even though methadone's half-life is 36 hours. This is because a minority of patients might decide to leave hospital after their full morning dose and obtain 'top up' heroin resulting in a high risk of respiratory depression. If someone misses methadone for more than 48 hours, then they may have reduced opioid tolerance, and it may be necessary to consider restarting dose titration.

Only senior doctors and specialists in addictions should prescribe methadone or buprenorphine (and never both in the same patient), given the risks of overdose and other harms. Just as benzodiazepines do not cure alcohol addiction, neither does OST provide a cure here. Psychoeducation is provided to patients about the risks of opioids use and treatment options available. Patients may be referred to specialist addiction services, perhaps to an inpatient rehabilitation unit. The uptake of Narcotics Anonymous is lower than AA, though it is effective, and people with cross addictions may attend both.

Other harm reduction strategies include needle exchange programmes to avoid sharing needles and prevent blood-borne infections, advice on barrier contraception, supervised 'shooting up rooms' to inject drugs with medical support but no police enforcement, and provision of home naloxone to reverse accidental overdose.

Novel psychoactive substances

These are synthetic compounds that produce similar effects to recreational drugs. There is limited data available, and the nature of the market makes it difficult to keep pace with the newest agents. Clinicians may look up the names online and try to match to known compounds. Most of them are not detected in urine/blood tests done for recreational drugs and this makes management more challenging. Novel psychoactive substances (NPSs) can be divided into stimulants, cannabinoids, hallucinogens, and depressants. Hallucinogens can be further divided into two categories as dissociative and psychedelics. Dissociative NPSs produce a 'dissociate' state—temporally altering a person's mood, thoughts, and perception. Patients frequently present to hospital emergency departments with dysphoric mood crashes or associated self-harm. Management is expectant, to treat withdrawal and keep the patient safe.

Further reading

Clinical Guidelines on Drug Misuse and Dependence Update 2017 Independent Expert Working Group. *Drug misuse and dependence: UK guidelines on clinical management*. London: Department of Health. 2017.
Psychiatry PRN. *Principles, Reality, Next Steps. Sarah Stringer, Laurence Church, Roxanne Keynejad and Juliet Hurn*. 2nd edn. Oxford University Press. 2020.
Semple, David, Roger Smyth. *Oxford Handbook of Psychiatry*, Oxford University Press, Incorporated, 2019.
Tracy DK, Wood DM, Baumeister D. Novel psychoactive substances: types, mechanisms of action, and effects. *Br Med J* 2017; 356.
Wesson DR, Ling W. The Clinical Opiate Withdrawal Scale (COWS). *J Psychoactive Drugs* 2003; 35(2):253–259.

PERSONALITY DISORDER

The term personality describes a person's unique character (traits) and ways of behaving. When taking a psychiatric history, it is important to ask about childhood temperament, and perhaps to which person in the family they are most alike, but there are no specific personality questions on the MSE. In treating anxiety, we try to define 'state anxiety' (episodic, often situational) as separate from 'trait anxiety' (habitual, and part of their usual personality). The same holds for people who have a constitutional mistrust of others: when might this become (if ever) a discrete episode of paranoia (by definition, a delusional belief system)? Personality dysfunction may become more obvious during physical and mental illness episodes: in a relapse of OCD, perfectionistic traits become obsessional behaviours such as rituals, while most aspects of character are exaggerated in mania. By their nature, personality disorders are fixed over an adult lifetime, though the behaviours and potential crises are modifiable with psychological treatments.

Box 30.37 The range of substances, taken by various means, and their harms

	Opiates	Cannabis	Stimulants	Hallucinogens	CNS depressants
Substances	Morphine, heroin Methadone Pethidine Codeine Dihydrocodeine NPS AH-7921, MT-45, novel fentanyl	Hash. Grass, Skunk, ganja, marijuana Main psychoactive contents are tetrahydrocannabinoid (THC) and cannabidiol (CBD) NPS: synthetic cannabinoid receptor agonists (SCRAs)	Cocaine, amfetamines Ecstasy (MDMA) NPS: Mephedrone, NBOMe, 2C-series	Two categories (1) dissociative, including ketamine, phencyclidine (PCP or 'angel dust') NPS: methoxetamine (2) Psychedelics, including lysergic acid diethylamide (LSD), gammahydroxybutyrate (GHB), psilocybin NPS: 5-MeO-DALT and the NBOMe- or 2C-series	Benzodiazepines Alcohol Barbiturates Sedatives Novel psychoactive substances: benzodiazepines—diclazepam, flubromazepam
Route	Smoking, snorted Tablets like codeine can be swallowed	Oral (cakes). Smoked (spliffs/joints) Inhaled (water pipe or 'bong') SCRAs sprayed on to herbal mixtures—called spice	Snorted, i.v. injection Amfetamines—oral Mephedrone (bombing—swallowed in wrapped paper)	Smoking Inhaled, swallowed, injected	Oral, i.v.
Intoxication	Rush is a brief intense pleasurable sensation, euphoria Other symptoms include pinpoint pupils, sedation, respiratory depression, analgesic effect	Euphoria, relaxation, distortion of time, disinhibition Psychiatric effect: Paranoia, over talkative giggling, anxiety Physical signs Tachycardia, conjunctival injection (red shot eye), dry mouth, poor coordination memory impairment	Abnormal excessive cheerfulness, over talkative, reduced appetite, increased frequency of micturition, poor concentration	Dissociative type—produce euphoric 'dissociated', state, catatonic state Psychedelics—perceptual alterations, quasi-mystical experiences MDMA can induce psychosis PCP ('angel dust')—can be fatal and cause violent reactions and brain damage LSD—exited, giggly. Euphoric, hallucinations, 'bad trip', flashback, visual distortions Tachycardia, hypertension, arrhythmias, fever, convulsions Formication—insects crawling under the skin (cocaine)	Lethargy Lack of concertation excessive sleepiness Fatal in high doses
Withdrawal symptoms	High resting pulse rate Sweating Restlessness Dilated pupils Bone and joint aches GI upset such as nausea, vomiting, loose stool Tremor Excessive yawning Anxiety or irritability Piloerection of skin Rhinorrhoea (runny nose) or lacrimation (tear production)	It may not cause physiological dependence Some people may experience irritability, anger, insomnia, vivid dreams, reduced appetite, weight loss, headache. Poor concentration, depression, stomach pan, sweating, chills	Drowsiness, fatigue, lack of interest, increased hunger		Withdrawal from barbiturates can lead to 'cold turkey' that can be fatal Benzodiazepines withdrawal symptoms include abdominal cramps, panic attacks, hallucinations, convulsions, and death
Complication	Overdose can cause respiratory depression Naloxone can reverse opioid overdose effects				Flumazenil can be used in benzodiazepine overdose
Management	Psychoeducation Methadone or buprenorphine substitution Harm reduction (avoid sharing needle, safe sex)				For benzodiazepine withdrawal—patient will need switching over from short-acting to long-acting benzodiazepines and then will slow tapering and stop

NPS, Novel psychoactive substances.

Box 30.38 Attachment styles in children and how these relate to adult attachment

Childhood attachment styles (prevalence)	Description of childhood behaviours observed in play rooms	Adult attachment
Secure attachment (60%–70%)	Upset when mother (carer) leaves their company, and play decreases. They settle, greet her on her return, and continue to explore their environment.	Autonomous
Insecure—avoidant (20%)	Infant appears independent when mother is present, and does not appear upset with separation; on her return seeks closeness but avoids gaze if picked up.	Dismissing
Insecure—ambivalent (10%)	Tend not to explore, preoccupied with mother and 'clingy'; upset by separation—may greet carer with ambivalence, e.g., reach out and push away at once.	Preoccupied
Unresolved (10%–15%)	No specific attachment style identified.	Disorganized

Box 30.39 ICD11 description of emotional, cognitive, and behavioural dysfunction in personality disorder

Emotional manifestations:
- Range and appropriateness of emotional experience and expression.
- Tendency to be emotionally over- or under-reactive.
- Ability to recognize and acknowledge emotions that are difficult or unwanted by the individual (e.g., anger, sadness).

Cognitive manifestations:
- Accuracy of situational and interpersonal appraisals, especially under stress.
- Ability to make appropriate decisions in situations of uncertainty.
- Appropriate stability and flexibility of belief systems.

Behavioural manifestations:
- Flexibility in controlling impulses and modulating behaviour based on the situation and consideration of the consequences.
- Appropriateness of behavioural responses to intense emotions and stressful circumstances (e.g., propensity to self-harm or violence).

Theoretical approaches

There are a multitude of personality theories, and many remain relevant to psychiatry (especially child psychiatry). Jean Piaget devised stages for each of cognitive and moral development. Social learning theorists point to the process of imitation of others (typically parents) as central to the development of personality, as it is in language development. Psychoanalysts (beginning with Sigmund Freud) have focussed on feelings and impulses (drives), for example, Erik Erikson's *Eight Stages of Man* as binary oppositions.

Building on many of these theories and drawing on anthropology, John Bowlby directly studied infant/child behaviour to devise *attachment theory* (see **Box 30.38**). According to attachment theory, around two-thirds of individuals were securely attached in childhood and are autonomous in relationships as adults. Adults in the 'insecure avoidant' group idealize their childhood, and tend to minimize ('dismiss') their own adverse, even traumatic, childhood events. Though the smallest proportion, the 'insecure ambivalent' or 'pre-occupied' group, is over-represented in psychiatric presentations: as adults, they often remain angry at their parents, wanting answers for their current unpleasant feelings.

The best definition of any psychotherapeutic intervention is that the therapist (who can never replace a parent or become the patient's friend) provides a *safe space from which to explore*. Once a person begins to understand variations in attachment, they can begin to imagine the effects of disrupted childhood development, and the multiple negative consequences of ACEs. In assessing patients with complex psychological symptoms, their early history (**Box 30.2**) can be a catalogue of childhood separation, loss, and uncertainty—even if there was no abuse.

Diagnosis

To establish this diagnosis, we must identify both the nature and degree (severity) of dysfunction in **functioning of aspects of the self** (identity, self-worth, capacity for self-direction) and/or problems in **interpersonal functioning** (developing and maintaining close and mutually satisfying relationships, understanding others' perspectives, managing conflict in relationships): see **Box 30.35**. There is a spectrum from 'normal' (poorly defined and subculture-dependent) to personality difficulty (traits present, but not a disorder) and then mild, moderate, and severe personality disorder [PD]). Just as Bowlby's prevalence rates (**Box 30.38**) are approximations, differences in measurement and standardization yield 'ball park' figures for personality difficulties (7%, but more prevalent in people in general hospitals due to the stress of being ill) and severe PD (4%).

Within psychiatric services, PD rates are higher still and greater levels of impact on others make the diagnosis more likely; PD can be context and culture specific, often manifesting at a time of crisis. Exclusions from diagnosing PD are ongoing substance misuse, effects of medications or medical illness (e.g., head injury causing enduring personality change), another active mental disorder, or specific developmental stages: for example, abnormal adolescent self-identity (of which, one example is gender incongruence). All five exclusions are also recognized co-morbidities in patients with established PD diagnoses. Severity and pervasiveness of core PD features (consistently present in adults for at least 2 years) merit documentation (**Box 30.39**). High scores in just one of five domains (**Box 30.40**) are unusual, but of itself, this is enough to confirm a PD diagnosis.

There is one clinical diagnosis of PD, but there are 93 separate ICD11 diagnostic possibilities from three levels of severity, five domains (plus a borderline identifier), and an additional 'unspecified PD' category: **Box 30.40**. Overlap is considerable, even before co-morbidities intervene. Approaches to managing an individual patient's difficulties differ widely: two people with *negative affectivity* may have either detachment (common in primary care) or dissociality (common in those in the criminal justice system). The former person is unhappy with self-blame and internalized anger; the latter may be of variable subjective mood but blaming of and angry at others.

Making a PD diagnosis relies on collateral history (**Box 30.6**): even with modern classifications, many of the words used *(fussy, rigid, manipulative)* are pejorative. We evaluate the behaviours (we might list instances of harm to others in dissociality; carefully noting the person's remorse for past actions) without judging the person as good or bad. For example, in the case of narcissistic PD (a 'cluster

B' diagnosis in the American DSM classification but not now part of ICD11), traits are predominantly viewed negatively by society: sense of entitlement, arrogance, displays of contempt to others, and more. Modern understandings of narcissistic traits lead us to see these patients (who ask for help) as insecure and having low self-esteem despite these traits. If that person engages in therapy, the focus might be to shift belief systems to the rewards of being a team player, that superiority to others is their value judgement or that we can be happy without needing affirmation and admiration by others.

Management

Only a minority will request treatment for PD of itself. Patients are more likely to present with a mood or anxiety disorder, perhaps where relationship difficulties dominate. Co-morbidities are treated in the usual ways, and changes in PD features as the crisis resolves are noted. Core treatment recommendations are structured assessments, engaging with carers, offering individual psychotherapy (or group-based for dissociality), and negotiating long-term goals. Medication is reserved for crises, and not given long term unless treating a key co-morbidity like depression. There is efficacy evidence for a specialist PD service within mental health services, providing therapies and multidisciplinary support for people with severe PD. Specific techniques showing promise include mentalization and dialectical behavioural therapy. The key is consistent support from a mental health professional within a team, with regular healthcare follow-up, and agreed 'safety net' plans in times of crisis.

Further reading

Bach B, Kramer U, Doering S et al. The ICD-11 classification of personality disorders: a European perspective on challenges and opportunities. *Borderline Personal Disord Emot Dysregul* 2022; 9(1):1–11.
Gabbard GO. *Psychodynamic Psychiatry in Clinical Practice*, 5th edn. American Psychiatric Association, 2014.
Goddard N. Human personality development (chapter) In: **Wright P, stern J, Phelan M (eds).** *Core Psychiatry*, 2nd edn. Elsevier: 2005.
Lee T, Solms M. Managing the clinical encounter with patients with personality disorder in a general psychiatry setting: Key contributions from neuropsychoanalysis. *BJPsych Advances* 2023; 1–8.
McAdams DP, Olson BD. Personality development: Continuity and change over the life course. *Annu Rev psychol* 2010; 61:517–542.
NICE borderline personality disorder https://www.nice.org.uk/guidance/cg78.
World Health Organization *ICD11* https://icd.who.int/browse11.
https://www.nice.org.uk/guidance/qs88.

EATING AND FEEDING DISORDERS

In ICD11 eating disorders are classified under 'feeding or eating disorders', and include anorexia nervosa (AN), bulimia nervosa (BN), binge eating disorder, avoidant or restrictive food intake disorder (ARFID), pica and ruminations—regurgitation disorders. Of these, AN is a severe mental illness with a high mortality and physical morbidity and is discussed in detail here.

Epidemiology

Eating disorders predominantly affect women but the prevalence rates are increasing in men too (in the USA, the lifetime prevalence for AN is 0.9% in women and 0.3% in men). Age at onset for AN for females is around 16–17 years (though more severe cases can start before or at puberty) and for males is around 12 years. Age at onset for BN is in females in mid-adolescent age. First-degree relatives of those with AN and BN have tenfold higher risk of developing eating disorders indicating they share similar genetic abnormalities.

Anorexia nervosa

The clinical features of AN are set out (**Box 30.41**). There are too many medical causes of unintentional weight loss to list here, and given the late presentations of AN, general hospitals are mostly the right setting to rule out medical illness as treatment begins. In BN there may be a similar concern about weight and body shape, with cycles of binge eating followed by purging, although body mass index (BMI) is often normal.

History

Patients with eating disorders are at various stages of motivation as described by Prochaska and DiClemente's transtheoretical model of change (**Box 30.34**). They may be reluctant to provide full history or may be ambivalent to engage in assessment. A collaborative, non-confrontational, and empathetic approach gathers a better history. If a patient says they have had three meals per day, then enquire in detail about the portion size and food contents to get an approximate idea about how many calories they are consuming. Record their explanations for weight loss and physical symptoms: these relate to insight and motivation. When asking about a

Box 30.40 ICD11 five personality domains, with additional specifier of borderline

Domains	Five-factor model	Description	Likely behaviours
Negative affectivity	High neuroticism	Lower self-esteem and self-confidence; tend to worry more/be more anxious; expect the worst	Avoid others and activities, at times become overdependent on others to regulate emotion
Detachment	Low extraversion	Low on trust for others, aloof; eccentric	No/few close relationships, tend to keep to themselves
Dissociality	Low agreeableness	Belief that rules are for other people, no care for consequences of their behaviours	Antagonistic, manipulative, or exploitative of other people; may become known to police
Disinhibition	Low conscientiousness	Cannot control emotions or temper	Impulsive, regret acts later; may be known for angry or unpredictable behaviours
Anankastic	High conscientiousness	Perfectionistic, high standards, obsessional by nature	Fixed routines, described by others as fussy or rigid
Borderline (also known as emotionally unstable PD)	Its name derives from the borderline of neurosis and psychosis: e.g., 'brief pseudo- or full hallucinations, paranoia in situations of high affective arousal'. PD with cycles of negative affectivity, high impulsivity (self-harm), instability of self-concept (emptiness), and relationships (fear of abandonment, lack of trust in others)		Long-term difficulties in work and relationships; shorter, intense relationships with denigration of person after; self-harm is common, and does not always present

patient's weight, now and over time, use open questions to check if there is any target weight (often pathological low weights) that the patient is aiming for.

Physical examination

Immediate physical examination is required in AN—do it systematically from head to toe, involving all the systems as included in **Box 30.10**. Patients can have signs of malnutrition (lanugo on arms), slow capillary filling, or proximal myopathy (**Box 30.42**). Check weight, height, BMI, pulse rate, and blood pressure, and check too for postural hypotension. You might use a squat test, asking the patient to stand and go low with fully bent knees before standing. When performing this test, support the patient as they may fall. Due to proximal myopathy, patients with AN are unable to rise from squatting position without supporting their upper limbs.

Box 30.41 ICD 11 criteria for anorexia nervosa

Extreme fear of weight gain
A persistent pattern of restrictive eating or other behaviours aimed at establishing or maintaining abnormally low body weight (reduced calorie intake, excessive exercise, or use of medications or laxatives)
Low body weight with a body mass index (BMI) less than 18.5 kg/m^2
Excessive preoccupation with body weight or shape/distorted body image
Endocrine disturbances (e.g., of the hypothalamic–pituitary–gonadal axis, which may result in amenorrhoea)

Investigations

Obtain an urgent ECG to confirm bradycardia or detect arrhythmias. Blood tests include nutritional markers (albumin, phosphate, magnesium). A DXA scan may be requested to assess bone density. Although rare in AN because of impaired physiology, always rule out pregnancy.

Management

Patients with AN may engage with medical services because they want a problem fixed (e.g., fainting, insomnia, secondary infections, etc.), and are usually ambivalent to engage in AN treatment that leads with weight gain. Initial assessment builds a therapeutic alliance: be empathetic (never blame them for not eating or overexercising) and use motivational interviewing techniques (**Box 30.33**) to work collaboratively with each patient. Medical Emergencies in Eating Disorders (MEED) is a new guideline to manage patients with eating disorders (**Box 30.43**). It provides a traffic light assessment tool based on a number of clinical features allowing clinicians to stratify clinical risk. In profoundly malnourished AN patients, the risk of refeeding syndrome should be considered and managed (see Chapter 25).

At an appropriate time, provide psychoeducation about healthy eating, health risks associated with starvation, exercise, and laxative use. Be sensitive with family members as they may be in a lot of

Box 30.42 Signs, symptoms, and medical complications of eating disorder

History/System	Starvation	Purging, laxative misuse, 'slimming pills' bought online
Hair	Thinning of hair, brittle hair, hair loss	
Central nervous system	Depression, insomnia, memory problems, fainting episodes, hypothermia	Generalized seizures EEG abnormalities Peripheral neuropathy
Haematological and Biochemical	Anaemia, pancytopenia Reduced ESR Compromised immune system Low magnesium, phosphates and more	Low potassium
Musculoskeletal	Proximal myopathy (Squat test) Reduced bone density leading to osteoporosis Joint pains Delayed growth	Worn upper teeth and marks on knuckles from purging Diarrhea may cause anal damage including rectal prolapse
Cardiovascular system	Bradycardia Low BP Reduced capillary filling Postural hypotension Symptoms of cardiac failure Mitral valve prolapse	Arrhythmia Cardiac failure Sudden death
Gastrointestinal	Dysphagia Delayed gastric emptying Acute gastric dilatation (superior mesenteric artery syndrome) Re-feeding pancreatitis Constipation, ileus	Parotid swelling Dental erosion Oesophageal erosion Gastric/duodenal ulcers
Endocrine and renal	Amenorrhoea Thyroid abnormalities Infertility Low glucose and phosphate Neurogenic diabetes insipidus Low sodium Renal stones	
Respiratory	Aspiration pneumonia Pneumothorax	

Box 30.43 MEED instrument to identify risk in eating disorders

MEED score example (Version 2)

	Red	Amber	Green
Weight loss	☐ Recent loss of weight of ≥1kg/week for 2 weeks (consecutive) in an undernourished patient 34 Rapid weight loss at any weight, e.g., in obesity or ARFID	☑ Recent loss of weight of 500-999g/week for 2 consecutive weeks in an undernourished patient	☐ Recent weight loss of <500g/week or fluctuating weight
BMI and weight	☐ Under 18 years: WFH <70%, Over 18: BMI <13	☑ Under 18 years: WFH 70%-80%, Over 18: BMI 13-14.9	☐ Under 18 years: WFH>80% Over 18: BMI >15
HR (awake)	☐ <40	☐ 40-50	☑ >50
Cardiovascular health click for blood pressure centile calculator	☑ Standing systolic BP below 0.4th centile for age or less than 90 if 18+, associated with recurrent syncope and postural drop in systolic BP of >20mmHg or increase in HR of over 30bpm (35bpm in <16 years)	☐ Standing systolic BP <0.4th centile or <90 if 18+ associated with occasional syncope; postrual drop in systolic BP of >15mmHg or increase in HR of up to 30bpm (35bpm in <16 years)	☐ Normal standing systolic BP for age and gender with reference to centile charts normal orthostatic cardiovascular changes, normal heart rhythm
Assessment of hydration status	☐ Fluid refusal, severe dehydration (10%): reduced urine output, dry mouth, postrual BP drop, decreased skin turgor, Sunken eyes, tachypnoea, tachycardia	☑ Severe fluid restriction, moderate dehydation (5-10%): reduced urine output, dry mouth, BP drop, normal skin turgor, some tachypnoea, some tachycardia, peripheral oedema	☐ Minimal fluid restriction, No more than mild dehydration (<5%): may have dry mouth or concerns about risk of dehydration with negative fluid balance
Temperature	☐ <35.5°C tympanic or 35.0°C axillary	☐ <36°C	☑ >36°C
Muscular function: Hand grip strength AND Muscular function: MUAC	☐ Male <30.5kg Female <17.5kg (3rd percentile) AND <18cm (approx BMI <13)	☑ Male <38kg Female <23kg (5th percentile) AND 18-20 cm (approx BMI <15.5)	☐ Male >38kg Female >23kg - AND >20 cm (approx BMI >15.5)
Other clinical state	☐ Life-threatening medical condition, e.g., severe haematemesis, acute confusion, severe cognitive slowing, diabetic ketoacidosis, upper gastrointestinal perforation, significant alcohol consumption	☐ Non-life-threatening physical compromise, e.g., mild haematemesis, pressure sores	☑ Evidence of physical compromise, e.g., poor cognitive flexibility, poor concentration
Activity and exercise	☐ High levels of dysfunctional exercise in the context of malnutrition (>2h/day)	☐ Moderate levels of dysfunctional exercise in the context of malnutrition (>1h/day)	☑ Mild levels of or no dysfunctional exercise in the context of malnutrition (<1h/day)
Purging behaviours	☑ Multiple daily episodes of vomiting and/or laxative abuse	☐ Regular (=>3x per week) vomiting and/or laxative abuse	
Self-harm and suicide	☐ Self-poisoning, suicidal ideas with moderate to high risk of completed suicide	☐ Cutting or similar behaviours, suicidal ideas with low risk of completed suicide	

Box 30.44 ICD11 criteria for bulimia nervosa

Frequent, recurrent episodes of binge eating (defined as a discrete period of time, e.g., 2 hours, during which the individual experiences a loss of control over their eating behaviour and eats notably more or differently than usual)
Extreme fear of weight gain
Excessive preoccupation with body weight or shape
Repeated inappropriate compensatory behaviours to prevent weight gain (excessive exercise, self-induced vomiting, use of laxatives or diuretics, omitting insulin if diabetic)

distress. Enteral (nasogastric) feeding is used as emergency rescue treatment—however, patients benefit from a *food first* approach as there is no physical reason they should not achieve nutrition by eating. A multidisciplinary approach including the medical team, LP, an eating disorders psychologist, and dietician helps support patients through difficult initial stages of treatment. Once they achieve a safe weight (note patients may employ bargaining strategies around weight targets, an indication of the severity of their AN), LP plans discharge with local eating disorders services.

Bulimia nervosa

Bulimia nervosa (BN) is a more common eating disorder than AN, but with lower morbidity and mortality. Diagnosis is by history (binges leading to self-induced vomiting, see **Box 30.44**), and there are commonalities with AN. If there is low BMI, physical examination and investigations are the same for AN. Differential diagnoses include rare medical diseases causing vomiting (upper gastrointestinal obstruction, raised intracranial pressure) and other psychiatric disorders (depression, PD, OCD).

Most patients with BN can be managed in the community as they are more forthcoming with their symptoms/cognitive distortions and more motivated to engage in treatment than low-weight AN patients. Treatment includes psychoeducation and treating any medical or nutritional complications. Psychological treatment comprises individual or group CBT adapted for eating disorders, or interpersonal therapy models. Family therapy is helpful in younger people. SSRIs in higher doses can be beneficial when patients have binging and purging episodes. BN has a better prognosis compared to AN, with a 30% remission rate with CBT.

Avoidant/restrictive food intake disorder

ARFID is an eating disorder, perhaps better described as disordered eating, where children/adults do not eat enough food (or do not eat a wide enough variety of foods), leading to periodic loss of weight. Children developing the condition before puberty develop significant, sometimes severe, nutritional deficiencies and/or growth problems. ARFID can happen at any age but is more common in younger children, more boys than girls. Children might present with history of food avoidance, reduced appetite, restrictive diet, dislike of the sensory aspects of hated foods, abdominal pains, or fear of vomiting. Children with ARFID are more likely to have coexisting difficulties such as anxiety, OCD, autism spectrum disorders, and attention deficit disorder. ARFID is different from AN and BN in that patients do not have body image disturbances, nor do they deliberately lose weight; induced vomiting is uncommon unless there is a severe aversion to a food type or its texture.

There are considerable overlaps between ARFID and both neurodevelopmental disorders (autism) and medical diagnoses (principally irritable bowel syndrome, now described as disorders of brain–gut interaction). To make the diagnosis, the restricted food intake and consequent weight loss must not be due to unavailability of food, or be a manifestation of another medical condition (e.g., food allergies, hyperthyroidism), must not be due to the effect of a substance or medication (e.g., amfetamines), including withdrawal, and not due to another mental disorder.

ARFID is best managed in an MDT outpatient setting consisting of a doctor, dietician, psychologist, and other professionals such as speech and language therapist (to address swallowing difficulties for certain foods). The main goals of treatment are to treat underlying nutritional deficiencies, to understand the reasons for avoiding certain type of food, and to provide psychoeducation with support towards healthy eating. Psychological treatment is based on CBT models where graded exposure can help patients to manage anxiety associated with food: increasing 'safe' foods and portion sizes are a good start. Family therapy can also help anxious or angry parents to support their children towards 'food first' healthy eating, create a healthy environment/routine around mealtimes, and thereby reduce anxiety associated with food. As ARFID is a relatively new diagnosis, it is unwise to speculate on prognosis.

Further reading

Hudson JI, Hiripi E, Pope HG Jr, Kessler RC. The prevalence and correlates of eating disorders in the National Comorbidity Survey Replication. *Biol Psychiatry* 2007; 61(3):348–358.
NICE guidelines on eating disorders; recognition and treatment. https://www.nice.org.uk/guidance/ng69/resources/eating-disorders-recognition-and-treatment-pdf-1837582159813.
RCPsych. *Medical Emergencies in Eating Disorders: Guidance on Recognition and Management*. RCPsych College Report [CR233]. 2022. Accessed online at https://www.rcpsych.ac.uk/improving-care/campaigning-for-better-mental-health-policy/college-reports/2022-college-reports/cr233.
https://digital.nhs.uk/data-and-information/publications/statistical/hospital-admitted-patient-care-activity/2021-22.

ADDITIONAL CONSIDERATIONS IN SELECTED PATIENT GROUPS

Box 30.45 sets out additional challenges in assessing a variety of other patient groups.

Neurodiversity and learning disability

Individuals with intellectual disability and/or neurodiversity are often treated by specialist mental health teams. If they are admitted to hospital, their developmental stage and communication differences impact on the investigation and treatment of medical and surgical conditions. Just as when there are specific physical barriers (e.g., in individuals with visual or hearing impairment or a language barrier), it is important to make *reasonable adjustments* to reduce uncertainties about diagnosis and treatment over the course of admission. Finding old notes *(what worked before)* and engaging carers are essential preparations, and although LP does not need to be involved in every case, the treating team benefit from direct contact with the community teams including general practitioner who knows them best. Many hospitals have specialist learning disability teams to promote equitable care for these individuals: this is necessary to support people with these conditions and to establish new diagnoses.

Box 30.45 Communication with specific patient populations

For all groups:
- be flexible in your interview style
- consider shorter interviews (come back later)
- is there a staff member who already gets on well with this patient? Consider a joint assessment
- think through the (potentially negative) consequences of giving your formulation/diagnosis there and then

Category	Challenges for clinician	How to approach interviews
Visually impaired or hearing-impaired patients, or those with a language barrier	Bias towards finding other deficits, assumptions that delayed responses indicate psychopathology; premature use of closed questions.	Check old notes, collateral history before interview, lower threshold to have friend or relative in the room, well-lit and quiet room, communication aids (hearing loop, cue picture cards), in-person interpreter; chaperone for physical exam.
Young adults (see Chapter 46)	Assessment must be appropriate to their developmental age; transitions (from paediatric to adult services) are difficult for all.	Trauma-informed approach, if possible embed mental health services within a clinical environment designed for young adults.
People with intellectual disability and/or autism spectrum disorder	Evaluate capacity to consent to assessment and treatment. Interview must be appropriate to developmental age. Consider IQ (intelligence quotient) testing; in autism, there may be a double empathy challenge contributing to patient–doctor misunderstandings: doctors assume the patient lacks empathy, and reduces their empathy towards that patient.	Check old notes, obtain collateral history. Consider locating the assessment in patient's home. Know the prevalence: 2% of the population have intellectual disability, only some of whom have mental disorders (although with 5 times the prevalence as in the general population); autism spectrum has a 1%–2% prevalence, though it is higher in females than previously thought. Seventy percent on the autistic spectrum have intellectual disability; autism has standardized assessment tools.
Attention deficit hyperactivity disorder (ADHD)	Patients present with long-term concentration and attention difficulties: key collateral comes from parents, employers and teachers/school records. ADHD is a collection of neurodevelopmental disorders: likely overdiagnosed in the USA, and underdiagnosed in UK adolescent and young adult populations.	ADHD diagnosis is specialist based and there are a number of validated instruments to support this. Amfetamine prescription is also best left to specialist clinics: there are risks of addiction and long-term use but clear benefits if the ADHD diagnosis is correct. A specialist ADHD clinic must then monitor progress closely.
Cognitive impairment	Capacity may be absent for the interview: consider safeguards for the patient; in traumatic brain injury, (post-traumatic) memory loss can be temporary; variation is also expected in delirium and post-ictal patients.	Check old notes for risk factors (smoking, hypertension, arrhythmias etc.), collateral history, and consider MOCA (Montreal Cognitive Assessment) before interview. Look for treatable pathology: depression is present in 50% of stroke patients and is common in Alzheimer's disease.
Forensic populations: people charged with or convicted of a crime	Think about consent: the role of medical practitioners is not to gather evidence of a crime, however, such details may arise in the interview. Criminal intent to harm others should be reported to the police, breaking confidentiality; potential risks to other patients or staff (e.g., from Impulsivity) should be recorded and mitigated. Previous violence to others predicts future violence: devise containment and contingencies plan.	Apply the same standards and ethics as to other patients; focus on the patient's symptoms. Mood, hope, and suicidality are key questions: suicide in prisoners on remand (not yet charged with a crime) is relatively high—especially in people who are innocent, or if the person has no one to advocate for them. Doctors have a role in advocating for patients' needs.

Attention deficit hyperactivity disorder

Currently one of the most frequent medical queries from adolescents and young adults at psychiatric clinics is '*do I have attention deficit hyperactivity disorder (ADHD)*?' Psychiatric co-morbidities are almost universal in ADHD, but patients may prefer ADHD medication (amfetamines) to treatments for depression, anxiety, eating disorder, addiction, trauma, etc. LP teams are not likely to start these but may continue their prescription in hospital if the patient has a plausible ADHD diagnosis.

Mental health problems in older people

Across the world, most hospital beds are occupied by people over 65. The community prevalence of dementia in this age group is 7% (with significant numbers undiagnosed); rates are higher in hospital due to multimorbidity. Most will be managed by geriatric medicine teams (see Chapter 49), though LP may be consulted for advice on management of *acute on chronic* confusion, specifically delirium. Depression may coexist with dementia and a specialist LP assessment may make a major difference to that person's quality of life in hospital and beyond.

Patients in the criminal justice system

Someone convicted or charged with a violent crime *may* pose risks to other patients or staff if they are admitted to hospital, and so this information should be communicated as necessary to staff in order to keep the ward safe. This should be done without discriminating against the patient, who has a right in law to equal treatment in healthcare settings. It is for the courts and prison systems to deal with offenses and punishment, not doctors. Mental capacity to consent to or refuse treatment is almost always present, and the limits of confidentiality should be discussed with patients, including in circumstances where reports on medical or psychiatric conditions might later be requested by judicial authorities. If the person's criminal offences have been linked to mental disorders, they are best managed by a specialist forensic psychiatry team, and agreeing adjustments to management between teams is an important part of care. The patient might be subject to prescribed restrictions to their liberty, and Court Orders take precedence over routine capacity decisions.

Further reading

Byrne P, Byrne N. *Psychiatric Assessment in Hutchinson's Clinical Methods*, 25th edn. 2022. **Glynn M, Drake W (eds).**

Kazda L, Bell K, Thomas R et al. Overdiagnosis of attention-deficit/hyperactivity disorder in children and adolescents: a systematic scoping review. *JAMA Netw Open* 2021; 4(4):e215335.

May T, Birch E, Chaves K et al. The Australian evidence-based clinical practice guideline for attention deficit hyperactivity disorder. *Aust N Z J Psychiatry* 2023; 57(8):1101–1116.

DIAGNOSES OF EXCLUSION: FACTITIOUS DISORDER AND FACTITIOUS DISORDER INDUCED BY ANOTHER

Sometimes patients do not fit neatly into the diagnostic categories of psychiatry and other medical specialties. Patients may continue to have high symptom levels, be unresponsive to treatments, or be assigned diagnoses that fluctuate as one is disproven over another.

Where there is concern about diagnosis, patients should be reassessed with colleagues in joint/interdisciplinary assessments, perhaps in a different setting than before. Information may be sought from additional sources, including historical notes and documentation from other hospitals. Contradictory investigations should be reviewed and alternative explanations considered. These patients take time and they often need the additional time investment by a multidisciplinary panel to collate findings.

There are three broad explanations when findings do not conform readily to recognized diagnostic categories:

- ***Wrong diagnosis:*** doctors don't always get it right. Perhaps a patient felt pressured to answer questions to give positive responses to closed questions, deny a key symptom 'in the moment', or the wrong assumptions, even bias, led the team in the wrong direction. Overlooking an outlier investigation which challenges a working assumption (usually a negative result) is readily done, though some critics have called this 'diagnostic greed'. If it is clear a mistake was made, there is a duty of candour to the patient to disclose this and formulate a new treatment plan. This is a variant of breaking bad news, although the patient may extract a positive from the exchange, as the wrong treatments will be stopped, with the hope of better alternatives. A good example of this is the late diagnosis of bipolar disease (**Box 30.22**) in a patient previously diagnosed with recurrent depression. The average time to diagnosis of bipolar remains 10–15 years; patience and an open mind reduce this.
- ***Lack of honesty during the medical or psychiatric evaluation.*** First impressions ('*the doctor was busy, they didn't listen or care*') impair the development of a trusting doctor-patient relationship. Disengagement is common in clinical practice, and sometimes dishonesty plays a role in the process, generally higher in that small proportion of patients who are dissatisfied with their medical care. In retrospect, it may be possible to explain inconsistencies in patients' histories: some might have felt they needed to exaggerate some symptoms to achieve faster or better help. *Primary gain* is the relief of symptoms; *secondary gain* refers to the indirect benefits of help-seeking behaviours, and does not result from a conscious process. Just as the sick child may get extra treats or get to skip school, so (rarely) there may be income generation from sickness in adults, such as a legal payout or access to income benefits. Studies of medical interviews have highlighted 'inconsistencies' in both what patients report, and what doctors say in response. In the same clinical scenario, one doctor may promise unrealistic cure, and a second might declare there is no hope for any recovery.

  With thought and time, these relationships are usually salvable—even when the patient has elaborated for the purposes of secondary gain. The best approach is *acknowledgement* (we now agree your symptoms are limited to X) then *action* (agreed treatment plan for X). If there are grains of truth within primary symptoms, these scenarios are not without hope and can usually be rescued in a redefined, open, and honest doctor–patient relationship, so that the patient's care remains within the treating hospital system. The key breakdown of the therapeutic relationship is when a patient has had a supportive challenge but continues to overstate or lie about symptoms.
- **Diagnoses of exclusion: malingering or factitious disorder (FD).** A malinger wants to miss work, an exam, or a court case, and consciously plans to achieve this. Perhaps they are bringing a legal case for compensation against a person or system, and think through what to say to actively simulate and dissimulate towards a payout. Malingering prevalence is thought to be less than 1% of all medical presentations, but up to 8% of people referred to occupational health programmes. Clinicians have duties to be honest (and not join a 'scam' to mislead), duties to their patients (to advise them not to risk reputation, health, and more by pursuing a false claim), and duties to the services that employ them. Confronting malingerers is difficult, and might end professional relationships with them.

  The (even) more challenging diagnosis FD describes simulation of physical and/or mental symptoms that are ***not*** present. There is no *apparent* gain for FD patients, unlike in those who are malingering. **Box 30.46** sets out common features from the literature. The 'gains' (why they do this) often relate to poor relationships with clinicians, grievances against others/institutions, and feeling 'special' as a patient.

Management of factitious disorder

Even with 'supportive confrontation', FD patients usually self-discharge and do not attend follow-up with that service. The risks to themselves are manifest; they consume considerable resources and may disrupt medical systems. Because there is no identifiable, treatable mental disorder, psychiatric treatment should not be forced on them. In some cases, there is a duty to share information with other hospitals—with both a best interest and a public interest defence. If a patient with a diagnosis of FD seeks to work

Box 30.46 Features of factitious disorder

Has attended many previous clinics/hospitals over a wide area, and (importantly) resists requests for shared information from these: they may appear cooperative but give wrong clinician contact details, etc.	History has inconsistencies in dates or severity of symptoms Physical examination/observations mismatch reported symptoms MSE: no findings, no disorder obviously explains symptoms	Illness course does not follow expected pattern: too many 'relapses' following brief responses to treatment. Sometimes a routine test, previously normal on many readings, comes back as abnormal—in this patient, in the absence of other causes, might s/he have interfered with the test?
Connections with health professionals: family members, medical/nursing students Up to 50% of FD cases in some series are in health professionals or students	Hospital admission reveals different patient accounts to different staff Nurses (mostly outside office hours) observe secretive/odd behaviours Patients try to read their notes or access forbidden hospital areas	Evidence emerges that some abnormal findings have been manipulated by the patient; i.v. lines/wounds might be wilfully contaminated Patients may have their own medication/insulin supply
May seek procedures that have been declined (for good medical reasons) by other hospitals: enteral feeding tubes, parenteral nutrition, etc.	Hospital admission also defined by arguments with staff and patients: disruptive of ward activity with 'self-aggrandizing behaviours'	Evidence of covert treatment refusal: supplies of tablets found hidden, medications/ parenteral nutrition flushed down the toilet
May seek euphoria-inducing drugs as part of complex sick role behaviours: opioids, benzodiazepines, cyclizine	May even interfere with other patients' medical care: 'I am only the one who can help—not these (bad) doctors and nurses'	Resist psychiatric evaluation. Inconsistencies when assessed Don't engage with support; don't attend follow-up

in a medical or nursing setting, objective evidence of dishonesty should be shared with their university and professional registration authorities.

Factitious disorder imposed on another

Factitious disorder imposed on another (FDIA) is a term assigned to an individual who is feigning, falsifying, or inducing the symptoms *in another person*, not to the person who is presented as having the symptoms. The term replaces the older much litigated diagnosis of 'Munchausen's syndrome by proxy'. The ill person is invariably unaware of subterfuge, and FDIA might be seen as a variant of *gaslighting*—where a more powerful (controlling) person convinces another that they are the one who is abnormal in some way. Occasionally the individual induces or falsifies symptoms in a pet rather than in another person. The commonest setting is on paediatric clinics and wards, where FDIA does happen (though it is thankfully rare). A parent, for whatever reason (often none can be found), *actively* makes their child sick, inviting more harmful investigations and treatments. Some parents seem validated by dramatic medical interactions or crises that most parents would dread.

Child protection procedures should be initiated: this process will gather evidence and get together as many professionals as possible (teachers, clinicians, social workers, and more) into the same case conference to decide how best to keep the child safe. Raising a tentative diagnosis of FDIA will involve conflict with the parent; however, professionals cannot be faulted for requesting a child protection conference if something does not seem right. By contrast, failing to act on a 'red flag' of unexplained injuries and/or harms to a child will have longstanding consequences for them (and the clinician involved). Doctors and other health professionals do not wilfully cause harm to their patients, but FDIA is an exception. If an FDIA diagnosis is proven in a healthcare professional, they cannot work again in a clinical setting, supervised or not.

Further reading

Bass C, Halligan P. Factitious disorders and malingering: challenges for clinical assessment and management. *Lancet* 2014; 383:1422–1432.
Eisendrath S, Rand D, Feldman M. Factitious disorders and litigation. In: **Feldman M, Eisendrath S (eds).** *The spectrum of factitious disorders*. American Psychiatric Press: Washington, DC 1996; 65–68.
World Health Organization *ICD11* https://icd.who.int/browse11.

PREVENTION OF MENTAL ILLNESS: PUBLIC MENTAL HEALTH

Rates of referral to psychiatry services are rising globally, and only a fraction of these are explained by increased mental health awareness, lower referral thresholds, and over-medicalization. Rather, there is evidence for a global mental health crisis, fuelled by worsening social and economic determinants of mental illness (**Box 30.15**) and a fall in protective factors. System reform of health services to make care more appropriate and accessible to people who need it is only part of the solution.

Public mental health applies a multisystem public health approach to the prevention of mental disorders, and advocates to reduce the mental health treatment gap (the proportion of people with mental disorders who get no treatment, currently 3 in 4). The alternative to adopting a more proactive, preventative approach to mental illness is unthinkable: a rising proportion of the population who are economically inactive due to illness and overwhelmed health services.

Prevention can be applied at three levels:

- *Tertiary prevention* is treating the symptoms of mental illness to restore good levels of function and prevent relapse, including much of the work traditionally carried out by specialist psychiatry services.
- *Secondary prevention* involves actively looking for illness and working with patients to remediate early ***(don't just screen, intervene).*** Early intervention (EI) teams in psychosis are a good

example of how providing rapid assessment, support, and treatment can avert a prolonged period of mental illness.

- *Primary prevention* involves actions to prevent onset of mental health disorders. **Box 30.15** illustrates how an evidence-based understanding of the social determinants of mental illness guides universal and selected interventions to support individuals to maintain good mental health.

Investment in public mental health programmes pays off in multiple domains, but political pressures that resist income distribution and commercial interests that benefit from activities that damage health provide challenges to implementation. Falling life expectancy in Western countries (despite medical advances and health spending) is frequently mediated by poor mental health. Future generations of doctors, working collaboratively, have a chance to change this.

Further reading

Kirkbride JB et al. The social determinants of mental health and disorder: evidence, prevention and recommendations. *World Psychiatry* 2024; 23(1):58–90.
Nason M, Petticrew M, Galea S, eds. *The Commercial Determinants of Health*. Oxford University Press, 2023.
Patel V et al. Transforming mental health systems globally: principles and policy recommendations. *Lancet* 2023; 402(10402):656–666.

Questions for this chapter can be found on Elsevier eBooks+.

31 Ophthalmology

Robin D. Hamilton

Background 1044
Applied anatomy and physiology 1044
Clinical approach to the patient with a disorder of the eye 1046
History 1046
Examination 1046
Ophthalmic emergencies 1048
Eye trauma 1048
Acute red eye 1049
Sudden onset sight loss 1049
Common disorders of the eye 1050
Refractive errors 1050
Disorders of the lids 1050
Conjunctivitis 1052
Corneal disorders 1054
Cataracts 1055
Glaucoma 1055
Uveitis 1057
Disorders of the retina 1059
Visual loss 1060

CORE SKILLS AND KNOWLEDGE

Ophthalmology often gains limited exposure during medical training. It remains the busiest specialty in medicine and although many diseases are minor and self-limiting, destructive diseases can lead to serious disability and may be caused by a range of systemic inflammatory disorders.

All doctors, especially those working in primary care, require a basic understanding of eye function and health. Key to this is the approach to the common symptoms and signs that may result in visual loss in both the short and long term. Care for more serious conditions is generally carried out in specialist ophthalmology centres.

Key knowledge and skills to obtain in ophthalmology include:

- Developing an appropriate diagnostic approach to examining the eye appropriately
- Identifying sight-threatening eye emergencies and understanding their urgent management
- Identifying urgent eye diseases that may be life-threatening and managing appropriately.

Learning opportunities include attending ophthalmology outpatient clinics and surgical lists as well as practising clinical skills including slit lamp examination and fundoscopy.

BACKGROUND

Ophthalmology is one of the busiest specialities in medicine. In the NHS, it accounted for 10% of all outpatient activity (nearly 8 million appointments) in 2019–2020, as well as 6% of all surgical interventions, mostly carried out as outpatient, day-case procedures. Cataract surgery is the most common operation offered on the NHS, and the second commonest worldwide after Caesarean section.

Most ophthalmologists are 'surgical ophthalmologists', who perform a range of surgical procedures including cataract surgery, squint and glaucoma surgery, retinal surgery, corneal transplantation, orbital surgery, nasolacrimal surgery, and oculoplastic surgery, often with the help of an operating microscope and with or without lasers.

However, over 50% of initial presentations of ophthalmic conditions are associated with underlying systemic disorders, requiring ophthalmologists to have a broad awareness of general medicine.

Medical ophthalmology is a growing medical subspecialty treating vascular, inflammatory, and neurological disorders of vision, including diabetic nephropathy, uveitis, eye disease in systemic conditions such as vasculitis or thyroid disorders, and neuroinflammatory conditions such as optic neuritis. Such '**ophthalmic physicians**' may deliver procedures such as laser therapy or intraocular injection, or may be involved in public health interventions such as retinal screening.

Most major types of eye disease are covered in this chapter; however, diabetic eye disease (Chapter 35) and hypertensive eye disease (Chapter 18) are discussed elsewhere.

APPLIED ANATOMY AND PHYSIOLOGY

The average length of the human eye is 24 mm. It is essentially made up of two segments:

- The smaller anterior segment is composed of the conjunctiva, anterior chamber, and iris and is coated by the transparent cornea, and measures approximately 4 mm in depth
- The larger posterior segment is composed of the vitreous, retina, choroid, optic nerve head and is coated by the opaque sclera, and measures approximately 20 mm in depth.

It is the cornea and the sclera that give the mechanical strength and shape to the exposed surface of the eye.

The cornea occupies the central aspect of the globe and is one of the most richly innervated tissues in the body. This clear, transparent, avascular structure, measuring 12 mm horizontally and 11 mm vertically, and provides 78% of the focusing power of the eye. The eyelids prevent the cornea from drying and becoming an irregular surface by distributing the tear film over the surface of the globe with each blink.

Anatomically, the cornea is made up of five layers:

- epithelium
- Bowman's layer (membrane)
- stroma
- Descemet's membrane
- endothelium.

The endothelial cells lining the inner surface of the cornea are responsible for maintaining the clarity of the cornea by continuously pumping fluid out of the tissue. Any factor that alters the function of these cells will result in corneal oedema and cause blurred vision.

The sclera is an opaque white structure covering four-fifths of the globe and is continuous with the cornea at the limbus. The six extraocular muscles responsible for eye movements are attached to the sclera, and the optic nerve perforates it posteriorly.

The conjunctiva covers the anterior surface of the sclera. This richly vascularized and innervated mucous membrane stretches from the limbus over the anterior sclera (where it is called the bulbar conjunctiva) and is then reflected on to the undersurface of the upper and lower lids (the tarsal conjunctiva). The area of conjunctival reflection under the lids makes up the upper and lower fornix.

The anterior chamber is the space between the cornea and the iris, and is filled with aqueous humour (Fig. 31.1). This fluid is produced by the ciliary body (2 μL/min) and provides nutrients and oxygen to the avascular cornea. The outflow of aqueous humour is through the trabecular meshwork and canal of Schlemm adjacent to the limbus. Any factor that impedes its outflow will increase the intraocular pressure (IOP). The upper range of normal for IOP is 21 mmHg.

The uveal tract is made up of the iris anteriorly, the ciliary body, and the choroid. The ***iris*** is the coloured part of the eye under the transparent cornea. The muscles of the iris diaphragm regulate the size of the pupil, thereby controlling the amount of light entering the eye. The muscles of the ***ciliary body*** control the accommodation of the lens, and the secretory epithelium produces the aqueous humour (see above). The highly vascular ***choroid*** lines the inner aspect of the sclera and upon this lies the retina.

The lens lies immediately posterior to the pupil and anterior to the vitreous humour. It is a transparent biconvex structure and is responsible for 22% of the refractive power of the eye. By changing its shape, it can alter its refractive power and help to focus objects at different distances from the eye. By the fourth decade of life, this ability to change shape starts to decline and, with time, the lens starts to become less transparent and cataracts begin to develop.

The vitreous humour fills the cavity between the retina and the lens.

The retina is a multilayered structure. The metabolically active region of the retina is represented in Fig. 31.2. There are two types of photoreceptors in the retina: rods and cones. There are approximately 6 million cones, mainly confined to the ***macula***, and these are responsible for detailed central vision and colour vision. The peripheral retina has around 125 million rods that are responsible for peripheral vision. The axons of the ganglion cells form the optic nerve (or disc) of the eye (Fig. 31.3).

The blood supply to the eye is via the ophthalmic artery; in particular, the central retinal artery is responsible for supplying the inner retinal layers. Venous return is through the central retinal and ophthalmic veins. Local lymphatic drainage is to the pre-auricular and submental nodes.

The sensory innervation of the eye is through the trigeminal (Vth) nerve. The six extraocular muscles are supplied by different nerves (see Chapter 28):

- oculomotor (IIIrd) nerve: medial, superior, inferior rectus, and inferior oblique
- trochlear (IVth) nerve: superior oblique
- abducens (VIth) nerve: lateral rectus.

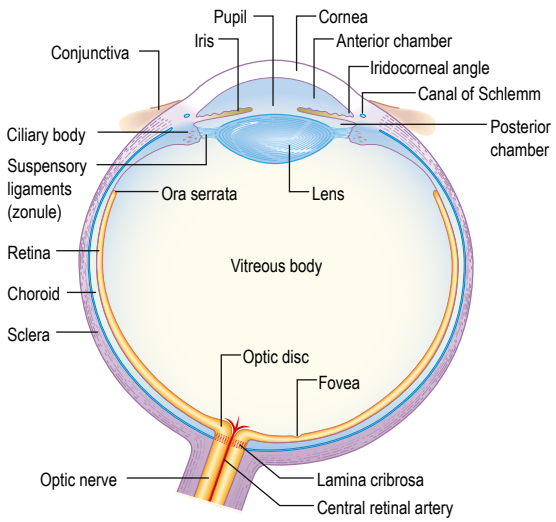


Fig. 31.1 **Cross-section of the eyeball.**

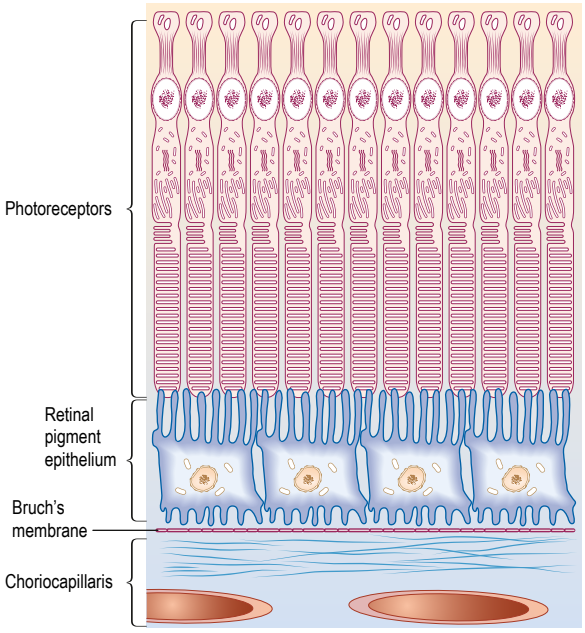


Fig. 31.2 **Schematic diagram of the retina.**

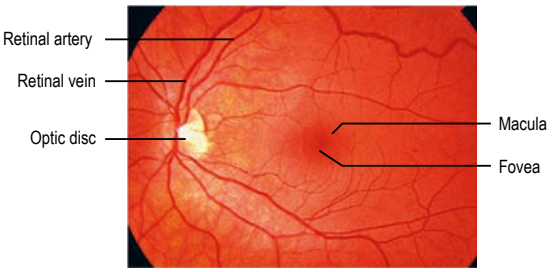


Fig. 31.3 **Fundus photograph of the left eye.**

The oculomotor (IIIrd) nerve also supplies the upper lid and, indirectly, the pupil (parasympathetic fibres are attached to it). The facial (VIIth) nerve supplies the orbicularis and other muscles of facial expression.

CLINICAL APPROACH TO THE PATIENT WITH A DISORDER OF THE EYE

History

A detailed history gives most of the facts needed to make a working diagnosis. The eye has limited mechanisms by which it can convey a diseased state.

In general, the following questions are basic guidelines to ophthalmic history taking:

- onset and duration of symptoms
- previous ocular history
- use of contact lenses
- red eye (see **Box 31.1**)
- photophobia
- eye pain (nature and character of the pain, association with eye movements)
- compromise of vision (onset of loss of vision: acute or gradual, permanent or transient, peripheral, altitudinal or central vision loss, abnormal colour vision)
- trauma—determine the involvement of foreign objects, type of objects (metallic or vegetation), and the mechanism of injury. Small, sharp, and high-velocity objects (e.g., from hammering metal or through the use of a grinder) can often lead to penetrating injuries
- diplopia (binocular diplopia resolves when one eye is covered; monocular diplopia, which is much rarer, persists with only one open)
- floaters
- loss of visual fields or scotomas.

A past medical and family history should be taken to identify autoimmune diseases, diabetes mellitus, hypertension, hypercoagulopathies, or thyroid disease. Symptoms in the history may point towards a wide range of systemic diseases with ophthalmic manifestations, as described in **Box 31.2**.

Examination

It is essential to adopt a systematic approach to the examination of the eye. Different approaches and instruments (including direct ophthalmoscope, slit lamp with or without Goldman or Volk lens) are necessary for examination of the lids and anterior and posterior segments, as well as extraocular movements. A slit lamp is a device combining a biomicroscope and a light source focused to produce a thin sheet of light, that can be adjusted to examine various features of the eye (**Fig. 31.4**). The sequence used for a full examination of the eye is described in **Box 31.3** (including the content below).

Box 31.1 Red flags for a red eye

The following symptoms require urgent referral:

- Severe pain
- Photophobia
- Reduced vision, particularly if sudden
- Coloured halos around point of light in a patient's vision
- Proptosis
- Smaller pupil in affected eye

Plus on medical assessment:

- High intraocular pressure
- Corneal epithelial disruption
- Shallow anterior chamber depth
- Ciliary flush

Clinical examination

Ophthalmic clinical examination involves using a vision chart and slit-lamp bio-microscope. An ophthalmic examination may include:

- Inspection of the eyes and surrounding structures:
  - Inspection for evidence of eye redness: generalized or localized, conjunctiva, episclera, or sclera. Causes of red eye are listed in **Box 31.1**
  - Inspection for evidence of thyroid eye disease: exophthalmos, lid lag, lid retraction, proptosis.
- Inspection of the cornea:
  - Keratic precipitates ('mutton fat' keratic precipitates with uveitis are seen in sarcoidosis or tuberculosis)
  - Loss of corneal sensation (herpetic eye disease)
  - Dry cornea or ulcers (exposure keratopathy)

Box 31.2 Symptoms and signs of systemic diseases associated with inflammatory eye conditions

Associated symptoms	Associated disease
Cardiac	
Chest pain	TB, RA, SLE, syphilis
Palpitations	Sarcoidosis, syphilis, RA, SLE, HIV
Respiratory	
Cough	TB, sarcoidosis, AAV, HIV, toxocariasis
Haemoptysis	TB, AAV, HIV, RA, SLE, sarcoidosis
Pleuritic chest pain	Sarcoidosis, TB, AAV, SLE, RA, HIV
Shortness of breath	Sarcoidosis, TB, AAV, SLE, RA, HIV
Gastrointestinal	
Diarrhoea	IBD, Behcet's, HIV
Blood in stool	IBD, Behcet's, HIV
Jaundice	IBD, HIV
Genitourinary	
Dysuria/discharge	Reiter's, syphilis, TB
Genital ulcers	Behcet's, syphilis, HLA-B27–related diseases
Haematuria	AAV, SLE, TB
Ear-nose-throat	
Oral ulcers	Behcet's, herpes, SLE
Deafness/tinnitus	VKH, AAV
Recurrent epistaxis	AAV
Musculoskeletal	
Joint pain and swelling	HLA B27 arthropathies, JIA, sarcoidosis, RA, SLE, Behcet's, AAV, Lyme
Lower back stiffness	HLA-B27, arthropathies, TB
Skin	
Rash	Sarcoidosis, Behcet's, TB, IBO, Herpes, syphilis, Lyme
Photosensitivity	SLE
Vitiligo	SLE, VKH, leprosy
Central nervous system	
Headache	Sarcoidosis, VKH, Behcet's TB, SLE, lymphoma
Weakness	Multiple sclerosis, sarcoidosis, Behcet's, HIV, leprosy, syphilis, lymphoma
Numbness/tingling	Multiple sclerosis, sarcoidosis, Behcet's, HIV, leprosy, syphilis, lymphoma
Behavioural change	VKH, sarcoidosis, Behcet's, SLE, AAV, HIV, TB, syphilis, lymphoma

AAV, ANCA-associated vasculitis; *IBO,* inflammatory bowel disease; *JIA,* juvenile idiopathic arthropathy; *RA,* rheumatoid arthritis; *SLE,* systemic lupus erythematosus; *TB,* tuberculosis; *VKH,* Vogt-Koyanagi-Harada syndrome.

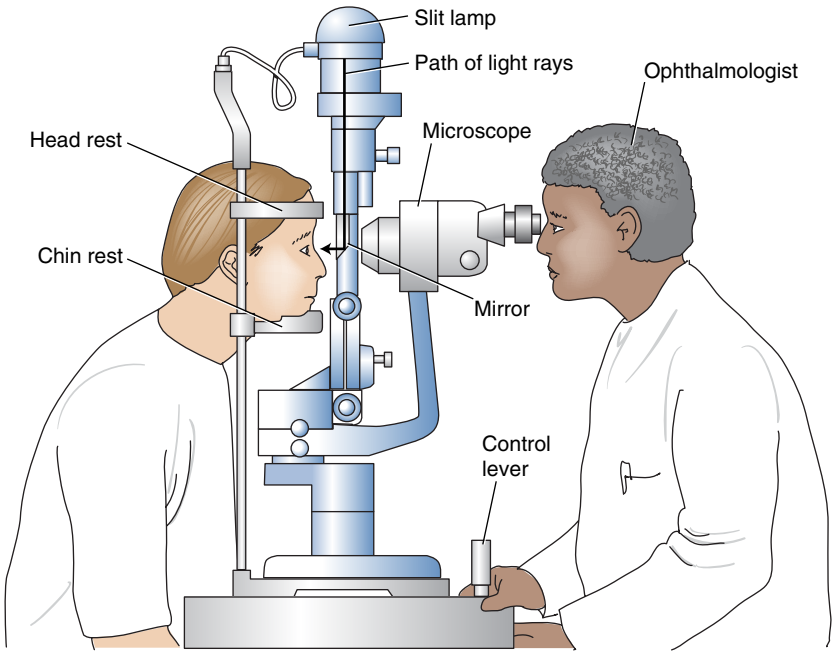


Fig. 31.4 Slit lamp. A slit lamp is a microscope with a light attached. The name is derived from the adjustable light beam.

Box 31.3 Differential diagnosis of the acute red eye

Cause	Conjunctival injection	Unilateral or bilateral	Pain	Photophobia	Vision	Pupil	Intraocular pressure
Conjunctivitis	Diffuse	Bilateral (often unilateral initially)	Gritty	Occasionally with adenovirus	Normal	Normal	Normal
Keratitis	Diffuse	Unilateral	Gritty	Yes	Reduced	Normal	Normal
Anterior uveitis	Circumcorneal	Unilateral	Painful	Yes	Reduced	Constricted	Normal or raised
Acute glaucoma	Diffuse	Unilateral	Severe pain	Mild	Reduced	Mid-dilated	Raised

  - Fluorescein staining and examination under a blue light to determine cornea epithelium integrity
  - Leak of aqueous humour (Seidel test) may be seen in cases of trauma, if full-thickness corneal injury is present.
- Inspection of the anterior chamber:
  - Anterior chamber: flare and inflammatory cells are seen in anterior uveitis, and in severe disease, hypopyon (pus in the anterior chamber) may be seen
  - Presence of blood (hyphaema) suggests trauma or a bleeding disorder, or pus (hypopyon) in the anterior chamber of the eye.
    - Integrity of eyelids, including any lumps and bumps or lacerations (partial or full thickness, involvement of canaliculi) in cases of trauma
    - Inspection for periorbital redness and swelling (periorbital cellulitis): if present, temperature measurement must be documented.
- Assessment of vision using a Snellen (or LogMAR) chart at 6 m
- Pupil examination, which involves useful objective tests to determine the afferent and efferent pathways:
  - Size of pupil
  - Pupil reactivity
  - Relative afferent pupillary defect.
- Fundal examination using cycloplegic eye drops such as tropicamide 1% or cyclopentolate 0.5% or 1%:
  - Vitreous examination: presence of vitritis in intermediate or posterior uveitis
  - Retinal examination:
    - dot/blot/flame-shaped haemorrhages in vascular diseases
    - perivascular sheathing, intraretinal infiltrates, cotton wool spots (microinfarcts of the retinal nerve fibre layer) in vasculitis and posterior uveitis
    - whitening of the retina with a cherry red spot in acute artery occlusion
    - presence of cholesterol or calcific embolus may be noted in the vessels.
  - Optic nerve swelling
- Visual field examination by confrontation or an automated visual fields (Humphreys) test
  - Enlarged blind spots in papilloedema
  - Bitemporal visual field loss in pituitary diseases
  - Visual field compromise involving the vertical midlines in various types of cerebrovascular disease.

Visual acuity

It is vital for an accurate assessment of visual acuity to be recorded in all people with an eye problem. The visual acuity of each eye is recorded in two ways: distance visual acuity and near visual acuity.

The Snellen visual acuity chart (Fig. 31.5) is most commonly employed to assess distance visual acuity, but use of the LogMAR

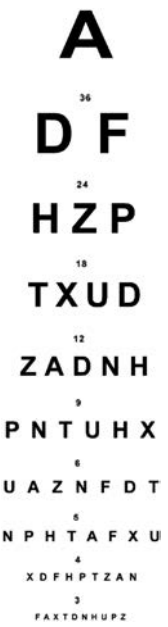


Fig. 31.5 Snellen chart.

Fig. 31.6 LogMAR chart (logarithm of the *M*inimum *A*ngle of *R*esolution).

chart (*log*arithm of the *M*inimum *A*ngle of *R*esolution; Fig. 31.6) is increasing, largely due to its necessity in studies or research, since it allows better statistical analysis of results. Unlike the Snellen and other visual acuity charts, the LogMAR chart has equal graduation between the letters on a line, as well as the space between lines. There is also a fixed number of letters—five—on each line. Research conducted using a logarithmic progression in size of letters on a test chart provides the most accurate visual acuity measurement. Snellen equivalents can be calculated from the LogMAR charts if necessary (Fig. 31.7).

The recording is given as an expression of the line of letters that can be discerned at a particular distance, usually 6 metres (20 feet): for example, 6/60, where 6 equals the distance of the chart from the eye in metres and 60 equals the distance at which the letter subtends 5′ at the nodal point.

OPHTHALMIC EMERGENCIES

The three main ophthalmic emergencies are:

- Eye trauma
- The painful red eye
- Sudden loss of vision.

An approach to these emergencies and the urgency of referral to specialist ophthalmological assessment is given in Fig. 31.8. The emergency assessment of these conditions is given below.

Eye trauma

Corneal abrasions

Trauma resulting in the removal of a focal area of epithelium on the cornea is very common. Abrasions usually occur when the eye is accidentally poked with a finger, a foreign body flies into the eye, or something brushes against the eye.

Symptoms include severe pain, due to exposure of the corneal nerve endings, lacrimation, and inability to open the eye (blepharospasm). Blinking and eye movement can aggravate the pain and foreign body sensation. The visual acuity is usually reduced. Most cases will need topical anaesthetic drops such as oxybuprocaine or tetracaine to be administered before it is possible to examine the eye. The cornea should be inspected with a blue light after instillation of fluorescein drops. The orange dye will stain the area of the abrasion. Under blue light, the abrasion lights up as green. Occasionally, foreign bodies can lodge on the undersurface of the upper lid and give rise to linear vertical abrasions. Eversion of the upper lid is necessary in all cases of abrasions (Fig. 31.9).

Treatment consists of a broad-spectrum topical antibiotic, such as chloramphenicol drops or ointment, four times a day for 5 days. The role of padding is controversial but common practice is to pad the affected eye for 24 hours once chloramphenicol ointment has been applied to the eye.

Corneal foreign body

Occasionally, when something flies into the eye, it sticks on the cornea (Fig. 31.10A). It may be associated with lacrimation and photophobia. Examination is best attempted following instillation of a topical anaesthetic and should include eversion of the upper lid (see Fig. 31.10B). Corneal foreign bodies can usually be seen directly with a white light.

The corneal foreign body should be removed. ***Treatment*** involves a topical antibiotic, such as chloramphenicol four times a day for 5 days, or fusidic acid twice a day for 5 days.

High-velocity trauma

In cases of high-velocity trauma, corneal perforation or an intraocular foreign body should be suspected. Examination may show a corneal laceration and a foreign body may also be embedded in the cornea. The foreign body may be present on the iris or in the lens or vitreous. Other clues pointing towards a penetrating injury include a large subconjunctival haemorrhage, a flat anterior chamber with low IOP, and the presence of blood in the anterior chamber (hyphaema). Urgent referral to the ophthalmologist is mandatory, ensuring that no drops are instilled into the eye and that a plastic shield is placed over the eye to minimize further risk of trauma.

Blunt trauma usually results in periorbital bruising and gross lid oedema, which can make examination to exclude perforating injury difficult. These patients should be referred to the ophthalmologist for a detailed ocular examination to exclude a perforation, retinal detachment, or a traumatic hyphaema (Fig. 31.11).

20 feet (Snellen)	6 metres (Snellen)	Decimal	4 metres	log MAR
20/200	6/60	0.1	4/40	+1.0
20/160	6/48	0.125	4/32	+0.9
20/125	6/38	0.16	4/25	+0.8
20/100	6/30	0.2	4/20	+0.7
20/80	6/24	0.25	4/16	+0.6
20/63	6/19	0.32	4/12.5	+0.5
20/50	6/15	0.4	4/10	+0.4
20/40	6/12	0.5	4/8	+0.3
20/32	6/9.5	0.63	4/6.3	+0.2
20/25	6/7.5	0.8	4/5	+0.1
20/20	6/6	1.0	4/4	+0.0
20/16	6/4.8	1.25	4/3.2	–0.1
20/12.5	6/3.8	1.60	4/2.5	–0.2
20/10	6/3	2.0	4/2	–0.3

Fig. 31.7 Snellen equivalents can be calculated from the LogMAR charts if necessary. Vision is often recorded at 4 metres in studies, and logMAR equivalents or decimals are used in statistics.

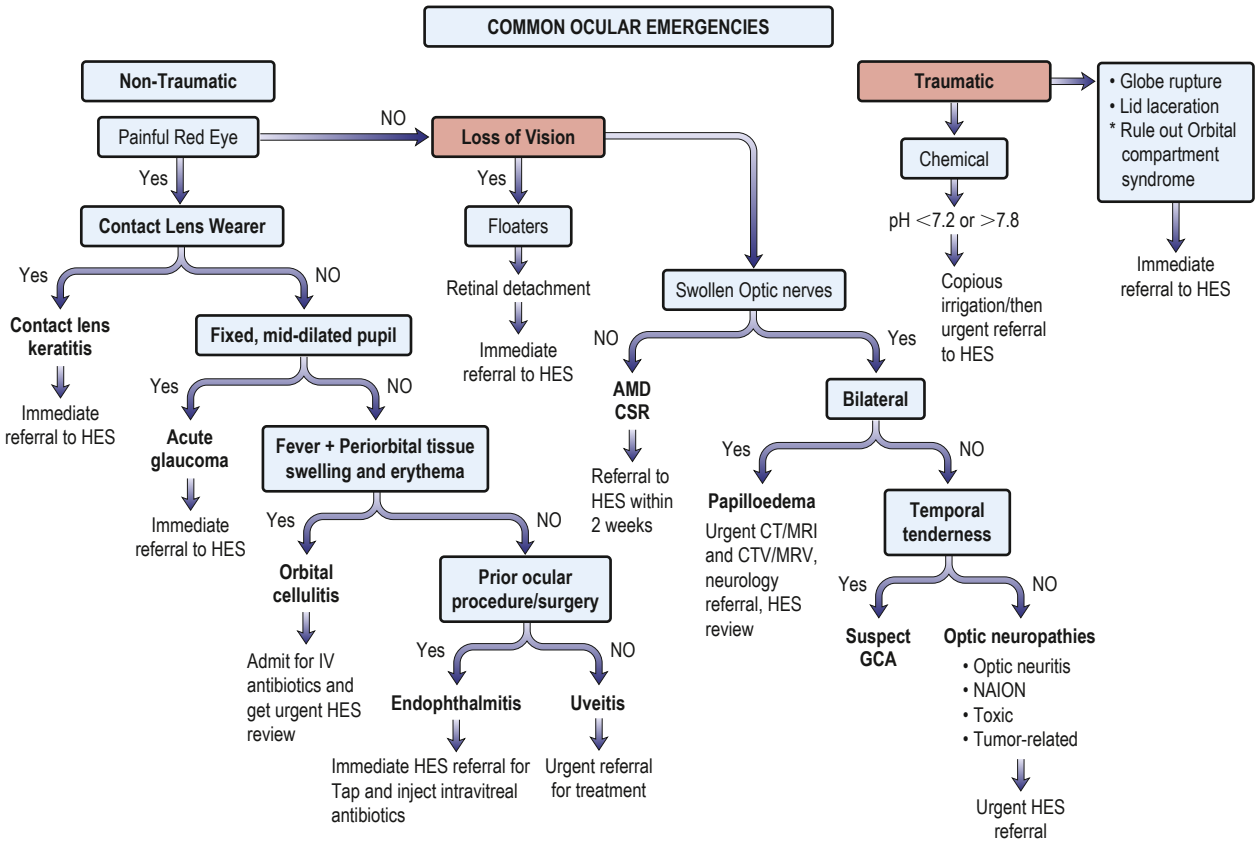


Fig. 31.8 Pathways and timing of referral for investigation and treatment in eye emergencies. *AMD,* Age-related macular degeneration; *CSR,* central serous retinopathy; *GCA,* giant cell arteritis; *HES,* hospital eye service; *NAION,* non-arteritic anterior ischaemic optic neuropathy.

Acute red eye

Patients may present complaining of acute red eyes with simple and self-limiting conditions (such as viral conjunctivitis or subconjunctival haemorrhage), or severe and sight-threatening ones. General doctors need to be adept in identifying 'red flag' features that suggest a serious diagnosis and require urgent specialist review (**Box 31.1**).

Featured suggestive of other causes of red eye are listed in **Box 31.3**.

Sudden onset sight loss

This is always a medical emergency, with a wide range of potential causes. Some of these require urgent specialist ophthalmology care (such as acute angle closure glaucoma), whilst others require urgent care by general medical or neurology teams (such as amaurosis fugax or optic neuritis).

The traditional divide is used to begin diagnostic evaluation based on whether or not the visual loss is associated with eye pain

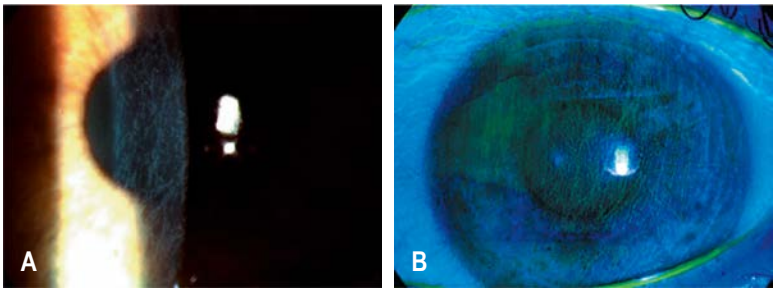


Fig. 31.9 Linear corneal abrasion. **(A)** Stained with fluorescein. **(B)** Stained with fluorescein and viewed with blue light.

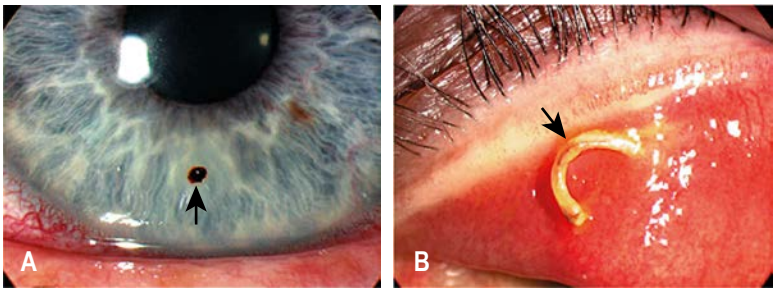


Fig. 31.10 Foreign bodies *(arrowed)*. **(A)** Corneal. **(B)** Subtarsal.

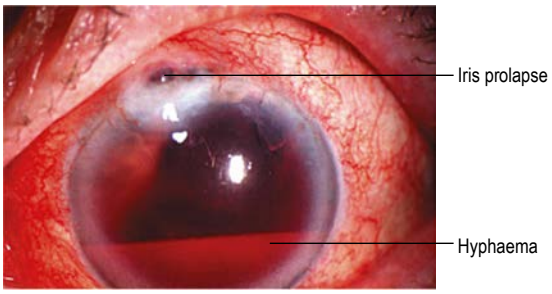


Fig. 31.11 Blunt trauma. Hyphaema and globe rupture with iris prolapse.

(Box 31.4). A guide to suggestive findings on history and examination is given in Box 31.5 and should be used to ensure patients are directly down appropriate emergency pathways.

COMMON DISORDERS OF THE EYE

Refractive errors

The eye projects a sharp and focused image on to the retina. Refractive errors refer to any abnormality in the focusing mechanism of the eye and not to any opacity in the system, such as a corneal or retinal scar.

The refraction of light in emmetropic (normal), myopic (short-sighted; negative lenses will correct), and hypermetropic (long-sighted; positive lenses will correct) eyes is shown in Fig. 31.12.

Astigmatism is a refractive error of the eye in which there is a different degree of refraction in the different meridians of curvature. It may be myopic in one plane and hypermetropic or emmetropic in the other plane. In this situation, the front surface of the eye is shaped more like a rugby ball than a football.

Presbyopia is the term used to describe the normal ageing of the lens, which leads to a change in the refractive state of the eye. As the lens ages, it becomes less able to alter its curvature and this causes difficulty with near vision, especially reading.

Management

Errors of refraction can be corrected by using spectacles or contact lenses. The latter often result in better-quality vision but carry the risk of infection. They may be the only option in some refractive states such as keratoconus, a degenerative disorder of the eye in which structural changes within the cornea cause it to thin and to take on a more conical shape than its normal gradual curve. A number of surgical techniques can correct these errors of refraction, with varying degrees of accuracy. Phakic intraocular lenses may be used to treat high degrees of myopia but the most popular method remains the excimer laser to re-profile the corneal curvature (using PRK, LASIK, and LASEK techniques). The laser either removes corneal tissue centrally to flatten the cornea in myopia or it removes tissue from the peripheral cornea to steepen it in hypermetropia.

Disorders of the lids

The lids afford protection to the eyes and help to distribute the tear film over the front surface of the globe. Excess tears are drained via the puncta and lacrimal system to the nose (Fig. 31.13). Malposition of the lids, factors that affect blinking and lacrimal drainage can all cause problems.

Entropion

The lid margin rolls inwards so that the lashes are against the globe (Fig. 31.14A). The lashes act as a foreign body and cause irritation, leading to a red eye that can mimic conjunctivitis. Occasionally, the constant rubbing of lashes against the cornea causes an abrasion. The most common cause is ageing and surgery is usually required.

Box 31.4 Causes of loss of vision

Painless loss of vision

- Cataract
- Open-angle glaucoma
- Retinal detachment
- Central retinal vein occlusion
- Central retinal artery occlusion
- Diabetic retinopathy
- Vitreous haemorrhage
- Posterior uveitis
- Age-related macular degeneration
- Optic nerve compression
- Cerebral vascular disease

Painful loss of vision

- Acute angle-closure glaucoma
- Giant cell arteritis
- Optic neuritis
- Uveitis
- Scleritis
- Keratitis
- Shingles
- Orbital cellulitis
- Trauma

Ectropion

The lid margin rolls outwards and is not apposed to the globe. As a result, the lacrimal punctum is not in the correct anatomical position to drain tears and patients usually complain of a watery eye. Underlying factors include age, VIIth nerve palsy, and cicatricial skin conditions. Surgery is usually required.

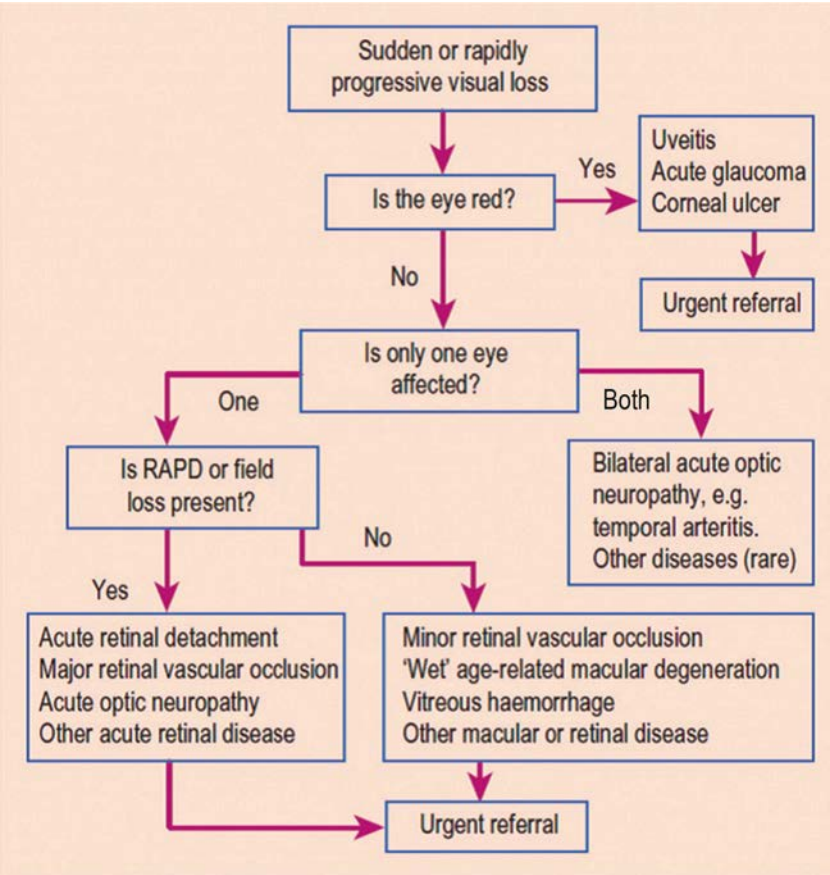
Dacryocystitis

Patients who have inflammation of the lacrimal sac usually present with a painful lump at the side of the nose adjacent to the lower lid (see Fig. 31.14B). This should be treated with oral broad-spectrum antibiotics such as a cephalosporin, and patients should be watched carefully for signs of cellulitis. All patients should be referred to the ophthalmologist, as some have an underlying mucocele or dilated sac, and will require surgery.

Blepharitis

This is an extremely common condition in which inflammation of the lid margins may involve the lashes and lash follicles (Fig. 31.15A), resulting in styes, or inflammation and blockage of meibomian glands (see Fig. 31.15B) leading to chalazion (see Fig. 31.15C). Common underlying causes of blepharitis include meibomian

Box 31.5 Initial history and examination in sudden loss of vision



RAPD, Relative afferent papillary defect.
(After Pane A, Simcock P. Practical Ophthalmology. *Edinburgh: Churchill Livingstone; 2005.)*

gland dysfunction, seborrhoea, and *Staphylococcus aureus* infection. Patients can be asymptomatic or complain of itchy, burning eyes because of tear film instability resulting from meibomian gland dysfunction. *S. aureus* is frequently responsible for chronic blepharoconjunctivitis and some patients may develop keratitis in the cornea (Fig. 31.16).

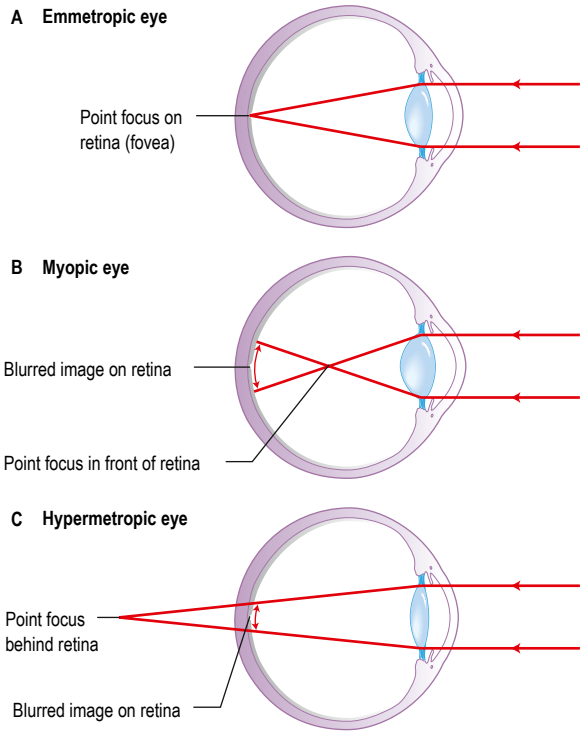


Fig. 31.12 Refraction. **(A)** Emmetropic (normal). **(B)** Myopic (short-sighted). **(C)** Hypermetropic (long-sighted).

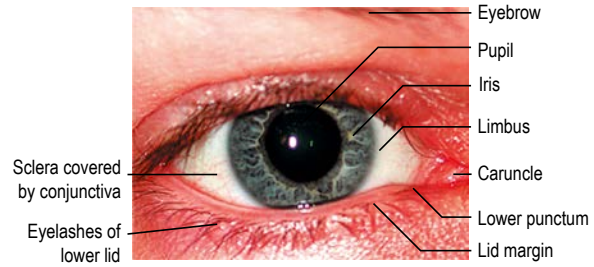


Fig. 31.13 The right eye.

Management

Lid hygiene is the mainstay of treatment for blepharitis, as it helps to reduce the bacterial load and unblock meibomian glands. A short course of topical chloramphenicol or fusidic acid is useful in chronic cases, but in severe cases or cases where acne rosacea is suspected, oral doxycycline is used. Some patients are left with a lump once the acute inflammatory phase has subsided. Most of these patients find the lump, or chalazion, cosmetically unacceptable and require incision and curettage. People with keratitis should be referred to the ophthalmologist for topical steroids.

Conjunctivitis

The most common cause of a red eye, inflammation of the conjunctiva can arise from a number of causes, viral, bacterial, and allergic being the most frequently encountered. Common features in all types include soreness, redness, and discharge; in general, the visual acuity is good. History should include the speed of onset of the inflammation, the colour and consistency of the discharge, whether the eye is itchy, and if there has been a recent history of a cold or sore throat. The differential diagnosis of conjunctivitis is shown in Box 31.6.

Bacterial conjunctivitis

Bacterial conjunctivitis is uncommon, making up 5% of all cases of conjunctivitis. In the vast majority of patients, it causes a sore or gritty eye in the presence of good vision. Bacterial conjunctivitis is invariably bilateral and should be suspected when conjunctival inflammation is associated with a purulent discharge.

Gonococcal conjunctivitis

This should be suspected when the onset of symptoms is rapid, the discharge is copious, and ocular inflammation includes chemosis (conjunctival oedema) and lid oedema. There may be a palpable pre-auricular node.

Less acute or subacute purulent conjunctivitis with moderate discharge can be attributed to organisms such as *Haemophilus influenzae* and *Streptococcus pneumoniae*. Chronic conjunctivitis is usually associated with mild conjunctival injection and scant purulent discharge. Common organisms include *S. aureus* and *Moraxella lacunata*.

Management

Prompt treatment with oral and topical penicillin is given in gonococcal conjunctivitis to ensure a reduced rate of corneal perforation. A Gram-stain of the conjunctival swab can quickly confirm the

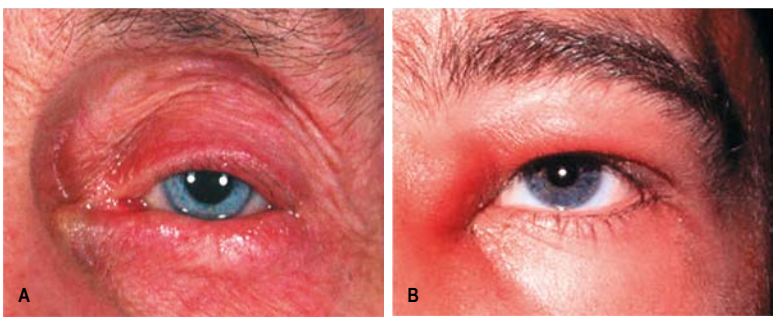


Fig. 31.14 Disorders of the eyelid. **(A)** Lid entropion. The lower lid appears inverted. **(B)** Acute dacryocystitis, showing a lump on the side of the nose.

Fig. 31.15 Blepharitis. **(A)** Crusty and scaly deposits on the lashes and lash bases. **(B)** Upper lid showing meibomian glands plugged with oily secretions. **(C)** Blockage of the meibomian glands leads to swelling of the lid (chalazion).

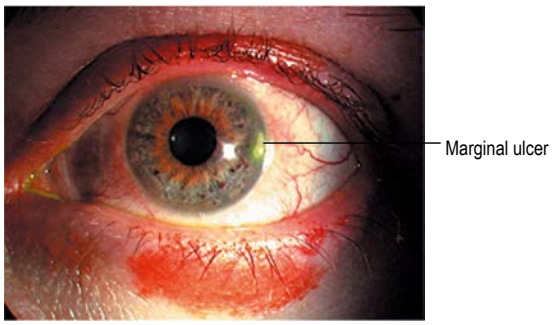


Fig. 31.16 Marginal keratitis.

presence of diplococci. Gonococcal conjunctivitis is a notifiable disease in the UK. Empirical treatment for both subacute and chronic conjunctivitis involves a topical broad-spectrum antibiotic, such as chloramphenicol. Swabs should be taken if these cases do not respond to this initial treatment. Antibiotic resistance is increasing.

Chlamydial conjunctivitis

Chlamydia trachomatis (see Chapter 43) in high-income countries causes a sexually transmitted infection that is most prevalent in sexually active adolescents and young adults. Direct or indirect contact with genital secretions is the usual route of infections but shared eye cosmetics can also be involved.

The onset of symptoms is slow, and patients may complain of mild discomfort for weeks. In these cases, the red eye is associated with a scanty mucopurulent discharge and a palpable pre-auricular lymph node. In chronic cases, it is not unusual to see superior corneal vascularization. Conjunctival swabs should be taken and a nucleic acid amplification test (NAAT; see Chapter 43) performed prior to commencement of treatment.

Management

Topical erythromycin twice daily is commenced and patients referred to the sexual health clinic.

Trachoma

This is a major cause of blindness worldwide. Clinical features are described in **Box 31.7**.

Viral conjunctivitis

Adenoviral conjunctivitis

This is highly contagious and can cause epidemics in communities. Transmission is through direct or indirect contact with infected individuals. The onset of symptoms may be preceded by a cold or influenza-like symptoms. Inflammation is commonly associated with chemosis, lid oedema, and a palpable pre-auricular lymph node. Some patients develop a membrane on the tarsal conjunctiva (**Fig. 31.17**) and haemorrhage on the bulbar conjunctiva. Viral conjunctivitis can cause deterioration in visual acuity owing to corneal involvement (focal areas of inflammation). In 50% of these patients, the conjunctivitis is unilateral.

The condition is largely self-limiting in the majority of cases. Lubricants, together with a cold compress, can be a soothing element of ***management*** for patients. Adhering to strict hygiene and keeping towels separate from those of the rest of the household go a long way towards reducing the spread of the infection. In people with corneal involvement or intense conjunctival inflammation, topical steroids are indicated.

Herpes simplex conjunctivitis

Primary ocular herpes simplex conjunctivitis is typically unilateral. It usually causes a palpable pre-auricular lymph node, and cutaneous vesicles develop on the eyelids and the skin around the eyes in the majority. Over 50% of these patients develop a dendritic corneal ulcer (**Fig. 31.18**). The organism responsible for this condition is the herpes simplex virus (HSV), usually HSV-1, although HSV-2 can give rise to ocular infection.

Primary ocular HSV infection is self-limiting but most clinicians choose ***treatment*** with topical aciclovir in order to limit the risk of corneal epithelial involvement.

Molluscum contagiosum conjunctivitis

This is typically unilateral; it produces a red eye that generally goes unrecognized and comes to the forefront because patients fail to improve and the cornea starts to become involved. On close inspection, pearly, umbilicated nodules, filled with the DNA poxvirus, can be seen on the lid margin.

Management includes curetting the central portion of the lesion, freezing the centre, or completely excising the lesion. If the corneal involvement is severe or the eye is very inflamed, a short course of topical steroids, such as prednisolone 0.5% or dexamethasone 0.1%, is helpful.

Phthiriasis palpebrarum

Phthiriasis palpebrarum (**Fig. 31.19**) is an eyelid infestation caused by *Phthirus pubis*, or the crab louse. Infestation of the cilia and eyelid is rare. It leads to blepharitis with marked conjunctival inflammation, pre-auricular lymphadenopathy and, rarely, secondary infection at the site of louse bite.

Management

Mechanical removal of the lice with fine forceps, physostigmine 1.25%, and pilocarpine gel 4% are all effective treatments.

Box 31.6 Conjunctivitis

Type	Discharge	Pre-auricular node	Corneal involvement	Comment
Bacterial	Mucopurulent	–ve (except gonococci)	+ve Gonococcus	Rapid onset
Viral	Watery	+ve	+ve Adenovirus	Cold and/or sore throat
Chlamydial	Watery	+ve	+ve	Genitourinary discharge
Allergic	Stringy	–ve	+ve	Itchiness

Box 31.7 Trachoma

- A very common cause of blindness worldwide (see Chapter 42)
- Found mainly in the tropics and the Middle East
- Caused by *Chlamydia trachomatis* but not usually sexually transmitted
- Chronic conjunctival inflammation causes progressive scarring, trichiasis, entropion, and subsequent corneal scarring
- The result may be severe visual impairment or blindness from corneal opacification or ulceration

Allergic conjunctivitis

There are five main types of allergic conjunctivitis: seasonal, perennial, vernal, atopic, and giant papillary. The last three are difficult to treat; they are chronic and can be sight-threatening, so should be referred to an ophthalmologist.

Seasonal/perennial conjunctivitis

Seasonal allergic conjunctivitis and perennial conjunctivitis, affecting 20% of the general population in the UK, are allergic reactions to grass and tree pollen and fungal spores. Seasonal allergic conjunctivitis occurs mainly in spring and summer. Perennial allergic conjunctivitis occurs all year round but peaks in the autumn; causes include allergens, such as house dust mites.

The main symptoms include itching, redness, soreness, watering, and a stringy discharge. Occasionally, the conjunctiva may become so hyperaemic that chemosis results. This is usually associated with swollen lids.

Lowering the allergen load (reducing dust) is helpful. Medical ***treatment*** includes the use of antihistamine drops such as azelastine and emedastine, together with topical mast-cell-stabilizing agents such as sodium cromoglicate and nedocromil. Olopatadine (twice daily) has dual action and is very effective. Corticosteroid drops should be avoided. Oral antihistamines and cold compresses help the itching.

Corneal disorders

Keratitis

This is a general term used to describe corneal inflammation. Common causes include HSV, contact lens-associated infection, and blepharitis. Symptoms include the sensation of a foreign body or pain (depending on the size and depth of the ulcer), photophobia, and lacrimation. Vision is reduced if the ulcer affects the visual axis.

Herpes simplex keratitis

Corneal epithelial cells infected with the virus eventually undergo lysis and form an ulcer, which is typically dendritic in shape (see Fig. 31.18). The ulcer stains with fluorescein and can be observed easily with a blue light. Topical immunosuppression, such as with steroid drops, or systemic immunosuppression, such as in AIDS, can lead to the centrifugal spread of the virus, such that the ulcer increases in area and is referred to as a geographic ulcer. Recurrent attacks of HSV keratitis can be triggered by ultraviolet light, stress, and menstruation. All these factors are responsible for activating the virus, which normally lies dormant in the ganglion of the Vth nerve.

Treatment consists of aciclovir ointment five times a day for 2 weeks; this is usually very effective.

Contact lens-related keratitis

A small number of contact lens wearers develop infective corneal ulcers, which are potentially sight-threatening (Fig. 31.20). The organisms usually responsible include Gram-positive and Gram-negative bacteria. Patients should be referred to an ophthalmologist for scraping of the ulcer and commencement of antibiotic treatment.

Keratoconus

Keratoconus is an eye condition in which the normally round, dome-shaped cornea progressively thins and causes a cone-shaped bulge to develop. Aetiology is uncertain but genetic factors play a role, and the condition is more common in people with allergic diseases such as asthma, in Down's syndrome and in some disorders of collagen such as Marfan's disease. Keratoconus affects up to 1 in 1000 people and is more common in individuals of Asian heritage. It is usually diagnosed in teenagers and young people.

Management

In the early stages, spectacles or soft contact lenses may be used to correct vision. As the cornea becomes thinner and steeper, rigid gas-permeable contact lenses may be necessary.

Corneal cross-linking is a new treatment that can stop keratoconus becoming worse. It is effective in more than 9 out of 10 patients, with a single 30-minute day-case procedure, but is only suitable when the corneal shape is continuing to deteriorate. In very advanced cases, where contact lenses fail to improve vision, a corneal transplant may be needed.

Corneal dystrophy

Corneal dystrophies may be classified anatomically as comprising:

- epithelial and subepithelial dystrophies
- epithelial–stromal *TGFBI* corneal dystrophies
- stromal dystrophies
- endothelial dystrophies.

The most common endothelial dystrophy, Fuchs' corneal dystrophy, is a genetically associated degenerative disorder leading to corneal oedema and vision loss. The gene involved is *TCF4*. The condition affects both eyes; it is more common in females and is of gradual onset, leading to blindness in the 40 to 60 age group. There is an accumulation of deposits (guttae) in the cornea with thickening of Descemet's membrane. Treatment is by corneal transplantation.

Corneal transplantation

A corneal transplant, often described as a graft or keratoplasty, is an operation to remove part or all of a damaged cornea and replace it with donor tissue. The type of keratoplasty depends on what part of the cornea is damaged:

- Penetrating keratoplasty—this is a full-thickness transplant, often used in patients with significant keratoconus or trauma.

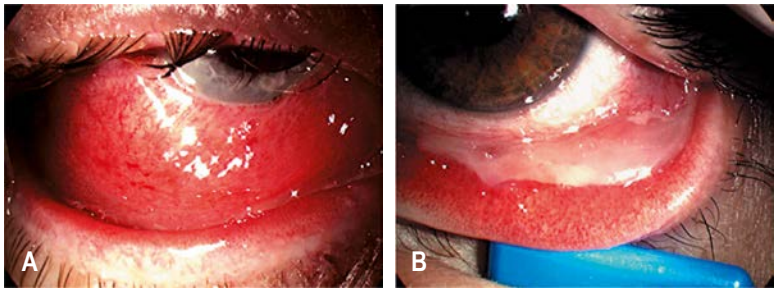


Fig. 31.17 Viral conjunctivitis. **(A)** Conjunctival chemosis. **(B)** Pseudomembrane on lower lid.

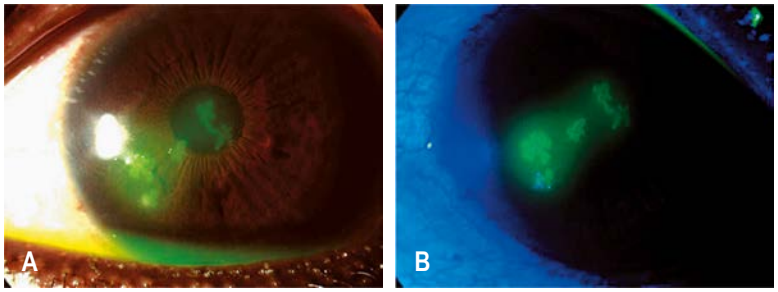


Fig. 31.18 Dendritic ulcers on the cornea. **(A)** Stained with fluorescein. **(B)** Stained with fluorescein and viewed with blue light.

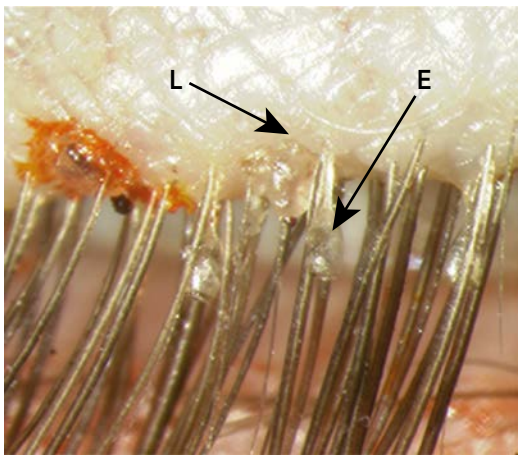


Fig. 31.19 Phthiriasis palpebrarum. The lice *(L)* and the nits (eggs) can be seen on the lid margins. *E,* Empty egg.

- Deep anterior lamellar keratoplasty (DALK)—this replaces or reshapes the outer (front) and middle layers of the cornea.
- Endothelial keratoplasty—this replaces the deeper (back) layers of the cornea.

Cataracts

Cataract (Fig. 31.21) is by far the most common cause of preventable blindness in the world, having an effective surgical treatment.

Aetiology

Age-related opacification of the lens (cataract) is the commonest cause of visual impairment, with 30% of people over 65 years having visual acuities below that required for driving (Snellen acuity less than 6/12). The common causes of cataracts are summarized in Box 31.8.

In young patients, familial or congenital causes should be excluded. Any history of ocular inflammation is noted.

Clinical features

Gradual painless deterioration of vision is the most common symptom. Other symptoms are dependent on the type of cataract: for example, a posterior capsular type would lead to glare and problems with night driving. Early changes in the lens are correctable by spectacles but eventually the opacification needs surgical intervention.

Investigations

Blood glucose, serum calcium, and liver biochemistry should be measured to diagnose metabolic disorders.

Management

Small-incision extracapsular or phacoemulsification cataract extraction with the insertion of an intraocular lens is the treatment of choice (see Fig. 31.20C). Recent advances have enabled surgeons to perform multiple steps in the surgical process with the excimer laser to enhance visual outcomes. Lens technology has also improved and toric lenses are available to treat astigmatism, and multifocal or accommodative lenses to overcome intraocular lens-induced presbyopia.

Glaucoma

Glaucoma is due to increased pressure inside the eye, which is sufficiently elevated to cause optic nerve damage and result in visual field defects, with loss of sight (Fig. 31.22). Normal IOP is 10 to 21 mmHg. Some types of glaucoma can result in an IOP exceeding 70 mmHg. Glaucoma is the second most common cause of blindness worldwide.

Primary open-angle glaucoma

Primary open-angle glaucoma (POAG) is the most common form of glaucoma. High IOPs result from reduced outflow of aqueous humour through the trabecular meshwork (Fig. 31.23A). Common risk factors include age (0.02% of 40 year olds vs. 10% of 80 year olds),

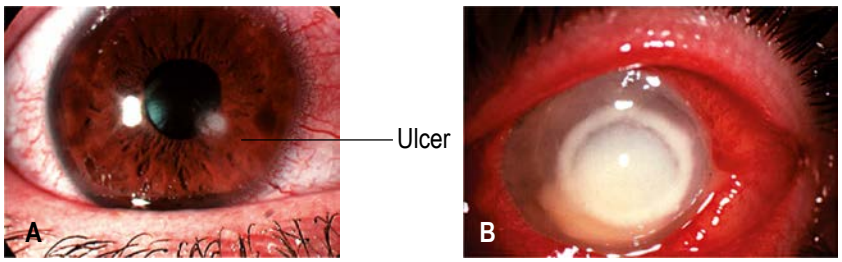


Fig. 31.20 Contact lens-related keratitis. **(A)** Corneal ulcer. **(B)** Severe keratitis with a corneal abscess and a hypopyon.

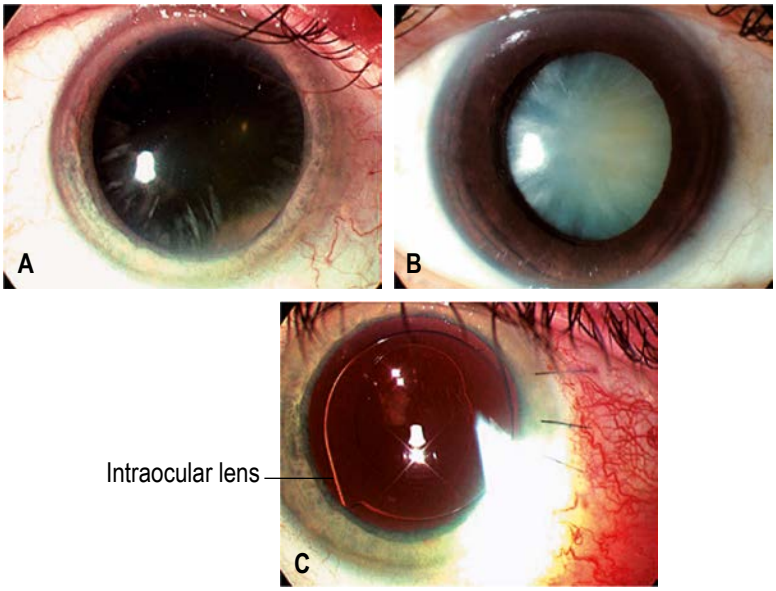


Fig. 31.21 Cataracts. **(A)** Early cataract. **(B)** White mature cataract. **(C)** Artificial intraocular lens following phacoemulsification.

race (black African patients are at five times greater risk than white patients), positive family history, and myopia.

Clinical features

POAG causes a gradual, insidious, painless loss of peripheral visual field, causing loss of vision. It is initially asymptomatic and the central vision remains good until the end-stage of the disease. Usually, glaucoma is identified during a routine ophthalmic examination. Diagnosis is only made if the IOP is measured. The optic disc is inspected and shows an enlarged cup with a thin neuroretinal rim. Visual fields are assessed and show a normal blind spot with scotomas.

Management

Treatment aims to reduce the IOP, either by reducing aqueous production or by increasing aqueous drainage:

- ***Beta-blockers***, such as timolol, carteolol, and levobunolol, reduce aqueous production and are the most commonly prescribed topical agents.
- ***Prostaglandin analogues***, such as latanoprost, bimatoprost, and travoprost, increase aqueous outflow and are also used (alone or in combination with beta-blockers) for POAG, as they can reduce IOP by 30%.
- ***Carbonic anhydrase inhibitors***, such as dorzolamide and acetazolamide, reduce aqueous production and are available in

? Box 31.8 Cataracts: aetiology

Congenital
- Maternal infection
- Familial

Advanced age

Metabolic
- Diabetes
- Galactosaemia
- Hypocalcaemia
- Wilson's disease

Drug-induced
- Corticosteroids
- Phenothiazines
- Miotics
- Amiodarone

Traumatic
- Post-intraocular surgery

Inflammatory
- Uveitis

Disorder associated
- Down's syndrome
- Dystrophia myotonica
- Lowe's syndrome

topical preparations. Acetazolamide is also available orally and, in this form, is the most potent drug for reducing IOP.

- ***Selective laser trabeculoplasty (SLT)*** is a form of laser surgery that can lower the IOP by about 30% when used as initial therapy. It is useful when eye-drop medications are not lowering the eye pressure enough or are causing significant side-effects. It may sometimes be used as initial treatment in glaucoma, although effects commonly last 1 to 5 years only.
- ***Surgical intervention*, such as trabeculectomy**, is designed to bypass the trabecular meshwork and create an alternative passage for aqueous drainage. This procedure has been largely superseded by the use of tube shunt surgery. More recent advances include minimally invasive glaucoma surgery, which involves the use of microstents to bypass the trabecular meshwork.

Acute angle-closure glaucoma

Acute angle-closure glaucoma (AACG) is an ophthalmic emergency. There is a sudden rise in IOP to levels over 50 mmHg. This occurs due to reduced aqueous drainage when the ageing lens pushes the iris forwards against the trabecular meshwork (see Fig. 31.23B). People most at risk of developing AACG are those with shallow anterior chambers, including women and hypermetropes. The attack is more likely to occur under reduced light conditions when the pupil is dilated.

Clinical features

AACG causes sudden onset of a red, painful eye and blurred vision. Patients become unwell, with nausea and vomiting, and complain of headache and severe ocular pain. The eye is injected and tender, and feels hard. The cornea is hazy and the pupil is semi-dilated (Fig. 31.24). Box 31.3 shows the differential diagnosis of the acute red eye. Box 31.1 shows features that require ***urgent referral*** to an ophthalmologist.

Management

Prompt treatment is required to preserve sight and includes:

- i.v. acetazolamide 500 mg (provided there are no contraindications) to reduce IOP, *and*
- instillation of pilocarpine 4% drops to constrict the pupil to improve aqueous outflow and prevent iris adhesion to the trabecular meshwork.

Other topical drops, such as beta-blockers and prostaglandin analogues, can also be instilled if available, provided there are no contraindications. Analgesia and antiemetics are given as required.

Patients must be referred to an ophthalmologist immediately so that reduction in IOP can be monitored and other agents, such as oral glycerol or i.v. mannitol, can be administered to non-responding patients. Definitive treatment involves making a hole in the periphery of the iris of both eyes either by laser or by surgical means.

Uveitis

Uveitis is inflammation of the uveal tract, which includes the iris, ciliary body, and choroid. It is classified according to the part of the uveal tract that the inflammation affects:

- ***Anterior uveitis*** is inflammation that affects the anterior part of the uveal tract. This can include the iris (iritis), or both the iris and the ciliary body (iridocyclitis). It is the most common type of uveitis.
- ***Intermediate uveitis*** is inflammation that affects the middle part of the uveal tract or eye, mainly the vitreous. It can also affect the underlying retina.
- ***Posterior uveitis*** is inflammation that affects the posterior part of the eye. It can affect the choroid, optic nerve head, and the retina (or any combination of these structures). It includes chorioretinitis, retinitis, and neuroretinitis.

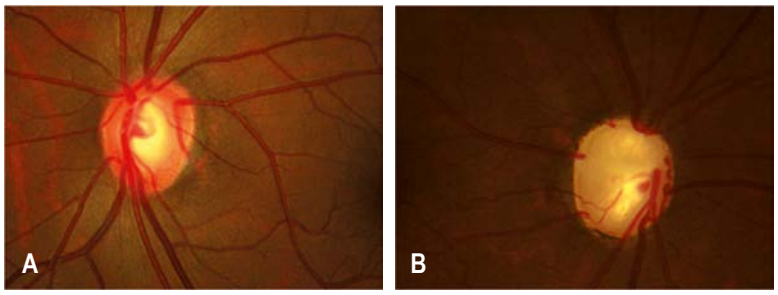


Fig. 31.22 The optic disc. **(A)** Normal optic disc. **(B)** Glaucomatous optic disc. The central cup is enlarged and deepened.

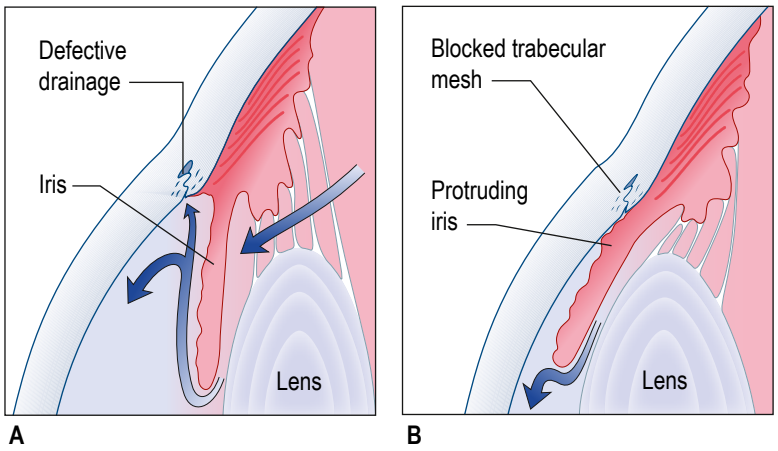


Fig. 31.23 Glaucoma. **(A)** Primary open-angle glaucoma. **(B)** Angle-closure glaucoma.

- ***Panuveitis*** is inflammation affecting the whole of the uveal tract.

The most common symptoms of uveitis are blurred vision, pain, redness, photophobia, and floaters. Each symptom is determined by the location of the inflammation, such that photophobia and pain are common features of iritis while floaters are commonly seen with posterior uveitis.

Anterior uveitis (iritis)

The classic presentation entails a triad of eye symptoms: redness, pain, and photophobia. Vision can be normal or blurred, depending on the degree of inflammation. The eye can be generally red or the injection can be localized to the limbus. The anterior chamber shows features consistent with inflammation, including cells, keratic precipitates on the corneal endothelium, fibrin or hypopyon (pus), and the pupil may have adhered to the lens (posterior synechiae) (Fig. 31.25). The IOP may be normal or raised, either due to cells clogging up the trabecular meshwork, or due to posterior synechiae causing aqueous humour to build up behind the iris and force the iris against the trabecular meshwork and so reduce aqueous drainage.

Management

This consists of reducing inflammation with the use of topical steroids such as dexamethasone 0.1% and dilating the pupil with cyclopentolate 1% to prevent formation of posterior synechiae. Dilation also allows fundoscopy to exclude posterior segment involvement. If the IOP is raised, this is treated with topical beta-blockers, prostaglandin analogues, or oral or i.v. acetazolamide. Referral should be made to the ophthalmologist.

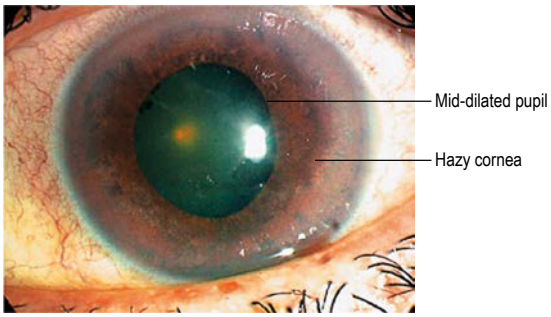


Fig. 31.24 Acute angle-closure glaucoma.

Intermediate uveitis

This usually causes painless blurred vision, most commonly associated with floaters. It is unusual to experience photophobia and redness. Both eyes are commonly affected in intermediate uveitis.

Management

This consists of a combination of treatment for anterior and posterior uveitis depending on the degree of anterior and posterior involvement.

Posterior uveitis

This commonly causes painless, blurred vision and can progress to severe visual loss. It is commonly associated with floaters and scotomata, or blind spots in the visual field.

Posterior uveitis is often found with systemic autoimmune diseases or infections; appropriate investigations should be performed and treatment given that is aimed at the cause.

Autoimmune diseases associated with uveitis include rheumatoid arthritis and Behçet's disease, ankylosing spondylitis and positive HLA-B27 (see Chapter 32, 'Ankylosing spondylitis' section), reactive arthritis, sarcoidosis, psoriasis and inflammatory bowel disease (Crohn's disease and ulcerative colitis; see Chapter 22). Infections, a rare cause of uveitis, include herpes simplex, herpes zoster, toxoplasmosis, cytomegalovirus, syphilis, tuberculosis, HIV infection, and Lyme disease. In a number of patients, no cause is found (idiopathic uveitis).

Management

Steroids are commonly given orally, or more locally by injection into or around the eye. If steroid treatment is needed to treat uveitis in the longer term, second-line immunosuppressive drugs, such as mycophenolate mofetil, ciclosporin, or azathioprine, are used. Biological agents, such as rituximab or adalimumab, are used in more severe cases, alongside treatment of any underlying systemic inflammatory disorder.

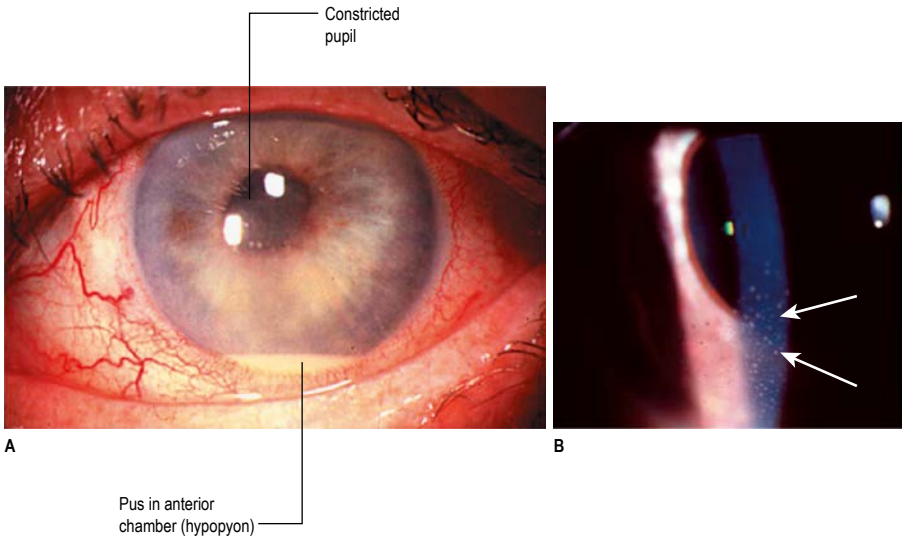


Fig. 31.25 Anterior uveitis. (A) Acute uveitis with hypopyon. **(B)** Keratic precipitates *(arrowed)* on the corneal endothelium.

Disorders of the retina

Central retinal vein occlusion

Central retinal vein occlusion (CRVO) usually leads to profound, sudden, painless loss of vision with thrombosis of the central retinal vein at or posterior to the lamina cribrosa, where the optic nerve exits the globe. The thrombus causes obstruction to the outflow of blood, leading to a rise in intravascular pressure. This results in dilated veins, retinal haemorrhage, cotton wool spots, and abnormal leakage of fluid from vessels, causing retinal oedema (**Fig. 31.26**). In severe cases, an afferent papillary defect (Chapter 28) is present and this suggests the ischaemic variant.

Predisposing factors include increasing age, hypertension and cardiovascular disease, diabetes, glaucoma and, in the younger age group, blood dyscrasias, and vasculitis.

Management

Treatment of any underlying medical condition is mandatory. Referral to an ophthalmologist is essential to monitor the eye, as some patients can develop retinal ischaemia with resulting neovascularization of the retina and iris. Panretinal photocoagulation should be commenced if there is neovascularization, and intravitreal steroid or anti-vascular endothelial growth factor (anti-VEGF) therapy is also used if there is macular oedema. Patients who develop iris neovascularization—rubeosis—where these new blood vessels block the drainage angle are at risk of developing rubeotic glaucoma.

Central retinal artery occlusion

Central retinal artery occlusion (CRAO) results in sudden, painless, severe loss of vision. Some patients may present with transient loss of vision or amaurosis fugax (see Chapter 29). Unrelieved retinal arterial occlusion results in infarction of the inner two-thirds of the retina. The arteries become narrow and the retina becomes opaque and oedematous. A cherry-red spot is seen at the fovea because the choroidal vasculature shows up through the thinnest part of the retina (**Fig. 31.27**). An afferent papillary defect is usually present.

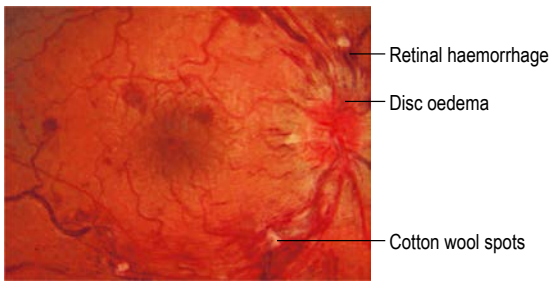


Fig. 31.26 **Central retinal vein occlusion.**

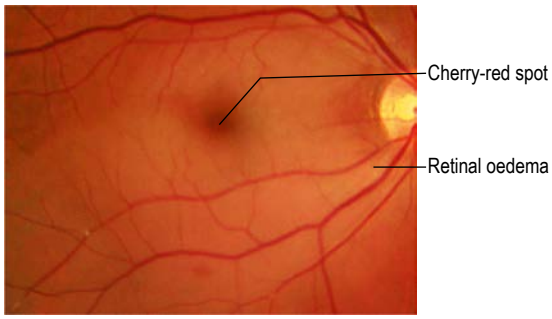


Fig. 31.27 **Central retinal artery occlusion.**

Arteriosclerosis-related thrombosis is the most common cause of CRAO. Emboli from carotid atheroma and diseased heart valves are other causes. Giant cell arteritis (see Chapter 32) must be excluded.

Management

CRAO is an ophthalmic emergency since studies have shown that irreversible retinal damage occurs *within 90 minutes of onset*. Ocular massage and 500 mg i.v. acetazolamide help to reduce ocular pressure and may assist in dislodging the emboli. Breathing into a paper bag allows a build-up of carbon dioxide, which acts as a vasodilator and so helps dislodge the emboli. Other options include making a corneal paracentesis to drain off some aqueous humour, thereby reducing the IOP.

People with CRAO should be referred for urgent investigation under stroke services to determine the aetiology of the emboli or thrombus (Chapter 29). All people with CRAO and amaurosis fugax should be started on oral aspirin if it is not medically contraindicated.

Retinal detachment

This causes a painless, progressive visual field loss. The visual 'shadow' corresponds to the area of detached retina. If the detachment affects the macula, central vision will be lost. Following a tear in the retina, fluid collects in the potential space between the sensory retina and the pigment epithelium (**Fig. 31.28**). Patients usually report a sudden onset of floaters, often associated with flashes of light (photopsia) prior to the detachment. These individuals should be referred to an ophthalmologist for a detailed fundal examination. Surgical treatments including vitrectomy (removal and replacement of vitreous humour) can prevent progressive visual loss.

Retinitis pigmentosa

This is a common chronic, inherited, degenerative disease of the retina, which can be primary or part of a syndrome, and leads to blindness. There is constriction of the peripheral vision, leading to tunnel vision and progressive loss of night vision.

Ophthalmoscopy shows bone spicule deposits and attenuated retinal vessels. Several genes are implicated.

There is no treatment but high-dose vitamin A supplementation may slow progression. Gene therapy is being investigated.

Fig. 31.28 **Retinal tear leading to detachment.**

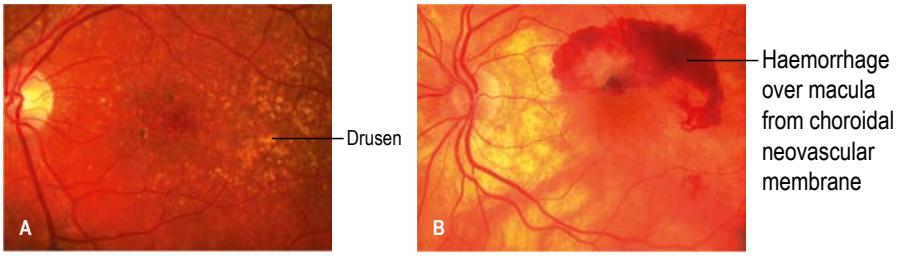


Fig. 31.29 Age-related macular degeneration. **(A)** With deposition of material (drusen) over the macula. **(B)** With choroidal neovascularization.

Age-related macular degeneration

Age-related macular degeneration (AMD) is the most common cause of visual impairment in patients over 50 years in the Western world, and the most common cause of blind registration in this age group. It affects 10% of people over 65 years and 30% over 80 years. Mutations in various genes have been reported: fibulin 5, complement factor H, and the Arg 80 Gly variant of complement C3.

The cause is unknown but suggested risk factors include increasing age, smoking, hypertension, hypercholesterolaemia and ultraviolet exposure.

There are two types:

- ***Non-exudative (dry) macular degeneration*** describes a painless and progressive loss of central vision. With age, lipofuscin deposits (drusen) are found between the retinal pigment epithelium (RPE) and Bruch's membrane (**Fig. 31.29A**; see **Fig. 31.2**). Drusen may be hard or soft, and there may be focal RPE detachment. Not all people with these changes will be affected visually but some develop distortion and blurring of their central vision. Extensive atrophy of RPE can occur (geographic atrophy).
- ***Exudative (wet) AMD*** (10% of cases) occurs with the development of abnormal subfoveal choroidal neovascularization in the region of the macula and causes severe central visual loss (see **Fig. 31.29B**).

Management

People with central distortion or with frank macular pathology should be referred *urgently* to the ophthalmologist for assessment of treatment. Anti-VEGF treatments, such as ranibizumab, aflibercept, and bevacizumab, are given by intravitreal injections with great success in wet AMD. The treatment course should be commenced *as a matter of urgency*, as vision is maintained in up to 95% of patients and improves in approximately one-third. Initial monthly monitoring with optical coherence tomography (OCT) is recommended (**Fig. 31.30**). Laser treatment and photodynamic therapy with verteporfin may also be used where there is extensive neovascularization.

There is no specific treatment for dry AMD. Severe visual loss is possible and low-vision aids, such as magnifying glasses, may help to improve a patient's independence.

The Age-Related Eye Disease Study (AREDS) has shown that vitamin supplementation containing vitamins C and E, β-carotene, zinc, and copper slow the progression of the disease, probably through their antioxidant effects, and may be used in selected cases.

Visual loss

The common causes of blindness are similar across the world (**Box 31.4**). In low- and middle-income countries, trachoma due to *C. trachomatis* (see Chapter 43) is also a major cause, accounting for 10% of global blindness; onchocerciasis (river blindness, due to *Onchocerca volvulus*; see Chapter 42) accounts for blindness in about 1 million people, although this figure is decreasing with community treatment programmes. In leprosy, 70% of patients have ocular involvement, and blindness occurs in 5% to 10% of these. Ocular involvement is common in cerebral malaria (see Chapter 42), although loss of vision is rare.

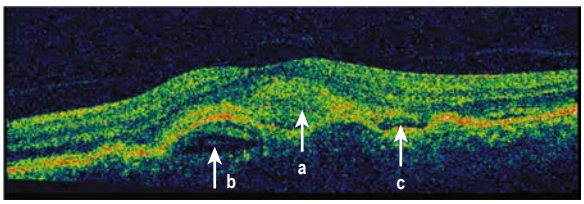


Fig. 31.30 Age-related macular degeneration seen by optical coherence tomography. **(a)** Area of thickened retina. **(b)** Pigment epithelial detachment. **(c)** Subretinal fluid.

HIV infection can produce uveitis but the major problem is severe opportunistic infection of the eye (e.g., with cytomegalovirus) in patients not of effective anti-retroviral treatment whose CD4 count is low.

Vitamin A deficiency and xerophthalmia affect millions each year; the World Health Organization (WHO) classification of xerophthalmia by ocular signs is shown in **Box 25.10**.

The WHO lists the most common causes of blindness across the world as cataract, glaucoma, acute macular degeneration, corneal opacity, diabetic retinopathy, and infections from bacteria or parasites.

Further reading

Amoaku Wm, Ghanch F, et al. Diabetic retinopathy and diabetic macular oedema pathways and management: UK Consensus Working Group. *Eye*. 2020;34:1–51.
Burkholder BM, Jabs DA. Uveitis for the non-ophthalmologist. *BMJ.* 2021;372:m4979.
Georgiou M, Fujinami K, Michaelides M. Inherited retinal diseases: Therapeutics, clinical trials and end points—A review. *Clin Exp Ophthalmol*. 2021;49(3):270–288.
Kim TI, Alió Del Barrio JL, Wilkins M, Cochener B, Ang M. Refractive surgery. *Lancet.* 2019;393(10185):2085–2098.
Liu YC, Wilkins M, Kim T, Malyugin B, Mehta JS. Cataracts. *Lancet.* 2017;390(10094):600–612.
Lynn WA, Lightman S. The eye in systemic infection. *Lancet.* 2004; 364:1439–1450.
Mitchell P, Liew G, Gopinath B, Wong TY. Age-related macular degeneration. *Lancet.* 2018;392(10153):1147–1159.
Quigley MA. Glaucoma. *Lancet.* 2011; 377:1367–1377.
Weiss JS, Møller HU, Aldave AJ et al. IC3D classification of corneal dystrophies—edition 2. *Cornea.* 2015; 34:117–159.

Questions for this chapter can be found on Elsevier eBooks+.

32 Rheumatology

Mohammed Abbas Anisur Rahman and Ian Philip Giles

Clinical skills 1062
Taking a musculoskeletal history 1062
Examination 1063
Anatomy and physiology of the normal joint 1064
Investigation of rheumatic disease 1065
Clinical conditions in rheumatology 1067
Common regional musculoskeletal problems 1067
Pain in the neck and shoulder 1067
Pain in the shoulder 1069
Pain in the elbow 1070
Pain in the hand and wrist 1070
Pain in the lower back 1071
Pain in the hip 1074
Pain in the knee 1075
Pain in the shin, calf, and ankle 1077
Pain in the foot 1077
Pain under the heel 1078
Pain in the chest 1078
Pain associated with sport and the performing arts 1079
Chronic pain syndromes 1079
Chronic widespread pain 1079
Chronic regional pain 1080
Analgesic and anti-inflammatory drugs for musculoskeletal problems 1081
Simple and compound analgesic agents 1081
Non-steroidal anti-inflammatory drugs 1081
Disorders of collagen 1082
Osteoarthritis 1083
Inflammatory arthritis 1087
Rheumatoid arthritis 1088
Spondyloarthritis 1099
Crystal arthritis 1103
Infection of joints 1106
Specific types of bacterial arthritis 1106
Arthritis in viral disease 1106
Fungal infections 1107
Autoimmune rheumatic diseases 1107
Systemic inflammatory vasculitis 1114
Large-vessel vasculitis 1115
Medium-sized vessel vasculitis 1116
Small-vessel vasculitis 1117
Behçet's disease 1117
Arthritis in children 1118
Juvenile idiopathic arthritis 1118
Rheumatological problems seen in other diseases 1118
Gastrointestinal and liver disease 1118
Respiratory disease 1118
Malignant disease 1118
Skin disease 1118
Neurological disease 1119
Blood disease 1119
Endocrine and metabolic disorders 1119
Miscellaneous arthropathies 1119

CORE SKILLS AND KNOWLEDGE

Musculoskeletal complaints are very common, comprising 25%–30% of consultations in primary care. The majority of patients have soft tissue complaints or osteoarthritis, with inflammatory arthritis and autoimmune rheumatic diseases having a much lower prevalence, of between 0.01% and 1% of the population.

Rheumatology is mainly an outpatient specialty, with care being delivered by multidisciplinary teams including nurse specialists, physiotherapists, and other allied health professionals. Apart from general rheumatology clinics, specialist clinics often exist (sometimes in tertiary referral centres) for patients with inflammatory conditions such as rheumatoid arthritis or systemic lupus erythematosus. Few patients require admission to hospital but many receive day-case infusions of biological disease-modifying drugs.

Key learning outcomes in musculoskeletal disease include:

- Learning to distinguish mechanical from inflammatory problems, and then identifying the number and pattern of joint involvement as well any extra-articular features in patients with inflammatory multisystem disease
- Becoming familiar with the main drug categories of analgesics and medication that alter the immune response
- Learning how to examine joints, and when to order autoantibody tests.

Opportunities for learning include sitting in general and specialist rheumatology clinics, observing musculoskeletal and joint injection lists, talking with and examining patients attending day units for infusions of medication, and attending imaging meetings.

The authors would like to dedicate this chapter to the memory of Dr Michael Shipley, who wrote the Rheumatology section of this textbook for many years and whose voice is still present in many of the sections below. He was a wonderful clinician and mentor and is greatly missed.

CLINICAL SKILLS

Taking a musculoskeletal history

In most rheumatological disorders (Box 32.1), the diagnosis is ***clinical:*** that is, made on the history and examination, and then supported by investigations. Taking a thorough history is therefore crucial, and the main points to consider are:

1. **Who is the patient?**
- ***Age:*** osteoarthritis, polymyalgia rheumatica, and giant cell arteritis present mainly in patients over the age of 50. Autoimmune rheumatic diseases and inflammatory arthritis frequently present in people under 50.
- ***Sex:*** rheumatoid arthritis (RA), Sjögren's syndrome, systemic lupus erythematosus (SLE), and systemic sclerosis are more common in females, whereas gout and ankylosing spondylitis are more common in males.
- ***Ethnicity:*** SLE is more common in people of black African or Caribbean background than in other ethnicities.
- ***Occupation:*** jobs that involve heavy use of one part of the body can lead to soft tissue pain and/or OA; for example, in the low back, hips, and knees of people who do a lot of heavy lifting. Musculoskeletal pain may also affect people's ability to do their job, or care for children or dependents. The effects of these on quality of life should be assessed.

2. **Has something happened to the patient?**
- ***Physical trauma:*** trauma to a particular region of the body may cause pain there and/or in related regions.
- ***Psychosocial stress and depression:*** physical, psychological, and social factors interact to contribute to symptoms. In musculoskeletal disorders, pain may cause distress and poor sleep, which in turn build up muscle tension and worsen the pain (see 'Chronic pain syndromes' section). Acute stressful events, such as bereavement, bullying, abuse, and redundancy, may exacerbate these symptoms.

3. **What are the symptoms?**
- **Pain:**
  - ***Where is it?*** The pattern of joint involvement is a useful clue to the diagnosis (Box 32.2).
  - ***Is it arising from joints, spine, muscles, or bone, with local tenderness?***
  - ***Could it be referred from another site?*** Joint pain is localized but may radiate distally—shoulder to upper arm; hip to thigh and knee.
  - ***Is it constant, intermittent, or episodic?***
  - ***Are there aggravating or precipitating factors?*** Is it made worse by activity and eased by rest (mechanical), or worse after rest (inflammatory)?
  - ***Are there any associated neurological features?*** Numbness, pins and needles, and/or loss of power suggest 'nerve' involvement.
- **Stiffness:**
  - Is it generalized or localized?
  - Does it affect the limb girdles or periphery?
  - Is it worse in the morning and relieved by activity?

  Joint stiffness for more than 30 minutes each morning—think of inflammatory arthritis. Morning spinal stiffness in younger adults—think of axial spondyloarthritis (e.g., ankylosing spondylitis) (see 'Axial spondyloarthritis' section). Shoulder and pelvic girdle stiffness and pain in a patient over 55 years may be polymyalgia rheumatica (see 'Polymyalgia rheumatica' section).
- **Swelling**
  - ***Does it affect one joint or several?*** Look for symmetry or asymmetry, and/or a peripheral or proximal pattern. An acute monoarthritis may be due to trauma, gout, pseudogout, or sepsis (fever or immunosuppression). A polyarthritis is more likely to be due to OA or RA.
  - ***Is it constant or does it come in short-lived or longer episodes?***
  - ***Is there associated inflammation*** (redness and warmth)?
- **Extra-articular symptoms:**
  - Osteoarthritis and soft tissue problems are not usually associated with symptoms outside the joints.
  - Features such as rash, breathlessness, neurological symptoms, or blood disorders suggest that the locomotor problems may be one facet of a more systemic disorder, such as RA, psoriatic arthritis, or SLE.

4. **What medication is the patient taking?**
- ***Could a drug be a cause?*** Diuretics may precipitate gout in men and older women. Steroids can cause avascular necrosis.
- ***Have the symptoms responded to medication?*** A good response to a trial of oral or intramuscular corticosteroid suggests an inflammatory or autoimmune problem.

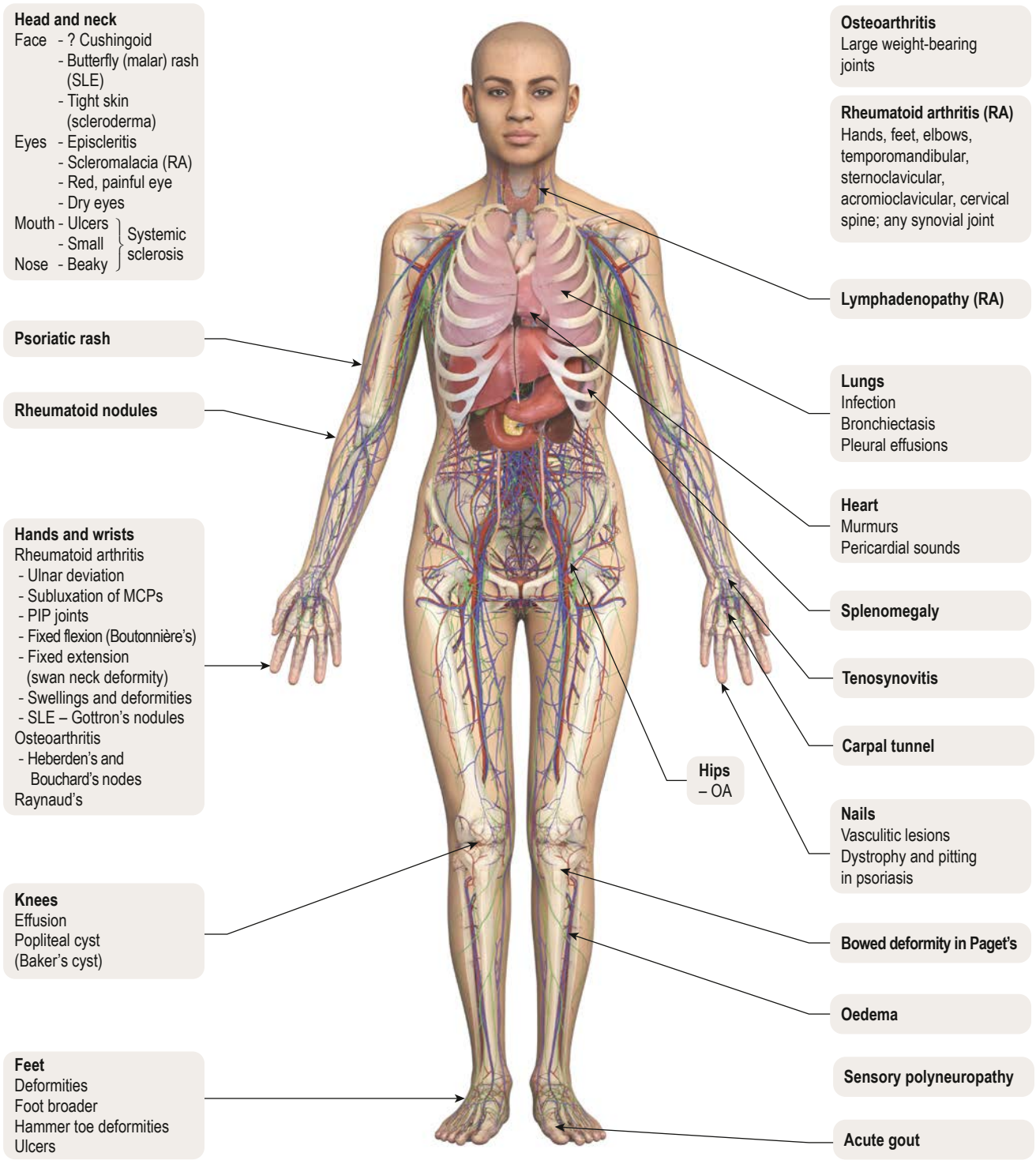
5. **What is the family history?**

i Box 32.1 The 'top 10' rheumatic conditions that students should understand

- Regional musculoskeletal conditions
- Chronic pain syndromes
- Osteoarthritis
- Rheumatoid arthritis
- Spondyloarthropathies
- Crystal arthropathies
- Infection-related arthritis
- Autoimmune rheumatic diseases
- Systemic vasculitides
- Juvenile idiopathic arthritis

? Box 32.2 Distinguishing patterns of rheumatic disease

Regional	Inflammatory	Multisystem
• Localized joint pain • Increased pain with activity • Short-lived stiffness • Distal radiation • Joint swelling • Functional limitation • Localized joint abnormalities on examination	• Widespread joint pain • Reduced pain with activity • Prolonged early morning stiffness • Multiple joint swelling • Functional limitation • Widespread joint abnormalities on examination	• Fever • Fatigue • Rash • Alopecia • Raynaud's phenomenon • Positive systems enquiry • Systems abnormalities on examination



Examination

Always observe patients, looking for disabilities, as they walk into the room and sit down. General and neurological examinations are often necessary, and may reveal some of the signs described in the figure above. Guidelines for rapid examinations of the limbs and spine are shown in **Box 32.3**.

Examining an individual joint involves three stages: looking, feeling, and moving (**Box 32.4**). A screening examination of the locomotor system, known by the acronym GALS (Gait, Arms, Legs, Spine), has been devised. Video demonstrations are available (see 'Further reading').

Further reading

https://www.versusarthritis.org/about-arthritis/healthcare-professionals/training-and-education-resources/clinical-assessment-of-patients-with-musculoskeletal-conditions/the-musculoskeletal-examination-gals/.

Box 32.3 Rapid examinations of the limbs and spine

Upper limbs

- Raise the arms sideways to the ears (abduction). Reach behind the neck and back. Difficulties with these movements indicate a shoulder or rotator cuff problem.
- Hold the arms forwards, with elbows straight and fingers apart, palm up and palm down. Fixed flexion at the elbow indicates an elbow problem. Examine the hands for swelling, wasting, and deformity.
- Place the hands in the 'prayer' position with the elbows apart. Flexion deformities of the fingers may be due to arthritis, flexor tenosynovitis, or skin disease. Painful restriction of the wrist limits the person's ability to move the elbows out with the hands held together.
- Make a tight fist. Difficulty with this indicates a loss of flexion or grip. Grip strength can be measured.

Lower limbs

- Ask the patient to walk a short distance away from and towards you, and to stand still. Look for abnormal posture or stance.
- Ask the patient to stand on each leg. Severe hip disease causes the pelvis on the non-weight-bearing side to sag (positive Trendelenburg test).
- Watch the patient stand and sit, looking for hip and/or knee problems.
- Ask the patient to straighten and flex each knee.
- Ask the patient to place each foot in turn on the opposite knee with the hip externally rotated. This tests for painful restriction of the hip or knee. Abnormal hips or knees must be examined with the patient lying down.
- Move each ankle up and down. Examine the ankle joint and tendons, medial arch, and toes while the patient is standing.

Spine

Stand behind the patient

- Ask the patient to (a) bend forwards to touch the toes with straight knees, (b) extend backwards, (c) flex sideways, and (d) look over each shoulder, flexing and extending and side-flexing the neck. Observe abnormal spinal curves—scoliosis (lateral curve), kyphosis (forward bending), or lordosis (backward bending). A cervical and lumbar lordosis and a thoracic kyphosis are normal. Muscle spasm is worse while standing and bending. Leg length inequality leads to a scoliosis that decreases on sitting or lying (the lengths are measured with the patient lying down).
- Ask the patient to lie supine. Examine any restriction of straight-leg raising (see disc prolapse, below).
- Ask the patient to lie prone. Examine for anterior thigh pain during a femoral stretch test (flexing knee while prone), which indicates a high lumbar disc problem.
- Palpate the spine and buttocks for tender areas.

Box 32.4 Examination of the joint

Look at the appearance of the joint

- Swelling—could be bony, fluid, or synovial
- Deformity:
  - Valgus, where the distal bone is deviated laterally (e.g., knock-knees or genu valgum)
  - Varus, where the distal bone is deviated medially (bow-legs or genu varum)
  - Fixed flexion or hyperextension
- Rash—especially psoriasis
- Muscle wasting—easier to see in large muscles like the quadriceps
- Scars—from surgery or trauma
- Signs of inflammation
- Symmetry:
  - Are the right and left joints (e.g., hips, knees, any other paired joint) the same?
  - If not, which do you think is abnormal?

Feel

- Swelling:
  - Fluid swelling (effusion) usually represents increased synovial fluid in inflammatory arthritis, but can be due to blood or pus
  - Synovial swelling is rubbery or boggy and usually occurs in inflammatory arthritis
  - Bony swelling, such as Heberden's nodes in the fingers, is usually seen in osteoarthritis
- Warmth—a warm joint may be inflamed or infected
- Tenderness—may represent joint inflammation, but many people have chronic tenderness all over the body (e.g., in fibromyalgia)

Move

- Active movement:
  - Is the range full and pain free?
  - Is the movement fluid?
  - In the hands, can the patient perform fine movements?
  - In the legs, can the patient walk properly?
- Compare movements on the right and left sides—are they symmetrical?
- Is there crepitus when the joint is moved?
- If active movement is limited, try passive movement:
  - In a joint problem both will usually be affected.
  - In a muscle or nerve problem, passive movement may remain full.

ANATOMY AND PHYSIOLOGY OF THE NORMAL JOINT

A joint can be defined as a place where two or more bones meet. There are three types of joints: fibrous, fibrocartilaginous, and synovial.

Fibrous and fibrocartilaginous joints

These include the intervertebral discs, the sacroiliac joints, the pubic symphysis, and the costochondral joints. Skull sutures are fibrous joints. Little movement occurs at such joints.

Synovial joints

Synovial joints (Fig. 32.1) include the ball-and-socket joints (e.g., hip) and the hinge joints (e.g., interphalangeal). They are designed to allow movement, which is restricted to a required range, and stability is maintained during use. The load is distributed across the surface, thus preventing damage by overloading or disuse. Each structural component of a synovial joint plays a key functional role, and different components are affected in different disease processes.

Juxta-articular bone

Bone structure and physiology are discussed in Chapter 33. The bone that abuts a joint (juxta-articular bone) is highly vascular and comprises a light framework of mineralized collagen enclosed in a thin coating of tougher, cortical bone. It withstands pressure poorly if the normal intra-articular covering of hyaline cartilage is lost, as in osteoarthritis (OA; see 'Osteoarthritis' section). This process can lead to abnormalities of bone growth and remodelling.

Articular cartilage

The hyaline cartilage lining the bones within a joint is called articular cartilage. It is avascular and derives nourishment from synovial fluid. It is predominantly composed of type II collagen, encoded by

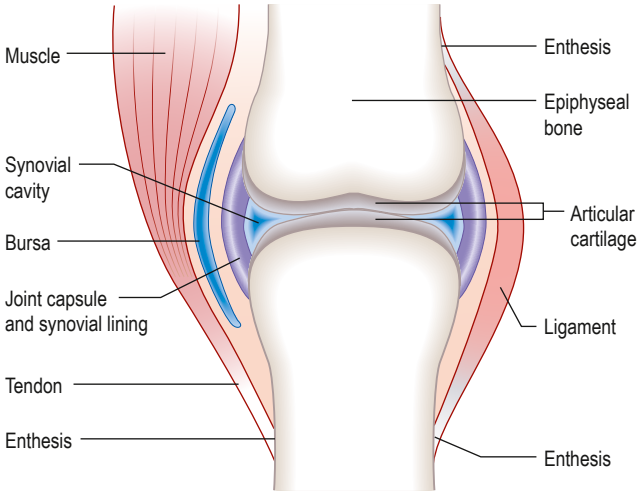


Fig. 32.1 Normal synovial joint.

the *COL2A1* gene, which forms a mesh-like network. Chondrocytes secrete collagen and proteoglycans, and are embedded in the cartilage. Defects in articular cartilage and underlying bone are features of ***osteoarthritis***.

Synovium and synovial fluid

The joint capsule, which is connected to the periosteum, is lined with synovium, which is a few cells thick and vascular. Its surface is smooth, non-adherent, and is permeable to proteins and crystalloids. Synovial fluid is a highly viscous fluid secreted by the synovial cells and has a similar consistency to plasma. Glycoproteins ensure a low coefficient of friction between the cartilaginous surfaces. Tendon sheaths and bursae are also lined by synovium. Inflammation of the synovium is a feature of ***inflammatory arthritis***.

Ligaments and tendons

These structures stabilize joints. Ligaments are variably elastic and this contributes to the stiffness or laxity of joints (see 'Hypermobile Ehlers-Danlos syndrome and hypermobility spectrum disorders' section). Tendons are inelastic and transmit muscle power to bones. The joint capsule is formed by intermeshing tendons and ligaments. The point where a tendon or ligament joins a bone is called an ***enthesis*** and may be the site of inflammation. Whereas most ligaments and tendons run outside the joints, some, like the supraspinatus tendon in the shoulder and the cruciate ligaments in the knee, run through the joint. Inflammation or trauma to these structures can cause severe joint symptoms.

Blood vessels and nerves

The ligaments, periosteum, synovial tissue, and capsule of the joint are richly supplied by blood vessels and nerves. Pain usually derives from inflammation of these sites because the synovial membrane is relatively insensitive.

Skeletal muscle

This tissue consists of bundles of myocytes containing actin and myosin molecules. These molecules interdigitate and form myofibrils, which cause muscle contraction in a similar way to myocardial muscle (see Chapter 17). Bundles of myofibrils (fasciculi) are covered by connective tissue, the perimysium, which merges with the epimysium (covering the muscle) and forms the tendon, which attaches to the bone surface (enthesis).

Though not strictly a component of the joint itself, muscles are so closely related to joints that strain and tension in muscles are commonly interpreted by patients as joint pain. Pain in muscles and ligaments (myofascial pain) is a very common cause of locomotor symptoms. Primary inflammatory disease of muscle (myositis) is far less common.

INVESTIGATION OF RHEUMATIC DISEASE

Investigations are unnecessary in many of the common musculoskeletal problems; the diagnosis is clear from the history and examination findings. Tests help to exclude another condition and to reassure the patient or their primary care physician.

Useful blood screening tests

- Full blood count:
  - ***Haemoglobin.*** Normochromic, normocytic anaemia suggests chronic inflammatory and autoimmune diseases. Hypochromic, microcytic anaemia indicates iron deficiency, sometimes due to non-steroidal anti-inflammatory drug (NSAID)-induced gastrointestinal bleeding.
  - ***White cell count.*** Neutrophilia is seen in bacterial infection (e.g., septic arthritis). It also occurs with corticosteroid treatment. Lymphopenia is found in viral illnesses or SLE. Neutropenia may reflect drug-induced bone marrow suppression. Eosinophilia is seen in eosinophilic granulomatosis with polyangiitis (EGPA) (see Chapter 20).
- ***Platelets.*** Raised platelets occur with any chronic inflammation. Thrombocytopenia is seen in drug-induced bone marrow suppression and may be a feature of SLE.
- ***Erythrocyte sedimentation rate (ESR) and C-reactive protein (CRP).*** An increase in these reflects inflammation. Plasma viscosity is also raised in inflammatory disease.
- ***Bone and liver biochemistry.*** A raised serum alkaline phosphatase may indicate liver or bone disease. A rise in liver enzymes is often seen with drug-induced toxicity. For other investigations of bone, see Chapter 33.

Other blood and urine tests

- Protein electrophoresis (and/or immunofixation), serum-free light chain testing, and urinary Bence Jones protein—to exclude myeloma as a cause of a raised ESR.
- Serum uric acid—for gout.
- Antistreptolysin O titre—in rheumatic fever.

Serum autoantibody studies

- ***Rheumatoid factors (RFs)*** (see also 'Rheumatoid arthritis' section). RFs are detected by enzyme-linked immunosorbent assay (ELISA). RFs are antibodies (usually immunoglobulin (Ig) M, but also IgG or IgA) against the Fc portion of IgG. They are detected in 70% of people with RA but are not diagnostic. RFs are detected in many autoimmune rheumatic disorders (e.g., SLE), in chronic infections, and in asymptomatic older people (**Box 32.5**).
- ***Anti-citrullinated peptide antibodies (ACPA).*** These antibodies are directed against citrullinated antigens, vimentin, fibrinogen, alpha enolase, and type II collagen. They are measured by an ELISA technique and are present in up to 80% of people with RA. They have a high specificity for RA (90%, with a sensitivity of 60%–80%). They are helpful in early disease when the RF is negative, to distinguish it from acute transient synovitis (see **Box 32.28**). Positivity for RF and/or ACPA is associated with a worse prognosis and an increase in the likelihood of bony erosions in people with RA.

- ***Antinuclear antibodies (ANAs).*** The gold-standard method for ANA detection is indirect immunofluorescence on human epithelial (Hep-2) cells. Different patterns reflect a variety of antigenic specificities that occur with different clinical pictures, for example the anti-centromere pattern seen in samples from patients with limited cutaneous systemic sclerosis and the speckled pattern suggesting antibodies to extractable nuclear antigens (see below). ANA is used as a screening test for SLE and systemic sclerosis—a negative ANA makes either condition highly unlikely—but low titres occur in RA and chronic infections and in normal individuals, especially the elderly (Box 32.6).
- ***Anti-double-stranded DNA (dsDNA) antibodies.*** These are usually detected by a precipitation test (Farr assay), ELISA, or an immunofluorescent test using *Crithidia luciliae* (which contains dsDNA). Raised anti-dsDNA is highly specific for SLE; the levels usually rise and fall in parallel with disease activity so can be used to monitor the level of treatment required.
- ***Anti-extractable nuclear antigen (ENA) antibodies*** (see Box 32.38). These produce a speckled ANA fluorescent pattern, and can be identified by ELISA. The most commonly measured ENAs are:
  - ***anti-Ro and anti-La***, which occur in Sjögren's syndrome and SLE
  - ***anti-Sm***, which is highly specific for SLE
  - ***anti-Jo-1***, which is the most common of the anti-tRNA synthetase enzymes that occur in some people with dermatomyositis or polymyositis
  - ***anti-topoisomerase I (anti–Scl-70)***, which is specific for systemic sclerosis
  - ***anti-RNA polymerase I and III***, which occur in systemic sclerosis and are associated with pulmonary fibrosis.
- ***Anti-neutrophil cytoplasmic antibodies (ANCAs)*** (see 'Small-vessel vasculitis' section). These are predominantly IgG autoantibodies directed against the primary granules of neutrophil and macrophage lysosomes. They are strongly associated with small-vessel vasculitis. Two major clinically relevant ANCA patterns are recognized on ***immunofluorescence:***
  - ***proteinase 3 (PR3-ANCA)***, also called cytoplasmic or cANCA, producing a granular immunofluorescence and seen in granulomatosis with polyangiitis (GPA)
  - ***myeloperoxidase (MPO-ANCA)***, also called perinuclear or pANCA, producing a perinuclear staining pattern and seen in microscopic polyangiitis and EGPA.
- ***Antiphospholipid antibodies*** These are detected in the antiphospholipid syndrome (see 'Antiphospholipid syndrome' section).
- ***Complement.*** Low complement levels indicate consumption and suggest an active disease process in SLE.

Box 32.5 Conditions in which rheumatoid factor (RF) is found in the serum

Condition	RF (IgM) %
Autoimmune rheumatic diseases	
Rheumatoid arthritis	70
Systemic lupus erythematosus	25
Sjögren's syndrome	90
Systemic sclerosis	30
Polymyositis/dermatomyositis	50
Juvenile idiopathic arthritis	Variable
Viral infections	**Hyperglobulinaemias**
Hepatitis	Chronic liver disease
Infectious mononucleosis	Sarcoidosis
	Cryoglobulinaemia
Chronic infections	**Normal population**
Tuberculosis	Elderly
Leprosy	Relatives of people with rheumatoid arthritis
Infective endocarditis	
Syphilis	

Box 32.6 Conditions in which serum antinuclear antibodies are found

Condition	%
Systemic lupus erythematosus	95
Systemic sclerosis	70
Sjögren's syndrome	80
Polymyositis and dermatomyositis	40
Rheumatoid arthritis	30
Juvenile idiopathic arthritis	Variable
Other diseases	
Autoimmune hepatitis	100
Drug-induced lupus	>95
Myasthenia gravis	50
Idiopathic pulmonary fibrosis	30
Diabetes mellitus	25
Infectious mononucleosis	5–10
Normal population	8

Joint aspiration and examination of synovial fluid

Examination of joint (or bursa) fluid is used mainly to diagnose septic, reactive, or crystal arthritis. The appearance of the fluid is an indicator of the level of inflammation. The procedure is often undertaken in combination with injection of a corticosteroid. Aspiration alone is therapeutic in crystal arthritis (Fig. 32.2 and Box 32.7).

Aspiration and analysis of synovial fluid are always indicated when septic or crystal-induced arthritis is suspected, particularly a monoarthritis. Normal fluid is clear and straw-coloured, and contains fewer than 3000 white blood cells (WCC)/mm^3. Inflammatory fluid is cloudy and contains more than 3000 WCC/mm^3. Septic fluid is opaque and less viscous, and contains up to 75 000 WCC/mm^3. There is much overlap.

Polarized light microscopy is performed for crystals:

- ***gout:*** negatively bi-refringent, needle-shaped crystals of sodium urate
- ***calcium pyrophosphate deposition arthropathy:*** rhomboidal, weakly positively bi-refringent crystals of calcium pyrophosphate.

Gram-staining is essential if septic arthritis is suspected and may identify the organism immediately. Joint fluid should be cultured and antibiotic sensitivities requested.

Diagnostic imaging and visualization

- ***X-rays*** can be diagnostic in certain conditions (e.g., established RA) and are the first investigation in many cases of trauma. X-rays can detect joint space narrowing, erosions in RA, calcification in soft tissue, new bone formation, e.g., osteophytes, and

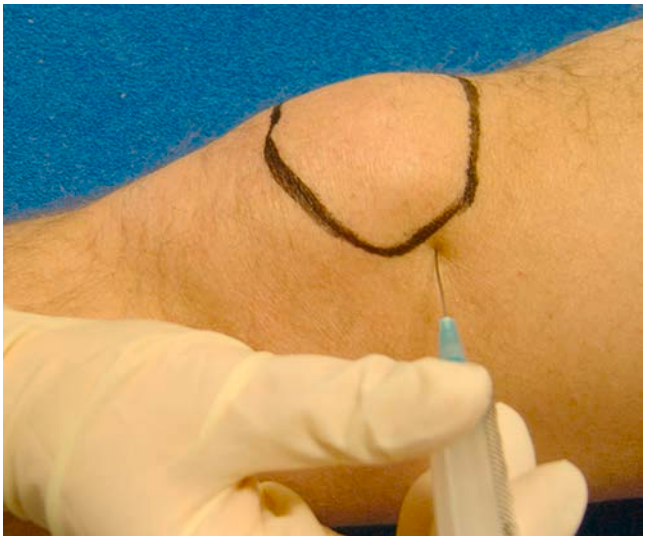


Fig. 32.2 Joint aspiration.

Box 32.7 Joint aspiration

This is a sterile procedure that should be carried out in a clean environment. Explain the procedure to the patient; obtain consent.

1. Decide on the site to insert the needle and mark it.
2. Clean the skin and your hands scrupulously.
3. Draw up local anaesthetic (and corticosteroid if it is being used).
4. Insert the needle, injecting local anaesthetic as it advances; if a joint effusion is suspected, attempt to aspirate as you advance it.
5. If fluid is obtained, change syringes and aspirate fully.
6. Examine the fluid in the syringe and decide whether or not to proceed with a corticosteroid injection (if fluid is clear or slightly cloudy) or send for microbiological tests.
7. Cover the injection site and advise the patient to rest the affected area for a few days. Warn the patient that the pain may increase initially but to report urgently if this persists beyond a few days, if the swelling worsens or if they become febrile, because this might indicate an infected joint.

Aspirated fluid is sent for microscopy, Gram-stain, and culture. Different types of crystals can be identified using polarized light microscopy. Fluid from septic arthritis is often excessively turbid.

decreased bone density (osteopenia) or increased bone density (osteosclerosis):

- In acute low back pain (see 'Pain in the lower back' section), X-rays are indicated only if the pain is persistent, recurrent, associated with neurological symptoms or signs, worse at night, or associated with symptoms such as fever or weight loss.
- Radiological changes are common in older people and may not indicate symptomatic OA or spondylosis, a term that implies OA within the spine.
- X-rays are of little diagnostic value in early inflammatory arthritis but are useful as a baseline from which to judge later change.

- ***Ultrasound*** is particularly useful for periarticular structures, soft tissue swellings, and tendons, and for detecting active synovitis in inflammatory arthritis. It is increasingly used to examine the shoulder and other structures during movement, e.g., shoulder impingement syndrome (see 'Pain in the shoulder' section). Doppler ultrasound measures blood flow and hence inflammation. Ultrasound may be used to guide local corticosteroid injections.
- ***Magnetic resonance imaging (MRI)*** shows bone changes and intra-articular structures in striking detail. Visualization of particular structures can be enhanced with different resonance sequences. T_1-weighted MRI is used for anatomical detail, T_2-weighted for fluid detection, and short tau inversion recovery (STIR) for the presence of bone marrow oedema. MRI is more sensitive than X-rays in the early detection of articular and periarticular disease. Gadolinium injection enhances inflamed tissue. MRI is especially useful for detection of damage to non-bony tissues in or near joints, e.g., meniscal tears in the knee and torn rotator cuff muscles in the shoulder; detection of nerve root compression in the spine; detection of inflamed muscle in myositis; and early detection of synovitis in inflammatory arthritis.
- ***Computerized axial tomography (CT)*** is useful for detecting changes in calcified structures but the dose of radiation is high.
- ***Bone scintigraphy*** utilizes radionuclides, usually ^{99m}Tc, and detects abnormal bone turnover and blood circulation; although non-specific, it helps in detecting areas of inflammation, infection, or malignancy. It is best used in combination with other anatomical imaging techniques.
- ***Dual-energy X-ray absorptiometry (DXA) scanning*** uses very low doses of X-irradiation to measure bone density and is used in the screening and monitoring of osteoporosis.
- ***Positron emission tomography (PET) scanning*** uses radionuclides, which decay by emission of positrons. ^{18}F-fluorodeoxyglucose uptake indicates areas of increased glucose metabolism. PET is used to locate tumours and demonstrate large-vessel vasculitis, e.g., Takayasu's arteritis (see Chapter 17). PET scans are combined with CT to improve anatomical details.

Other tests to investigate rheumatological disease

- ***Arthroscopy*** is a direct means of visualizing a joint, particularly the knee or shoulder. Biopsies can be taken, surgery performed in certain conditions (e.g., repair or trimming of meniscal tears), and loose bodies removed.
- ***Nerve conduction studies and electromyography*** are used to diagnose nerve entrapment syndromes (such as carpal or tarsal tunnel) and to distinguish myositis from neuropathies.

CLINICAL CONDITIONS IN RHEUMATOLOGY

These can be divided into three types:

- ***Pain arising from soft tissues around joints***, such as muscles and tendons. The joints themselves are normal. These are the most common locomotor problems, and are generally self-limiting or respond to simple analgesia, physiotherapy, or exercise.
- ***Diseases of the joints themselves.*** Osteoarthritis is more common than inflammatory arthritis, especially in the elderly.
- ***Autoimmune rheumatic diseases.*** These include disorders such as SLE, where the locomotor symptoms are just one manifestation of a systemic disorder.

COMMON REGIONAL MUSCULOSKELETAL PROBLEMS

See Fig. 32.3.

Pain in the neck and shoulder

See Box 32.8.

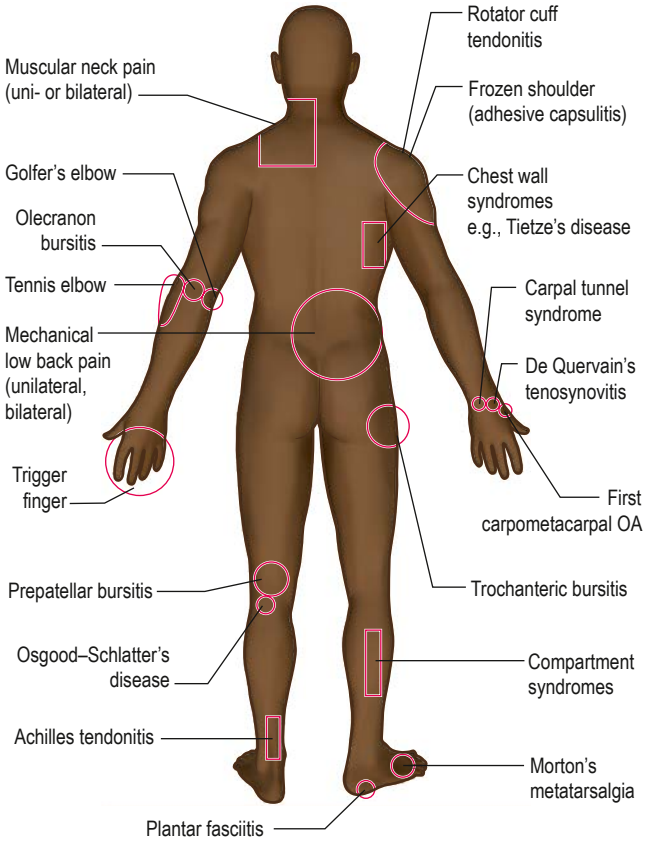


Fig. 32.3 Common regional musculoskeletal problems. *OA,* Osteoarthritis.

? Box 32.8 Pain in the neck and shoulder

- Trauma (e.g., a fall)
- Mechanical or muscular neck pain
- Whiplash injury
- Disc prolapse—nerve root entrapment
- Ankylosing spondylitis
- Shoulder lesions:
  - Rotator cuff tendonitis
  - Calcific tendonitis or bursitis
  - Impingement syndrome or rotator cuff tear
  - Adhesive capsulitis (true 'frozen' shoulder)
  - Inflammatory arthritis or osteoarthritis
- Polymyalgia rheumatica
- Fibromyalgia (chronic widespread pain)
- Chronic (work-related) upper-limb pain syndrome
- Tumour

Mechanical or muscular neck pain (shoulder girdle pain)

Unilateral or bilateral muscular-pattern neck pain is common and usually self-limiting. Chronic burning neck pain occurs because of muscle tension from anxiety and stress.

Spondylosis seen on X-ray increases after the age of 40 years but it is not always related to pain. Spondylosis can, however, cause stiffness and increases the risk of mechanical or muscular neck pain. Muscle spasm is palpable and tender, and may lead to abnormal neck posture (e.g., acute torticollis). Muscular-pattern neck pain is not localized but affects the trapezius muscle, the C7 spinous process, and the paracervical musculature (shoulder girdle pain). Pain often radiates upwards to the occiput and is commonly associated with tension headaches. These features are also seen in chronic widespread pain (see 'Chronic widespread pain' section).

i Box 32.9 Cervical nerve root entrapment: symptoms and signs

Nerve root	Sensory changes	Reflex loss	Weakness
C5	Lateral arm	Biceps	Shoulder abduction
			Elbow flexion
C6	Lateral forearm	Biceps	Elbow flexion
	Thumb and index finger	Supinator	Wrist extension
C7	Middle finger	Triceps	Elbow extension
C8	Medial forearm	None	Finger flexion
	Little and ring fingers		
T1	Medial upper arm	None	Finger abduction and adduction

Management

Patients are given short courses of analgesic therapy, along with reassurance and explanation. Physiotherapists can help to relieve spasm and pain, teach exercises and relaxation techniques, and improve posture. An occupational therapist can advise about the ergonomics of the workplace if the problem is work-related (see 'Chronic [work-related] upper-limb pain syndrome' section).

Nerve root entrapment

This is caused by an acute cervical disc prolapse or pressure on the root from spondylotic osteophytes narrowing the root canal.

Acute cervical disc prolapse presents with unilateral pain in the neck, radiating to the interscapular and shoulder regions. This diffuse, aching, dural pain is followed by sharp, electric shock-like pain down the arm, in a nerve root distribution, often with pins and needles, numbness, weakness, and loss of reflexes (**Box 32.9**).

Cervical spondylosis occurs in the older patient with posterolateral osteophytes compressing the nerve root and causing root pain (see **Fig. 28.43**, commonly at C5/C6 or C6/C7; it is seen on oblique radiographs of the neck. An MRI scan clearly distinguishes facet joint OA, root canal narrowing, and disc prolapse.

Management

A support collar, rest, analgesia, and sedation are used initially as necessary. Patients should be advised not to carry heavy items. They usually recover in 6–12 weeks. MRI is the investigation of choice if surgery is being considered or the diagnosis is uncertain (**Fig. 32.4**). A cervical root block, administered under direct vision by an experienced pain specialist, may relieve pain while the disc recovers. Neurosurgical referral is recommended if the pain persists or if the neurological signs of weakness or numbness are severe or bilateral. Bilateral root pain, with or without long tract symptoms or signs, is a neurosurgical emergency because a central disc prolapse may compress the cervical spinal cord. Posterior osteophytes may cause spinal claudication and cervical myelopathy.

Whiplash injury

Whiplash injury results from acceleration–deceleration forces applied to the neck, usually in a road traffic accident when the car of a person wearing a seat belt is struck from behind. A simple decision plan based on clinical criteria helps to distinguish those most

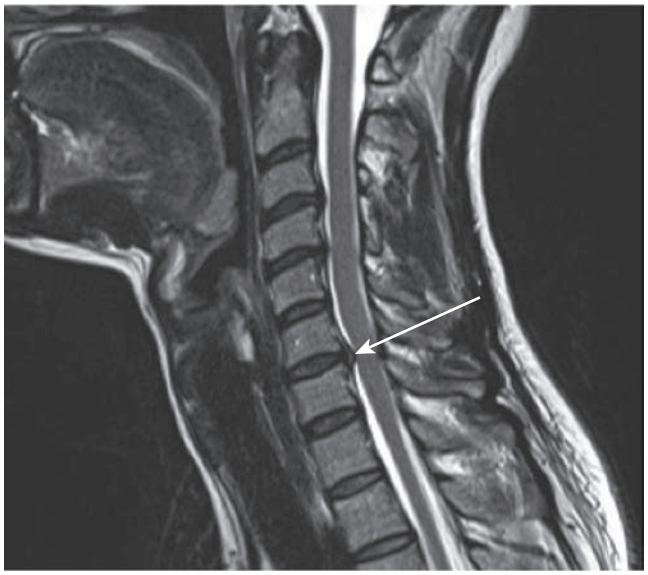


Fig. 32.4 Magnetic resonance image of the cervical spine. A large central disc prolapse *(arrowed)* is shown at the C6/7 level, and smaller disc bulges at C3/4 and C4/5.

at risk and who warrant radiography. There is a low probability of serious bony injury if there is:

- no midline cervical tenderness
- no focal neurological deficit
- normal alertness
- no intoxication
- no other painful distracting injury.

CT scans are reserved for those with bony injury. MRI scans occasionally show severe soft tissue injury. Whiplash injuries may lead to litigation or insurance claims.

Whiplash injury is a common cause of chronic neck pain, although most people recover within a few weeks or months. Delayed recovery depends, in part, on the severity of the initial injury. The pattern of chronic neck pain is often complex, involving pain in the neck, shoulder, and arm. It may be accompanied by subjective symptoms, such as headache, dizziness, and poor concentration. The subjective nature of these symptoms has led to controversy about their cause.

Management

Management is with reassurance (the patient is often distressed and anxious), analgesia, a short-term support collar and physiotherapy. Pain may take a few weeks or months to settle and the patient should be warned of this. Clinical trial evidence shows no benefit with prolonged physiotherapy.

Pain in the shoulder

The shoulder is a shallow joint with a large range of movement. The humeral head is held in place by the rotator cuff, which is part of the joint capsule. It comprises the tendons of infraspinatus and teres minor posteriorly, supraspinatus superiorly, and teres major and subscapularis anteriorly. The rotator cuff (particularly supraspinatus) prevents the humeral head from blocking against the acromion during abduction; the deltoid pulls up and the supraspinatus pulls in to produce a turning movement, and the greater tuberosity glides under the acromion without impingement. Shoulder pathology restricts or is made worse by shoulder movement. Specific diagnoses are difficult to make clinically but may be clearer on imaging.

? Box 32.10 Differential diagnosis of 'shoulder' pain

- Rotator cuff tendonitis pain is worse at night and radiates to the upper arm.
- Painful shoulders produce secondary muscular neck pain.
- Muscular neck pain (also known as shoulder girdle pain) does not radiate to the upper arm.
- Cervical nerve root pain is usually associated with pins and needles or neurological signs in the arm.

Pain in the shoulder can sometimes be due to problems in the neck; the differential diagnosis is shown in **Box 32.10**. Early inflammatory arthritis and polymyalgia rheumatica (PMR) in the elderly may present with shoulder pain.

Rotator cuff (supraspinatus) tendonosis

This condition is a common cause of shoulder pain at all ages. It follows trauma in 30% of cases and is bilateral in under 5%. The pain radiates to the upper arm and is made worse by arm abduction and elevation, which are often limited. The pain is often worse during the middle of the range of abduction, reducing as the arm is raised fully; a so-called 'painful arc syndrome'. When examined from behind, the scapula rotates earlier than usual during elevation. Passive elevation reduces impingement and is less painful. Severe pain virtually immobilizes the joint, although some rotation is retained (compare adhesive capsulitis; see below). There is also painful spasm of the trapezius. There may be an associated subacromial bursitis. Isolated subacromial bursitis occurs after direct trauma, falling on to the outstretched arm or elbow. Acromioclavicular osteophytes increase the risk of impingement and may need to be removed surgically.

X-ray is often normal but ultrasound is very useful to distinguish bursitis, tendonitis, and partial or complete tendon tears. Corticosteroid injections can be guided by ultrasound.

Management

Analgesics, NSAIDs, and/or physiotherapy may suffice, but severe pain responds to an injection of corticosteroid into the subacromial bursa (see **Fig. 32.5**). Patients should be warned that 10% will develop worse pain for 24–48 hours after injection. Some 70% improve over 5–20 days and mobilize the joint themselves. Physiotherapy helps persistent stiffness. If injection in the clinic is unsuccessful ultrasound-guided corticosteroid injections may be more effective.

Torn rotator cuff

This condition is caused by trauma but also occurs spontaneously in the elderly and in RA. It prevents active abduction of the arm but patients learn to initiate elevation using the unaffected arm. Once elevated, the arm can be held in place by the deltoid muscle. In younger people, the tear is repaired surgically but this is rarely possible in the elderly or in RA. Some patients require arthroscopic surgery.

Calcific tendonosis and bursitis

Calcium pyrophosphate deposits in the tendon are visible on X-ray but they are not always symptomatic. The pathogenesis is unclear, although ischaemia may play a part. The deposit is usually just proximal to the greater tuberosity. It may lead to acute or chronic recurrent shoulder pain and restriction of movement. A local

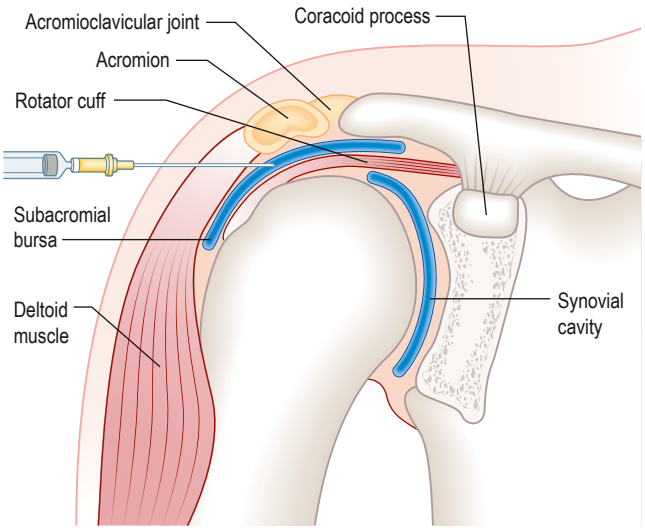


Fig. 32.5 The shoulder region, showing site of injection and subacromial space.

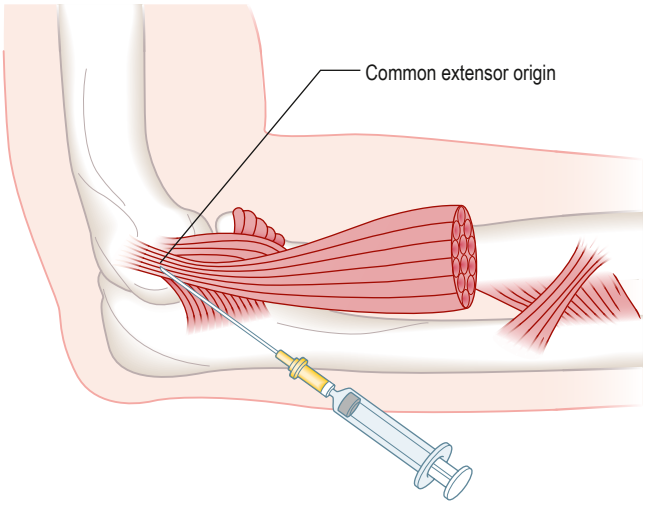


Fig. 32.6 Injection for tennis elbow.

corticosteroid injection may relieve the pain. The calcification may persist or resolve. Aspiration or breaking up of the deposit under ultrasound control may be required for persistent pain. Rarely, arthroscopic removal is necessary.

Shedding of crystals into the subacromial bursa causes a bursitis with severe pain and shoulder restriction. The shoulder feels hot and swollen, and an X-ray shows a diffuse opacity in the bursa. The differential diagnosis of calcific bursitis is gout, pseudogout, or septic arthritis. Aspiration and injection with corticosteroid can help.

Adhesive capsulitis (true 'frozen' shoulder)

This condition can develop with rotator cuff lesions, or following hemiplegia, chest or breast surgery, or myocardial infarction and in diabetics. Initially, it causes severe shoulder pain and a gradually reducing range of movement, leading to a 'frozen' phase where there is loss of all shoulder movements but little pain. NSAIDs and intra-articular injections of local anaesthetic and corticosteroids are helpful in the painful phase. Hydrodistension, where 20–30 mL of a mixture of saline, steroid, and local anaesthetic is injected into the shoulder to stretch the capsule, is effective and carried out under ultrasound guidance. Subsequently, there is usually a gradual improvement in function over weeks to months. Therapeutic exercises and physiotherapy help in later phases. Once the pain settles, arthroscopic release speeds functional recovery but is not necessary in most cases.

Pain in the elbow

Pain in the elbow can be due to epicondylitis, inflammatory arthritis, or, occasionally, OA.

Epicondylitis

Two common sites where the insertions of tendons into bone become inflamed (enthesitis) are the common wrist extensor origin at the lateral humeral epicondyle ('tennis elbow') and the common wrist flexor origin at the medial epicondyle ('golfer's elbow'). Despite the names, both conditions are usually unrelated to either sport!

There is local tenderness. Pain radiates into the forearm on using the affected muscles—typically, gripping or holding a heavy bag in tennis elbow or carrying a tray in golfer's elbow. Pain at rest also occurs.

Management

Advise rest and arrange review by a physiotherapist. A local injection of corticosteroid at the point of maximum tenderness is helpful when the pain is severe (Fig. 32.6) but needs physiotherapy follow-up to prevent recurrences. Avoid the ulnar nerve when injecting golfer's elbow. Both conditions settle spontaneously eventually, but occasionally persist and require surgical release.

Pain in the hand and wrist

See Box 32.11. Hand pain is commonly caused by injury or repetitive, work-related activities. When associated with pins and needles or numbness, it suggests a neurological cause arising at the wrist, elbow, or neck. Pain and stiffness that are worse in the morning are due to tenosynovitis or inflammatory arthritis. The distribution of hand pain often indicates the diagnosis.

Tenosynovitis

The finger flexor tendons run through synovial sheaths and under loops that hold them in place. Inflammation occurs with repeated or unaccustomed use, or in inflammatory arthritis. The thickened sheaths are often palpable.

Flexor tenosynovitis causes finger pain when gripping and stiffness of the fingers in the morning. Occasionally, a tendon causes a trigger finger, when the finger remains flexed in the morning or after gripping, and has to be pulled straight. A tender tendon nodule is palpable, usually in the distal palm. Trigger finger or thumb is more common in diabetic patients.

Dorsal tenosynovitis is less common, except in RA. The hourglass swelling extends from the back of the hand and under the extensor retinaculum.

De Quervain's tenosynovitis causes pain and swelling around the radial styloid, where the abductor pollicis longus tendon is held in place by a retaining band. There is local tenderness, and the pain at the styloid is worsened by flexing the thumb into the palm.

Management

Resting, splinting, and NSAIDs may help. Local corticosteroids injected alongside the tendon under low pressure (not into the

Box 32.11 Causes of pain in the hand and wrist

All ages
- Trauma/fractures
- Tenosynovitis:
  - Flexor (with/without triggering)
  - Dorsal
  - De Quervain's
- Carpal tunnel syndrome
- Ganglion
- Inflammatory arthritis
- Raynaud's syndrome (see Chapter 17)
- Complex regional pain syndrome type I (see 'Complex regional pain syndrome' section)

Older patients
- Nodal OA:
  - DIPs (Heberden's nodes)
  - PIPs (Bouchard's nodes)
  - First carpometacarpal joint
- Trauma—scaphoid fracture
- Pseudogout
- Gout:
  - Acute
  - Tophaceous

tendon itself) are helpful. Occasionally, surgery is needed if symptoms persist.

Carpal tunnel syndrome

This condition arises due to median nerve compression in the limited space of the carpal tunnel. Thickened ligaments, tendon sheaths, or bone enlargement can cause it, but it is usually idiopathic. (Causes are discussed in Chapter 28.) The history is usually typical and diagnostic, with the patient waking with numbness, tingling, and pain in a median nerve distribution. The pain may radiate to the forearm. Wasting of the thenar eminence muscles develops with sensory loss in the radial three-and-a-half fingers. The pain may be produced by tapping the nerve in the carpal tunnel (Tinel's sign) or by holding the wrist in flexion (Phalen's test).

Management

Management is with a splint to hold the wrist in dorsiflexion overnight, which relieves the symptoms and is diagnostic; used nightly for several weeks, it may produce full recovery. If it does not, a corticosteroid injection into the carpal tunnel (avoid the nerve!) helps in about 70% of cases, although it may recur. Persistent symptoms or nerve damage produce prolonged latency across the carpal tunnel on nerve conduction studies and require surgical decompression.

Other conditions causing pain

Inflammatory arthritis

This may present with pain, swelling, and stiffness of the hands. In RA, the wrists, proximal interphalangeal (PIP) joints, and metacarpophalangeal (MCP) joints are affected symmetrically. In psoriatic arthritis and reactive arthritis, a finger may be swollen (dactylitis), or the distal interphalangeal (DIP) joints and nails are affected asymmetrically.

Nodal osteoarthritis

This affects the DIP and, less commonly, PIP joints, which are initially swollen and red. The inflammation and pain settle but bony swellings remain (see 'Nodal OA' section).

First carpometacarpal osteoarthritis

This causes pain at the base of the thumb when gripping, or painless stiffness at the base of the thumb, often in people with nodal OA.

Scaphoid fractures

These cause pain in the anatomical snuffbox. They are not seen immediately on X-ray; if there is clinical suspicion, a cast is necessary. Untreated scaphoid fractures can eventually cause pain because of failed union.

Ganglion

A ganglion is a jelly-filled, often painless swelling caused by a partial tear of the joint capsule or tendon sheath. The wrist is a common site. Treatment is not essential, as many resolve or cause little trouble; otherwise, surgical excision is the best option.

Dupuytren's contracture

This condition is a painless, palpable fibrosis of the palmar aponeurosis, with fibroblasts invading the dermis. It causes puckering of the skin and gradual fixed flexion, usually of the ring and little fingers. It is more common in males, white people, individuals with diabetes mellitus, and those who overuse alcohol. A similar fibrosis occurs in the feet and is often more aggressive. Intralesional steroid injections may help in early disease and some advocate transcutaneous needle aponeurotomy. Percutaneous collagenase injection into the lesion has been shown to be effective in several studies and is now first-line treatment used before surgery. Surgical release of the contracture is restricted to those with severe deformity of the fingers.

Pain in the lower back

Low back pain is a common symptom. It is often traumatic and work-related, although lifting apparatus, other mechanical devices, and improved office seating help to avoid it. Episodes are generally short-lived and self-limiting, and patients attend a physiotherapist or osteopath more often than a doctor. Chronic back pain is the cause of 14% of long-term disability in the UK. The causes are listed in **Box 32.12**.

The majority of cases (about 90%) of low back pain are uncomplicated and arise from mechanical causes. Management of this type of back pain is summarized in **Box 32.13**.

One should be alert to clues in the history and examination that could suggest more serious causes of low back pain. These ***'red flag'*** clues (summarized in **Fig. 32.7**) include the following:
- pain has persisted for more than 6 weeks
- is worse at night or in the morning, when an inflammatory arthritis (e.g., ankylosing spondylitis), infection, or a spinal tumour may be the cause
- is associated with a systemic illness, fever, or weight loss
- is associated with neurological symptoms or signs, especially new onset incontinence, which may suggest cauda equina compression.

Investigations

- ***Spinal X-rays*** are required only if the pain is associated with 'red flag' symptoms or signs.
- ***MRI*** (**Figs. 32.8 and 32.9**) is preferable to CT scanning when neurological signs and symptoms are present. CT scans demonstrate bony pathology better. Interpretation of the relevance of the findings may require a specialist opinion.
- ***Bone scans*** are useful in infective and malignant lesions but are also positive in degenerative lesions.
- ***Full blood count, ESR, and biochemical tests*** are required only when the pain is likely to be due to malignancy, infection, or a metabolic cause. Normal ESR and CRP distinguish mechanical back pain from PMR, a likely differential in the elderly.

Mechanical low back pain

Mechanical low back pain starts suddenly, may be recurrent and is helped by rest. It is often precipitated by an injury and may be unilateral or bilateral. It is usually short-lived.

Box 32.12 Causes of pain in the back (lumbar region)

Mechanical
- Trauma
- Muscular and ligamentous pain
- Lumbar spondylosis
- Facet joint osteoarthritis
- Lumbar disc prolapse
- Spinal and root canal stenosis
- Spondylolisthesis
- Disseminated idiopathic skeletal hyperostosis (DISH)
- Fibromyalgia, chronic widespread pain

Inflammatory
- Infective lesions of the spine
- Axial spondyloarthritis/sacroiliitis

Metabolic (see Chapter 33)
- Osteoporotic spinal fractures
- Osteomalacia
- Paget's disease

Neoplastic (see Chapter 41)
- Metastases
- Multiple myeloma
- Primary tumours of bone

Referred pain

Box 32.13 Management of back pain

- Most back pain presenting to a primary care physician needs no investigation
- Mechanical back pain is managed with analgesia, brief rest if necessary, and physiotherapy
- Patients should stay active within the limits of their pain
- Early treatment of the acute episode, advice, and exercise programmes reduce long-term problems and prevent chronic pain syndromes
- Physical manipulation of uncomplicated back pain produces short-term relief and enjoys high patient satisfaction ratings
- Psychological and social factors may influence the time of presentation
- Appropriate early management reduces long-term disability

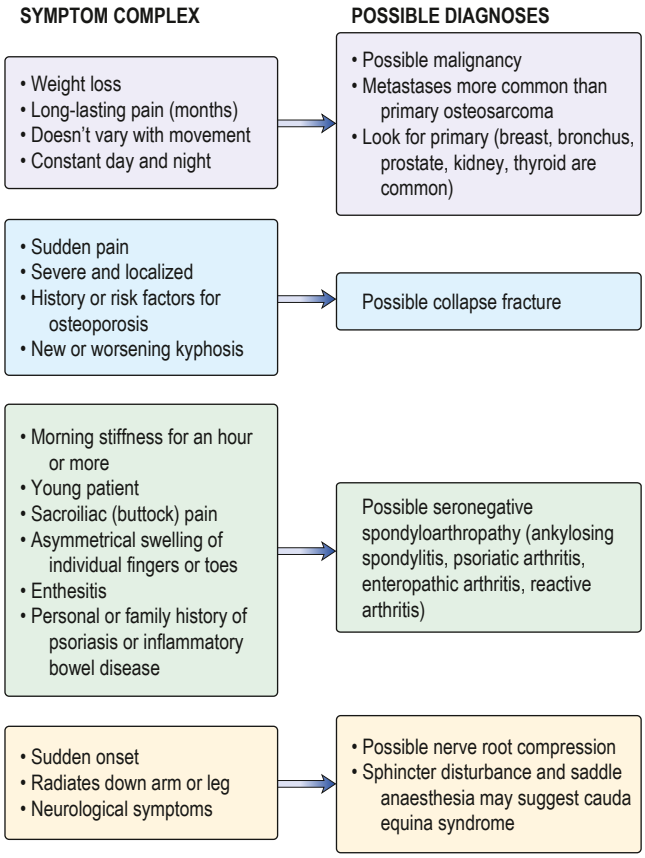


Fig. 32.7 Diagnoses for serious causes of back pain.

Examination

The back is stiff and a scoliosis may be present when the patient is standing. Muscular spasm is visible and palpable, and causes local pain and tenderness. It lessens when sitting or lying. Excessive rest should be avoided. Once a patient develops low back pain, although the episode itself is usually self-limiting, there is a significantly increased risk of further back pain episodes. Risk factors for recurrent back pain include:

- female sex
- increasing age
- pre-existing chronic widespread pain (fibromyalgia)
- psychosocial factors, such as high levels of psychological distress, poor self-rated health, and dissatisfaction with employment.

Chronic low back pain is a major cause of disability and time off work and is reduced by appropriate early management.

Spinal movement occurs at the disc and the posterior facet joints, and stability is normally achieved by a complex mechanism of spinal ligaments and muscles. Any of these structures may be a source of pain. An exact anatomical diagnosis is difficult but some typical syndromes are recognized (see below). They are often associated with, but not necessarily caused by, radiological spondylosis (see Chapter 28).

Postural back pain develops in individuals who sit in poorly designed, unsupportive chairs.

Lumbar spondylosis

The fundamental lesion in spondylosis occurs in an intervertebral disc, a fibrous structure whose tough capsule inserts into the rim of the adjacent vertebrae. This capsule encloses a fibrous outer zone and a gel-like inner zone. The disc allows rotation and bending.

Fig. 32.8 Spinal metastasis in L2. The patient had severe back pain and weight loss, and prior carcinoma of the breast.

Changes in the discs occasionally start in teenage years or early twenties and often increase with age. The gel changes chemically, breaks up, shrinks, and loses its compliance. The surrounding fibrous zones develop circumferential or radial fissures. In the majority, this

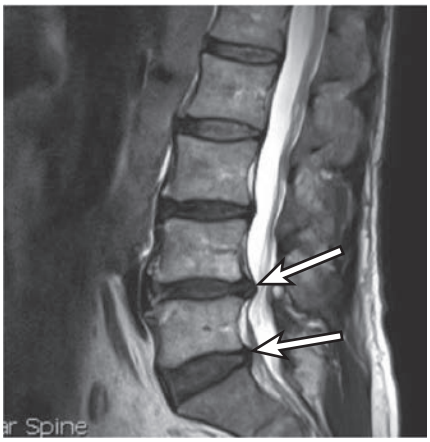


Fig. 32.9 Magnetic resonance image of the lumbar spine. A large disc prolapse is shown at the L4/L5 level and a smaller prolapse at L5/S1 *(arrowed)*. The signal from all the lumbar discs indicates dehydration.

is initially asymptomatic but visible on MRI as decreased hydration. Later, the discs become thinner and less compliant. These changes cause circumferential bulging of the intervertebral ligaments.

Reactive changes develop in adjacent vertebrae; the bone becomes sclerotic and osteophytes form around the rim of the vertebra (see Fig. 32.9). The most common sites of lumbar spondylosis are L5/S1 and L4/L5.

In young people, disc prolapse through an adjacent vertebral end-plate produces a Schmorl's node on X-ray. This process is painless but may accelerate disc degeneration.

Spondylosis may be symptomless but can cause:

- episodic mechanical spinal pain
- progressive spinal stiffening
- facet joint pain
- acute disc prolapse, with or without nerve root irritation
- spinal stenosis
- spondylolisthesis.

Facet joint syndrome

Lumbar spondylosis also causes secondary OA of the misaligned facet joints. Pain is typically worse on bending backwards and when straightening from flexion. It is lumbar in site, unilateral or bilateral, and radiates to the buttock. The facet joints are well seen on MRI and may show OA, an effusion, or a ganglion cyst. Direct corticosteroid injections into the joints under imaging may help but their long-term value is unclear. Physiotherapy to reduce hyperlordosis and weight reduction are helpful.

Management of mechanical back pain

Adequate analgesia to allow normal mobility and avoidance of bed rest is best, combined with physical treatments such as physiotherapy, back muscle training regimens, and manipulation. Manipulation produces more rapid pain relief in some patients. Acupuncture may help. Re-education in lifting and exercises helps to prevent recurrent attacks of pain. Most episodes recover, irrespective of the treatment given. A positive approach probably reduces the development of chronic pain. A comfortable sleeping position should be adopted using a mattress of medium (not hard) firmness.

Acute lumbar disc prolapse

The central disc gel may extrude into a fissure in the surrounding fibrous zone and cause acute pain and muscle spasm. These events are often self-limiting. A disc prolapse occurs when the extrusion extends beyond the limits of the fibrous zone (see Fig. 32.9). The weakest point is posterolateral, where the disc may impinge on emerging spinal nerve roots in the root canal.

Box 32.14 Lumbar nerve root entrapment: symptoms and signs

Nerve root	Sensory changes	Reflex loss	Weakness	Usual disc prolapse
L2	Front of thigh	None	Hip flexion/ adduction	L1/2
L3	Inner thigh and knee	Knee	Knee extension	L2/3
L4	Inner calf	Knee	Knee extension	L3/4
L5	Outer calf	None	Inversion of foot	L4/5
	Upper, inner foot		Dorsiflexion of toes	
S1	Posterior calf	Ankle	Plantar flexion of foot	L5/S1
	Lateral border of foot			

The episode often starts dramatically during lifting, twisting, or bending, and produces a typical combination of low back pain and muscle spasm, and severe, lancinating pains, paraesthesia, numbness, and neurological signs in one leg (rarely both). The back pain is diffuse and usually unilateral, and radiates into the buttock. The muscle spasm leads to a scoliosis that reduces when lying down. The nerve root pain develops with, or soon after, the onset. The site of the pain and other symptoms are determined by the root affected (Box 32.14). A central high lumbar disc prolapse may cause spinal cord compression and long tract signs (i.e., upper motor neurone). Below L2/L3, it produces lower motor neurone lesions.

Examination

On examination, the back often shows a marked scoliosis and muscle spasm. The straight-leg-raising test, while the patient is lying, is positive in a lower lumbar disc prolapse—raising the straight leg beyond 30 degrees produces pain radiating down the leg further than the knee. Slight limitation or pain in the back limiting this movement is seen with mechanical back pain. Pain in the affected leg produced by a straight raise of the other leg suggests a large or central disc prolapse. Look for perianal sensory loss and urinary retention or incontinence, which indicate a cauda equina lesion—a neurosurgical emergency (see Chapter 28). An upper lumbar disc prolapse produces a positive femoral stretch test: pain in the anterior thigh when the knee is flexed in the prone position.

Management

Advise a short period (2–3 days) of bed rest, lying flat for a lower disc but semi-reclining for a high lumbar disc, and prescribe analgesia and muscle relaxants. Once the pain is tolerable, encourage the patient to mobilize and refer them to a physiotherapist for exercises and preventative advice. The investigation of choice is MRI, which identifies the abnormal disc and any compressed nerves. An imaging-guided epidural or nerve root canal injection reduces pain rapidly, although the evidence that it speeds resolution or prevents surgery is unclear. Caudal epidural injections are less effective than lumbar ones. Resuscitation equipment must be available for these procedures. Referral to a surgeon for possible microdiscectomy or hemi-laminectomy is necessary if the neurological signs are severe,

if the pain persists and is severe for more than 6–10 weeks, or if the disc is central. If bladder or anal sphincter tone is affected, it becomes a neurosurgical emergency.

Spinal and root canal stenosis

Progressive loss of disc height, OA of the facet joints, posterolateral osteophytes, and buckling of the ligamentum flavum all contribute to root canal stenosis. These changes cause nerve root pain or spinal root claudication—pain and paraesthesiae in a root distribution brought on by walking and relieved slowly by rest. The associated sensory symptoms, slower recovery when the patient rests, and presence of normal foot pulses distinguish this from peripheral arterial claudication. Severe cervical spondylosis may also produce spinal claudication, often with arm symptoms and signs.

The investigation of choice to identify the location and extent of stenosis is MRI. Spinal canal stenosis at more than one level is often associated with severe spondylosis and/or a congenitally narrow spinal canal. It causes buttock and bilateral leg pain, 'heaviness', paraesthesiae, and numbness when walking. Rest helps, as does bending forwards, a manoeuvre that opens the spinal canal. Specialist surgical advice is necessary.

Spondylolisthesis

This condition occurs in adolescents and young adults when bilateral congenital pars interarticularis defects cause instability and permit a vertebra to slip, with or without preceding injury. Rarely, a cauda equina syndrome develops, with loss of bladder and anal sphincter control, and saddle-distribution anaesthesia. It is diagnosed radiologically and can be seen on plain radiographs, though MRI may be necessary if entrapment of nerve roots or cauda equina syndrome is suspected. Low back pain in adolescents warrants investigation, and spondylolisthesis requires orthopaedic assessment. It needs monitoring during the growth spurt.

A degenerative spondylolisthesis may also develop in older people with lumbar spondylosis and OA of the facet joints.

Diffuse idiopathic skeletal hyperostosis

Diffuse idiopathic skeletal hyperostosis (DISH, or Forestier's disease) affects the spine and extraspinal locations. It causes bony overgrowths and ligamentous ossification, and is characterized by flowing calcification over the anterolateral aspects of the vertebrae. The spine is stiff but not always painful, despite the dramatic changes seen on plain radiographs, which are the imaging investigation of choice. Ossification at muscle insertions around the pelvis produces radiological 'whiskering'. Similar changes occur at the patella and in the feet. It is more common in people with metabolic syndrome (high body mass index [BMI]), diabetes mellitus, hypertension, and dyslipidaemia; see Chapter 37).

Management is with analgesics or NSAIDs for pain, and exercise to retain movement and muscle strength.

Osteoporotic crush fracture of the spine

Osteoporosis is asymptomatic but leads to an increased risk of fracture of peripheral bones, particularly neck of femur and wrist, and thoracic or lumbar vertebral crush fractures. Such vertebral fractures may develop with no or minimal trauma. They may develop painlessly or cause agonizing localized pain that radiates around the ribs and abdomen. Multiple fractures lead to an increased thoracic kyphosis. They cause disability and reduced quality of life. The diagnosis is confirmed by X-rays, showing loss of anterior vertebral body height and wedging, with sparing of the vertebral endplates and pedicles. Bone oedema on MRI indicates that a fracture is recent. An underlying tumour and pathological fracture need to be excluded.

Management

Advise bed rest and analgesia until the severe pain subsides over a few weeks, then gradual mobilization. It may warrant hospitalization, and the prescription of intravenous bisphosphonates or subcutaneous or nasal calcitonin to relieve pain. There may be some residual pain and deformity.

The role of percutaneous vertebroplasty and balloon kyphoplasty remains unclear. Both involve inserting a needle through a pedicle into the affected vertebral body under CT guidance with the aim of stabilizing the fracture. Kyphoplasty involves inflating a balloon filled with methyl methacrylate cement in order to restore vertebral shape. Vertebroplasty is the injection of cement alone, without restoring vertebral shape. Pain relief is usual with both but the risks are higher with vertebroplasty.

Bone density measurement and preventative treatment of osteoporosis are essential (see Chapter 33).

Septic discitis

Septic discitis may cause severe pain and rapid adjacent vertebral destruction. It is seen on MRI and requires urgent neurosurgical referral.

Ankylosing spondylitis

Buttock pain and low back stiffness in a young adult suggest axial spondyloarthritis (e.g., ankylosing spondylitis) (see 'Axial spondyloarthritis' section), especially if they are worse at night and in the morning.

Pain in the hip

'Hip' refers to a wide area between the upper buttock, trochanter, and groin. It is useful to ask the patient to point to the site of pain and its field of radiation (**Box 32.15**).

Osteoarthritis of the hip

OA (see 'Osteoarthritis' section) is the most common cause of hip joint pain in a person over the age of 50 years. It gives rise to pain in the buttock and groin on standing and walking. Stiff hip movements cause difficulty in putting on a sock and may produce a limp. Sudden-onset pain may be associated with an effusion on

? Box 32.15 Pain in the hip: causes

Hip region problems	Main sites of pain
Osteoarthritis of hip	Groin, buttock, front of thigh to knee
Trochanteric bursitis (or gluteus medius tendinopathy)	Lateral thigh to knee
Meralgia paraesthetica	Anterolateral thigh to knee
Referred from back	Buttock
Facet joint pain	Buttock and posterior thigh
Fracture of neck of femur	Groin and buttock
Inflammatory arthritis	Groin, buttock, front of thigh to knee
Sacroiliitis (ankylosing spondylitis)	Buttock(s)
Avascular necrosis	Groin and buttocks
Polymyalgia rheumatica	Lumbar spine, buttocks, and thighs

MRI and can be treated by an ultrasound-guided steroid injection. Severe hip OA is characterized by pain and limitation even at rest and abnormal gait. In severe cases, total hip replacement is the only successful therapy.

Lateral hip pain syndrome: trochanteric bursitis and gluteus medius tendinopathy

This syndrome may be due to trochanteric bursitis and caused by trauma or unaccustomed exercise. It also occurs in inflammatory arthritis. The pain over the trochanter is worse on going up stairs, lying on that side in bed, and crossing the legs. The best management is unclear but exercises help, as may a local corticosteroid injection, although the evidence base for treatment is poor. Surgery is rarely necessary. Lateral hip pain may be referred from the upper lumbar spine. A tear of the gluteus medius tendon at its insertion into the trochanter causes a similar syndrome but does not respond to injection.

Meralgia paraesthetica

This condition causes numbness and burning dysaesthesia (increased sensitivity to light touch) over the anterolateral thigh, and may be precipitated by a sudden increase in weight, an injury, or pelvic surgery. It is usually self-limiting but can be helped by amitriptyline or gabapentin at night.

Fracture of the femoral neck

This fracture usually occurs after a fall, occasionally spontaneously. There is pain in the groin and thigh, weight-bearing is painful or impossible, and the leg is shortened and externally rotated. Occasionally, a fracture is not displaced and remains undetected. X-rays are diagnostic. Anyone with a hip fracture, especially after minimal trauma, should be reviewed for osteoporosis (see Chapter 33).

Avascular necrosis (osteonecrosis) of the femoral head

This condition is uncommon but occurs at any age. (Risk factors are discussed in Chapter 33.) There is severe hip pain. X-rays are diagnostic after a few weeks, when a well-demarcated area of increased bone density is visible at the upper pole of the femoral head. The affected bone may collapse. Early, the X-ray is normal but bone scintigraphy or MRI demonstrates the lesion and shows bone marrow oedema.

Inflammatory arthritis of the hip

Inflammatory arthritis produces pain in the groin and stiffness, which is worse in the morning. RA rarely presents with hip pain, although the hip is involved eventually in severe RA. Ankylosing spondylitis and other seronegative spondyloarthropathies cause inflammatory hip arthritis in younger people.

Polymyalgia rheumatica

Bilateral hip, buttock, and thigh pain and stiffness that are worse in the morning in an elderly patient may be attributable to PMR (see 'Polymyalgia rheumatica' section). Neck and shoulder pain and stiffness are usually also present.

Pain in the knee

The knee depends on ligaments and quadriceps muscle strength for stability. It is frequently injured, particularly during sports. Trauma or overuse of the knee leads to a variety of peri- and intra-articular problems. Some are self-limiting; others require physiotherapy, local corticosteroid injections, or surgery.

? Box 32.16 Pain in the knee: causes

Problems arising within the joint
- Osteoarthritis
- Inflammatory arthritis
- Meniscal tear
- Cruciate ligament tear
- Chondromalacia
- Osteochondritis dissecans
- Spontaneous osteonecrosis of the knee

Problems arising in tissues around the joint

Medial
- Medial ligament strain
- Anserine bursitis

Anterior
- Pre-patellar bursitis
- Infrapatellar bursitis
- Quadriceps tendon enthesitis
- Osgood–Schlatter disease

Posterior
- Popliteal (Baker's) cyst

Other
- Hypermobility syndrome
- Referred from hip joint

Trauma and overuse

Painful knee problems can be divided into those that arise within the joint and those that arise in the soft tissues around the joint (Box 32.16). Only problems arising within the joint are likely to cause accumulation of fluid (effusion). Identification and aspiration of fluid from effusions are relevant diagnostic procedures.

Knee joint effusions

An effusion of the knee causes swelling, stiffness, and pain. The pain is more severe with an acute onset and with increasing inflammation because the capsule that contains the pain receptors is stretched. A full clinical history must include a past medical, family, and drug history.

Examination

A large and tense effusion is easily seen and felt on each side of the patella and in the suprapatellar pouch, and is fluctuant. The effusion delays the patella tapping against the femur when it is pressed firmly and quickly with the knee held straight and relaxed (the 'patellar tap' sign). Small effusions also demonstrate the 'bulge' sign when the patient is lying with the quadriceps relaxed. For this test, apply a gentle sweeping pressure, first to the medial side of the joint and then, watching the medial dimple, to the lateral side. Slightly delayed bulging of the medial dimple indicates fluid in the joint.

Investigations

These are:
- blood tests (urate, blood cultures)
- aspiration (see Fig. 32.2) and examination of the knee effusion.

The basic technique of aspiration is described in Box 32.7. If the fluid obtained is very cloudy, septic arthritis (or occasionally gout) is the likeliest diagnosis. Slightly cloudy or blood-stained fluid that does not clot is likely to indicate pseudogout. Frank blood that clots suggests a haemarthrosis. Fluid should be sent for Gram-stain and culture if sepsis is suspected. It should be sent for polarized light microscopy if looking for crystals (see 'Crystal arthritis' section).

Pain arising from within the knee

Osteoarthritis of the knee

Minor radiographic changes of OA (see Fig. 32.15) are very common in those aged over 50 years and do not usually cause pain, which is more likely to arise from surrounding soft tissues. X-ray appearances and the degree of pain felt in OA of the knee are not closely correlated. Marked valgus, varus, or fixed flexion deformities suggest severe OA. For severe cases, surgery, usually total knee

replacement, is the treatment of choice. In milder cases, pain relief, physiotherapy and, sometimes, intra-articular corticosteroid injections may help.

Inflammatory arthritis of the knee

Monoarthritis of the knee, associated with severe pain and marked redness, may be due to septic arthritis, gout, or pseudogout. A cool, clear, viscous effusion is seen in elderly people with moderate or severe symptomatic OA. RA rarely presents with knee involvement alone, though seronegative spondyloarthritis may do so.

Haemarthrosis of the knee

This condition is caused by:

- trauma: meniscal, cruciate, or synovial lining tear
- clotting or bleeding disorders, e.g., haemophilia, sickle cell disease, or von Willebrand's disease.

Torn meniscus

The menisci are partially attached fibrocartilages that stabilize the rounded femoral condyles on the flat tibial plateaux. In the young, they are resilient, but this decreases with age. They can be torn by an injury, commonly in sports that involve twisting and bending. The history is usually diagnostic. There is immediate medial or lateral knee pain and swelling within a few hours. The affected side is tender. If the tear is large, the knee may lock flexed. The immediate treatment is to apply ice. MRI demonstrates the tear (Fig. 32.10). In most circumstances, especially in active sportspeople, early arthroscopic repair or trimming of the torn meniscus is essential. Surgical intervention reduces recurrent pain, swelling, and locking but not the risk of secondary OA. In older patients with OA plus meniscal tear, surgery is no more effective than physiotherapy alone. The long-term benefit of early repair of tears is not yet known. Post-surgical quadriceps exercises aid a return to sport and other activities.

Torn cruciate ligaments

Torn cruciate ligaments account for around 70% of knee haemarthroses in young people. They often coexist with a meniscal tear. Partial cruciate tears are difficult to diagnose clinically. On flexing the knee to 90 degrees, a torn anterior cruciate allows the tibia to be pulled forwards on the femur. MRI is the investigation of choice. Such injuries need urgent orthopaedic referral, reconstructive surgery usually being necessary in young active adults. There is a significant incidence of secondary OA.

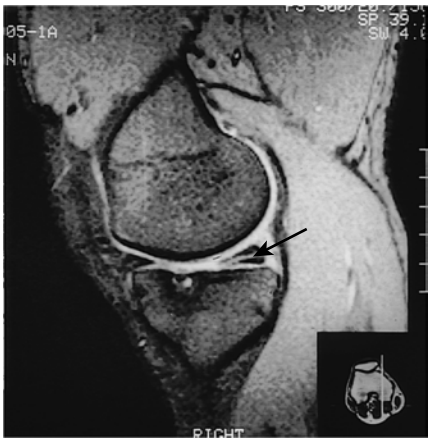


Fig. 32.10 Magnetic resonance image of a knee. A complete tear of the posterior horn of the medial meniscus is shown, extending to its lower surface *(arrowed)*.

Chondromalacia patellae

This diagnosis is made arthroscopically. The retropatellar cartilage is fibrillated. In most cases, the pain settles eventually. When there is patellar misalignment, it may need surgery, as does recurrent patellar dislocation in adolescent females.

Osteochondritis dissecans

This condition occasionally causes knee pain and swelling in adolescents and young adults, more commonly males. It is probably traumatic, possibly with hereditary predisposing factors. A fragment of bone and its attached cartilage detach by shearing, most commonly from the lateral aspect of the medial femoral condyle.

There is aching pain after activity and, if the fragment becomes loose, locking or 'giving way' occurs. The lesion is seen on a tunnel-view X-ray but MRI is more sensitive, especially if the fragment is undisplaced. Undisplaced lesions are treated with rest, then isometric quadriceps exercises. Loose fragments can be fixed arthroscopically or removed. A similar lesion affecting the lateral femoral condyle occurs in older people.

Spontaneous osteonecrosis of the knee

Osteonecrosis may occur spontaneously or after injury. There is local pain and there are marked bone marrow changes on MRI (Fig. 32.11A) or single-photon emission computed tomography (SPECT; Fig. 32.11B). In particular, weight-bearing must be avoided. Pamidronate by infusion is sometimes used. Spontaneous osteonecrosis of the knee (SONK) may progress to bone infarction and require replacement surgery.

Pain from structures around the knee

Medial knee pain

There may be medial or lateral ligament strain but the medial ligament is more commonly affected. There is pain at the ligament's insertion into the upper medial tibia, which is worsened by standing or stressing the affected ligament.

Anserine bursitis causes pain and localized tenderness 2–3 cm below the posteromedial joint line in the upper part of the tibia at the site of the bursa. It occurs in obese females, often with valgus deformities, and in breast-stroke swimmers.

Management is with physiotherapy and a local corticosteroid injection.

Anterior knee pain

Anterior knee pain is common in adolescence. In many cases, no specific cause is found, despite investigation. This is called 'anterior knee pain syndrome' and settles with time. Isometric quadriceps exercises and avoidance of high heels both help the condition. Patients and parents often need firm reassurance. Abnormal patellar tracking may be a cause and may need surgical treatment. Hypermobility of joints causes joint pain, maltracking and, rarely, recurrent patellar dislocation (see also 'Hypermobile Ehlers–Danlos syndrome and hypermobility spectrum disorders' section).

Pre- and infrapatellar bursitis

This can occur in patients whose jobs involve frequent kneeling ('housemaid's knee', 'clergyman's knee', 'carpet-layer's knee'). There is local pain, tenderness, and sometimes swelling. Avoidance of kneeling and a local corticosteroid injection are helpful. Septic bursitis can occur.

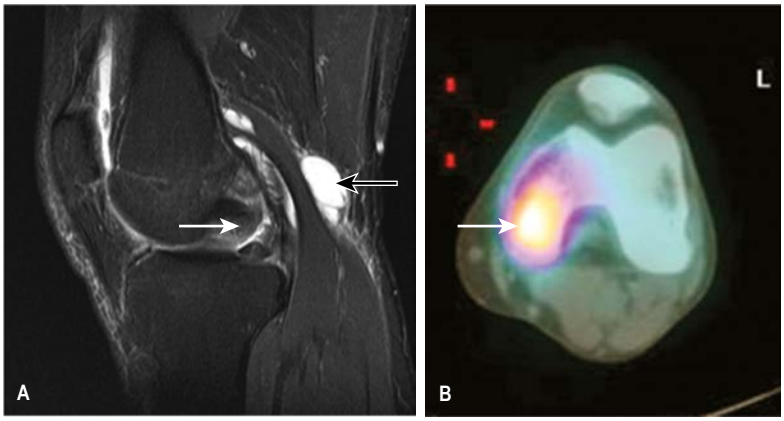


Fig. 32.11 Spontaneous osteonecrosis of the knee. **(A)** Magnetic resonance image showing a high signal in the posterior aspect of the femoral condyle *(white arrow)*, a small effusion and a popliteal cyst *(black arrow)*. **(B)** Single-photon emission computed tomography (SPECT) image showing a high signal in the posterior medial femoral condyle.

Osgood–Schlatter disease

Osgood–Schlatter disease causes pain and swelling over the tibial tubercle. It is a traction apophysitis of the patellar tendon and occurs particularly in teenage sports players.

Enthesitis

This may occur at the patellar end of the quadriceps tendon (jumper's knee).

Posterior knee pain

Popliteal cyst (Baker's cyst)

In approximately 5% of people with a knee effusion, a swollen, painful popliteal cyst develops. The semi-membranosus bursa in some individuals has a valve-like connection to the knee, allowing the effusion to flow into the bursa but not back. The cyst is best seen and felt in the popliteal fossa with the patient standing.

Ruptured popliteal cyst

Fluid escapes into the soft tissue of the popliteal fossa and upper calf, causing sudden and severe pain, swelling, and tenderness of the upper calf.

A history of previous knee problems and the sudden onset of pain and tenderness high in the calf suggest a ruptured cyst rather than a deep vein thrombosis (DVT). A diagnostic ultrasound examination distinguishes a ruptured cyst from a DVT (see Chapter 8), which can avoid inappropriate treatment with anticoagulants. Analgesics or NSAIDs, rest with the leg elevated, and aspiration and injection with corticosteroids into the knee joint are required.

Pain in the shin, calf, and ankle

Sever's disease

This is a traction apophysitis of the Achilles tendon in young people (compare Osgood–Schlatter disease, see above).

Pain at the insertion of the Achilles tendon into the calcaneum is an enthesitis. This is traumatic or it can complicate spondyloarthritis. Raising the shoe heel reduces pain. Occasionally, a low-pressure corticosteroid injection near the enthesis is necessary.

Achilles tendonosis

This causes a painful, tender swelling a few centimetres above the tendon's insertion. Advise against walking barefoot and jumping. Tendon damage or rupture can occur with quinolone, for example, ciprofloxacin therapy. Therapeutic ultrasound is helpful. (*Caution:* a local injection may cause the tendon to rupture.) Autologous platelet concentrates are used but evidence of efficacy is poor. Procedures to remove or sclerose new blood vessels in the inflamed areas may help.

Achilles bursitis

This lies clearly anterior to the tendon and can be safely injected with corticosteroid.

Compartment syndromes

The muscles of the lower leg are enclosed in fascial compartments, with little room for expansion to occur. Compartment syndromes can be acute and severe, such as following exercise.

In ***anterior tibial syndrome*** there is severe pain in the front of the shin, occasionally with foot drop. Immediate surgical decompression to prevent muscle necrosis is sometimes required.

Chronic compartment syndrome produces pain in the lower leg that is aggravated by exercise and may therefore be mistaken for a vascular or neurological disorder.

Pain in the foot

See **Box 32.17**. The feet are subjected to extreme pressures by weight-bearing and inappropriate shoes. They are commonly painful. Broad, deep, thick-soled shoes are essential for sporting activities, prolonged walking or standing, and in people with congenitally flat or arthritic feet.

There are two common types of foot deformity:

- ***Flat feet.*** These stress the ankle and throw the hindfoot into a valgus (everted) position. A flat foot is rigid and inflexible.
- ***High-arched feet.*** These place pressure on the lateral border and ball of the foot.

The foot is affected by a variety of inflammatory arthritic conditions. After the hand, the foot joints are the most commonly affected by RA. The diagnosis depends on assessment of the distribution of the joints affected, the pattern of other joint problems, or the finding of an associated condition (e.g., psoriasis; see Chapter 45).

Hallux valgus

The big toe migrates laterally. In the congenital form, the first metatarsal bone is displaced medially (metatarsus primus varus). The

Box 32.17 Pain in the foot and heel: causes

Foot pain
- Structural—flat (pronated) or high-arched (supinated)
- Hallux valgus/rigidus (± osteoarthritis)
- Metatarsalgia
- Morton's neuroma
- Stress fracture
- Inflammatory arthritis
- Acute, monoarticular—gout
- Chronic, polyarticular—rheumatoid arthritis
- Chronic, pauciarticular—spondyloarthritis
- Tarsal tunnel syndrome

Heel pain
- Plantar fasciitis: below heel
- Plantar spur: below heel
- Achilles tendonitis/bursitis: behind heel
- Sever's disease: behind heel
- Arthritis of ankle/subtaloid joints

shape of modern shoes causes later onset of hallux valgus. It is a common complication of RA.

Hallux rigidus

OA of the first metatarsophalangeal (MTP) joint in a normally aligned or valgus joint causes hallux rigidus: a stiff, dorsiflexed, and painful big toe. Careful choice of footwear and the help of a podiatrist suffice for most cases but some require surgery.

Metatarsalgia

This is common, especially in people who wear high-heeled shoes, after trauma and in those with hammer toes. The ball of the foot is painful to walk and stand on. Callosities and pressure-induced bursae develop under the metatarsal heads. RA causes misalignment of the metatarsal bones and severe metatarsalgia.

Management is with podiatry and the wearing of appropriate shoes. Surgery is occasionally needed, particularly in the rheumatoid forefoot.

Morton's neuroma

This usually occurs between the third and fourth metatarsal heads. It causes pain, burning, and numbness in the adjacent surfaces of the affected toes when walking. It is helped by wearing wider, cushion-soled shoes. Occasionally, a steroid injection under ultrasound guidance or excision is necessary.

Stress fractures of the metatarsals

These cause sudden, severe, weight-bearing pain in the distal shaft of the fractured metatarsal bone. They occur after unaccustomed walking or with new shoes. There is local tenderness and swelling, but initially X-rays are normal and diagnosis delayed (Fig. 32.12). A radioisotope bone scan or MRI reveals the fracture earlier than X-rays. Reduced weight-bearing for a few weeks usually suffices. Underlying osteoporosis may be a cause.

Tarsal tunnel syndrome

This is an entrapment neuropathy of the posterior tibial nerve at the medial malleolus. It produces burning, tingling, and numbness of the toes, sole and medial arch. The nerve is tender below the malleolus and, when tapped, produces a shock-like pain (Tinel's sign). A local steroid injection under the retinaculum, between the medial malleolus and calcaneum, is helpful.

Pain under the heel

See Box 32.17.

Fig. 32.12 Callus forming around a stress fracture. The image shows a stress fracture of the second metatarsal at 4 weeks. X-rays initially were normal.

Plantar fasciitis

This is an enthesitis at the insertion of the plantar fascia into the calcaneum. It produces localized pain under the heel when standing and walking, as well as local tenderness. It occurs alone or in spondyloarthritis. Obesity, particularly in flat-footed people who walk a lot, can predispose to plantar fasciitis.

Plantar spurs

These are traction lesions at the insertion of the plantar fascia in older people and are usually asymptomatic. They become painful after trauma.

Calcaneal bursitis

This is a pressure-induced (adventitious) bursa that produces diffuse pain and tenderness under the heel. Compression of the heel pad from the sides is painful, which distinguishes it from plantar fascia pain.

Clinical features and management of heel pain

Whatever the cause, heel pain is always worse in the morning as soon as weight is placed on the foot.

All these lesions are treated with heel pads, and reduced walking; they are often self-limiting. A dorsiflexion splint at night to stretch the plantar fascia is worth trying. When an injection is necessary, a medial approach is used, rather than advancing through the heel pad, often under ultrasound guidance.

Pain in the chest

Musculoskeletal conditions are sometimes a cause of chest pain. An example is Tietze's disease. In this condition, pain arises from the costosternal junctions. It is usually unilateral and affects one, two, or three ribs. There is local tenderness, which helps to make the diagnosis. The condition is benign and self-limiting. It often responds well to anti-inflammatory drugs. Other causes of chest wall pain include rib fractures due to trauma, osteoporosis, or a malignant deposit. Costochondral pain occurs in ankylosing spondylitis (see 'Ankylosing

spondylitis' section). In people with heart disease, costochondral pain may cause severe anxiety but it is not like angina and the patient should be reassured.

Pain associated with sport and the performing arts

Pain in muscles and soft tissues is common after sport or associated with performing arts, such as dancing or playing musical instruments. General advice, such as warming up properly and using appropriate supportive footwear for running, can help. However, in cases of prolonged pain or where the person suffering pain is a professional sportsperson or performer, referral to a sports medicine specialist is advisable.

Further reading

Chiarotto A, Koes BW. Non-specific low back pain. *N Engl J Med* 2022; 386:1732–1740.
Coombes BK, Bisset L, Vicenzino B. Efficacy and safety of corticosteroid injections and other injections for management of tendinopathy: a systematic review of randomised controlled trials. *Lancet* 2010; 376:1751–1767.
Ho E, Ferreira M, Chen LX et al. Psychological interventions for chronic, non-specific low back pain: systematic review with network meta-analysis. *Br Med J* 2022; 376:e067718.
Klazan CA, Lohle PN, de Vries J et al. Vertebroplasty versus conservative treatment in acute osteoporotic vertebral compression fractures: an open-labelled study. *Lancet* 2010; 376:1085–1092.
Ropper AH, Zafonte RD. Sciatica. *N Engl J Med* 2015; 372:1240–1248.
Whelan G, Yeowell G, Littlewood C. The impact of introducing hydrodistension as a treatment for frozen shoulder in a primary care musculoskeletal service: a retrospective audit. *Musculoskeletal Care* 2023.

CHRONIC PAIN SYNDROMES

Chronic pain is defined as pain lasting more than 3 months (the natural tissue-healing time). Many rheumatological illnesses (e.g., RA and OA) cause pain of such duration but are not considered as chronic pain syndromes. In chronic pain syndromes, the pain generally has no clear structural cause or curative treatment and is often combined with psychological distress, poor sleep, and altered use of the muscles due to the pain (called fear-avoidance behaviour or abnormal pain behaviour). This altered behaviour can lead to more stress and tension in the muscles, exacerbating the pain (Fig. 32.13).

The pain suffered by these patients is neither imaginary nor artefactual, even though there may be no structural abnormality in the painful area. There is a problem with the pain processing system in the nervous system, leading to amplification of pain signals. Psychological factors often contribute to this amplification, and the pain can make psychological distress worse, in a vicious circle. The combination of physical pain and psychological distress makes these syndromes difficult to manage.

Living with chronic pain is difficult. Patients may become anxious, depressed, or socially isolated, and their quality of life is reduced. There are often adverse effects on employment, personal relationships, and dependence on others. Consultations need to address these issues, as well as the pain itself, to be effective. In chronic pain syndromes, patients need help to lead a more normal life despite their pain, and are best referred to a specialist, multidisciplinary pain service. Medications alone are not the answer.

These syndromes can be conveniently subdivided into chronic widespread pain (above and below the waist and on both sides of the body) and chronic regional pain (any other distribution). Both are very common. Epidemiological studies show that the prevalence of chronic widespread pain is 10%–11% and that of chronic regional pain is 20%–30%.

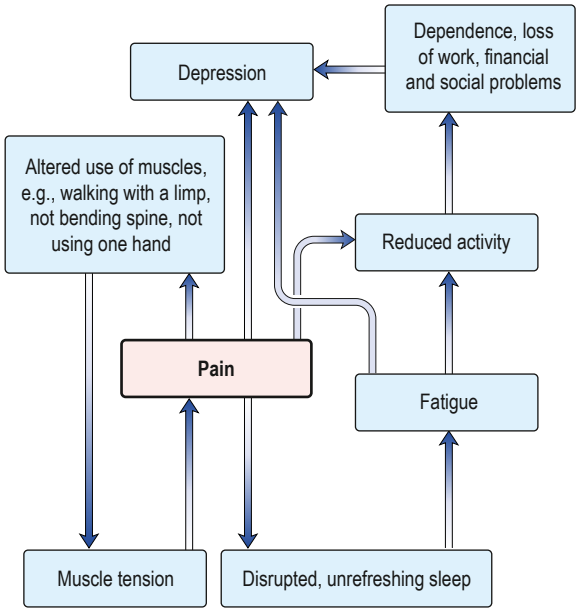


Fig. 32.13 Interaction of physical and psychological factors in chronic widespread pain.

Chronic widespread pain

Pain that starts in a single area can spread to other areas of the body, as more muscles become tense and tender. Conversely, patients with chronic widespread pain can develop new, localized causes of pain (e.g., appendicitis, OA) that can be treated successfully. It should not be assumed that every pain in these patients is always due to the same chronic problem.

Two of the main causes of chronic widespread pain are fibromyalgia and hypermobility spectrum disorders (HSDs; see 'Hypermobile Ehlers–Danlos syndrome and hypermobility spectrum disorders' section).

Fibromyalgia

Fibromyalgia is not a diagnosis of exclusion. It can occur in patients who have other illnesses like RA and SLE. Individuals suffer chronic widespread pain with disrupted and unrefreshing sleep, constant tiredness, and tender points detectable on pressing their muscles (Fig. 32.14). Multiple other symptoms, such as irritable bowel syndrome (IBS), tension headaches, dysmenorrhoea, atypical facial or chest pain, and forgetfulness, often coexist. It occurs in about 1 in 50 people, can develop at any age and affects females more than males (2:1). The diagnosis is clinical, and blood tests and imaging are normal. These tests may be requested to exclude other causes of pain.

Management

See also Chapter 30. A clear explanation of the diagnosis is vital. While being honest about the fact that there is no cure for fibromyalgia, it is also necessary to reassure the patient that it is not arthritis and that the pain is not causing damage to joints or muscles. Many patients have never had an explanation of the cause of their symptoms, which leads to fear and doubt. Treatment options that can be offered to help (not cure) are described below. Management guidelines stress use of non-drug measures instead of or alongside drugs.

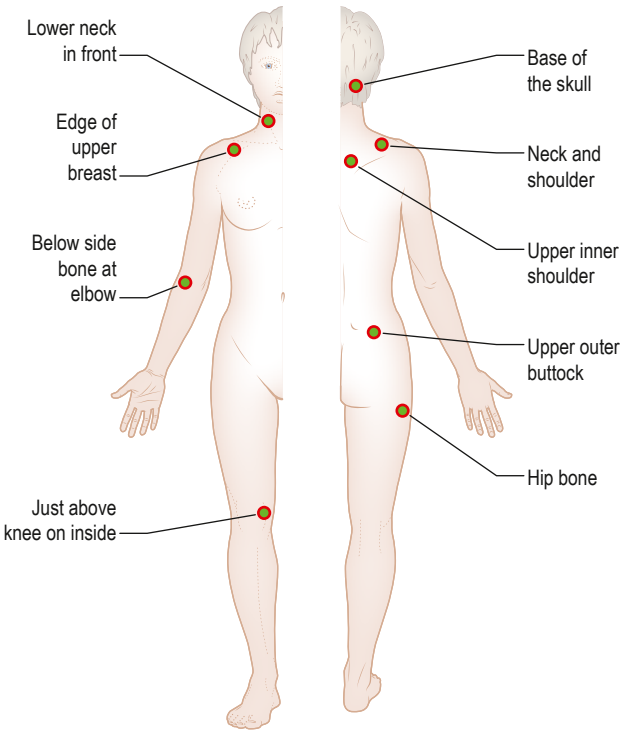


Fig. 32.14 **Trigger points in fibromyalgia.**

Drugs

There is evidence from clinical trials for use of analgesics, anticonvulsants, and antidepressants. Commonly used drugs include amitriptyline (which may also improve sleep), pregabalin, paracetamol, tramadol, duloxetine, and gabapentin. Benefits, however, are often short-term and adverse effects common.

Non-drug therapies

A sympathetic, psychosocial, multidisciplinary approach is appropriate. A graded, supervised aerobic exercise regimen over 3 months is safe and effective. When depression is present, it should be treated. Cognitive behavioural therapy can help the person to pace their life more effectively and to cope better. Pain management programmes, combining psychology with physiotherapy, are designed to improve physical function and quality of life but do not reduce pain intensity. Acupuncture can lessen pain in some cases but the effect is usually transient.

Chronic regional pain

Chronic (work-related) upper-limb pain syndrome

This term is preferred to 'repetitive strain injury' (RSI). The predominant symptoms are pain in all or part of one or both arms. A specific lesion, such as tennis elbow or carpal tunnel syndrome, or muscular-pattern neck pain often develops first, and early recognition and treatment may prevent chronicity. After a variable period, the pain becomes more diffuse and no longer simply work-related, and there is often severe distress. It is seen in keyboard workers and in musicians. When it arises at work, it is often at a time of stress, such as changing work practices, shortage of staff, or disharmony.

Management

If possible, there should be a brief period off work and a gradual return to activity as the pain settles. Use of analgesia and NSAIDs, with physiotherapy, is helpful during the initial phase to prevent a vicious circle developing. Amitriptyline or pregabalin is helpful for some patients.

A review of working practices and the positioning of screen, keyboard, and chair are essential, as is support of the patient by their manager. Musicians are helped by expert advice on playing technique and should reduce playing times temporarily, but not stop completely.

Temporomandibular pain dysfunction syndrome

This pain syndrome is a disorder of the temporomandibular joint that is associated with nocturnal tooth grinding or abnormalities of bite. It occurs most commonly in anxious people. It gives rise to pain in one or both temporomandibular joints.

Dental correction of the bite helps a few, but when no dental cause is found, low-dose tricyclic antidepressant therapy is used. Many patients are made worse by unnecessary dental treatment.

Complex regional pain syndrome

Complex regional pain syndrome (CRPS) is a rare condition (prevalence 20:100 000), in which regional neuropathic pain is associated with abnormal sensory, autonomic, motor, and/or trophic changes. There are two types. In CRPS type I, there is no identifiable overt nerve lesion, whereas such a lesion is identifiable in CRPS type II. CRPS usually occurs after trauma but the pain is disproportionate in time or intensity to that usually caused by such an injury. It may also develop after central nervous system (CNS) lesions (e.g., strokes) or without cause.

Its features are pain and other sensory abnormalities, including hyperaesthesia; and autonomic vasomotor dysfunction, leading to abnormal blood flow and sweating; and motor system abnormalities, leading to structural changes of superficial and deep tissues (trophic changes). Not all components need be present. The sensory, motor, and sympathetic nerve changes are not restricted to the distribution of a single nerve and may be remote from the site of injury. The early phase, with pain, swelling, and increased skin temperature, is difficult to diagnose but potentially reversible.

After a period of weeks or months, a second, still painful, dystrophic phase develops, characterized by articular stiffness, cold skin, and trophic changes, often with localized osteoporosis.

A late phase involves continued pain, skin and muscle atrophy, and muscle contractures, and is extremely disabling.

Diagnosis is clinical—a high index of suspicion and recognition of the unusual distribution of the pain. There are no specific or sensitive investigations, though demineralization may be seen on X-ray. Bone scan and MRI are not required to make the diagnosis.

Management

Management is difficult and the problem is often very disabling. The evidence base for treatment is poor. Early diagnosis, effective pain relief, and general care of the patient are essential. NSAIDs, tricyclic antidepressants, serotonin–norepinephrine reuptake inhibitors, and pregabalin or gabapentin are used in the early phase, together with active exercise of the limb, encouraged by a physiotherapist. Intravenous bisphosphonates can also be effective. Referral to a specialist pain clinic is essential. Nerve blocks were used in the past but are less common now. Intravenous immunoglobulins have been used in clinical trials but are not yet an accepted therapy.

Further reading

Johnson S, Cowell F, Gillespie S, Goebel A. Complex regional pain syndrome what is the outcome?—a systematic review of the course and impact of CRPS at 12 months from symptom onset and beyond. *Eur J Pain* 2022; 26:1203–1220.
Macfarlane GJ, Kronisch C, Dean LE et al. EULAR revised recommendations for the management of fibromyalgia. *Ann Rheum Dis* 2017; 76:318–328.
Rahman A, Underwood, M, Carnes D. Fibromyalgia. *Br Med J* 2014; 348:g1224.
Royal College of Physicians. *The Diagnosis of Fibromyalgia Syndrome*. UK clinical guidelines. London: RCP. 2022.

ANALGESIC AND ANTI-INFLAMMATORY DRUGS FOR MUSCULOSKELETAL PROBLEMS

The key to using drugs, particularly in chronic disorders and the elderly, is to balance risk and benefit and to review their appropriateness constantly. **Box 32.18** shows the main drugs available.

Simple and compound analgesic agents

Simple agents, such as paracetamol, aspirin, or codeine compounds (or combination preparations), used when necessary or regularly, relieve pain and improve function. Sleep may also be improved. Side-effects are relatively infrequent, although drowsiness and constipation occur with codeine preparations, especially in the elderly.

Stronger analgesics, such as dihydrocodeine, tramadol, or morphine derivatives, should be used only for severe pain.

Non-steroidal anti-inflammatory drugs

Non-steroidal anti-inflammatory drugs (NSAIDs) have anti-inflammatory and centrally acting analgesic properties. They inhibit cyclo-oxygenase (COX), a key enzyme in the formation of prostaglandins, prostacyclins, and thromboxanes). There are two specific COX enzymes:

- ***COX-1*** is the constitutive form present in many normal tissues.
- ***COX-2*** is the form mainly induced in response to pro-inflammatory cytokines and is not found in most normal tissues (except the kidney). It is associated with oedema and the nociceptive and pyretic effects of inflammation.

i Box 32.18 Analgesics and NSAIDs

Drug	Dose	Frequency
Analgesics[a]		
To be taken only if needed. Maximum doses are indicated.		
Paracetamol	500–1000 mg	6-hourly
Paracetamol (500 mg) and codeine (8–30 mg)	1–2 tablets	6-hourly
Dihydrocodeine	30–60 mg	Every 6–8 h
Paracetamol with dihydrocodeine (e.g., 'Co-dydramol' preparations)	1–2 tablets	Every 6–8 h
Non-steroidal anti-inflammatory drugs (NSAIDs)		
Always to be taken with food. Slow-release preparations are used in inflammatory conditions or if more regular pain control is needed. Examples are shown.		
Ibuprofen	200–400 mg	Every 6–8 h
Ibuprofen slow-release	600–800 mg	12-hourly
Diclofenac	25–50 mg	8-hourly
Diclofenac slow-release	75–100 mg	×1–2 daily
Naproxen	250 mg	×3–4 daily
Celecoxib[b]	100–200 mg	×2 daily

[a]In order of potency.
[b]COX-2-specific NSAID (coxib).

Effects and side-effects

Most of the older NSAIDs are non-specific and block both enzymes but with variable specificity ('non-specific NSAIDs', or nsNSAIDs). Their therapeutic effect depends on blocking COX-2 and their side-effects mainly on blocking COX-1. COX-1 protects the gastric mucosa and blocking it accounts for the majority of upper gastrointestinal side-effects.

The most common side-effects with non-specific NSAIDs are indigestion or skin rashes. More serious upper gastrointestinal side-effects are gastric erosions and peptic ulceration with perforation and bleeding. These occur more frequently in the elderly, in whom mortality is higher, in long-term use, and in those with high-risk factors: a history of ulcers, *Helicobacter pylori*, and concurrent corticosteroid or anticoagulant therapy. Ibuprofen, in combination with low-dose aspirin, significantly increases the risk of severe gastrointestinal bleeding. Practice guidelines recommend proton pump inhibitors in high-gastrointestinal-risk patients on non-specific NSAIDs. H_2 blockers are less effective as gastroprotective agents. Prostaglandin E_2 analogues, such as misoprostol, reduce ulcer complications and are popular, but may cause nausea and diarrhoea. Lower gastrointestinal side-effects of non-specific NSAIDs are becoming more common.

COX-2 inhibitors ('coxibs') produce fewer gastrointestinal side-effects but these still occur. Coxibs are used in patients who have a high risk of gastrointestinal disease but are avoided in patients with cardiovascular risks. In fact, all NSAIDs can cause an increase in risk of cardiovascular events, such as myocardial infarction or stroke. NSAIDs and coxibs also increase the risk of heart failure.

Coxibs and NSAIDs may reduce renal function, especially in the elderly (see **Box 26.21**).

Uses

- ***In musculoskeletal pain and in OA and spondylosis***, short courses of NSAIDs or coxibs are used but simple analgesia is often more appropriate.
- ***In crystal synovitis***, NSAIDs and coxibs have a true anti-inflammatory effect.
- ***In chronic inflammatory synovitis***, NSAIDs and coxibs do not alter the chronic inflammatory process or decrease the risk of joint damage, but they do reduce pain and stiffness.
- ***In inflammatory arthritis and situations where more constant pain control is needed***, slow-release preparations are useful.
- ***In chronic arthritis***, NSAID gels have some value.

The standard advice is to use NSAID in the lowest dose possible for the shortest time necessary to control pain. Be aware of the patient's gastrointestinal, renal, and cardiac risks before prescribing NSAIDs or coxibs.

Further reading

Coxib and Traditional NSAID Trialists Collaboration. Vascular and upper gastrointestinal effects of non-steroidal anti-inflammatory drugs: meta-analyses of individual participant data from randomized trials. *Lancet* 2013; 382:769–779.
Schmidt M, Sorensen HT, Pedersen L. Diclofenac use and cardiovascular risks: series of nationwide cohort studies. *BMJ* 2018; 362:k3426.

Box 32.19 The major types of collagen

Collagen structure	Type number	Encoding gene	Associated conditions
Fibrillar	I, II, III, V, XI	*COL1A1–2, COL2A2, COL3, COL5, COL11*	Osteogenesis imperfecta, Ehlers–Danlos syndrome (subtypes)
Basement membrane	IV	*COL4A1–5*	Alport's syndrome (see Chapter 26)
Fibril-associated collagen with interrupted triple helix (FACIT)	IX, XII, XIV	*COL9, 12, 14*	
Filament-producing	VI	*COL6A1–3*	
Network-forming	VIII, X	*COL8A1, 10A1*	
Anchoring fibril	VII	*COL7A1*	Epidermolysis bullosa (see Chapter 45)

Box 32.20 Beighton hypermobility score[a] and diagnostic criteria for joint hypermobility syndrome

Joint	Finding	Score
Little (fifth) finger	Passive dorsiflexion >90 degrees	1 for each side
Thumb	Passive dorsiflexion to the flexor aspect of the forearm	1 for each side
Elbow	Hyperextends >10 degrees, extends ≤10 degrees	1 for each side
Knee	Hyperextends >10 degrees, extends ≤10 degrees	1 for each side
Forward flexion of trunk with the knees fully extended	Palms can rest flat on the floor	1

[a]In this 9-point system, the higher the score, the higher the laxity. Young adults can score 4–6. Diagnostic criteria for joint hypermobility syndrome: Beighton score ≥4; arthralgia for ≥3 months in ≥4 joints.

DISORDERS OF COLLAGEN

Collagen is responsible for many of the structural, tensile, and load-bearing properties in the various tissues where it is found. Thirty or more dispersed genes encode for more than 19 different types of collagens (**Box 32.19**).

Hypermobile Ehlers–Danlos syndrome and hypermobility spectrum disorders

Hypermobility is very common (15%–20% of adults); it may be demonstrated by showing increased flexibility in the thumbs, little fingers, elbows, knees, and lumbar spine, and quantified using the Beighton score (**Box 32.20**). The majority of hypermobile people suffer no adverse effects from the hypermobility but in some it can cause recurrent subluxations of individual joints and/or persistent widespread musculoskeletal pains. According to the 2017 classification, these people are classified into two groups; hypermobile Ehlers–Danlos syndrome (hEDS) for those who fulfil specific classification criteria and hypermobility spectrum disorder (HSD) for those who do not. Both groups may suffer other symptoms such as bowel disturbance, easy scarring, and faintness on standing (postural orthostatic tachycardia syndrome; POTS), and the combination can be very disruptive to normal life. Treatment of the musculoskeletal symptoms relies on specialist physiotherapy to advise on exercises that take the patient's hypermobility into account. Pain management programmes can be helpful in more severe cases. It is usually better to avoid surgery due to difficulties in healing.

Other forms of Ehlers–Danlos syndrome

Whereas hEDS is the most common form of EDS and has no known genetic basis, there are 12 rarer forms that have all been linked to particular genes encoding collagen or collagen modifiers. These rarer forms are named according to their clinical phenotype for example classical EDS, classical-like EDS, cardiac valvular EDS, vascular EDS, and kyphoscoliotic EDS. Classical, vascular, arthrochalasia, and periodontal EDS are autosomal dominant whereas the others are autosomal recessive (myopathic EDS can be either).

Vascular EDS can be particularly severe and the most common form is associated with an autosomal dominant mutation in the *COL3A1* gene. These patients may suffer rupture of arteries, the sigmoid colon, or the uterus and have hypermobility in small rather than large joints.

Marfan's syndrome

This condition is described in Chapter 17.

Osteogenesis imperfecta

This is a heterogeneous group of disorders inherited mainly in autosomal dominant fashion with mutations in *COL1A1* and *COL1A2* genes. There are four main types of osteogenesis imperfecta and clinical subtypes are also described (V, VI, and VII). The major clinical feature is bone fragility but other collagen-containing tissues are also involved, such as tendons, skin, and eyes.

- ***Type I:*** mild bony deformities, blue sclerae, defective dentine, early-onset deafness, hypermobility of joints and heart valve disorders
- ***Type II:*** death in the perinatal period
- ***Type III:*** severe bone deformity and blue sclerae
- ***Type IV:*** fewer fractures, normal sclerae, normal lifespan but can also be severe, as in type III.

Management with daily oral risedronate in children improves BMD and reduces fracture risk. Intravenous pamidronate is another option. Prognosis is variable, depending on the severity of the disease. Stem cell therapy is being used. Types I and IV are associated with normal lifespan and people with these forms continue to have increased fracture risk into adulthood.

Achondroplasia

Achondroplasia ('dwarfism') is diagnosed in the first years of life. The disease is inherited in an autosomal dominant manner and is caused by a defect in the fibroblast growth factor receptor-3 gene. The trunk is of normal length but the limbs are very short and broad due to abnormal endochondral ossification. The vault of the skull is enlarged, the face is small and the nose bridge is flat. Intelligence is normal.

Further reading

Castori M, Tinkle M, Levy H et al. A framework for the classification of joint hypermobility and related conditions. *Am J Med Genet C Semin Med Genet* 2017; 175C:148–157.
Malfait F, Francomano C, Byers P et al. The 2017 international classification of the Ehlers-Danlos syndromes. *Am J Med Genet C Semin Med Genet* 2017; 175C:8–26.

OSTEOARTHRITIS

Osteoarthritis (OA) is the most common type of arthritis. It results from a loss of articular cartilage induced by an interaction of genetic, metabolic, and biomechanical factors, affecting cartilage, subchondral bone, ligaments, menisci, synovium, and capsule. Despite intense investigation, disease-modifying therapies remain elusive for this condition.

Epidemiology

Globally approximately 1 in 5 females and 1 in 10 males aged over 60 years have symptomatic OA, such that 80% have limitations in movement and 25% cannot perform major daily activities of life. Worldwide it has a variable pattern of presentation. For instance, hip OA is less common and knee OA is more common in Asians than in Europeans. Beyond 55 years of age, females are affected more commonly than males by nodal and primary generalized forms of OA, with strong familial predisposition. Resulting disabilities have major socioeconomic resource implications, particularly in the developed world where OA is one of the ten most disabling diseases in older adults.

Aetiology and pathogenesis

Cartilage is a matrix of water (>70%) and organic extracellular matrix components, mainly type II collagen and aggrecan or other proteoglycans; it has a smooth surface and shock-absorbing properties. OA is now considered a disease of the whole joint with alterations in subchondral bone, ligaments, capsule, synovial membrane as well as articular cartilage. Consequent structural changes include surface fibrillation and ulceration with loss of cartilage that exposes underlying bone to increased stress, producing microfractures and cysts leading to abnormal sclerotic subchondral bone and overgrowths at the joint margins, called osteophytes (Fig. 32.15). Inflammatory changes may also occur. This process produces a spectrum of OA, ranging from atrophic disease in which cartilage destruction occurs without any subchondral bone response, to hypertrophic disease with massive new bone formation at the joint margins.

Various mechanisms have been suggested:

- ***Abnormal stress and loading***, leading to mechanical cartilage damage.
- ***Obesity*** is a risk factor for developing OA of the hand and knee and thought to trigger metabolic inflammation via adipokines, released from adipose tissue, inducing pro-inflammatory cytokines that cause cartilage and bone damage.
- ***Matrix degradation*** by matrix metalloproteinase (MMP)-1, MMP-2, MMP-3, MMP-9, and MMP-13, which are secreted by chondrocytes in an inactive form. Extracellular activation then leads to the degradation of both collagen and proteoglycans around chondrocytes.
- ***Tissue inhibitors of metalloproteinases*** (TIMPs) regulate the MMPs. Disturbance of this regulation may lead to an increase in cartilage degradation over synthesis and contribute to the development of OA.

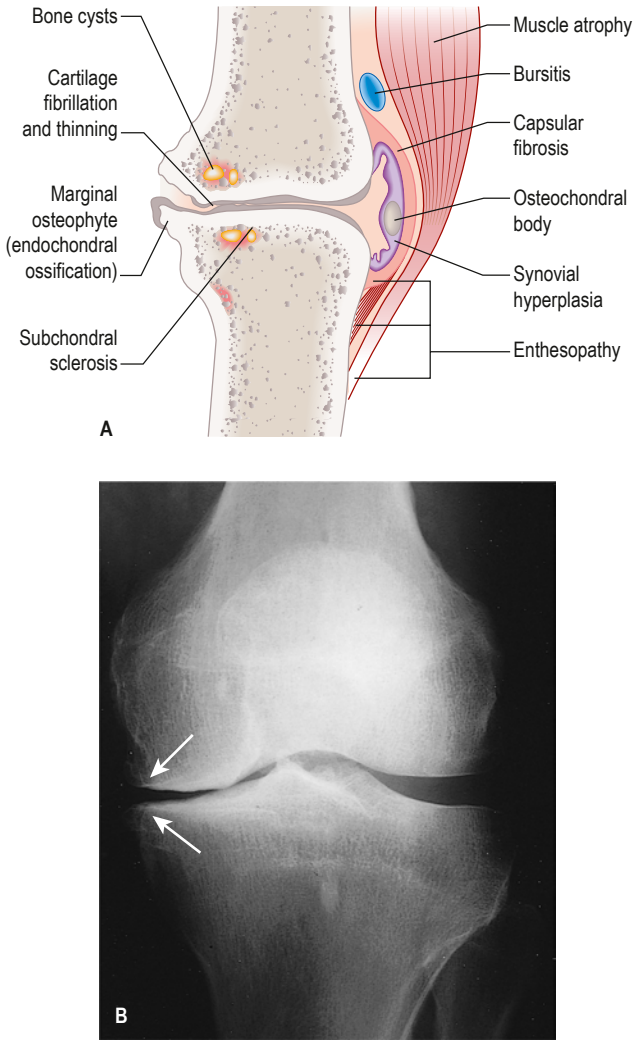


Fig. 32.15 Early osteoarthritis of a knee. (A) Pathology. (B) On the X-ray, there is medial compartment narrowing owing to cartilage thinning with subarticular sclerosis and marginal osteophyte formation *(arrowed)*.

- ***Osteoprotegerin (OPG), RANK, and RANK ligand (RANKL)*** control subchondral bone remodelling. Their levels are significantly different in OA chondrocytes.
- ***Aggrecanase*** production is stimulated by pro-inflammatory cytokines, and aggrecan (the major proteoglycan) levels fall.
- ***Inflammatory mediator release*** (including interleukin-1, IL-1 and tumour necrosis factor-alpha, TNF-α) stimulates metalloproteinase production, and IL-1 inhibits type II collagen production. IL-6 and IL-8 may also be involved. The production of cytokines by macrophages and that of MMPs by chondrocytes in OA are dependent on the transcription factor nuclear factor kappa B (NF-κB).
- ***Growth factors***, including insulin-like growth factor 1 (IGF-1) and transforming growth factor beta (TGF-β), are involved in collagen synthesis, and their deficiency may play a role in impairing matrix repair. Paradoxically, increased TGF-β may also cause increased subchondral bone density.
- ***Cell derived and/or cartilage breakdown products***, such as alarmins, fibronectin fragments, hyaluronic acid fragments, collagen fragments, proteoglycan fragments, and high mobility group box proteins, act to further deregulate chondrocyte function and contribute to cartilage degradation.

- ***Vascular endothelial growth factor (VEGF)*** from macrophages is a potent stimulator of angiogenesis and may contribute to inflammation and neovascularization in OA. Innervation can accompany vascularization of the articular cartilage.
- ***A strong hereditary element*** underlying OA is suggested by twin studies. The influence of genetic factors is estimated at 35%–65%. To date, genome-wide association studies (GWAS) have identified several common variants associated with knee or hip OA, although the individual risk alleles exert only moderate to small effects. Loci that are associated with OA include genes encoding components of the TGF-β and bone morphogenetic protein signalling pathways: growth/differentiation factor 5 (GDF5); type II iodothyronine deiodinase (DIO2), which regulates synthesis of triiodothyronine that is involved in cartilage maintenance and repair; proteins involved in apoptosis and mitochondrial damage; molecules that regulate the turnover of extracellular matrix components; and proteins that are associated with inflammation and immune responses.
- ***Sex.*** In females, weight-bearing sports produce a two- to threefold increase in risk of OA of the hip and knee. In males, there is an association between hip OA and certain occupations: farming and labouring. OA may flare after the female menopause or after cessation of hormone replacement therapy.
- ***Periarticular enthesitis*** is recognized in nodal generalized OA (NGOA) and may be difficult to distinguish from psoriatic arthritis at DIP joints.

The term ***primary OA*** may be used when various predisposing factors (**Box 32.21**), occur in the absence of other conditions that may cause secondary OA (**Box 32.22**).

Box 32.21 Factors predisposing to osteoarthritis (OA)

- Obesity: This predicts a later risk of radiological and symptomatic OA of the hip and hand in population studies
- Heredity: There is a familial tendency to develop nodal and generalized OA
- Gender: Polyarticular OA is more common in females; a higher prevalence after the menopause suggests a role for sex hormones
- Hypermobility (see 'Hypermobile Ehlers–Danlos syndrome and hypermobility spectrum disorders' section): Increased range of joint motion and reduced stability lead to OA
- Diseases: See **Box 32.22**
- Trauma: A fracture through any joint predisposes. Meniscal and cruciate ligament tears cause OA of the knee
- Congenital joint dysplasia: This alters joint biomechanics and leads to OA. Mild acetabular dysplasia is common and leads to earlier onset of hip OA
- Joint congruity: Congenital dislocation of the hip or a slipped femoral epiphysis or Perthes' disease predispose; osteonecrosis of the femoral head (see Chapter 33) in children and adolescents causes early-onset OA
- Occupation: Miners develop OA of the hip, knee and shoulder, cotton workers OA of the hand, and farmers OA of the hip
- Sport: Repetitive use and injury in some sports cause a high incidence of lower-limb OA

Box 32.22 Causes of osteoarthritis

Primary OA
- No known cause

Secondary OA

Pre-existing joint damage
- Rheumatoid arthritis
- Gout
- Spondyloarthritis
- Trauma
- Overuse/abnormal use, e.g., athletes
- Septic arthritis
- Paget's disease
- Avascular necrosis, e.g., corticosteroid therapy

Metabolic disease
- Cartilage calcification
- Hereditary haemochromatosis
- Acromegaly

Systemic disease
- Haemophilia—recurrent haemarthrosis
- Haemoglobinopathies, e.g., sickle cell disease
- Neuropathies

Clinical features

OA affects many joints, in diverse clinical patterns, typically causing mechanical pain with movement and/or loss of function (**Fig. 32.16**). Hip and knee OA are major causes of disability. Early OA is rarely symptomatic, however, unless accompanied by a joint effusion, while advanced radiological and pathological OA is not always symptomatic.

Symptoms are usually gradual in onset and progressive. Episodic disease flare-ups may be inflammatory in nature, although typically ESR and/or CRP are normal. Radiological OA is not inevitably progressive. Radiological improvement is uncommon but has been observed, suggesting that repair is possible.

Symptoms

- Joint pain with movement and/or weight-bearing
- Short-lived morning joint stiffness
- Functional limitation

Signs

- Crepitus
- Restricted movement
- Bony enlargement
- Joint effusion and variable levels of inflammation
- Bony instability and muscle wasting

Clinical subsets

Localized OA

Nodal OA

In nodal OA (**Box 32.23**), joints of the hand are usually affected one at a time over several years, with the DIPs more frequently involved than the PIPs. Nodal OA often begins around the female menopause, with inflammation causing painful, tender, swollen interphalangeal joints and impairment of hand function. At this stage, enthesitis can be seen on MRI. Intra-articular corticosteroid injections may be helpful at this stage. The inflammatory phase settles after some months or years, leaving painless bony swellings posterolaterally: ***Heberden's nodes (DIPs)*** and ***Bouchard's nodes (PIPs)***, along with stiffness and deformity (**Fig. 32.17**). Functional impairment is usually limited, although PIP OA restricts grip more than DIP involvement. On X-ray, the nodes are marginal osteophytes and there is joint space loss.

Thumb-base OA coexists with nodal OA causing pain and disability, which decrease as the joint stiffens. The 'squared' hand in OA (see **Fig. 32.17**) is caused by bony swelling of the first carpometacarpal joint and fixed adduction of the thumb. Function is rarely severely compromised.

Polyarticular hand OA is associated with a slightly increased frequency of OA at other sites.

Hip OA

Hip OA affects 7%–25% of adult white people but is significantly less common in black and Asian populations. There are two major subgroups defined by the radiological appearance. The most common is ***superior-pole hip OA***, where joint space narrowing and

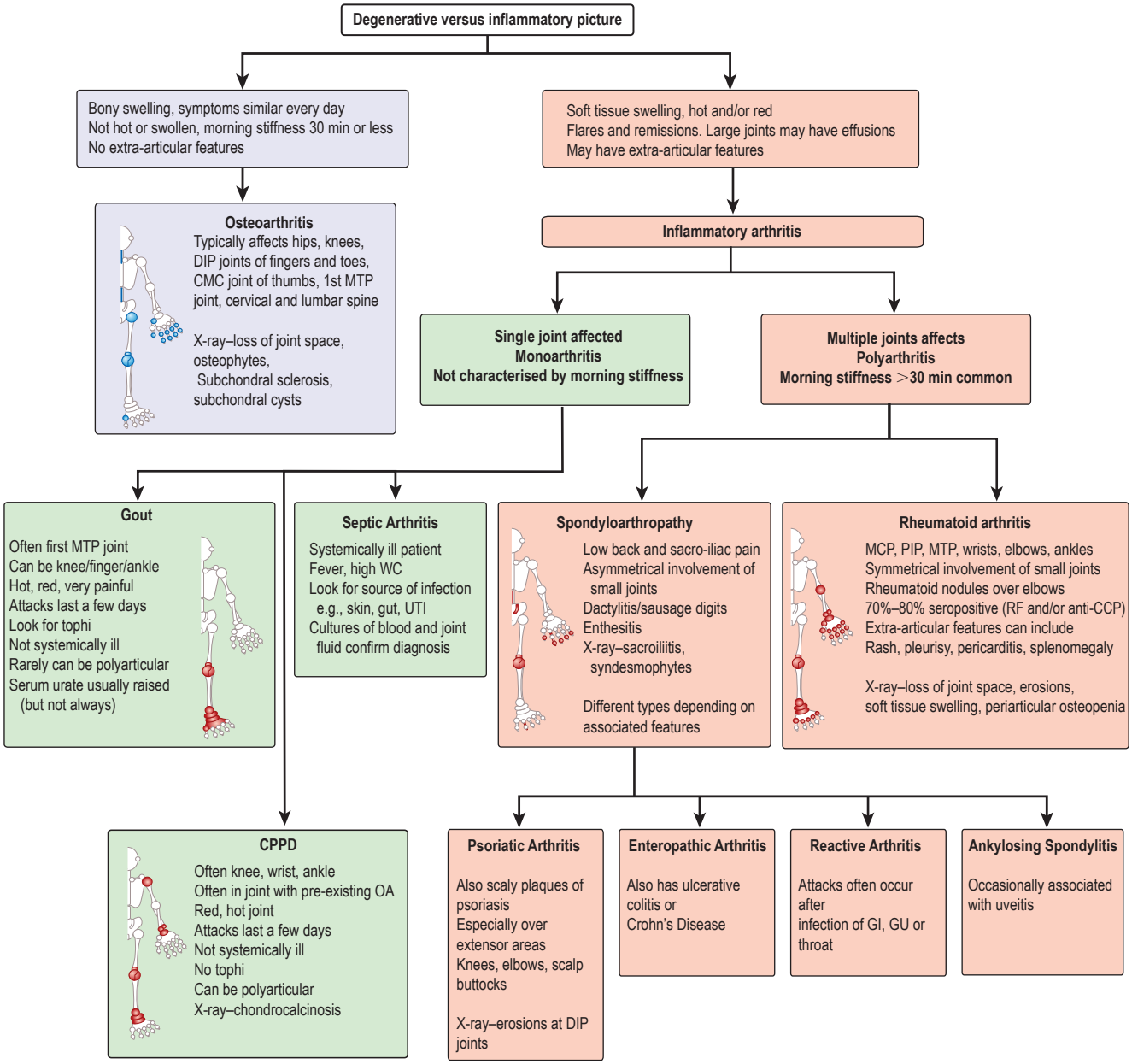


Fig. 32.16 Patterns of joint involvement in degenerative and inflammatory arthritis. *CMC,* Carpometacarpal; *CPPD,* calcium pyrophosphate deposition; *DIP,* distal interphalangeal; *MCP,* metacarpophalangeal; *MTP,* metatarsophalangeal; *PIP,* proximal interphalangeal; *RF,* rheumatoid factor; *WC,* white cell count.

i Box 32.23 Features of nodal osteoarthritis

- Demonstrates a familial nature
- Has a higher incidence in females
- Shows a typical pattern of polyarticular involvement of the hand joints
- Develops in late middle age and around female menopause
- Has a generally good long-term functional outcome
- Is associated with osteoarthritis of the knee, hip, and spine (nodal generalized osteoarthritis)

sclerosis predominantly affect the weight-bearing upper surface of the femoral head and adjacent acetabulum. This finding is most common in males and unilateral at presentation, although both hips may become involved in progressive disease. Early onset of hip OA is associated with acetabular dysplasia or labral tears. Less commonly, ***medial cartilage loss*** occurs. This appearance is seen most commonly in females and is associated with hand involvement (NGOA), and is usually bilateral. It is more rapidly disabling.

Knee OA

The prevalence of symptomatic knee OA is 40% in individuals over 75 years of age and is more common in females and black compared with white people. There is a strong relationship with obesity. The disease is generally bilateral and strongly associated with nodal OA of the hand in elderly females, or as part of generalized OA. The medial compartment is most commonly affected, leading to a varus (bow-legged) deformity. Often, retropatellar OA is also present. Previous trauma and meniscal and cruciate ligament tears are

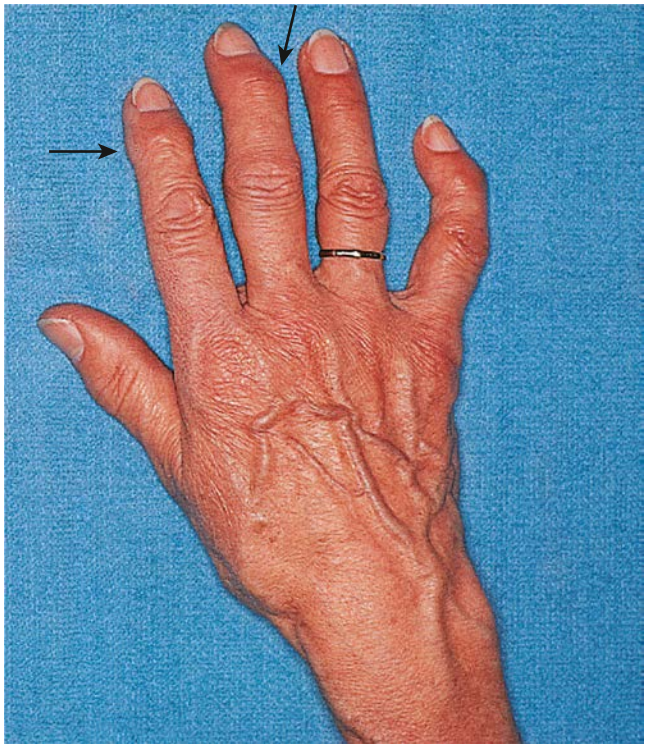


Fig. 32.17 Severe nodal osteoarthritis. The distal interphalangeal joints (DIPs) demonstrate Heberden's nodes *(arrowed)*. The middle finger DIP joint is deformed and unstable. The thumb is adducted and the bony swelling of the first carpometacarpal joint is clearly shown—'the squared hand of nodal osteoarthritis'.

risk factors for developing knee OA. Bone marrow lesions seen on MRI predict disease progression and eventual joint replacement.

Primary generalized OA

This condition is rare and usually seen in combination with NGOA. Other affected areas include the knees, first MTP, hip, and intervertebral (spondylosis) joints. Its onset is often sudden and severe. There is a female preponderance and a strong familial tendency. Periarticular ligamentous pathology may have an important role in the phenotypic expression of NGOA.

Erosive OA

In this rare subgroup, the DIPs and PIPs are inflamed and equally affected, with a poor functional outcome. Radiologically, there is marked osteolysis. Destructive phases are followed by phases of remodelling.

Crystal-associated OA

This condition most commonly occurs with calcium pyrophosphate deposition (CPPD) in the cartilage. It increases in frequency with age and causes cartilage calcification (CC) on over 40% of knee X-rays in the over-eighties, but is usually asymptomatic. The joints most frequently affected are the knees (hyaline cartilage and fibrocartilage) and wrists (triangular fibrocartilage); see Fig. 32.16. There is patchy linear CC on X-ray (Fig. 32.18).

A chronic arthropathy (pseudo-OA) occurs, predominantly in elderly females with severe CC. There is a florid inflammatory component and marked osteophyte and cyst formation visible on X-rays that predominantly affects the knees, then wrists and shoulders. CC is associated with CPPD crystal-induced arthritis (see 'Calcium pyrophosphate dihydrate deposition arthropathy' section).

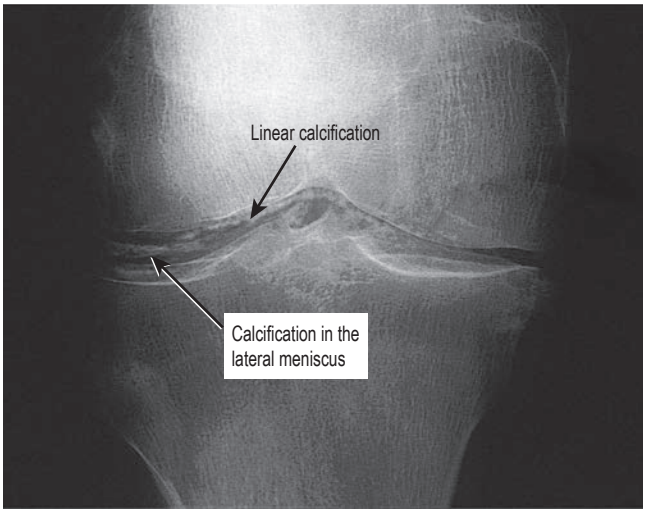


Fig. 32.18 Cartilage calcification of the knee. Note the linear calcification in the hyaline cartilage and calcification of the lateral meniscus (plus mild secondary osteoarthritis).

A rare, rapidly destructive arthritis in elderly females, affecting shoulders, hips, and knees, is associated with the finding of crystals of calcium apatite in a bloody joint effusion. The outlook is poor and joints require early surgical replacement.

Investigations

- ***Blood tests*** lack specificity; the ESR is usually normal, although high-sensitivity CRP may be slightly raised. RFs and antinuclear antibodies are negative.
- ***X-rays*** are abnormal only when the damage is advanced. They are useful in preoperative assessments. For knees, a standing X-ray (stressed) is used to assess cartilage loss, and 'skyline' views in flexion are used for patello-femoral OA.
- ***MRI*** demonstrates meniscal tears, early cartilage injury, and subchondral bone marrow changes (osteochondral lesions).
- ***Arthroscopy*** reveals early fissuring and surface erosion of the cartilage.
- ***Aspiration of synovial fluid*** if there is a painful effusion shows a viscous fluid with few leucocytes (see 'Joint aspiration and examination of synovial fluid' section).

Management

The guiding principle is to treat the symptoms and disability, not the radiological appearances; depression and poor quadriceps strength are better predictors of pain than radiological severity in OA of the knee. Patient education about the disease and its effects reduces pain, distress, and disability, and increases compliance with treatment. Psychological or social factors alter the impact of the disease.

Physical measures

Weight management is recognized to reduce the risk of development of OA and slow its progression. Females who lost an average of 5 kg decreased their risk for knee OA by 50% in the Framingham study. As little as 1% change in body weight in obese adults modifies the rate of knee cartilage loss. Overall, maintenance of muscle mass and reduction of adiposity is central to OA management. Exercises for strength and stability are useful and land-based exercise programmes for the hip and knee improve physical function and pain. Hydrotherapy helps, especially in lower-limb OA. Insoles for flat feet and a walking stick held on the contralateral side to the

affected lower limb joint are useful. Transcutaneous electrical nerve stimulation, acupuncture, and thermotherapy may be adjuncts for treating OA but are not universally recommended due to their limited evidence of efficacy.

Medication

Topical, oral, and injectable therapies are available and potential benefit must be balanced against potential side-effects, especially in the elderly. First-line therapies include topical or oral NSAIDs. Paracetamol has limited efficacy in OA (see **Box 32.18**). NSAIDs or coxibs should be used intermittently when possible due to their toxicities. Topical rubefacients, such as capsaicin, may be used as a supplementary analgesic. Opioids are a last resort and should be used cautiously in older patients.

Intra-articular corticosteroid injections may produce short-term (~6 weeks) improvement when there is a painful joint effusion in elderly patients unsuitable for joint replacement, but symptoms are likely to return. There are concerns of possible cartilage degradation and worsening symptoms with repeat injections over 2 years. Heterogeneity of relevant studies of intra-articular hyaluronan preparations displays limited efficacy.

Glucosamine and chondroitin (sold as food supplements) have no clinically relevant effect on joint pain or joint space narrowing.

There are no proven agents that halt or reverse OA, although they are greatly needed.

Surgery

Arthroscopy for knee OA is rarely beneficial. Replacement arthroplasty, however, has transformed the management of severe OA to reduce pain, improve function, restore quality of life, and regain independence. Increasing numbers of arthroplasties, particularly hip and knee, are being performed worldwide each year with shortened recovery times, improved durability and mortality. However, an increasing trend for hip or knee arthroplasty in patients less than 60 years of age is associated with an increased lifetime risk of revision surgery. Serious complications occur in less than 1%, with venous thromboembolism and infection being the most serious. These slight but definite risks make it essential for the patient to be certain that surgery is necessary. Resurfacing hip surgery has become popular but may have higher complication rates in women, particularly with metal-on-metal resurfacing. Uni-compartmental knee replacement is a less major procedure and may be appropriate in some cases.

Other surgical procedures include re-alignment osteotomy of the knee or hip, excision arthroplasty of the first MTP and base of the thumb, and fusion of a first MTP joint.

Further reading

Martel-Pelletier J, Barr AJ et al. Osteoarthritis. *Nat Rev Dis Primers* 2016; 2:16072.
Katz JN, Arant KR, Loeser RF. Diagnosis and treatment of hip and knee osteoarthritis. *JAMA* 2021; 325:568.
Orchard JW. Is there a place for intra-articular corticosteroid injections in the treatment of knee osteoarthritis? *BMJ* 2020; 368:l6923.

INFLAMMATORY ARTHRITIS

Inflammatory arthritis is characterized by synovial inflammation. The three main subgroups of inflammatory arthritis are rheumatoid arthritis (RA), spondyloarthritis, and crystal arthritis (**Box 32.24**). The diagnosis of these conditions is helped by distinguishing:

- the number and pattern of joint involvement (symmetrical or asymmetrical, large or small) (**Box 32.25** and **Fig. 32.16**)
- presence of any non-articular disease
- a past and family history
- periodicity of the arthritis (single acute, relapsing, chronic, and progressive).

Box 32.24 The three main subgroups of inflammatory arthritis

1. Rheumatoid arthritis (associated with antibodies)
2. Spondyloarthritis (associated with human leucocyte antigen (HLA)-B27)
3. Metabolic arthritis (e.g., associated with crystals)

Box 32.25 Pattern of joint involvement in inflammatory arthritis

Monoarthritis
- Crystal arthritis
- Septic arthritis
- Palindromic rheumatism
- Traumatic ± haemathrosis
- Juxta-articular bone tumour
- *PsA*
- *ReA*
- *RA*

Oligoarthritis
- ReA
- Palindromic rheumatism
- Crystal arthritis
- Septic arthritis
- *PsA*
- *ReA*

Polyarthritis
- RA
- PsA, AxSpA
- ReA
- Postviral
- Lyme arthritis
- EA
- Arthritis associated with EN

AxSpA, Axial spondyloarthropathy; *EA*, enteropathic arthritis; *EN*, erythema nodosum; *PsA*, psoriatic arthritis; *RA*, rheumatoid arthritis; *ReA*, reactive arthritis. Italics indicate less common pattern of joint involvement.

Certain non-articular diseases, such as psoriasis, iritis, inflammatory bowel disease, non-specific urethritis, or recent dysentery, suggest spondyloarthritis. There may be evidence of recent viral illness (rubella, hepatitis B, or erythrovirus), rheumatic fever, or a tick bite and skin rash (Lyme disease). In early arthritis, it may not be possible to make a specific diagnosis until the disease has evolved from an undifferentiated arthritis into a chronic form.

There is a distinct genetic separation of rheumatoid-pattern synovitis and spondyloarthritis; RA (see below) is associated with a genetic marker in the class II major histocompatibility complex (MHC) genes, while spondyloarthritis shares certain alleles in the B locus of class I MHC genes, usually B27 (see 'Spondyloarthritis' section).

In general, the pain and stiffness of inflammatory arthritis are worse in the morning, often lasting for several hours and improving with activity, in contrast with the much shorter morning stiffness and mechanical pain with activity of OA. Inflammatory markers (ESR and CRP) are typically raised in inflammatory arthritis. Specific types of arthritis are discussed below.

Early inflammatory polyarthritis

Undifferentiated polyarthritis requires urgent referral to a rheumatologist for diagnosis and treatment, including the early introduction of disease-modifying agents when indicated (see 'Disease-modifying anti-rheumatic drugs' section). In persistent inflammatory arthritis, sustained remission depends on rapid diagnosis and intensive treatment. Poor prognostic features for undifferentiated polyarthritis are:

- polyarticular onset
- positive anti-citrullinated peptide antibodies (ACPAs)
- positive rheumatoid factor
- joint erosion on X-ray at presentation
- disease for longer than 3–6 months.

RHEUMATOID ARTHRITIS

RA is a chronic inflammatory rheumatic disease. It has a multifactorial pathogenesis, with various genetic and environmental factors being implicated, giving rise to immune dysregulation with consequent joint inflammation and tissue damage. Patients typically develop a chronic symmetrical polyarthritis with systemic inflammation.

Epidemiology

RA has a worldwide distribution affecting 0.5%–1% of the population (with a female preponderance of 3:1). The prevalence is low in black people of African background and Chinese people. The incidence is falling. RA is the 42nd highest contributor to global disability just below malaria and just above iodine deficiency. It presents from early childhood (when it is rare) to late old age. The most common age of onset is between 30 and 50 years.

Aetiology and pathogenesis

There has been a greater understanding of genetic and environmental factors in the last two decades.

- ***Genetic factors.*** There is an increased incidence in first-degree relatives and a high concordance among monozygotic twins of up to 15%. Over 100 loci are associated with disease risk and progression. There is a strong association between susceptibility to RA and certain human leucocyte antigen (HLA) haplotypes: HLA-DR4, which occurs in 50%–75% of patients and correlates with a poor prognosis, as does possession of certain shared alleles of HLA-DRB1*04. The possession of these shared epitope alleles in HLA-DRB1 (S2 and S3P) increases susceptibility to RA and may predispose to anti-cyclic citrullinated peptide antibody (ACPA) development against citrullinated antigens. Citrullination is a process that modifies antigens, allowing them to fit into the shared epitope on HLA alleles. A genome-wide association study in ACPA-positive RA found an association with loci near HLA-DRB1 and PTPN22 in people of European descent. These genes affect the presentation of autoantigens *(HLA-DRB1)*, T-cell receptor signal transduction *(PTPN22)*, and targets of ACPA *(PAD14)*.
- ***Environment.*** Smoking and other forms of bronchial stress increase the risk of RA with HLA-DR4 and act synergistically with HLA-DRB1 to increase the risk of having ACPA. Mucosal surfaces of the oral cavity, upper respiratory tract, and gut are colonized by commensal microorganisms; a population known as the microbiome. Alterations in the microbiome may facilitate developments in innate and adaptive immunity that predispose to RA, such as the association of periodontal disease and alterations in the gut microbiome with RA.
- ***Autoantibodies.*** Rheumatoid factor (RF), autoantibodies to the Fc portion of immunoglobulin G, are found in 75%–80% of RA patients and are relatively specific for diagnosis of RA compared with other forms of inflammatory arthritis but have little predictive value in the general population. In contrast, ACPA is more specific and sensitive to RA.

Immunology

RA is primarily a synovial disease, and synovitis occurs when chemoattractants produced in the joint recruit circulating inflammatory cells. Over-production of tumour necrosis factor-alpha (TNF-α) leads to synovitis and joint destruction. Interaction of macrophages and T and B lymphocytes drives this over-production. TNF-α stimulates over-production of IL-6, as well as other cytokines. The increased understanding of the immunopathogenesis of this disease has informed the development of targeted biological therapies (**Fig. 32.19**).

Dysfunction of certain cell types is significant in immunopathogenesis:

- ***Synovial cells*** in chronic rheumatoid synovitis are predominantly fibroblast-like synoviocytes and macrophage-like synoviocytes that produce pro-inflammatory cytokines.
- ***Osteoclasts*** cause bone and cartilage destruction.
- ***Synovial B cells***, activated by cytokine-activated macrophages and T cells, produce autoantibodies, of which IgM and IgA RF are the most typical in RA. As RFs bind the Fc portion of IgG, they have the potential for self-aggregation and immune complex formation in the synovium. These may then trigger macrophages via IgG Fc receptors to produce even more cytokines, including IL-1, IL-8, TNF-α, and granulocyte–macrophage colony-stimulating factor, and fibroblasts to produce IL-6.
- ***Synovial fibroblasts*** have high levels of the adhesion molecule, vascular cell adhesion molecule (VCAM-1, a molecule that supports B-lymphocyte survival and differentiation), decay accelerating factor (DAF, a factor that prevents complement-induced cell lysis), and cadherin II (which mediates cell-to-cell interactions). These molecules may facilitate the formation of ectopic lymphoid tissue in synovium.
- ***T cells*** can be a part of the destructive process. The pro-inflammatory cytokine IL-17 is produced by a specific subset, Th17 helper cells (see Chapter 6), which produce IL-17, 21 and 22, and TNF-α. The normal regulatory T cells are suppressed by TGF-β and interleukins (produced by macrophages and dendritic cells), allowing Th17 helper cells to increase.

The ***triggering antigen***, which leads to self-maintained inflammation in RA, remains unclear. Triggers for ACPA production include filaggrin, type II collagen, and vimentin. There is little evidence that type II collagen is the triggering antigen, although it is a cause of arthritis in animal models of RA. Smoking is a potential trigger, particularly in ACPA-positive RA.

Pathology

RA is typified by widespread, persistent synovitis of joints, tendon sheaths, or bursae. Normal synovium comprises a thin lining layer that contains fibroblast-like synoviocytes and macrophages overlying loose connective tissue. In RA, the synovium becomes greatly thickened, causing 'boggy' swelling around joints and tendons, with proliferation of the synovium into folds and fronds, and infiltration by a variety of inflammatory cells, including polymorphs, which transit through the tissue into the joint fluid, and lymphocytes and plasma cells. The normally sparse surface layer of lining cells becomes hyperplastic and thickened (**Fig. 32.20**). There is marked vascular proliferation. Increased permeability of blood vessels and the synovial lining layer leads to joint effusions containing lymphocytes and dying polymorphs.

The hyperplastic synovium spreads from the joint margins to the cartilage surface. This 'pannus' of inflamed synovium damages the underlying cartilage by blocking its normal route for nutrition and by the direct effects of cytokines on the chondrocytes. The cartilage becomes thinned and the underlying bone exposed. Local cytokine production and joint disuse combine to cause juxta-articular osteoporosis during active synovitis.

Fibroblasts from the proliferating synovium also grow along the course of blood vessels between the synovial margins and the epiphyseal bone cavity, and damage the bone. This process is

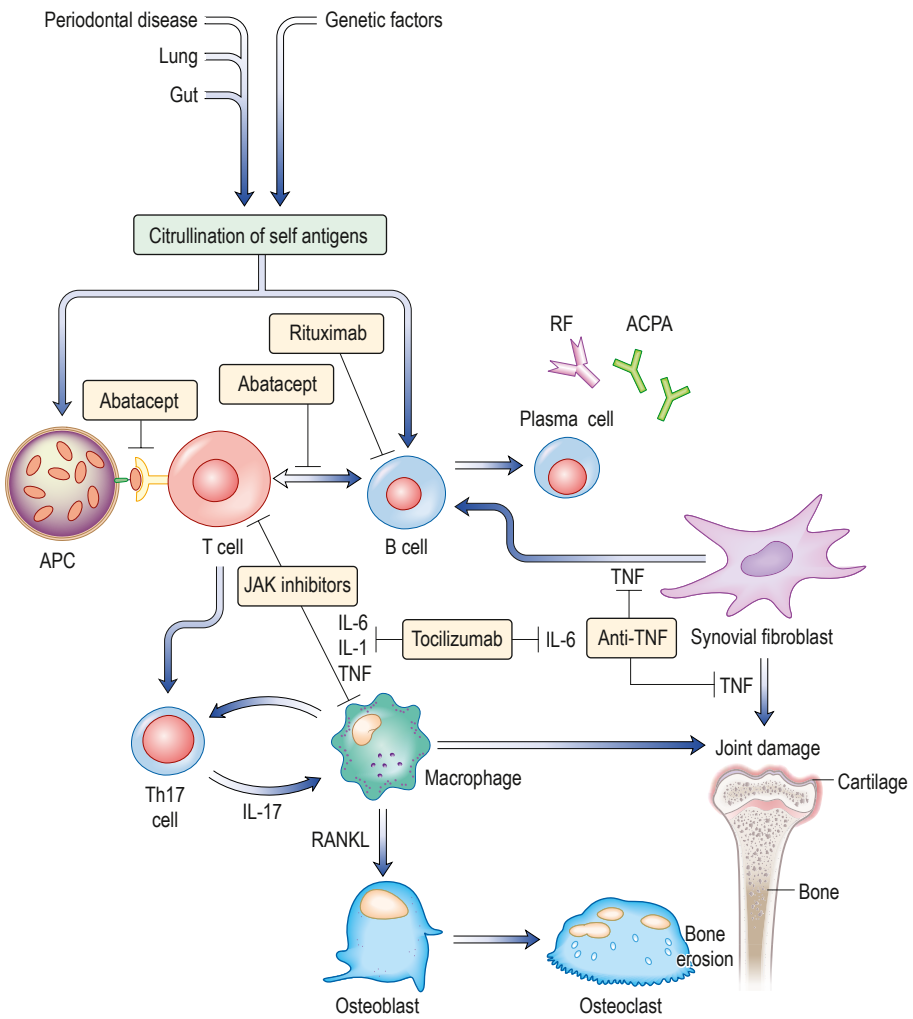


Fig. 32.19 Pathogenesis of rheumatoid arthritis. Environment–gene interactions promote citrullination of self-proteins, which can then be detected by T and B cells; this leads to a loss of tolerance and promotion of the inflammatory response, resulting in joint damage. Targeted therapy is shown in yellow. *ACPA,* Anti-citrullinated peptide antibody; *APC,* antigen-presenting cell; *IL,* interleukin; *RANKL,* receptor activator of nuclear factor kappa B ligand; *RF,* rheumatoid factor; *Th,* T helper; *TNF,* tumour necrosis factor.

shown by MRI (Fig. 32.21) to occur in the first 3–6 months following onset of the arthritis before the diagnostic, ill-defined, juxta-articular bony 'erosions' appear on X-ray (Fig. 32.22). This early damage justifies the use of disease-modifying anti-rheumatic drugs (DMARDs; see 'Disease-modifying anti-rheumatic drugs' section) within 3 months of onset of the arthritis to try to induce disease remission and prevent erosion formation. Erosions lead to a variety of deformities and contribute to long-term disability.

Rheumatoid factors and anti-citrullinated peptide antibodies (ACPAs)

RFs (see 'Serum autoantibody studies' section) are of any immunoglobulin class (IgM, IgG, or IgA), but the most common tests employed clinically detect IgM RF. The term 'seronegative RA' is used when the standard tests for IgM RF are persistently negative. These patients tend to have a more limited pattern of synovitis.

IgM RF is not diagnostic of RA and its absence does not rule the disease out; however, it is a useful predictor of prognosis. A persistently high titre in early disease implies more persistently active synovitis, more joint damage leading to greater disability, and thus justifies earlier use of DMARDs.

ACPAs (see 'Serum autoantibody studies' section) are usually present with RF in RA. They are better predictors of a transition from early transient inflammatory arthritis to persistent synovitis and early RA. RF and ACPA together are even more specific.

Clinical features

Typical presentation

RA typically presents (≈70%) as a progressive, symmetrical, peripheral polyarthritis, evolving over a period of a few weeks or months in patients between 30 and 50 years of age, although the disease can occur at any age (Fig. 32.16). Less commonly (≈15%), a rapid onset can occur over a few days (or explosively overnight), with a severe symmetrical, polyarticular involvement, especially in the elderly. Factors indicating a poor prognosis are listed in Box 32.26. The differential diagnosis of early RA is shown in Box 32.27.

Current classification criteria from 2010 are more suitable for assessing and diagnosing early arthritis than previous versions

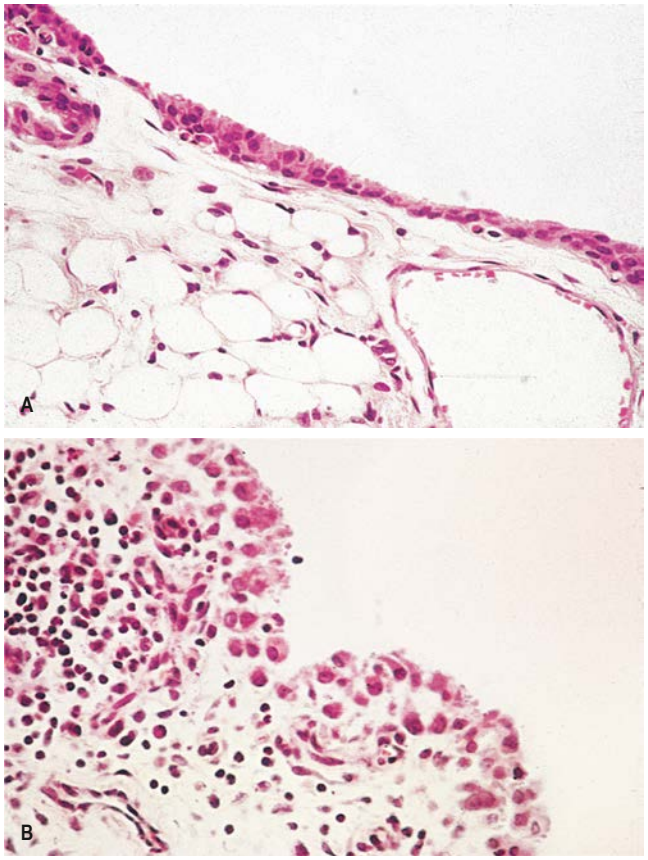


Fig. 32.20 Histological appearance of synovium in rheumatoid arthritis (RA). **(A)** Normal synovium. **(B)** Synovial appearances in established RA, showing marked hypertrophy of the tissues with infiltration by lymphocytes and plasma cells. (*From Shipley M.* Colour Atlas of Rheumatology, *3rd edn. London: Mosby–Wolfe; 1993, with permission.*)

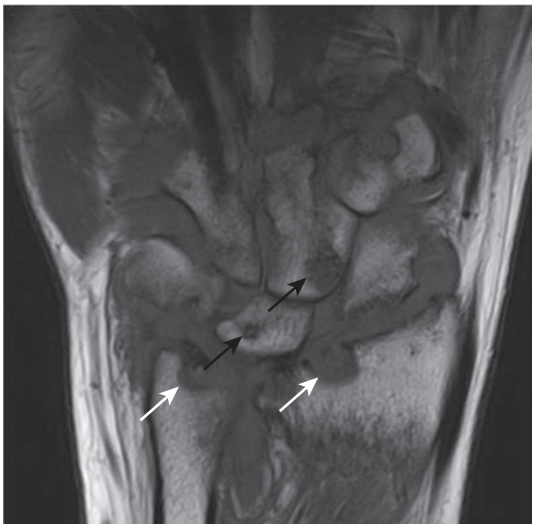


Fig. 32.21 Magnetic resonance imaging of erosions at the wrist in rheumatoid arthritis. Typical erosions may be seen at the ulnar styloid, radial head *(white arrows)*, and carpal bones *(black arrows)*. (*From Elias-Jones C.* Crash Course Orthopaedics and Rheumatology, *3rd edn. Edinburgh: Elsevier, 2015.*)

Fig. 32.22 X-ray of early rheumatoid arthritis. Typical erosions may be seen at the thumb *(black and white arrows on the left)*, middle metacarpophalangeal joints *(central black arrow)*, and ulnar styloid *(white arrow on the right)*.

Box 32.26 Factors predicting a poor prognosis for progression in early rheumatoid arthritis

- Older age
- Female sex
- Symmetrical small joint involvement
- Morning stiffness >30 min
- >4 swollen joints
- Cigarette smoking
- Co-morbidity
- C-reactive protein >20 g/dL
- Positive rheumatoid factor and anti-citrullinated peptide antibodies

Box 32.27 Differential diagnosis of early rheumatoid arthritis

- Postviral arthritis: rubella, hepatitis B, or erythrovirus
- Seronegative spondyloarthropathies
- Polymyalgia rheumatica
- Acute nodal osteoarthritis (proximal and distal interphalangeal joints involved)

because they do not rely on later changes, such as erosions and extra-articular disease to distinguish RA (**Box 32.28**).

In early RA, the combination of at least one swollen joint for more than 6 weeks with no prior injury, no associated history or family history of spondyloarthritis or associated conditions such as psoriasis (see Chapter 45), and a positive ACPA test is the best way to select patients for earlier treatment to avoid joint damage. This

Box 32.28 ACR/EULAR 2010 criteria for rheumatoid arthritis[a]

Criteria	Points
1. Joint involvement	**0–5**
1 medium to large joint	0
2–10 medium to large joints	1
1–3 small joints (large joints not counted)	2
4–10 small joints (large joints not counted)	3
>10 joints; at least one small joint	5
2. Serology	**0–3**
Negative RF and negative ACPA	0
Low-positive RF or low-positive ACPA	2
High-positive RF or high-positive ACPA	3
3. Acute-phase reactants	**0–1**
Normal CRP and normal ESR	0
Abnormal CRP or abnormal ESR	1
4. Duration of symptoms	**0–1**
<6 weeks	0
≥6 weeks	1

[a]The cut-off point for RA is at ≥6 points. Patients can also be classified as having RA if they have both typical erosions and longstanding disease previously satisfying the classification criteria.
ACPA, Anticitrullinated peptide antibody; *ACR*, American College of Rheumatology; *CRP*, C-reactive protein; *EULAR*, European League Against Rh.
Adapted from Scott DL, Wolfe F, Huizinga TW. Rheumatoid arthritis. Lancet *2010; 376: 1094–1108.*

evidence-based approach has been shown to reduce the risk of the development of damage and permanent joint deformities.

Symptoms and signs of early RA

Most patients complain of pain and stiffness of the small joints of the hands (MCPs, PIPs) and feet (MTPs). The DIPs are usually spared. Wrists, elbows, shoulders, knees, and ankles are also affected. In most cases, multiple joints are involved, but 10% of patients present with a monoarthritis of the knee or shoulder or with carpal tunnel syndrome. The pain and stiffness are significantly worse in the morning. Sleep disturbance and fatigue are common complaints.

The joints are usually warm and tender with some joint swelling. There is limitation of movement and muscle wasting. Deformities and non-articular features develop if the disease cannot be controlled (see below).

RA in older patients may mimic PMR; the synovitis becomes apparent as the corticosteroid dose is reduced.

Other presentations

The variable presentations and progression of RA are shown in **Box 32.29**. Relapses and remissions occur either spontaneously or on drug therapy. Active disease, producing progressive joint damage, requires targeted treatment to induce remission.

Seronegative RA initially affects the wrists more often than the fingers and has a less symmetrical joint involvement. It has a better long-term prognosis but some cases progress to severe disability. This form can be confused with psoriatic arthropathy, which has a similar distribution (see 'Psoriatic arthritis' section).

Palindromic rheumatism is unusual (5%) and consists of 24–48-hour episodes of acute monoarthritis. The joint becomes acutely painful, swollen, and red but resolves completely. Further attacks occur in the same or other joints. About 50% of patients develop typical chronic rheumatoid synovitis after a delay of months

Box 32.29 Presentations of rheumatoid arthritis

- ***Palindromic.*** Palindromic monoarticular attacks last 24–48 h; 50% progress to other types of RA.
- ***Transient.*** Disease is self-limiting, lasting <12 months and leaving no permanent joint damage. It is usually seronegative for IgM RF and ACPA. Some of these patients may have undetected postviral arthritis.
- ***Remitting.*** There is a period of several years during which the arthritis is active but then remits, leaving minimal damage.
- ***Chronic, persistent.*** This is the most typical form; it may be seropositive or seronegative for IgM RF. The disease follows a relapsing and remitting course over many years. Seropositive (plus ACPA) patients tend to develop greater joint damage and long-term disability. They warrant earlier and more aggressive treatment with disease-modifying agents.
- ***Rapidly progressive.*** The disease progresses remorselessly over a few years and leads rapidly to severe joint damage and disability. It is usually seropositive (plus ACPA), has a high incidence of systemic complications, and is difficult to treat.

ACPA, Anti-citrullinated peptide antibody; *RF*, rheumatoid factor.

or years. The rest remit or continue with acute episodic arthritis. Detection of RF or ACPA predicts conversion to chronic, destructive synovitis.

Complications

Septic arthritis (see also Chapter 11)

This serious complication has significant morbidity and mortality. In immunosuppressed patients, affected joints may not display the typical signs of inflammation with accompanying fever found in patients with an intact immune system. Treatment is with systemic antibiotics and drainage.

Amyloidosis

Amyloidosis (see Chapter 26) is found in a very small number of people with uncontrolled RA. RA is the most common cause of secondary AA amyloidosis. AL amyloidosis causes a polyarthritis that resembles RA in distribution and is also often associated with carpal tunnel syndrome and subcutaneous nodules.

Joint involvement in RA

These changes are seen in established disease or when early drug treatment has been ineffective.

Hands and wrists

In early disease, the fingers are swollen, painful, and stiff. Inflamed flexor tendon sheaths increase functional impairment and may cause a carpal tunnel syndrome. Joint damage causes:

- ***A combination of ulnar drift and palmar subluxation of the MCPs.*** This change (**Fig. 32.23**) leads to unsightly deformity, but function may be preserved once the patient has learned to adapt and pain is controlled.
- ***Fixed flexion*** (buttonhole or boutonnière deformity) or ***fixed hyperextension*** (swan-neck deformity) of the PIP joints, which impairs hand function.
- ***Swelling and dorsal subluxation of the ulnar styloid***, which causes wrist pain. It may also cause rupture of the finger extensor tendons, leading to a sudden drop of the little and ring fingers that requires urgent surgical repair.

Shoulders

RA commonly affects the shoulders. Initially, the symptoms mimic rotator cuff tendonosis (see 'Rotator cuff [supraspinatus] tendonosis'

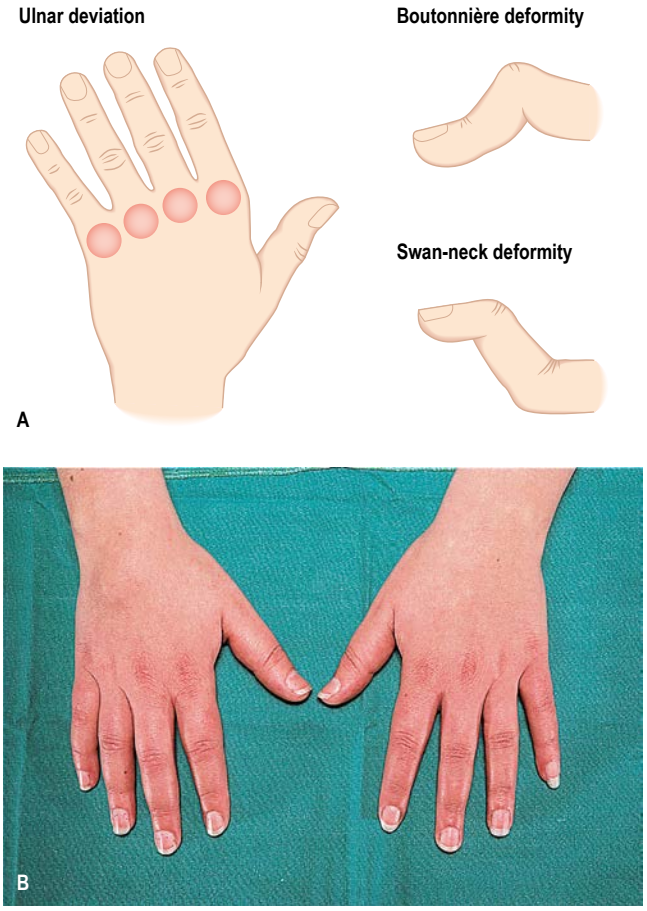


Fig. 32.23 Rheumatoid arthritis. (A) Characteristic hand deformities. (B) Early rheumatoid arthritis—dorsal tenosynovitis of the right wrist and small joints of both hands with spindling of the fingers.

section) with a painful arc syndrome and pain in the upper arms at night. Further joint damage leads to global stiffening; rotator cuff tears become more common and interfere with dressing, feeding, and personal toilet.

Elbows

Synovitis of the elbows causes swelling and a painful fixed flexion deformity. In late disease, flexion may be lost and severe difficulties with feeding result, especially combined with shoulder, hand, and wrist deformities.

Feet

One of the earliest manifestations of RA is painful swelling of the MTP joints.

- The foot becomes broader and a hammer-toe deformity develops.
- Exposure of the metatarsal heads to pressure by the forward migration of the protective fibrofatty pad (Fig. 32.24) causes pain.
- Ulcers or calluses may develop under the metatarsal heads and over the dorsum of the toes.
- Mid- and hindfoot RA causes a flat medial arch and loss of flexibility of the foot.
- The ankle often assumes a valgus position.

Appropriate broad, deep, cushioned shoes are essential but rarely wholly adequate, and walking is often painful and limited. Podiatry helps and surgery may be required.

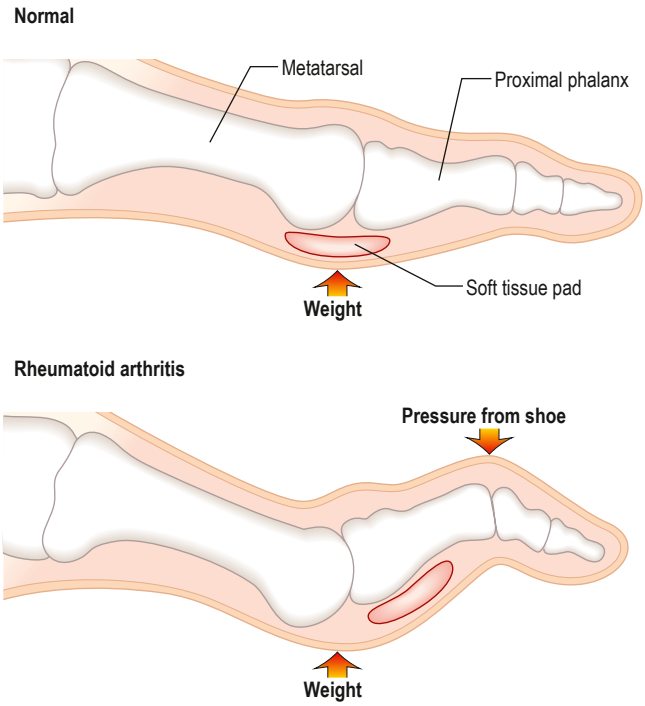


Fig. 32.24 The toes in rheumatoid arthritis. There is exposure of the metatarsal heads with forward migration of the soft tissue pad.

Knees

Massive synovitis and knee effusions occur but respond well to aspiration and steroid injection. A persistent effusion increases the risk of popliteal cyst formation and rupture (see 'Popliteal cyst [Baker's cyst]' section). In later disease, erosion of cartilage and bone causes loss of joint space on X-ray and damage to the medial and/or lateral and/or retropatellar compartments of the knees. Depending on the pattern of involvement, the knees may develop a varus or valgus deformity. Secondary OA follows. Total knee replacement may be required to restore mobility and relieve pain.

Hips

The hips are occasionally affected in early RA but less commonly so than the knees at all stages of the disease. Pain and stiffness are accompanied by radiological loss of joint space and juxta-articular osteoporosis. The latter may permit medial migration of the acetabulum (protrusio acetabulae). Later, secondary OA develops. Hip replacement may be necessary.

Cervical spine

Painful stiffness of the neck in RA is often muscular, but it may be due to rheumatoid synovitis affecting the synovial joints of the upper cervical spine and the bursae, which separate the odontoid peg from the anterior arch of the atlas and its retaining ligaments. This synovitis leads to bone destruction, damages the ligaments and causes atlantoaxial or upper cervical vertebral instability. Subluxation and local synovial swelling may damage the spinal cord, producing pyramidal and sensory signs. MRI is the imaging of choice, but lateral flexed and extended neck X-rays can demonstrate instability. In late RA, difficulty walking that cannot be explained by articular disease, weakness of the legs, or loss of control of bowel or bladder may be due to spinal cord compression and is a neurosurgical emergency.

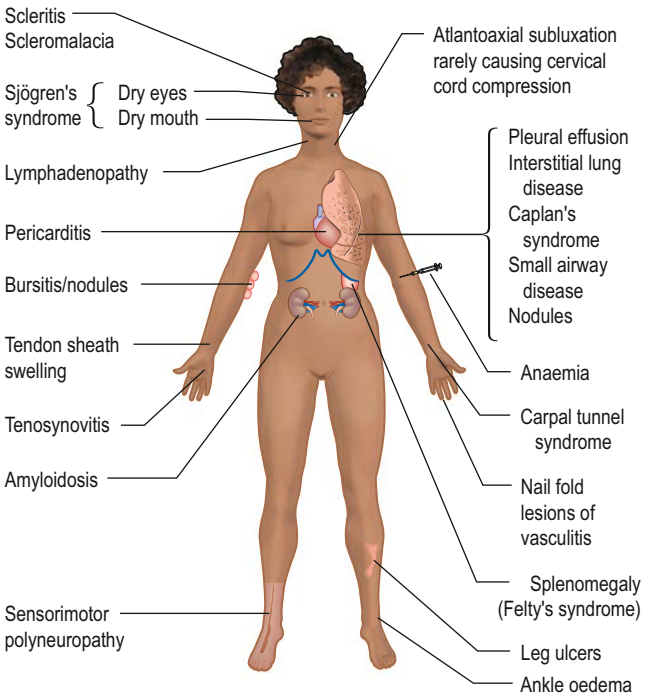


Fig. 32.25 Non-articular manifestations of rheumatoid arthritis.

Imaging of the cervical spine in flexion and extension is recommended in patients with RA before surgery or upper gastrointestinal endoscopy to check for instability and reduce the risk of cord injury during intubation.

Other joints

The temporomandibular, acromioclavicular, sternoclavicular, cricoarytenoid and any other synovial joint can be affected.

Non-articular manifestations

See Fig. 32.25.

Soft tissue surrounding joints

Subcutaneous nodules are firm and intradermal, generally occurring over pressure points: typically, the elbows, the finger joints, and the Achilles tendon in patients with seropositive erosive disease. They can be removed surgically but they tend to recur.

The olecranon and other bursae may be swollen (bursitis).

Tenosynovitis of flexor tendons in the hand can cause stiffness and occasionally a trigger finger. Swelling of the extensor tendon sheath over the dorsum of the wrist is common.

Muscle wasting around joints is common. Corticosteroid-induced myopathy occurs. Osteoporosis is more common in poorly controlled RA.

Less common non-articular manifestations

Non-articular complications are less common, probably because of more effective disease control.

Lungs

Findings in the lungs (see Chapter 20) include:

- airways disease: from predominant bronchiectasis (cough and daily sputum) to predominant obliterative bronchiolitis (progressive breathlessness)
- pleural disease: pleural effusion (asymptomatic to mildly breathless) and thickening
- interstitial lung disease: a combination of inflammation and basal lung fibrosis
- peripheral, intrapulmonary nodules: asymptomatic but may cavitate, especially with pneumoconiosis (Caplan's syndrome)
- infective lesions, e.g., tuberculosis in patients on biological DMARDs.

Vasculitis

Vasculitis (see 'Systemic inflammatory vasculitis' section) is uncommon. Risk factors include high titres of RF and active extra-articular disease elsewhere. Findings include:

- widespread cutaneous vasculitis with necrosis of the skin
- mononeuritis multiplex (see Chapter 28).

Heart and peripheral vessels

Poorly controlled RA with a persistently raised CRP and high cholesterol is a cardiovascular risk factor, independent of traditional risk factors (i.e., high cholesterol and hypertension). Other cardiovascular problems include:

- pericarditis, which is rarely symptomatic
- endocarditis and myocardial disease, rarely symptomatic, found at postmortem in approximately 20% of cases.

Nervous system

- Peripheral sensory neuropathies: mononeuritis multiplex or symmetrical peripheral neuropathy—due to vasculitis of the vasa nervorum.
- Compression neuropathies: carpal or tarsal tunnel syndrome—due to synovitis.
- Cord compression: due to atlantoaxial subluxation (see earlier).

Eyes

- Sicca syndrome causes dry mouth and eyes (see Sjögren's syndrome, 'Sjögren's syndrome' section).
- Scleritis and episcleritis occur in severe, seropositive disease, resulting in painful red eye.
- Scleromalacia perforans is a rare feature.

Kidneys

Amyloidosis causing proteinuria, nephrotic syndrome, and chronic kidney disease occurs rarely in severe, longstanding RA, due to the deposition of highly stable serum amyloid A protein (SAP) in the intercellular matrix of a variety of organs (see Chapter 26).

Spleen, lymph nodes, and blood

Felty's syndrome is splenomegaly and neutropenia in a patient with RA. Leg ulcers and sepsis are complications. HLA-DR4 is found in 95% of such patients, compared with 50%–75% of people with RA alone.

Anaemia is almost universal and is usually normochromic and normocytic. It may be iron-deficient owing to gastrointestinal blood loss from NSAID ingestion, or rarely haemolytic (Coombs-positive). There may be a pancytopenia due to hypersplenism in Felty's syndrome or as a complication of DMARD treatment. A high platelet count occurs with active disease.

Diagnosis and investigations

Diagnosis relies on the clinical features described earlier. The predictors of poor prognosis arthritis are listed in Box 32.26. Initial investigations include:

- ***Blood count*** may show a normochromic, normocytic anaemia.

Fig. 32.26 MRI of chronic rheumatoid arthritis of the knee. There is inflamed synovium throughout joint and spreading into the popliteal bursa *(white arrows)*.

- ***ESR and/or CRP*** are raised in proportion to the activity of the inflammatory process and are useful in monitoring treatment.
- ***Serology*** reveals ACPA positivity (see 'Serum autoantibody studies' section) that may be present early in the disease (and even predate it by many years), and in early inflammatory arthritis indicates the likelihood of progressing to RA. RF is present in approximately 75%–80% of cases and ANA at low titre in 30%.
- ***X-rays*** show soft tissue swelling in early disease, but ultrasound and MRI (**Fig. 32.26**) are useful to demonstrate synovitis and early erosions.
- ***Aspiration of the joint*** may be needed if an effusion is present. The aspirate looks cloudy owing to white cells. In a suddenly painful joint, septic arthritis should be suspected (see Chapter 11).
- ***Musculoskeletal ultrasound*** is a very effective way of demonstrating persistent synovitis when deciding on the need for DMARDs or assessing their efficacy.
- **MRI** can detect RA erosive change, particularly in early disease and allows direct visualization and assessment of synovitis and of bone oedema.

Other investigations will depend on the clinical picture, as outlined above. In severe disease, extensive imaging of joints may be required.

Management

See **Box 32.30**. The diagnosis of RA inevitably causes concern and fear in the patient and requires explanation and reassurance. Anyone with persistent arthritis should be referred urgently to a rheumatologist if small joints of hands or feet are affected, more than one joint is affected, or there has already been a delay of more than 3 months from symptom onset.

The doctor should have a positive approach and remind the patient that, with the help of drugs, most people continue to lead a more or less normal life, as the aim of therapy is disease remission. Furthermore, early therapy targeted at remission within 3 months of symptom onset is more likely to achieve sustained remission with subsequent drug tapering. People should be helped and encouraged to stay at work, as 30% lose their job within 2 years of diagnosis. Patients often adjust and cope remarkably with time and support from the specialist team in a rheumatology unit (including doctors, nurses, physiotherapists, podiatrists, and psychologists) and from leaflets, websites, and local patient groups.

Patients from socially deprived backgrounds and smokers have a worse prognosis. Due to the increased prevalence of cardiovascular risk in RA, aggressive management of traditional cardiac risk factors and suppression of disease are required.

Box 32.30 Management of rheumatoid arthritis

- Establish the diagnosis clinically.
- Use non-steroidal anti-inflammatory drugs (NSAIDs) and analgesics to control symptoms.
- Consider intramuscular depot methylprednisolone 80–120 mg if synovitis persists beyond 6 weeks.
- Refer to a rheumatologist to start csDMARDs, typically beginning with methotrexate monotherapy; if lack of response consider switch to other csDMARD.
- Refer for physiotherapy and general advice through a specialist team.
- As improvement occurs, as measured by less pain, less morning stiffness and reduced acute-phase response, tail off steroids, and possibly reduce drugs.
- If no better after two csDMARDs use bDMARD or tsDMARD to induce remission/low disease activity.

bDMARD, Biologic DMARD; *csDMARD,* conventional synthetic DMARD; *DMARDs,* disease-modifying anti-rheumatic drugs; *tsDMARD,* targeted synthetic DMARD.

Drug therapy

No curative agent exists for RA but early recognition and an intensified treat to target (T2T) regimen, with regular review and escalation of therapy until remission occurs, improves outcomes. Less frequent review is then continued to assess disease activity, damage, function, and co-morbidities.

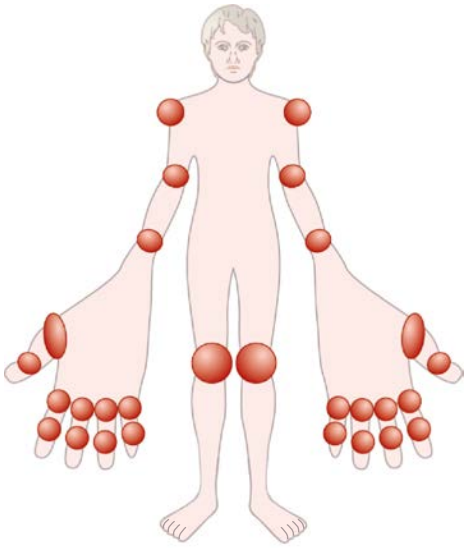
The disease activity score DAS28 is widely used to measure disease activity in RA by counting the number of tender and swollen joints in 28 joints in the upper limbs and knees, combining these values with the ESR and the patient's assessment of their general health on a visual analogue scale to generate a numerical score (**Fig. 32.27**). A DAS28 score of greater than 5.1 implies active disease, less than 3.2 low disease activity, and less than 2.6 remission.

For newly diagnosed patients, treatment begins with conventional synthetic (cs)DMARDs, typically methotrexate, at diagnosis, with consideration of short-term use of corticosteroids when initiating or changing csDMARD, with tapering and discontinuation as rapidly as clinically feasible to give benefit with minimal side-effects. With regular and frequent review, the T2T approach enables optimization and intensification of therapy with effective doses of csDMARDs, often in combination to prevent the long-term irreversible damaging effects of joint inflammation.

For patients who fail to achieve low disease activity or remission with serial csDMARDs, typically methotrexate and at least one other such as sulfasalazine or leflunomide, biological therapies (bDMARDS) are then used to achieve the desired target. If efficacy is not achieved, or subsequently lost, then switching to an alternative bDMARD is indicated. A new therapeutic class of DMARDs that inhibit intracellular Janus kinase (JAK) enzymes is now licensed in RA. These drugs, known as targeted synthetic (ts)DMARDs, have oral bioavailability, rapid onset of action, efficacy similar to bDMARDs and may be used with consideration of pertinent risk factors discussed below.

Non-steroidal anti-inflammatory drugs and coxibs

Most people with RA require an NSAID to relieve night pain and morning stiffness in addition to DMARDs. The individual response



Number of tender joints = t
Number of swollen joints = sw
Erythrocyte sedimentation rate in mm/h = ESR
Visual analogue scale of overall health in mm = GH

0 mm = Best — Worst = 100 mm

Calculate on online calculator (www.4s-dawn.com/DAS28)
or calculate on equation below:

DAS28 = 0.56*√(t) + 0.28*√(sw) + 0.70*Log(ESR) + 0.014*GH

Fig. 32.27 Calculation of disease activity score from 28 joints (DAS28).

to NSAIDs varies greatly, so several different drugs may have to be tried for at least a week to find the most effective (see **Box 32.18**). The major side-effects of NSAIDs and the use of coxibs are discussed earlier (see 'Non-steroidal anti-inflammatory drugs' section). If gastrointestinal side-effects are prominent, or the patient is over 65 years of age, add a proton pump inhibitor. For additional relief, a simple analgesic is taken as required (e.g., paracetamol or a combination of codeine or dihydrocodeine and paracetamol).

Corticosteroids

The early use of corticosteroids slows down the course of the disease but intensive short courses in very early arthritis do not appear to stop progressive disease. Recent guidelines recommend avoidance of long-term (≥3 months) corticosteroids due to their established toxicity profile and to increase focus on disease remission with escalation of DMARDs. Corticosteroids are a common cause of secondary osteoporosis; therefore, concomitant vitamin D and bisphosphonates are necessary to reduce fracture risk in patients who may require long-term (≥3 months) corticosteroid therapy.

Intra-articular injections with semi-crystalline steroid preparations have a powerful but sometimes only short-lived effect.

Intramuscular depot injections (40–120 mg depot methylprednisolone) are used to help induce remission while waiting for DMARDs to work and to control severe disease flares.

Oral corticosteroids are powerful disease-controlling drugs but cause a number of problems (**Box 32.31**; see also **Box 34.20**); so are not recommended long term because side-effects are inevitable. Short-term (<3 months) regimens may be considered to help induce remission. Corticosteroids are invaluable to people with severe disease who have extra-articular manifestations such as vasculitis.

Box 32.31 Problems associated with the use of corticosteroids

- Awareness of the potential side-effects of corticosteroids is increasing both among clinicians and patients. Guidelines recommend minimizing corticosteroid use and considering the risks posed by the drugs themselves.
- Patients should be advised to reduce consumption of sugars and saturated fats, and to be proactive in adjusting their diet to reduce the risk of weight gain.
- The skin becomes thin and easily damaged.
- Patients should be monitored for diabetes and hypertension.
- Cataract formation may be accelerated.
- Osteoporosis develops within 3 months on doses above 7.5 mg daily. Monitor with DXA scan and treat with calcium, vitamin D, and bisphosphonate (see Chapter 33).

DXA, Dual-energy absorptiometry.

Disease-modifying anti-rheumatic drugs

Various DMARDs are listed in **Box 32.32**. Their beneficial effect is not immediate, taking 2–3 months to become apparent. As monotherapy, conventional synthetic DMARDs (csDMARDs) often have only a partial effect, achieving between 20% and 50% improvement by ACR criteria for disease remission (**Box 32.33**), and are frequently used in combination.

Early intervention with csDMARDs within 3 months of disease onset improves the outcome. Although DMARD combinations (sulfasalazine, methotrexate, and/or hydroxychloroquine) may be used, a switch to bDMARDs is preferred when at least two csDMARDs have failed. Several DMARDs are contraindicated in pregnancy (see **Box 48.11**). Effective treatment with DMARDs including TNF-α blockers reduces the increased cardiovascular risk in RA.

Methotrexate

Methotrexate remains the anchor drug in RA therapy, although it should not be used in pregnancy. A screening history, chest X-ray, and an interferon-gamma release assay (IGRA) in high-risk patients are performed to exclude tuberculosis. Initial pneumococcal and annual influenza vaccinations are given. The starting *weekly* dose of 10 mg orally is increased up to 15–25 mg as necessary to T2T. It is well tolerated, although nausea or poor absorption may limit its efficacy, in which case it is given by subcutaneous injection. Oral folic acid reduces side-effects but may also lessen efficacy. Full blood counts and liver biochemistry should be monitored. Methotrexate usually works within 1–2 months. More patients remain on this agent than on most other DMARDs, indicating that it is effective and has relatively few side-effects.

Sulfasalazine

Sulfasalazine is well tolerated and can be used during pregnancy. The usual starting dose of 500 mg per day is increased to a maintenance dose of 2–3 g per day. Around 50% of patients respond in the first 3–6 months, but efficacy can be lost.

Hydroxychloroquine

A dose of 200–400 g daily is well tolerated. It is used alone in mild disease or commonly as an adjunct to other DMARDs. Due to the risk of irreversible retinopathy with increasing dose and duration of

Box 32.32 Disease-modifying anti-rheumatic drugs (DMARDs)

Drug	Dose	Side-effects	Monitoring
Sulfasalazine (enteric coated)	500 mg daily after food, increasing to 2–3 g daily	Nausea Skin rashes and mouth ulcers Neutropenia and/or thrombocytopenia Abnormal liver biochemistry	Baseline FBC, serum creatinine and electrolytes, LFTs, then at 2, 4, 6, and 10 weeks, then 3-monthly
Methotrexate[a]	7.5–10 mg increasing to max. 25 mg weekly, orally or s.c.	Nausea, mouth ulcers, and diarrhoea Abnormal liver biochemistry Neutropenia and/or thrombocytopenia Renal impairment Rare—pulmonary fibrosis	Baseline FBC, serum creatinine and electrolytes, LFTs, then at 2, 4, 6, and 10 weeks, then 3-monthly
Leflunomide[a]	10–20 mg daily, occasionally with initial loading dose	Diarrhoea Neutropenia and/or thrombocytopenia Abnormal liver biochemistry Alopecia Hypertension	Baseline FBC, serum creatinine and electrolytes, LFTs, then at 2, 4, 6, and 10 weeks, then 3-monthly
Biological drugs—TNF-α blockers[a]			
Etanercept (alone or with methotrexate)	s.c. 25 mg × 2-weekly or 50 mg weekly		Monitor as per concomitant DMARD; if monotherapy repeat FBC, serum creatinine and electrolytes, and LFTs 3–6 monthly
Adalimumab (with methotrexate)	s.c. 40 mg alternate weeks	Injection site reactions Heart failure Reversible lupus-like syndrome Infections Hypersensitivity reactions Rare—demyelination and autoimmune syndromes	As per etanercept
Infliximab (with methotrexate)	i.v. 3–10 mg/kg every 4–8 weeks		As per etanercept
Certolizumab pegol (alone or with methotrexate)	s.c. 400 mg in weeks 0, 2, 4, then 200 mg fortnightly		As per etanercept
Golimumab	s.c. monthly 50 mg		As per etanercept
Biological drugs—non-TNF-α blockers[a]			
Rituximab	i.v. 500–1000 mg	Hypo-/hypertension Pruritus and skin rash Back pain Hypogammaglobulinaemia Rare—toxic epidermal necrolysis	FBC, LFTs, and immunoglobulins at baseline and prior to each subsequent infusion
Abatacept	i.v. 10 mg/kg on days 1, 15, 30, and then monthly s.c. preparations available	Nausea, vomiting Headache Rare—hypersensitivity	Monitor as per concomitant DMARD; if monotherapy repeat FBC, serum creatinine and electrolytes, and LFTs 3–6 monthly
Tocilizumab	i.v. 8 mg/kg infusion every 4 weeks or s.c. 162 mg weekly	Upper respiratory tract infections, dyslipidaemia, neutropenia, increased liver enzymes, skin reactions	As per abatacept plus baseline and 3-month check of lipids with repeat at physician's discretion
Sarilumab	s.c. 200 mg alternate weeks	As per tocilizumab	As per tocilizumab
Ustekinumab	s.c. 45 mg (body weight <101 kg, or 90 mg ≥101 kg) on day 1 and 28 then every 12 weeks	Nasal congestion, sore throat, asthenia, nausea, skin reactions Rare—cutaneous lupus erythematosus, hypersensitivity vasculitis	As per abatacept
Guselkumab	s.c. 100 mg on day 1 and 28 then every 4–8 weeks	Arthralgia, diarrhoea, headache, increased risk of infection Uncommon—skin reactions	As per abatacept
Risankizumab	s.c. 150 mg on day 1 and 28 then every 12 weeks	As per guselkumab	As per abatacept
Secukinumab	s.c. 150 mg weekly for 5 doses then 150–300 mg every 4 weeks	Sore throat or dry cough, runny nose, fatigue, diarrhoea, skin reactions Rare—anaphylactic, hypersensitivity vasculitis	As per abatacept

Box 32.32 Disease-modifying anti-rheumatic drugs (DMARDs)—cont'd

Drug	Dose	Side-effects	Monitoring
Ixekizumab	s.c. 160 mg on day 1 then 80 mg 2 weekly for 6 doses then 80 mg every 4 weeks	Conjunctivitis, nausea, oropharyngeal pain, skin reactions Rare—anaphylactic reaction	As per abatacept
Targeted small-molecule inhibitors			
JAK inhibitors[a]			
Tofacitinib	5 mg oral twice daily	Anaemia, cough, neutropenia, diarrhoea, dyslipidaemia, dyspnoea, fatigue, gastritis, headache, hypertension, skin reactions Uncommon—decreased leucocytes, deep vein thrombosis, myocardial infarction, neoplasms, thromboembolism	Baseline as per etanercept, then monitor FBC, serum creatinine and electrolytes, and LFTs at 4 and 8 weeks then 3-monthly. Check lipids at baseline and after 8 weeks
Baricitinib	4 mg oral once daily in adults 18–74 years; 2 mg oral once daily in adults ≥75 years	Dyslipidaemia, herpes zoster, increased risk of infection, nausea, oropharyngeal pain, thrombocytosis, acne, neutropenia, weight gain	As per tofacitinib
Filgotinib	200 mg oral once daily in adults 18–74 years; 100 mg oral once daily in adults ≥75 years	Abdominal pain, anaemia, cough, dyslipidaemia, fever, headache Unknown frequency of embolism and thrombosis, malignancy, reactivation of infections	As per tofacitinib
Upadacitinib	15 mg oral once daily	As per filgotinib	As per tofacitinib
Phosphodiesterase type-4 (PDE4) inhibitor[a]			
Apremilast	10 mg oral once a day increasing to by 10 mg a day to maintenance 30 mg twice daily on day 6	Reduced appetite, back pain, cough, depression, diarrhoea, increased risk infection Rare—gastrointestinal haemorrhage, suicidal behaviours	No routine laboratory monitoring

All patients administered immunosuppressive medications (labelled[a]) should be given pneumococcal, COVID-19, and annual influenza vaccinations.
CD19, Cluster of differentiation 19; *FBC,* full blood count; *i.v.,* intravenous; *JAK,* Janus kinase inhibitor; *LFTs,* liver function tests; *s.c.,* subcutaneous; *TNF-α,* tumour necrosis factor-alpha.

Box 32.33 The 2011 ACR/EULAR definitions for disease remission in rheumatoid arthritis clinical trials

- Tender joint count (including feet and ankles) ≤1
- Swollen joint count (including feet and ankles) ≤1
- C-reactive protein ≤1 mg/L
- Patient global assessment ≤1 (on a 10-cm visual analogue scale)

ACR, American College of Rheumatology; *EULAR,* European League Against Rheumatism.

use, a dose ceiling of 5 mg/kg/day is recommended and current guidance is for annual eye assessment (ideally including optical coherence tomography) after 5 years of therapy are completed or earlier if patient has eGFR <60 mL/min, concomitant tamoxifen or dose >5 mg/kg/day.

Leflunomide

This csDMARD prevents pyrimidine production in proliferating lymphocytes through blockade of the enzyme dihydro-orotate dehydrogenase, thus blocking clonal expansion of T cells. It has a long half-life of 4–28 days. A dose of 20 mg daily (10 mg if diarrhoea is a problem) is used. Diarrhoea diminishes with time. Blood monitoring is obligatory (full blood count, platelets, liver biochemistry). The onset of action is 4 weeks with some further improvement sustained at 2 years. Leflunomide works in some patients who have failed to respond to methotrexate. Its long half-life means that it is best avoided in females planning a family.

Biological therapies (bDMARDs)

These drugs are more expensive than csDMARDs and rheumatologists should consider the cost of therapies that are being reduced by the introduction of biosimilar medicines. These medicines are highly similar to another (originator) drug and have been shown to have no clinically meaningful differences in quality, safety or efficacy from the originator compound.

TNF-α blockers

These agents are used after at least two csDMARDs (usually sulfasalazine and methotrexate) have failed. They are usually given in combination with methotrexate to reduce loss of efficacy due to anti-drug antibody (ADA) formation.

- ***Etanercept*** is a fully humanized p75 TNF-α receptor IgG1 fusion protein given by self-administered subcutaneous injection. Around 65% of patients respond well. Some develop an injection reaction.
- ***Adalimumab*** is a fully human monoclonal antibody against TNF-α, given along with methotrexate.
- ***Infliximab*** is a monoclonal antibody against TNF-α, given intravenously and co-prescribed with methotrexate to prevent loss of efficacy because of antibody formation.
- ***Certolizumab pegol*** is a Fab fragment of a humanized TNF-α inhibitor monoclonal antibody that has polyethylene glycol (PEG) groups attached to reduce immunogenicity and prolong half-life. The presence of PEG and lack of Fc portion ensure minimal transfer across the placenta and into breast milk. It is useful with or without methotrexate for severe active RA.

- ***Golimumab*** is a human IgG1-κ monoclonal antibody against TNF, which is given by subcutaneous injection once monthly for severe RA.

These products slow or halt erosion formation in up to 70% of people with RA and produce healing in a few. Malaise and tiredness improve in a manner that is not seen with csDMARDs. Secondary failure may occur with all in the first year; changing to another anti-TNF agent is justified and often regains control of the disease. Failure to respond to one does not predict failure to others. Currently, biosimilars exist for etanercept, infliximab, and adalimumab.

Safety data. Various national registries exist to study adverse events, long-term outcomes, and problems arising from biological therapies. To date, the results are reassuring. Infection rates are increased with TNF-α blockers, particularly in the first few months of treatment, but the rate then declines. There is a known association between RA and non-Hodgkin's lymphoma (NHL) but, overall, registers have not shown an increased risk of NHL in patients with RA treated with TNF-α blockers, to date. There is no convincing evidence of any increased risk of other cancers. Reactivation of old tuberculosis may occur but is probably less common with etanercept. A pre-treatment chest X-ray is recommended, with IGRA and specialist review for high-risk groups. Tuberculosis should be treated before these agents are prescribed, and a course of prophylaxis is used in latent disease. Hepatitis B and C infection requires risk analysis and regular aminotransferase monitoring if anti-TNF agents are prescribed. They should not be used in patients who have severe cardiac failure. To date, there is no evidence of an adverse effect on pregnancy outcome but care is essential (see **Box 48.11**).

Other biological agents

- ***Rituximab*** is a chimeric monoclonal antibody (see Chapter 6), directed against the CD20 receptor expressed on pre-B and mature B cells. Rituximab produces significant improvement in RF-positive RA for 8 months to several years when used alone or in combination with corticosteroids and/or methotrexate. This clinical improvement is associated with a 6–9-month B-cell lymphopenia. A re-flare is often accompanied by a return of peripheral B lymphocytes and rise in CRP. Repeated courses, with disease flare, are well tolerated and around 80% of RF-positive patients respond, with 50%–60% showing persistent disease control; however, immunoglobulin levels should be monitored, as they may fall with repeated treatments. Rituximab is mainly used in patients who have failed to respond to anti-TNF agents and a biosimilar exists.
- ***Abatacept*** is a recombinant fusion protein of CTLA4 and the Fc portion of IgG1, which selectively modulates T-cell activation by co-stimulation blockade. It may be used in patients who do not respond to anti-TNF regimens.
- ***Tocilizumab and sarilumab*** are monoclonal antibodies against the anti-IL-6 receptor that may be used with methotrexate for moderate to severe RA, usually after at least one other biologic has failed, but may also be considered as a first-line biological therapy.

Targeted synthetic DMARDs (tsDMARDs)

Baricitinib, filgotinib, tofacitinib, and upadacitinib are oral JAK inhibitors that may be used in combination with methotrexate, or as monotherapy in patients who cannot take methotrexate, instead of a bDMARD. An RCT studied patients with RA >50 years of age who had cardiovascular risk factors, comparing two groups: those on tofacitinib and those treated with TNF-α blockers. There was an increased risk of major adverse cardiovascular events and malignancy in the tofacitinib group. Therefore, all JAK inhibitors are considered with caution in patients >65 years and/or with risk factors for cardiovascular disease, thromboembolism, or malignancy.

Switching between DMARDs and tapering therapy

Patients who fail to respond initially (primary failure) or subsequently lose efficacy (secondary failure) to a bDMARD over time will usually be switched to another bDMARD. In patients who lose response to a TNF-α blocker, it is helpful to measure drug levels and for presence of neutralizing ADAs, where possible. If the drug level is therapeutic, then a non-TNF-α blocker is preferred. Alternatively, if the drug level is subtherapeutic, non-adherence or the presence of ADA are likely explanations and switching to another TNF-α blocker or a different class of bDMARD are viable alternatives. If primary or secondary failure occurs to a subsequent bDMARD the switching process may be repeated multiple times. Switching to tsDMARDs may now be considered.

In patients who achieve sustained remission (>1 year) it is now commonplace to attempt dose reduction of DMARDs by either dose reduction or prolongation of dosing interval. Potential biomarkers of responsiveness are being explored to ensure a precision medicine approach to each individual patient.

Physical measures

Input is required from the multidisciplinary team. Physiotherapists advise a combination of rest for active arthritis and exercises to maintain joint range and muscle power. Exercise in a hydrotherapy pool is popular and effective. Occupational therapists help to manage activities of daily living despite the arthritis and provide functional adaptations in the home or at work. Family and friends should be involved. Podiatry, footwear advice, and psychological support should also be offered to all people with RA.

Surgery

Surgery has a useful role in the long-term approach to patient management but is less frequently needed as therapeutic disease control becomes more effective. Its main objectives are prophylactic, to prevent joint destruction and deformity, and reconstructive, to restore function.

Surgical options in RA include: surgical synovectomy; tendon repair; osteotomy to realign weight-bearing surfaces; joint fusion to stabilize damaged joints that are not easily replaced; small joint implant arthroplasty, for example, MCP joints; excision arthroplasties of the metatarsal heads to reduce metatarsal pain; and total joint replacement arthroplasty of the shoulder, elbow, wrist, hip, knee, and ankle. The main aim of joint arthroplasty is to reduce pain and improve function, requiring careful planning with explanation of risks and outcomes to the patient.

Prognosis

A poor prognosis is indicated by:

- ***a clinical picture*** of an insidious rather than an explosive onset of RA, female sex, increasing number of peripheral joints involved, and the level of disability at the onset
- ***blood tests*** showing a high CRP/ESR, normochromic normocytic anaemia, and high titres of ACPA and of RF
- ***X-rays*** with early erosive damage (*Note:* ultrasound and MRI can show cartilage and bone damage prior to conventional X-rays).

Prognosis can be altered dramatically with early DMARD therapy and T2T regimen (see 'Drug therapy' section) under expert supervision.

SPONDYLOARTHRITIS

The term spondyloarthritis (SpA) describes a group of conditions affecting the spine and peripheral joints with familial clustering and a link to certain type 1 HLA antigens. These conditions share genetic factors and increasing implication of the IL-17 pathway in pathogenesis. Joint involvement differs from RA, with a different distribution and associated extra-articular features (**Fig. 32.16**). These diseases occasionally present in childhood.

They may also be categorized according to their predominant clinical manifestation of axial (sacroiliac and/or spine) or peripheral (arthritis, enthesitis, and/or dactylitis) disease with possible overlap. Distinct SpA conditions exist (see **Box 32.34**).

Histologically, synovitis is similar to that of RA, but there is no production of RF or ACPA. Inflammation of the enthesis (junction of ligament or tendon and bone) and joint ankylosis develop more commonly than in RA. All are associated with an increased frequency of sacroiliitis and HLA-B27.

Aetiology

These disorders have a striking association with HLA-B27, particularly ankylosing spondylitis (AS). HLA type B27 is present in more than 90% of Caucasians with AS but only 8% of controls. HLA-B27 exhibits a number of unusual characteristics, including a high tendency to misfold. The types of arthritis that follow a precipitating infection are called reactive arthritis (see 'Reactive arthritis' section).

Axial spondyloarthritis

Axial spondyloarthritis (AxSpA) occurs in 1% of the general population. It is an inflammatory disorder primarily affecting sacroiliac joints or fibrous and synovial joints of the spine that are detectable on MRI. When radiographic changes at the sacroiliac joints are present, the term ankylosing spondylitis (AS) is used.

Ankylosing spondylitis

It is now recognized that AS forms part of the spectrum of AxSpA. It presents with inflammatory back pain and sacroiliac inflammation. It is seen in 0.2%–0.5% of the population in Northern Europe and approximately 0.5% in the USA, affecting mainly young adults (late teens to early thirties) and occurring worldwide, with a male to female ratio of 3:1. Diagnostic delay remains a problem, particularly in females who are under-diagnosed. The frequency of AxSpA in different populations is roughly paralleled by the incidence of HLA-B27; Africans and Japanese have a low incidence of both HLA-B27 and AS, while the North American Haida Indians have a high incidence of both.

Aetiology

A combination of genetic and environmental features:

- Among white individuals with AS, 95% carry HLA-B27, in contrast to 8% of most white populations. Twin studies indicate a much higher disease concordance in HLA-B27-positive monozygotic twins (up to 70%) than in dizygotic twins (about 20%–25%).
- Other genes lying within the MHC (the *IL-1* gene cluster and *CYP2D6*) also influence susceptibility to AS but the disease is polygenic and GWAS have identified significant associations with the endoplasmic reticulum aminopeptidase *(ERAP)*-1 and *IL-23* genes.
- The association of *ERAP-1* with AxSpA may support the arthritogenic peptide hypothesis, whereby disease is triggered by presentation of peptide by HLA-B27 to CD8-positive T cells. Furthermore, misfolding of HLA-B27 leads to the production of IL-23, and T cells resident in entheses promote inflammation characteristic of SpA in response to IL-23.
- The finding of increased frequency of Th17 cells and various genetic association studies have implicated the IL-23-IL-17 axis in pathogenesis and this knowledge has been translated into therapies that target this pathway.
- Reports of dysbiosis in the gut microbiome of AS patients support previous work indicating a link between gut microbiota, intestinal inflammation, and AxSpA.
- Lymphocyte and plasma cell infiltration occurs with local erosion of bone at the attachments of the intervertebral and other ligaments (enthesitis), which heals with new bone (syndesmophyte) formation.

Clinical features

Initially, the diagnosis is often missed because the patient is asymptomatic between episodes and radiological abnormalities are absent.

- ***Back pain*** with episodic inflammation of the sacroiliac joints in the late teenage years or early twenties is the first manifestation of AxSpA.
- ***Pain in one or both buttocks*** and low back pain and stiffness are typically worse in the morning and relieved by exercise.
- ***Retention of the lumbar lordosis during spinal flexion*** is an early sign. Later, paraspinal muscle wasting develops.

Criteria for classifying inflammatory back pain as AxSpA are shown in **Box 32.35**. These criteria encompass the whole spectrum of disease, with evidence of sacroiliitis being based on radiographic or MRI evidence of disease. Spinal stiffness can be measured by Schober's test; skin over the midline of the spine is marked 5 cm below and 10 cm above the dimples of Venus and the increase in distance between those marks during flexion is recorded. An increase of less than 5 cm implies spinal stiffness.

Non-spinal complications (uveitis or costochondritis) suggest the diagnosis of SpA (**Box 32.36**):

- ***Costochondral junction inflammation*** causes anterior chest pain. Measurable reduction of chest expansion is due to costovertebral joint involvement.
- ***Peripheral joint involvement*** is asymmetrical and affects a few, predominantly large, joints. Hip involvement leads to fixed

Box 32.34 Spondyloarthritis

- Axial spondyloarthritis, including ankylosing spondylitis
- Psoriatic arthritis
- Reactive arthritis (sexually acquired)
- Post-dysenteric reactive arthritis
- Enteropathic arthritis (ulcerative colitis/Crohn's disease)

Box 32.35 Back pain criteria for diagnosing axial/ankylosing spondylitis

- Age of onset <45 years
- Insidious onset
- Improvement of back pain with exercise
- No improvement of back pain with rest
- Pain at night with improvement on getting up

The presence of four of the five criteria suggests ankylosing spondylitis with 80% sensitivity. All criteria have high sensitivity.

Box 32.36 Non-articular problems in spondyloarthritis

- Uveitis, in all types
- Cutaneous lesions in reactive arthritis (keratoderma blennorrhagica), histologically identical to pustular psoriasis
- Nail dystrophy, in psoriasis and reactive arthritis
- Aortitis, occasionally in ankylosing spondylitis and reactive arthritis

flexion deformities of the hips and further deterioration of posture. Young teenage males occasionally present with a lower-limb monoarthritis, which later develops into AS.

- ***Acute anterior uveitis*** occurs in approximately 30% of patients with AxSpA and related diseases, and is occasionally the presenting complaint. Severe eye pain, photophobia, and blurred vision are an emergency (see Chapter 31).
- ***Other extra-articular features*** include:
  - cardiovascular disease—aortic incompetence occurs in up to 1% of patients with established AS and cardiac conduction abnormalities in around 5%
  - respiratory disease—rarely, chest wall rigidity is associated with interstitial lung disease resulting in pulmonary fibrosis with an upper lobe predilection
  - renal impairment—reported in 10%–35% of patients with AS and most commonly linked to chronic NSAID use
  - axial osteoporosis—occurs in approximately 25%, with vertebral fracture in 10%.

Overall clinical assessment of disease activity is based on a 1–10 score for levels of fatigue, spinal pain, arthralgia, swelling, localized tenderness, inflammation of tendons/enthesis, and duration and severity of morning stiffness. It is measured using instruments such as the Bath Ankylosing Spondylitis Disease Activity Index and increasingly the Ankylosing Spondylitis Disease Activity Score that includes laboratory measures as well as patient perspective (see Further reading).

Investigations

- ***Blood.*** ESR and CRP are usually raised.
- ***HLA-B27 testing.*** This test is not definitive due to the high frequency of HLA-B27 in the population, but may strengthen or confirm suspected diagnoses in the absence of sacroiliitis on imaging.
- ***X-rays.*** The medial and lateral cortical margins of both sacroiliac joints lose definition owing to erosions and eventually become sclerotic (**Fig. 32.28**). Radiological appearances in the spine of blurring of the upper or lower vertebral rims at the thoracolumbar junction are caused by an enthesitis at the insertion of the intervertebral ligaments and may eventually affect the whole spine causing bony spurs (syndesmophytes), bony ankylosis and permanent stiffening. The sacroiliac joints eventually fuse, as may the costovertebral joints, reducing chest expansion. Calcification of the intervertebral ligaments and fusion of the spinal facet joints and syndesmophytes leads to a 'bamboo' spine (**Fig. 32.29**).
- ***MRI.*** MRI with gadolinium demonstrates sacroiliitis before it is seen on X-rays, as well as persistent enthesitis.

Management

- Effective management of AS requires early diagnosis so that a regimen of preventative exercises is started before syndesmophytes have formed. Morning exercises aim to maintain spinal mobility, posture, and chest expansion; regular NSAIDs to improve symptoms and signs of SpA are often required to achieve this goal.

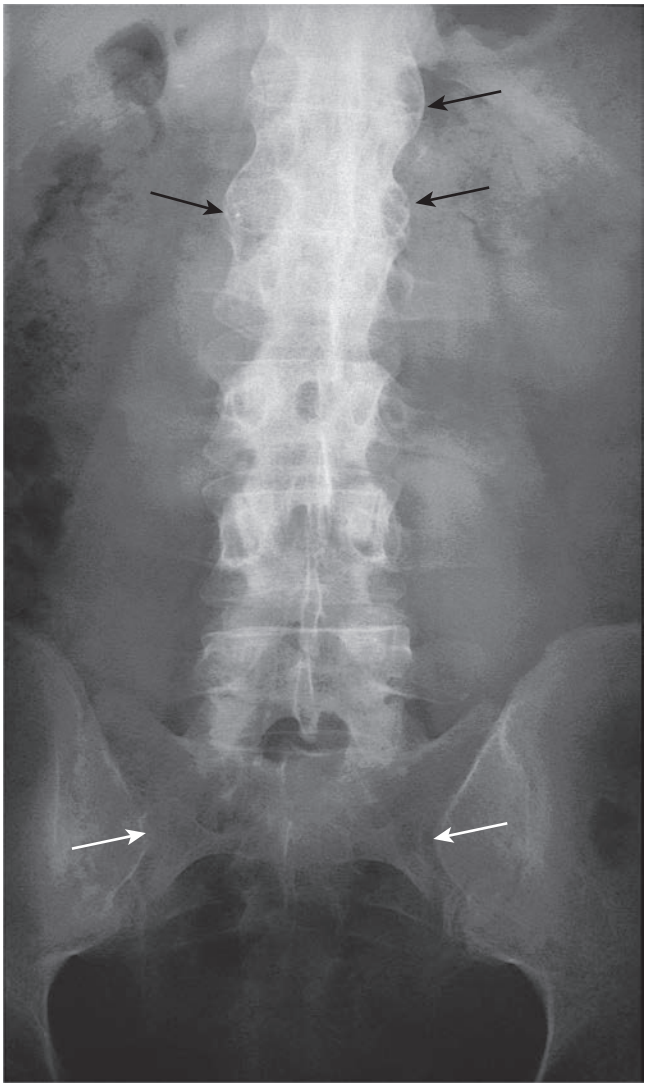


Fig. 32.28 X-ray of ankylosing spondylitis. The sacroiliac joints are eroded and show marginal sclerosis *(white arrows)*. There is bridging syndesmophyte formation at the thoracolumbar junction *(black arrows)*.

- Failure to control pain and to encourage regular spinal and chest exercises leads to an irreversible dorsal kyphosis and wasted paraspinal muscles, which, along with stiffening of the cervical spine, makes forward vision difficult.
- When the inflammation is active and the morning pain and stiffness are too severe to permit effective exercise, an evening dose of a long-acting or slow-release NSAID or an NSAID suppository improves sleep, pain control, and exercise compliance.
- Sulfasalazine, methotrexate, and leflunomide may help peripheral arthritis but not spinal disease.
- When NSAIDs have failed, the TNF-α-blocking drugs adalimumab, etanercept, golimumab, certolizumab, and infliximab (see **Box 32.32**) all reduce symptoms of spinal and peripheral joint inflammation substantially and improve function, as well as quality of life. There are some reports of reduction of bony progression. Relapse occurs on stopping therapy but may be delayed by several months making intermittent treatment feasible.
- Other bDMARDs including the IL-17 blockers ixekizumab and secukinumab or JAK inhibitors (tofacitinib and upadacitinib) are

Fig. 32.29 X-ray of bamboo spine in ankylosing spondylitis. In advanced disease, there is calcification of the interspinous ligaments and fusion of the facet joints, as well as syndesmophytes at all levels. The sacroiliac joints fuse.

alternative options in patients with persistently high disease activity despite conventional treatments.

Prognosis

With exercise and pain relief, the prognosis is excellent. The advent of bDMARD and tsDMARD therapies has reduced the morbidity of severe disease, lowering the risk of permanent spinal stiffness and progressive peripheral joint disease.

Psoriatic arthritis

The prevalence of psoriasis is 2%–3% worldwide; in this population, around 10% have arthritis, which precedes skin disease in around 15% of cases. A family history of psoriasis may be a clue to the diagnosis. The aetiology and pathogenesis are described in Chapter 45.

Clinical features

Patterns of psoriatic arthritis include:

- ***Mono- or oligoarthritis***.
- ***Polyarthritis:*** often begins with an asymmetrical pattern and progresses to be virtually indistinguishable from RA.
- ***Spondylitis:*** unilateral or bilateral sacroiliitis and early cervical spine involvement; only 50% are HLA-B27-positive.
- ***Distal interphalangeal arthritis:*** the most typical pattern of joint involvement in psoriasis, often with adjacent nail dystrophy (see Chapter 45) reflecting enthesitis extending into the nail root. Dactylitis, in which an entire finger or toe is swollen, with joint and tendon sheath involvement, is characteristic of this condition.
- ***Arthritis mutilans:*** affects about 5% of patients who have psoriatic arthritis and causes marked periarticular osteolysis and bone shortening ('telescopic' fingers) (Fig. 32.30).

Fig. 32.30 Hand showing psoriatic arthritis mutilans. All the fingers are shortened and the joints unstable, owing to underlying osteolysis.

Fig. 32.31 X-ray of psoriatic arthritis. There is osteolysis of the metatarsal heads and central erosion of the proximal phalanges to produce the 'pencil in cup' appearance *(circle)*. All the lesser toes are subluxed.

Radiologically, psoriatic arthritis is erosive but the erosions are central in the joint, not juxta-articular, and produce a 'pencil in cup' appearance (Fig. 32.31). The skin and nail disease can be mild and may develop after the arthritis.

Management and prognosis

NSAIDs are recommended to reduce pain with consideration of the benefit-to-risk ratio in patients with potential cardiovascular comorbidities. Local synovitis responds to intra-articular corticosteroid injections. A csDMARD (preferably methotrexate) is recommended in patients with persistent peripheral joint synovitis, although none are proven to slow the development of joint damage. Hydroxychloroquine is best avoided because it may exacerbate psoriatic skin reactions. Similarly, oral corticosteroids may destabilize skin disease; they are best avoided but corticosteroids are valuable when injected into a single inflamed joint.

Various bDMARDs and tsDMARDs are now available and highly effective for severe skin and joint disease in patients where csDMARDs, principally methotrexate, have failed. The bDMARDs include anti-TNF-α agents (see 'TNF-α blockers' section); IL-17 inhibitors, secukinumab and ixekizumab; the IL-12/23 inhibitor, ustekinumab; and IL-23 inhibitors, guselkumab and risankizumab. The tsDMARDs include an oral PDE4 inhibitor (apremilast) and JAK inhibitors (tofacitinib and upadacitinib). The choice of drug depends on disease manifestations, with a preference for TNF-α blockers or IL-17 inhibitors in predominantly axial disease.

Reactive arthritis

Reactive arthritis is a sterile synovitis, which occurs following an infection. Spondyloarthritis develops in 1%–2% of patients after an acute attack of dysentery, or after a sexually acquired infection—non-specific urethritis (NSU) in the male, non-specific cervicitis in the female. In males, positivity for HLA-B27 increases the risk of developing reactive arthritis after such an infection by thirty- to fiftyfold but not all patients are HLA-B27-positive. Females are less commonly affected.

Aetiology

A variety of organisms can be the trigger, including strains of *Salmonella* or *Shigella* spp. in bacillary dysentery. *Yersinia enterocolitica* causes diarrhoea and a reactive arthritis. In NSU, the organisms are *Chlamydia trachomatis* or *Ureaplasma urealyticum*.

People with reactive arthritis are not more susceptible to infection but appear to respond differently. Bacterial antigens or bacterial DNA have been found in the inflamed synovium of affected joints, so persistent antigenic material may drive the inflammatory process. HLA-B27 increases susceptibility to reactive arthritis by:

- T-cell receptor repertoire selection
- molecular mimicry causing autoimmunity against HLA-B27 and/or other self-antigens
- mode of presentation of bacteria-derived peptides to T lymphocytes.

There are other organisms that also trigger reactive arthritis but have a different genetic basis; see post-streptococcal arthritis (see Chapter 42), gonococcal arthritis (see 'Gonococcal arthritis' section), and brucellosis (see 'Brucellosis' section). In these conditions, the line between reactive arthritis and septic arthritis is more indistinct and they can cause both.

Clinical features

Typically an acute, asymmetrical, lower-limb arthritis, occurring days to weeks after the infection. The arthritis may be the presenting complaint if the infection is mild or asymptomatic. Enthesitis is common, causing plantar fasciitis or Achilles tendon enthesitis (see 'Plantar fasciitis' and 'Achilles tendonosis' sections), and dactylitis may also occur (**Fig. 32.32**); 70% of patients recover fully within 6 months but many have a relapse.

In susceptible individuals with reactive arthritis, sacroiliitis and spondylitis may also develop. Sterile conjunctivitis occurs in 30%. Acute anterior uveitis complicates more severe or relapsing disease but is not synchronous with the arthritis.

The skin lesions resemble psoriasis:

- ***Circinate balanitis*** in the uncircumcised male causes painless superficial ulceration of the glans penis. In the circumcised male, the lesion is raised, red, and scaly. Both heal without scarring.
- ***Keratoderma blennorrhagica*** involves the skin of the feet and hands, which develops painless, red, and often confluent raised plaques and pustules that are histologically similar to pustular psoriasis.
- ***Nail dystrophy*** occurs.

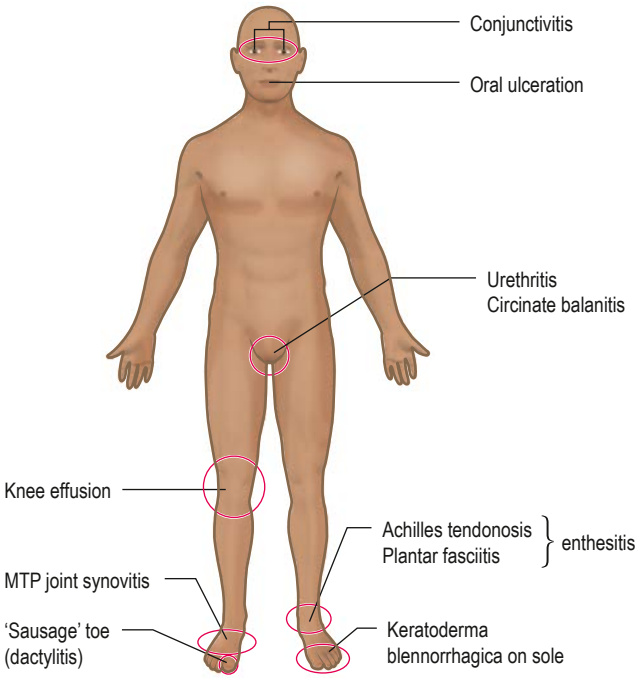


Fig. 32.32 Clinical features of reactive arthritis. *MTP,* Metatarsophalangeal.

Management

Treating persisting infection with antibiotics alters the course of the arthritis, once it has developed. Cultures should be taken and any infection treated. Sexual partners must be screened.

Pain responds well to NSAIDs and locally injected or oral corticosteroids. The majority of individuals with reactive arthritis have a single attack that settles, but a few develop a disabling, relapsing, and remitting arthritis. Relapsing cases are sometimes treated with sulfasalazine or methotrexate (see **Box 32.32**).

Enteropathic arthritis associated with inflammatory bowel disease

Enteropathic synovitis occurs in up to 10%–15% of patients who have ulcerative colitis or Crohn's disease (see Chapter 22). The link between the bowel disease and the inflammatory arthritis is not clear. The arthritis is asymmetrical and predominantly affects lower-limb joints. An HLA-B27-associated sacroiliitis or spondylitis also occurs. The joint symptoms may predate the development of bowel disease and lead to its diagnosis.

Remission of ulcerative colitis or total colectomy usually leads to remission of the joint disease, but arthritis can persist even in well-controlled Crohn's disease.

Management

Inflammatory bowel disease should be treated (see Chapter 22). In all cases of enteropathic arthritis, the symptoms are helped by NSAIDs, although they may make diarrhoea worse. A monoarthritis is best treated by intra-articular corticosteroids. Sulfasalazine is more frequently prescribed than mesalazine, as it may help both bowel and joint disease. The bDMARDs that are used to treat arthritis and inflammatory bowel disease are the TNF-α-blocking drugs infliximab, adalimumab, golimumab, and certolizumab, and ustekinumab, the IL-12/23 inhibitor. The JAK inhibitors baricitinib, filgotinib, tofacitinib, and upadacitinib may also be used.

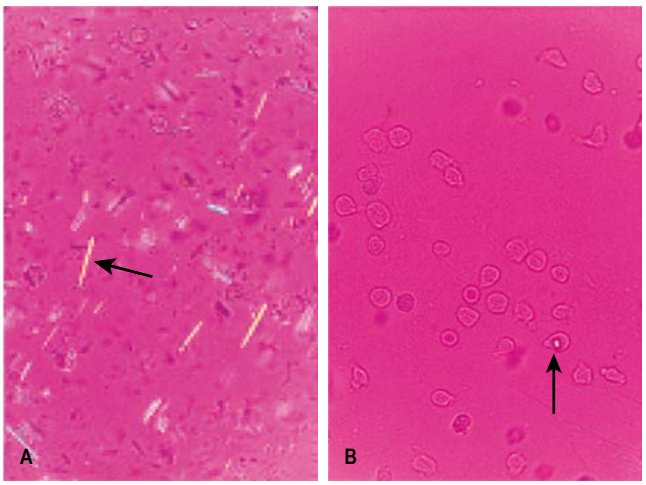


Fig. 32.33 Crystal arthritis. **(A)** Needle-shaped urate crystals *(arrowed)*. **(B)** A small, intracellular pyrophosphate crystal *(arrowed)*. Both are viewed under polarized light with a red filter.

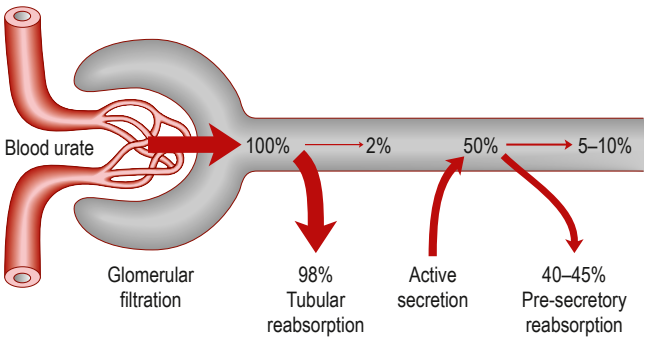


Fig. 32.34 Urate renal transport. The net result is that about 5%–10% of the glomerular load is excreted in the urine under normal circumstances.

CRYSTAL ARTHRITIS

The two main types of crystal-induced arthritis are caused by sodium urate and calcium pyrophosphate crystals, which are distinguished by their different shapes and refringence properties under polarized light with a red filter (Fig. 32.33). Rarely, crystals of calcium apatite (see 'Crystal-associated OA' section) or cholesterol cause acute synovitis.

Gout and hyperuricaemia

Gout is an inflammatory arthritis associated with hyperuricaemia and intra-articular sodium urate crystals.

Epidemiology

The prevalence of gout has increased in recent decades to 2.5% in the UK and 3.9% in the USA. Asian populations are also increasingly at risk as their diet becomes more Western. This rising prevalence is due to changing diets with purine-rich foods, high saturated fats, and fructose-containing drinks; alcohol misuse; increasing comorbidities that promote hyperuricaemia; and suboptimal management. Gout is more common in males than females (5:1); it rarely occurs before young adulthood (when it suggests a specific genetic defect), and seldom in premenopausal females. Some 85%–90% of cases are idiopathic. Hyperuricaemia is common in certain ethnic groups (e.g., Maoris).

The last two steps of purine metabolism in humans are the conversion of hypoxanthine to xanthine, and of xanthine to uric acid, catalyzed by the enzyme xanthine oxidase. Serum uric acid (SUA) levels are higher in males than in females. Pathological hyperuricaemia is defined as an SUA level of 408 μmol/L above which monosodium urate crystal formation occurs in vitro at physiological pH and temperature, although it may occur in peripheral joints with lower tissue pH and temperature, for example, 360 μmol/L at 35°C and 300 μmol/L at 30°C. Hyperuricaemia is mostly asymptomatic, although OA joints are more prone to attacks of gout. The range of SUA for individuals with gout is higher than in healthy controls. SUA levels increase with age, obesity, a 'Western' diet (see earlier) and combined hyperlipidaemia, diabetes mellitus, ischaemic heart disease, and hypertension (metabolic syndrome; see Chapter 37). Gout is often familial.

Pathogenesis of hyperuricaemia and gout

Uric acid is the final product of endogenous and dietary purine metabolism in humans, and SUA depends on the balance between purine synthesis, ingestion of dietary purines, and the elimination of urate by the kidney (66%) and intestine (33%).

Some 90% of people with gout have impaired excretion of uric acid (10% have increased production due to high cell turnover and <1% due to an inborn error of metabolism). Renal excretion is coordinated by a group of secretory and reabsorptive renal tubular urate transport molecules, some of which are targets of urate-lowering drugs. These molecules can be grouped into reabsorptive urate anion exchangers (URAT1/SLC22A12, OAT4/SLC22A11, and OAT10/SLC22A3), the resorptive GLUT9/SLC2A9 urate transporter, secretory anion-exchange transporters (OAT1, OAT2, and OAT3) and sodium-phosphate transporter proteins (NPT1/SLC17A1 and NPT4/SLC17A3), and the ATP-driven secretory pump MRP4/SLC17A3 (Fig. 32.34). The secretory transporter ABCG2 is significant in the gut, with reduced functioning contributing to extrarenal under-excretion. GWAS have identified prominent loci encoding urate transporters: primarily SLC2A9 and ABCG2; with a second tier dominated by genes encoding SLC22A11, SLC22A12, SLC17A1, and SLC17A3. Causes of hyperuricaemia are shown in Box 32.37.

Gout as an autoinflammatory disease

The involvement of the innate immune system and inflammasomes indicates that gout is an autoinflammatory disease, similar to the hereditary periodic fevers (see 'Miscellaneous arthropathies' section). This response is initiated when monosodium urate crystals interact with resident macrophages to form and activate NLRP3 inflammasome. Caspase 1 is recruited by the activated inflammasome (see Chapter 6) leading to activation of IL-1β and cellular activation, which triggers an IL-8-mediated influx of neutrophils. Ingestion by polymorphonuclear leucocytes of monosodium urate crystals causes the release of pro-inflammatory cytokines, particularly IL-1β and complement. Colchicine inhibits the microtubule formation that is necessary for this process to occur.

Clinical features

Hyperuricaemia may be asymptomatic or may cause:

- ***Acute gout***, followed by an asymptomatic intercritical phase; a second acute attack is likely within 2 years.
- ***Chronic interval gout***, with acute attacks superimposed on low-grade inflammation and potential joint damage.

Box 32.37 Causes of hyperuricaemia

Impaired excretion of uric acid
- Chronic renal disease (clinical gout unusual)
- Drug therapy, e.g., thiazide diuretics, low-dose aspirin
- Hypertension
- Lead toxicity
- Primary hyperparathyroidism or hypothyroidism
- Increased lactic acid production from alcohol, exercise, starvation
- Glucose-6-phosphatase deficiency (interferes with renal excretion)

Increased production of uric acid

Increased purine synthesis de novo due to:
- Hypoxanthine–guanine–phosphoribosyl transferase (HGPRT) reduction (an X-linked inborn error causing the Lesch–Nyhan syndrome)
- Phosphoribosyl-pyrophosphate synthase overactivity
- Glucose-6-phosphatase deficiency with glycogen storage disease type 1

Increased turnover of purines due to:
- Myeloproliferative disorders, e.g., polycythaemia vera
- Lymphoproliferative disorders, e.g., leukaemia
- Others, e.g., carcinoma, severe psoriasis

- ***Chronic polyarticular tophaceous gout***, which is rare and characterized by chronic joint pain, activity limitation, structural joint damage, and frequent flares.
- ***Urate renal stone*** formation (see Chapter 26).

Acute gout presents typically in a middle-aged male with a sudden onset of agonizing pain, swelling, and redness of the first MTP joint (**Fig. 32.16**). The attack may be precipitated by excess food, alcohol, dehydration, or diuretic therapy. Untreated attacks last about 7 days. Recovery is typically associated with desquamation of the overlying skin. In 25% of attacks, a joint other than the great toe is affected. Triggers for flares include acute medical or surgical illness, dehydration, or dietary factors such as alcohol intake or purine-rich foods.

In severe attacks, overlying crystal cellulitis makes gout difficult to distinguish clinically from infective cellulitis. A family or personal history of gout and the finding of a raised SUA suggest the diagnosis but, if in doubt, blood and joint fluid cultures should be taken to exclude sepsis.

Chronic tophaceous gout is described below.

Investigations

The clinical picture is often diagnostic, as is the rapid response to NSAIDs or colchicine.

- ***Joint fluid microscopy*** is the most specific and diagnostic test but is technically difficult.
- ***Serum uric acid*** is usually raised (>600 μmol/L). If it is not, it should be rechecked several weeks after the attack, as levels fall immediately after an acute episode. Acute gout rarely occurs with SUA in the lower half of the normal range below the saturation point of 360 μmol/L.
- ***Serum urea, creatinine***, and estimated glomerular filtration rate (eGFR) are monitored for signs of renal impairment.

Management

The use of NSAIDs or coxibs in high doses rapidly reduces the pain and swelling. The first dose should be taken at the first indication of an attack:

- ***Naproxen:*** 750 mg immediately, then 500 mg every 8–12 h
- ***Diclofenac:*** 75–100 mg immediately, then 50 mg every 6–8 h.

After 24–48 hours, reduced doses are given for a further week.

Caution: NSAIDs may cause renal impairment. In individuals with renal impairment or a history of peptic ulceration, alternative treatments include:

- ***Colchicine:*** loading doses will cause diarrhoea or colicky abdominal pain, so 500 μg 2–3 times per day is usually sufficient to terminate attacks without side-effects.
- ***Corticosteroids:*** oral prednisolone or intramuscular or intra-articular depot methylprednisolone is used.

Dietary advice

The first attacks may be separated by up to 2 years and are managed symptomatically. Recognition and prompt treatment, including lifestyle modifications, are paramount. Advice to reduce alcohol intake, especially beer, which is high in purines and fructose, and consumption of non-diet carbonated soft drinks, also high in fructose, is essential; other dietary change include reduction of total calorie and cholesterol intake, and avoidance of purine-rich foods, such as offal, red meat, shellfish and spinach. These modifications can reduce serum urate by 15% and delay the need for drugs that reduce serum urate levels. Dietary advice is readily available on the Internet.

Treatment with agents that reduce serum uric acid levels

The aim of treatment is to reduce and maintain the uric acid level below 360 μmol/L in all patients, and in those with severe gout, such as tophi or frequent attacks, a lower level below 300 μmol/L is required to achieve clinical remission.

Allopurinol

Allopurinol should only be used when the attacks are frequent and severe (despite dietary changes), or associated with renal impairment or tophi, or when the patient finds NSAIDs or colchicine difficult to tolerate. Allopurinol is a xanthine oxidase inhibitor, which reduces SUA levels rapidly; it is relatively non-toxic but should be used at low doses (50–100 mg) in renal impairment. It can be started during an acute attack and always under cover of NSAIDs or colchicine for the first 2–4 weeks before and 4 weeks after starting allopurinol, as it may induce acute gout. Theories to explain allopurinol-induced exacerbation of gout are controversial and lack evidence, generally implicating urate concentration flux and remodelling/destabilization of intra-articular microscopic tophi. The dose can be increased gradually from 100 mg every few weeks until the uric acid level is below the 360 μmol/L level. Skin rashes and gastrointestinal intolerance are the most common side-effects. A hypersensitivity reaction is the most serious but rare adverse event, as is bone marrow suppression. Xanthine oxidase inhibitors (including both allopurinol and febuxostat) may provoke life-threatening agranulocytosis and pancytopaenia if inadvertently prescribed to patients taking the anti-metabolite immunosuppressant drugs azathioprine or 6-mercaptopurine at conventional doses.

Febuxostat

Febuxostat (80–120 mg) is a non-purine analogue inhibitor of xanthine oxidase that is well tolerated and is an effective alternative to allopurinol. It is safer in renal impairment, as it undergoes hepatic metabolism rather than renal excretion, and is helpful in patients who cannot tolerate allopurinol. All-cause mortality and cardiovascular mortality, however, are higher with febuxostat than with allopurinol.

Pegloticase

Pegloticase, a pegylated recombinant uricase given intravenously, lowers urate levels dramatically and is recommended in US

guidance for patients with severe, refractory gout in whom target serum urate concentrations are not achieved or who cannot tolerate oral urate-lowering therapy (ULT). The drug was withdrawn from use in the European Union in 2016 due to reports of infusion reactions or anaphylaxis with concomitant use of oral ULT.

Uricosuric agents

These also lower the SUA but their use is restricted throughout Europe by the very rare occurrence of serious hepatotoxicity. ***Benzbromarone*** acts on the URAT1 transporter and is well tolerated. ***Sulfinpyrazone*** and ***probenecid*** are best avoided in renal impairment. Availability of these drugs varies between countries—in the UK, benzbromarone and probenecid can be obtained for treating named patients.

Losartan

Losartan is an angiotensin I receptor antagonist and is uricosuric in hypertensive patients with gout. It may reduce the risk of gout in patients with the metabolic syndrome.

Targeted therapies

The IL-1 receptor inhibitor anakinra, canakinumab (an anti-human IL-1β monoclonal antibody), and rilonacept (a dimeric fusion protein consisting of portions of IL-1R and the IL-1R accessory protein linked to the Fc portion of IgG1) have shown efficacy. The high cost of these agents, however, means they may only be considered in patients resistant to standard therapy.

Chronic tophaceous gout

Individuals with persistently high levels of uric acid can present with chronic tophaceous gout, as sodium urate forms smooth white deposits (tophi) in skin and around joints, on the ear, fingers (Fig. 32.35) or the Achilles tendon. Large deposits are unsightly and ulcerate. There is chronic joint pain and sometimes superimposed acute gouty attacks.

Periarticular deposits lead to a halo of radio-opacity and clearly defined ('punched out') bone cysts on X-ray.

Tophaceous gout is often associated with renal impairment and/or the long-term use of diuretics. There may be acute or chronic urate nephropathy or renal stone formation. Whenever possible, diuretics should be stopped or changed to less urate-retaining ones, such as bumetanide. Treatment is with allopurinol and/or uricosuric agents (see above).

Calcium pyrophosphate dihydrate deposition arthropathy

Deposition of calcium pyrophosphate dehydrate (CPP) crystals in hyaline and fibrocartilage is the most common cause of CC.

Fig. 32.35 Gout with tophi. *(Courtesy of Dr Varkey Alexander, Al-Sabah Hospital, Kuwait.)*

Previous terms, including pseudogout, have been unified into calcium pyrophosphate dihydrate deposition (CPPD) arthropathy. CPPD-associated arthritis is the third most common inflammatory arthritis, which increases with age, OA, joint trauma/injury, and metabolic disease (hyperparathyroidism, haemochromatosis, hypomagnesaemia), and has a familial predisposition. Shedding of crystals into a joint precipitates acute synovitis that resembles gout, except that it is more common in elderly females and usually affects the knee, ankle, or wrist (Fig. 32.16).

Diagnosis

Diagnosis is made by detecting rhomboidal, weakly positively birefringent crystals in joint fluid (see Fig. 32.33), or is deduced from the presence of CC on X-ray. Joint fluid looks purulent so should be sent for culture to exclude septic arthritis, as the attacks may also be associated with fever and a raised white blood cell count.

Management

There is no specific treatment to eliminate calcium pyrophosphate crystals. Therefore, management is focused on symptom control and treatments overlap with those for gout and OA. An evidence base is lacking in this condition, however, and there have been no randomized controlled clinical trials of acute management. Aspiration of the joint reduces the pain dramatically but it is usually necessary to use an NSAID or colchicine, as for gout. If infection can be excluded, an intra-articular injection of a corticosteroid helps. CPP crystals activate the NLRP3 inflammasome-IL-1 pathway (see Chapter 6) and although benefits have been shown with anakinra in small studies, it was not superior to prednisolone.

Basic calcium phosphate deposition disease

Basic calcium phosphate (BCP) crystals, including hydroxyapatite, tricalcium phosphate, and octacalcium phosphate, can be deposited in any tissues, particularly at intra-articular and periarticular locations. Crystal formation is thought to be partly regulated by the effects of extracellular inorganic phosphate on chondrocytes and is linked with the development of OA. BCP crystals cause joint damage through induction of fibroblast proliferation, inflammatory cytokines (IL-1β and TNF-α), nitric oxide, and metalloproteinases.

BCP crystal deposits in periarticular soft tissue may be asymptomatic or can cause acute calcific periarthritis (particularly in the supraspinatus tendon), tendonitis, bursitis, and enthesitis. Less frequently, they can cause arthritis. In acute attacks, periarticular tissues or joints can be swollen, tender, and hot. This presentation may mimic cellulitis, gout, CPPD disease, and septic arthritis. Periarticular or articular BCP crystal deposits are also found in an extremely destructive chronic arthropathy of the elderly that occurs most often in shoulders (Milwaukee shoulder) or other large joints, such as the hips and knees, and in erosive OA of the fingers.

Diagnosis

The diagnosis is made clinically and septic arthritis should be excluded. A neutrophilia with elevation of ESR and CRP may occur during an acute attack. Intra- and/or periarticular calcifications, with or without erosive, destructive, or hypertrophic changes, may be seen on X-ray. Alizarin red S staining is not specific for BCP crystals in synovial fluid and other calcium-containing particulates are also stained by this method.

Management

Management of acute attacks is similar to that of CPPD disease with NSAIDs, oral colchicine, aspiration of effusions, and/or

intra-/periarticular injection of steroids to try to shorten the duration of symptoms. In patients with underlying progressive articular changes, the response to medical therapy is usually less rewarding. Total joint replacement may be required for patients with severe destructive arthropathy in large joints.

Further reading

Andres M, Sivera F, Pascual E. Therapy for CPPD: options and evidence. *Curr Rheumatol Rep* 2018; 20:31.
Dalbeth N, Gosling AL, Gaffo G, Abhishel A. Gout. *Lancet* 2021; 397:1843.
Gossec L, Baraliakos X, Kerschbaumer A et al. EULAR recommendations for the management of psoriatic arthritis with pharmacological therapies: 2019 update. *Ann Rheum Dis* 2020; 79:700.
Raine C, Keat A. Axial spondylarthritis. *Medicine* 2018; 46:231–236.
Ramiro S, Nikiphorou E, Sepriano A et al. ASAS-EULAR recommendations for the management of axial spondyloarthritis: 2022 update. *Ann Rheum Dis* 2023; 82:19.
Ritchlin CT, Colbert RA, Gladman D. Psoriatic arthritis. *N Eng J Med* 2017; 376:957–970.
Smolen JS, Aletaha D, Barton A et al. Rheumatoid Arthritis. *Nat Rev Dis Primers* 2018; 4:18001.
Smolen JS, Landewè RBM, Bergstra SA et al. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2022 update. *Ann Rheum Dis* 2023; 82:3.
http://basdai.com *Bath Ankylosing Spondylitis Disease Activity Index*.

INFECTION OF JOINTS

For general principles in diagnosis and management of the acutely septic joint and discussion of common pathogens, see Chapter 11.

Specific types of bacterial arthritis

Gonococcal arthritis

Disseminated gonococcal infection occurs in 0.5%–3% of patients infected with *Neisseria gonorrhoeae*. It is a common cause of acute arthritis in previously fit young adults, with increasing rates in males in general and men who have sex with men.

Patients present with a fever and characteristic pustules on the distal limbs, often in association with polyarthralgia and tenosynovitis. Blood cultures are reportedly positive in up to 70% of cases, particularly those with skin lesions, tenosynovitis, and polyarthralgia. Nucleic acid amplification tests may be positive even when cultures are negative. Patients should be screened for other STIs. Later, large-joint mono- or pauciarticular arthritis may follow. Culture is usually positive from the genital tract, although the joint fluid may be sterile. It is not clear whether this joint inflammation is simply a septic arthritis (which responds rapidly to antibiotics), or whether there is also a reactive element to a variety of virulence and growth factors specific to *N. gonorrhoeae*.

Management consists of initial i.v. ceftriaxone (or cefotaxime or ceftizoxime) repeatedly until clinical improvement occurs, plus a single oral dose of azithromycin or alternatively a week of doxycycline. In the absence of purulent arthritis after 2–3 days of improvement a 7-day course of therapy may be completed with intramuscular ceftriaxone. Purulent arthritis, however, often requires 7–14 days of i.v. therapy and joint drainage.

Tuberculous arthritis

Around 1% of people with tuberculosis develop joint and/or bone involvement. It occurs as the primary disease in children. In adults, it is usually due to haematogenous spread from pulmonary or renal lesions. The onset is insidious and diagnosis often delayed.

The organism invades the synovium or intervertebral disc. There are caseating granulomas and rapid destruction of cartilage and adjacent bone. Some patients develop a reactive polyarthritis (Poncet's disease).

Osteoarticular TB most commonly affects the spine (50%), hip (12%–15%), knee (10%), and ribs (10%). Patients become febrile with night sweats, anorexia, and weight loss. The usual risk factors for tuberculosis apply—debility, excess alcohol use, or immunosuppression. People with untreated or poorly controlled HIV infection or AIDS are at particular risk.

Investigations should include culture of fluid, and culture and biopsy of the synovium. *Mycobacterium tuberculosis* is the usual organism but atypical mycobacteria are occasionally implicated. A chest X-ray should be performed. Initially, joint or spinal X-rays may be normal but joint-space reduction and bone destruction develop rapidly if treatment is delayed. MRI shows the abnormality earlier in the spine and CT-guided biopsy from the affected disc is often necessary to obtain cultures.

Management is as for pulmonary tuberculosis with 6 months therapy in people with active TB without CNS involvement and 12 months in the presence of CNS involvement. The joint should be rested and the spine immobilized in the acute phase.

Meningococcal arthritis

Joint inflammation may complicate meningococcal septicaemia, presenting as a migratory polyarthritis. Organisms are rarely cultured from the joint and most cases are due to immune-complex deposition. Treatment is urgent with immediate penicillin therapy.

Infective endocarditis

This condition may present with arthralgia, PMR-like symptoms, or an infective arthritis (see Chapter 17).

Lyme arthritis

About 25% of people with Lyme disease develop arthralgia, less commonly an acute pauciarticular arthritis (see Chapter 42); this usually resolves but 10% of untreated cases go on to develop a chronic arthritis. There are no positive markers of an ongoing infection in these patients).

Diagnosis is by the detection of IgM antibodies against the spirochaete *Borrelia burgdorferi*.

Treatment with antibiotics (amoxicillin or doxycycline) is highly effective in early disease.

Brucellosis

Brucellosis (see Chapter 42) has a worldwide distribution. The most common cause of chronic brucellosis and of arthritis is *Brucella melitensis*. There is usually a migratory large-joint mono- or oligoarticular arthritis, which is septic or reactive. Arthritis is more common in chronic infections of more than 6 months.

Leprosy

Acute or chronic symmetrical polyarthritis resembling RA, swollen hands and feet due to lepra reactions, tenosynovitis, and thickened nerves with or without cutaneous manifestations are seen in leprosy (see Chapter 42).

Arthritis in viral disease

A transient polyarthritis or arthralgia can occur before, during, or after many viral illnesses. These include infectious mononucleosis, chickenpox, mumps, adenovirus, rubella, erythrovirus B19, hepatitis B and C, arboviral infections, and HIV. In most of these, it is due to a direct toxic effect or immune-complex deposition.

Rubella

In rubella (see Chapter 42), the virus can occasionally be isolated from the joint. This arthritis is rare in countries where rubella vaccination is routine. It occurs in up to 50% of young adult females a few days after rubella infection (6% of males). It is a symmetrical polyarthritis involving the MCP or PIP joints most commonly, but many joints can be affected. It closely resembles RA. IgM rubella antibodies are present. It resolves within a few weeks in most cases. A mild arthritis occurs rarely 2–4 weeks after rubella vaccination.

Erythrovirus B19

Erythrovirus B19 (see Chapter 42) causes an acute, self-limiting arthritis and is associated with erythema infectiosum ('slapped cheek disease').

Hepatitis

In hepatitis B infection (see Chapter 23), a sudden, symmetrical polyarticular arthritis of the small joints of the hands occurs in approximately one-third of patients, often in the prodromal phase, and mostly resolves before the onset of jaundice. Hepatitis C infection causes type II mixed cryoglobulinaemia (see Chapter 26).

Arbovirus

Arbovirus infections (see Chapter 42), which are endemic in many parts of the world, give rise to an arthralgia and/or arthritis. For example, the Ross River virus causes an epidemic polyarthritis in Australia and the South Pacific; it involves the small joints of the hands and clears in 2–4 weeks. Other viral infections causing epidemic arthritis include chikungunya and o'nyong-nyong.

Musculoskeletal aspects of infection with HIV and AIDS

Musculoskeletal manifestations are common in these patients and are often caused by triggers such as opportunistic infections and drug therapy rather than HIV itself. Infective arthritis seen in these immunosuppressed patients often has minimal symptoms and signs. Certain of the antiviral agents cause an acute arthritis, possibly because of crystallization in the joint.

Arthralgia is common in AIDS. There is a ***seronegative, predominantly lower-limb arthritis***, similar to psoriasis or reactive arthritis. ***Spondylitis*** also occurs but is not associated with HLA-B27. ***Avascular necrosis***, possibly associated with corticosteroids or alcohol, is seen.

Non-articular diseases, such as Sjögren- and lupus-like syndromes, systemic vasculitis of the necrotizing and hypersensitivity types, and myositis also occur.

Fungal infections

Fungal infections of joints occur rarely. Bone abscesses may be seen. Destructive joint lesions can also occur with blastomycosis. A benign polyarthritis accompanied by erythema nodosum occasionally occurs in coccidioidomycosis and histoplasmosis. Culture of purulent synovial fluid and skin tests for fungi may help the diagnosis.

Further reading

Coakley G, Mathews CJ. Septic arthritis. *BMJ Best Prac* 2018; 1–31.

AUTOIMMUNE RHEUMATIC DISEASES

Autoimmune diseases are conditions in which the immune system attacks tissues of the body. The antigens can be present in multiple organs so the clinical manifestations are systemic and diverse. In some diseases, such as Graves' disease, Hashimoto's thyroiditis, and type 1 diabetes mellitus, only a single organ is affected. Each individual autoimmune rheumatic disease (ARD) has a characteristic pattern of symptoms and signs, which are used to make the diagnosis. In some ARDs, there are also characteristic autoantibodies (i.e., antibodies that recognize antigens that are normal constituents of the body, such as DNA and phospholipids). Positive blood tests for autoantibodies are useful but not essential in the diagnosis of ARDs (Box 32.38).

Systemic lupus erythematosus

Systemic lupus erythematosus (SLE) is an inflammatory, multisystem autoimmune disorder with arthralgia and rashes as the most common clinical features, and cerebral and renal disease as the most serious problems.

Epidemiology

SLE occurs worldwide and is about nine times as common in females as in males, with a peak age of onset between 20 and 40 years. The prevalence varies between ethnic groups, being highest (at 1:250) in females of African or Caribbean background. In other populations, the prevalence varies between 1:1000 and 1:10000.

Aetiology

The cause is unknown but there are several predisposing factors:

- ***Heredity.*** There is a higher concordance rate in monozygotic twins (up to 25%) compared with dizygotic twins (3%).

Box 32.38 Antinuclear autoantibodies and disease associations

Antibody	Disease	Prevalence
ds-DNA	SLE	70%
Anti-histone	Drug-induced lupus	–
Anti-centromeric	Limited scleroderma	70%
Anti-Ro (SS-A)	SLE	40%–60%
	Primary Sjögren's	60%–90%
Anti-La (SS-B)	SLE	15%
	Primary Sjögren's	35%–85%
Anti-Sm	SLE	10%–25% (Caucasian)
		30%–50% (black African)
Anti-UI-RNP	SLE	30%
	Overlap syndrome	
Anti-Jo-1 (antisynthetase)	Polymyositis	30%
	Dermatomyositis	
Anti-topoisomerase-1 (Scl-70)	Diffuse cutaneous SSc	30%

ds-DNA, Double-stranded DNA; *RNP,* ribonucleoprotein; *Ro, La,* first two letters of name of patients; *SLE,* systemic lupus erythematosus; *Sm,* Smith, patient's name; *SS-A, SS-B,* Sjögren's syndrome A and B; *SSc,* systemic scleroderma.

First-degree relatives have a 3% chance of developing the disease but approximately 20% have autoantibodies.

- ***Genetics.*** Over 50 whole-genome analyses have led to the identification of a large number of genes linked to the development of SLE. These include some HLA genes, as well as genes involved in T- and B-lymphocyte function and genes related to autophagy and interferon function. Homozygous deficiencies of the complement genes C1q, C2, or C4 are very rare but convey a high risk of developing SLE.
- ***Sex hormone status.*** Premenopausal females are most frequently affected.
- ***Drugs.*** Drugs such as hydralazine, isoniazid, procainamide, and penicillamine can induce a form of SLE that is usually mild, in that the kidneys and CNS are not affected.
- ***Ultraviolet light.*** This can trigger flares of SLE, especially in the skin.
- ***Exposure to Epstein–Barr virus.*** This has been suggested as a trigger for SLE.

Pathogenesis

When cells die by apoptosis, the cellular remnants appear on the cell surface as small blebs that carry self-antigens. These antigens include nuclear constituents (e.g., DNA and histones), which are normally hidden from the immune system. In people with SLE, removal of these blebs by phagocytes is inefficient, so that they are transferred to lymphoid tissues, where they can be taken up by antigen-presenting cells. The self-antigens from these blebs can then be presented to T cells, which in turn stimulate B cells to produce autoantibodies directed against these antigens (see Chapter 6). It has been shown that, in some patients, the autoantibodies are present in stored blood samples that were taken years before the patient developed clinical features of SLE. The combination of availability of self-antigens and failure of the immune system to inactivate B cells and T cells that recognize these self-antigens (i.e., a breakdown of *tolerance*; see Chapter 6) leads to the following immunological consequences:

- Development of autoantibodies that either form circulating complexes or deposit by binding directly to tissues.
- Activation of complement and influx of neutrophils, causing inflammation in those tissues.
- Abnormal cytokine production: increased blood levels of IL-10 and interferon-alpha are particularly closely linked to high activity of inflammation in SLE, and anifrolumab—a monoclonal antibody that targets the interferon receptor—has been shown to be effective in the treatment of some patients with SLE.

More recently, other cells have been implicated in the pathogenesis of SLE. These include dendritic cells, which are activated by immune complexes containing nucleic acids to produce interferon, and invariant natural killer T cells (iNKT cells), which are reduced in number and function in patients with SLE.

Pathology

SLE of the skin and kidneys is characterized by deposition of complement and IgG antibodies, and by influx of neutrophils and lymphocytes. Biopsies of other tissues are carried out less frequently but can show vasculitis affecting capillaries, arterioles, and venules. The synovium of joints can be oedematous and may contain immune complexes. Haematoxylin bodies (rounded, blue, homogeneous haematoxylin-stained deposits) are seen in inflammatory infiltrates and are thought to result from the interaction of antinuclear antibodies and cell nuclei.

The pathology of lesions in other organs is described in the appropriate chapters.

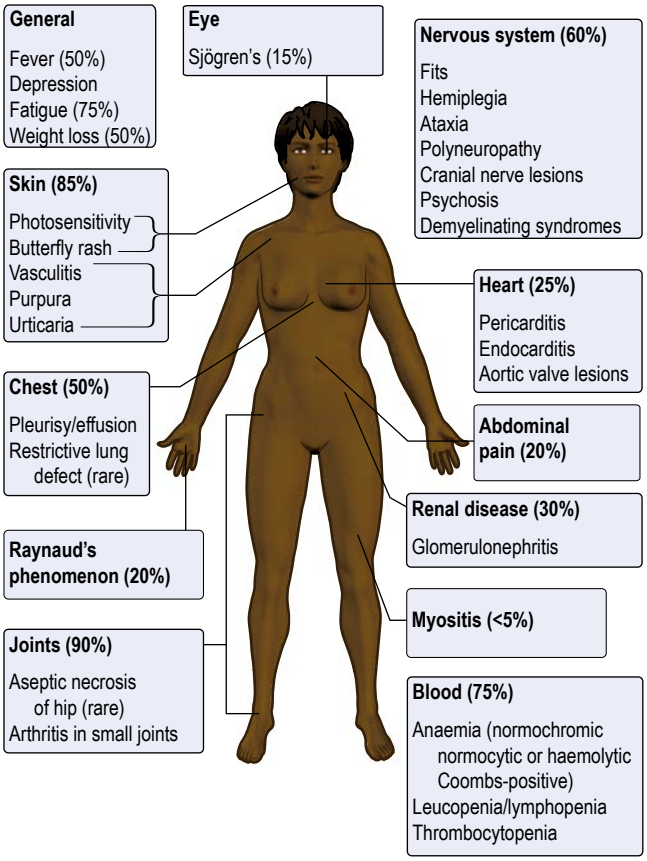


Fig. 32.36 Clinical features of systemic lupus erythematosus.

Clinical features

The manifestations of SLE vary greatly between patients. Most patients suffer fatigue, arthralgia, and/or skin problems. Involvement of major organs is less common but more serious (Fig. 32.36).

General features

Fever is common in exacerbations. Patients complain of marked malaise and tiredness, and these symptoms do not correlate with disease activity or severity of organ-based complications.

Joints and muscles

Joint involvement is the most common clinical feature (>90%). Patients often present with symptoms resembling RA with symmetrical small joint arthralgia. Joints are painful but characteristically appear clinically normal, although sometimes there is slight soft tissue swelling surrounding the joint. Deformity because of joint capsule and tendon contraction is rare, as are bony erosions. Rarely, major joint deformity resembling RA (known as ***Jaccoud's arthropathy***) is seen. These deformities are reducible, unlike those seen in severe RA. Avascular necrosis affecting the hip or knee is a rare complication of the disease or of treatment with corticosteroids.

Myalgia is present in up to 50% of patients but a true myositis is seen in only less than 5%. If myositis is prominent, the patient may well have an overlap ARD with both polymyositis and SLE.

Skin

The skin (see Chapter 45) is affected in 85% of cases. Erythema, in a 'butterfly' distribution on the cheeks of the face and across the bridge of the nose (see Fig. 45.42), is characteristic. Vasculitic lesions on the fingertips and around the nail folds, purpura, and urticaria

occur. In 40%–50% of cases, there is photosensitivity (especially in patients positive for anti-Ro antibodies). Prolonged exposure to sunlight can lead to exacerbations of the disease. Livedo reticularis, palmar and plantar rashes, pigmentation, and alopecia are seen. Scarring alopecia can lead to irreversible bald patches. Raynaud's phenomenon (see Chapter 17) is common and may precede the development of other clinical problems by years.

Discoid lupus is a benign variant of lupus in which only the skin is involved. The rash is characteristic and appears on the face as well-defined erythematous plaques that progress to scarring and pigmentation (see Chapter 45). Subacute cutaneous lupus erythematosus, a rare variant, is also described in Chapter 45.

Lungs

Up to 50% of patients will have lung involvement at some time during the course of the disease (see Chapter 20). Recurrent pleurisy and pleural effusions (exudates) are the most common manifestations and are often bilateral. Pneumonitis and atelectasis are seen; in some patients a restrictive lung defect develops, with loss of lung volumes and raised hemidiaphragms. This 'shrinking lung syndrome' is poorly understood but may have a neuromuscular basis. Rarely, pulmonary fibrosis occurs, more commonly in overlap syndromes (see "Overlap' syndromes and undifferentiated autoimmune rheumatic disease' section). Intrapulmonary haemorrhage associated with vasculitis is a rare but potentially life-threatening complication.

Heart and cardiovascular system

The heart is involved in 25% of cases. Pericarditis, with small pericardial effusions detected by echocardiography, is common, although rarely large effusions may occur resulting in cardiac tamponade. A mild myocarditis also occurs, giving rise to arrhythmias. Aortic valve lesions and a cardiomyopathy can rarely be present. A non-infective endocarditis involving the mitral valve (Libman–Sacks syndrome) is very rare. Raynaud's, vasculitis, and arterial and venous thromboses can occur, especially in association with the antiphospholipid syndrome (see below). There is an increased frequency of ischaemic heart disease and stroke in people with SLE, which is partly due to altered levels of common risk factors such as hypertension and lipid levels, but the presence of chronic inflammation over many years may also be contributory. It is not known whether intensive treatment of cardiovascular risk factors or use of statins in SLE will alter the risk of developing coronary disease or stroke.

Kidneys

A classification of types of nephritis is given in Chapter 36. Postmortem studies suggest that histological changes are very frequent, but clinical renal involvement occurs in only approximately 30% of cases. All patients should have regular screening of urine for blood and protein. An asymptomatic patient with proteinuria may be in the early stages of lupus nephritis, and treatment may prevent progression to renal impairment. Proteinuria should be quantified and haematuria should prompt examination for urinary casts or dysmorphic red cells that suggest glomerulonephritis. Renal biopsy is used to define the type and severity of nephritis. Renal vein thrombosis can occur in nephrotic syndrome or in association with antiphospholipid antibodies.

Nervous system

Involvement of the nervous system occurs in up to 60% of cases and symptoms often fluctuate. There may be a mild depression, but occasionally, more severe psychiatric disturbances occur. Epilepsy, migraines, cerebellar ataxia, aseptic meningitis, cranial nerve lesions, cerebrovascular disease, or a polyneuropathy may be seen. The pathogenic mechanism for cerebral lupus is complex. Lesions may be due to vasculitis or immune complex deposition, thrombosis, or non-inflammatory microvasculopathy. In people with cerebral lupus, infection should be excluded, or treated in parallel with administration of corticosteroids and immunosuppression.

Eyes

Retinal vasculitis can cause infarcts, which appear as hard exudates, and haemorrhages. There may be episcleritis, conjunctivitis, or optic neuritis, but blindness is uncommon. Secondary Sjögren's syndrome is seen in about 15% of cases.

Gastrointestinal system

Mouth ulcers are common and may be a presenting feature. These may be painless or become secondarily infected and painful. Mesenteric vasculitis can produce inflammatory lesions involving the small bowel (infarction or perforation). Liver involvement and pancreatitis are uncommon.

Investigations

Blood

- ***A full blood count*** may show a leucopenia, lymphopenia, and/or thrombocytopenia. Anaemia of chronic disease or autoimmune haemolytic anaemia also occurs. The ESR is raised in proportion to the disease activity. In contrast, the CRP is usually normal but may be high when the patient has lupus pleuritis or peritonitis, arthritis, or a coexisting infection.
- ***Urea and creatinine*** only rise when renal disease is advanced. Low serum albumin, high urine albumin/creatinine ratio, or an active urinary sediment are earlier indicators of lupus nephritis.
- ***Autoantibodies*** of many different types may be present in SLE but the most significant are ANA, anti-dsDNA, anti-Ro, anti-Sm, and anti-La (see **Box 32.38**). Antiphospholipid antibodies are present in 25%–40% of cases but not all of these patients develop antiphospholipid syndrome (see later).
- ***Serum complement*** C3 and C4 levels are often reduced during active disease. The combination of high ESR, high anti-dsDNA, and low C3 may herald a flare of disease. All these markers tend to return towards normal as the flare improves, but in some patients, anti-dsDNA levels will remain high even during clinical remission.

Histology

Characteristic histological and immunofluorescent abnormalities (deposition of IgG and complement) are seen in biopsies from the kidney and skin.

Diagnostic imaging

CT scans of the brain sometimes show infarcts or haemorrhage with evidence of cerebral atrophy. MRI can detect lesions in white matter that are not seen on CT. However, it can be very difficult to distinguish true vasculitis from small thrombi.

Management

General measures

The disease and its management should be discussed with the patient, particularly the effect upon the patient's lifestyle: for example, appearance and debility due to fatigue. Patients are advised to avoid excessive exposure to sunlight and it is also necessary to reduce cardiovascular risk factors.

Symptomatic treatment

Arthralgia, arthritis, fever, and serositis all respond well to standard doses of NSAIDs (see 'Non-steroidal anti-inflammatory drugs' section). Topical corticosteroids are effective and widely used in cutaneous lupus. Antimalarial drugs (chloroquine or hydroxychloroquine) help mild skin disease, fatigue, and arthralgias that cannot be controlled with NSAIDs, but patients require regular eye checks because of potential retinal toxicity. It is now thought that retinal changes occur in up to 7.5% of patients on long-term treatment with hydroxychloroquine, with increased risk in those taking more than 5 mg/kg per day.

Corticosteroids and immunosuppressive drugs

Single intramuscular injections of long-acting corticosteroids or short courses of oral corticosteroids are useful in treating severe flares of arthritis, pleuritis, or pericarditis. In some cases, these symptoms can only be kept under control using long-term oral corticosteroids.

Renal or cerebral disease and severe haemolytic anaemia or thrombocytopenia must be treated with high-dose oral corticosteroids, and the first two of these require immunosuppressive drugs in addition. Cyclophosphamide was most commonly used to achieve remission in these severe forms of lupus but has largely been replaced by mycophenolate mofetil, which has fewer side-effects. Azathioprine is also used to maintain remission. Calcineurin inhibitors (ciclosporin, tacrolimus, or voclosporin) and methotrexate are also used, but in fewer cases than azathioprine or mycophenolate. Newer agents, which specifically target cells or cytokines in the immune system, are coming into use, especially in refractory cases. These include rituximab (anti-CD20) and belimumab (anti-BLyS), which are both monoclonal antibodies acting against B lymphocytes, and the anti-interferon receptor agent anifrolumab. There have been case reports of successful use of chimeric antigen receptor T cells (CAR-T cells) targeted against B cells in refractory SLE. Currently biologics remain second-line options where standard immunosuppressives have failed and CAR-T cells have only been used in a small number of cases of severe refractory disease.

Prognosis

An episodic course is characteristic, with exacerbations and complete remissions that may last for long periods. However, SLE can also be a chronic persistent condition. The mortality rate in SLE has fallen dramatically over the last 50 years; the 10-year survival rate is about 90%, but this is lower if major organ-based complications are present. Deaths early in the course of disease are mainly due to renal or cerebral disease or infection. Later, coronary artery disease and stroke become more prevalent. Chronic progressive destruction of joints, as seen in RA and OA, occurs rarely, but a few patients develop deformities such as ulnar deviation. People with SLE have an increased long-term risk of developing some cancers, especially lymphoma.

Pregnancy and SLE

Fertility is usually normal, except in severe disease, and there is no major contraindication to pregnancy. Recurrent miscarriages can occur, especially in women with antiphospholipid antibodies. Exacerbations can arise during pregnancy and are frequent postpartum. The patient's medications should be reviewed. Mycophenolate is teratogenic and should be stopped at least 6 weeks in advance of conception, whereas azathioprine, hydroxychloroquine, and low-dose oral corticosteroids are safe. Hypertension must be controlled.

People with anti-Ro or anti-La antibodies have a 2% risk of giving birth to babies with neonatal lupus syndrome (rash, hepatitis, and fetal heart block).

Antiphospholipid syndrome

Patients who have thrombosis (arterial or venous) and/or recurrent miscarriages (see Chapter 47), and who also have persistently positive blood tests for antiphospholipid antibodies (aPL), have the antiphospholipid syndrome (APS). Antiphospholipid antibodies can be detected by several different tests:

- The ***anticardiolipin test***, which detects antibodies (IgG or IgM) that bind the negatively charged phospholipid, cardiolipin.
- The ***lupus anticoagulant test***, which detects changes in the ability of blood to clot in a test tube. Despite the name, it is not a test for lupus. It is a test for APS. The ***anticoagulant*** effect caused by aPL in the test tube causes an opposite ***procoagulant*** effect inside the body because the balance of factors stimulating thrombosis is different there.
- The ***anti-β_2-glycoprotein I test***, which detects antibodies that bind β_2-glycoprotein I, a molecule that interacts closely with phospholipids.

A persistently positive test (i.e., positive on at least two occasions ≥12 weeks apart) in one or more of these assays is needed to diagnose APS. However, some people who test positive for aPL will never develop APS; that is, not all aPLs are harmful. APS can present in patients who already have another ARD, especially SLE. APS can also occur on its own (primary APS).

New diagnostic tests for APS are being developed and include anti-prothrombin assays, tests for IgA aPL, and tests for antibodies to the N-terminal domain of β_2-glycoprotein I. Although not currently used in clinical practice, these tests may soon be helpful for assessing the overall thrombosis risk of a patient with APS, or for diagnosing patients with non-criteria APS who have typical clinical features but test negative for anticardiolipin, lupus anticoagulant and anti–β_2-glycoprotein I.

Pathogenesis

Negatively charged phospholipids and β-glycoprotein I are present on the outer surface of apoptotic blebs, and so aPLs are believed to arise by a similar mechanism to the lupus autoantibodies described above. Pathogenic aPLs bind to the N-terminal domain of β_2-glycoprotein I. Changes in the oxidation state of β_2-glycoprotein I in patients with APS enhance the availability of this domain to bind aPLs, and this interaction is facilitated when the protein is bound to phospholipid on the surface of cells such as endothelial cells, platelets, monocytes, and trophoblasts. This change alters the functioning of those cells, leading to thrombosis and/or miscarriage.

Clinical features

Ischaemic strokes occur in about 20% of patients and DVT in about 40%. Unlike most causes of thrombophilia, APS can cause either arterial or venous thrombosis (though rarely both in the same patient). Some 27% of women who have had two or more spontaneous miscarriages have APS.

Large studies, however, show that people with APS can also have many other features, including:

- thrombocytopenia
- chorea, migraine, and epilepsy
- valvular heart disease
- cutaneous manifestations (e.g., livedo reticularis)
- positive Coombs test

- renal impairment due to ischaemia in the small renal vessels.

Catastrophic APS is a rare variant (about 1% of cases) in which multiple infarcts in different organs of the body cause failure of multiple organs simultaneously. There is a high mortality from catastrophic APS.

Management

Guidelines for managing patients with APS and people with aPL who have never had thrombosis have been published by the European Alliance of Associations for Rheumatology. In people with APS who have had one thrombosis or more, the recommended treatment to prevent further thrombosis is long-term anticoagulation with warfarin, although the guidelines say that indefinite anticoagulation may not be necessary if a particular event clearly provoked the thrombosis. The target INR is lower in people with a history of venous compared to arterial thrombosis. Pregnant women with APS, in whom warfarin is contraindicated, are given oral aspirin and subcutaneous low-molecular-weight heparin from early in gestation. This therapy reduces the chance of a miscarriage but pre-eclampsia and poor fetal growth remain common. Aspirin or clopidogrel are sometimes given prophylactically in aPL-positive people who have never had a thrombosis, especially if they are triple positive in all three of the aPL-tests—which is a known predictor of high thrombotic risk. Warfarin is given much more rarely in these circumstances, but may be indicated in patients at particularly high risk; for example, those who have other risk factors for thrombosis as well as aPL. Direct oral anticoagulants (DOACs, including factor Xa and direct thrombin inhibitors) are not routinely prescribed in APS in light of clinical trials which did not favour their use in comparison with warfarin.

Systemic sclerosis (scleroderma)

Systemic sclerosis (SSc, or scleroderma; see Chapter 45) is a multisystem disease and is distinct from localized scleroderma syndromes, such as morphea, that do not involve internal organ disease and are rarely associated with vasospasm (Raynaud's phenomenon). SSc has the highest case-specific mortality of any of the autoimmune rheumatic diseases. It occurs worldwide but there may be racial or ethnic differences in clinical features. For example, renal involvement is less frequent in Japanese people.

The incidence of SSc is 10/million population per year with a 3:1 female to male ratio. The peak incidence is between 30 and 50 years of age. It is rare in children.

Environmental risk factors for scleroderma-like disorders include exposure to vinyl chloride, silica dust, adulterated rapeseed oil, and trichloroethylene. Drugs such as bleomycin also produce a similar picture. Although unusual, familial cases are reported and twin cohorts suggest higher concordance in monozygotic pairs, consistent with genetic determinants of aetiology.

Pathology and pathogenesis

Vascular features

Widespread vascular damage involving small arteries, arterioles, and capillaries is an early lesion. There is initial endothelial cell damage with release of cytokines, including endothelin-1, which causes vasoconstriction. There is continued intimal damage with increasing vascular permeability, leading to cellular activation and activation of adhesion molecules (E-selectin, VCAM, intercellular adhesion molecule 1 [ICAM-1]), with migration of cells into the extracellular matrix. Migrating lymphocytes are IL-2-producing cells, expressing surface antigens such as CD3, CD4, and CD5. All these factors cause release of other mediators (e.g., IL-1, 4, 6, and 8, TGF-β and platelet-derived growth factor) with activation of fibroblasts. Plasma levels of the chemokine CXCL4 are elevated in SSc and correlate with skin and lung fibrosis.

The damage to small blood vessels also produces widespread obliterative arterial lesions and subsequent chronic ischaemia.

Fibrotic features

Fibroblasts synthesize increased quantities of collagen types I and III, as well as fibronectin and glycosaminoglycans, producing fibrosis in the lower dermis of the skin as well as the internal organs. It is possible that antibodies to platelet-derived growth factor receptor, which have been found in blood of people with SSc, stimulate fibroblasts to cause fibrosis.

Clinical features

See Fig. 32.37.

Raynaud's phenomenon

Raynaud's phenomenon is seen in almost all cases and can precede the onset of the full-blown disease by many years.

Limited cutaneous scleroderma (LcSSc): 70% of cases

This condition usually starts with Raynaud's phenomenon many years (up to 15) before any skin changes. The skin involvement is limited to the hands, face, feet, and forearms. The skin is tight over the fingers and often produces flexion deformities of the fingers. Involvement of the skin of the face produces a characteristic 'beak'-like nose and a small mouth (microstomia). Painful digital ulcers and telangiectasia with dilated nail-fold capillary loops are seen. Digital ischaemia may lead to gangrene. Gastrointestinal tract involvement is common. Pulmonary hypertension develops

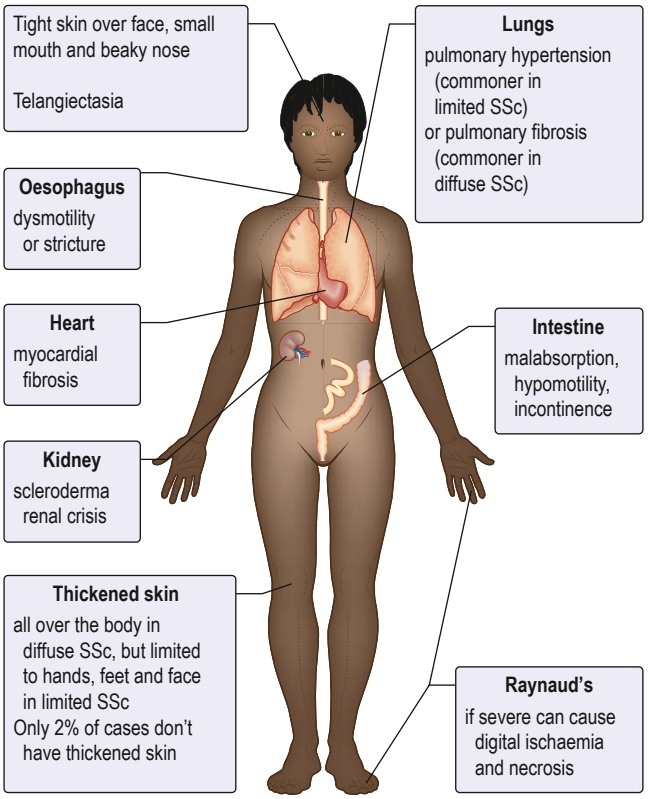


Fig. 32.37 Clinical features of systemic sclerosis (SSc).

in 21% of people with LcSSc and pulmonary interstitial disease also occurs.

Diffuse cutaneous scleroderma (DcSSc): 30% of cases

Initially oedematous in onset, skin sclerosis rapidly follows. Raynaud's phenomenon usually starts just before or concomitant with the oedema.

Diffuse swelling and stiffness of the fingers is rapidly followed by more extensive skin thickening, which can involve most of the body in the severest cases. Later, the skin becomes atrophic. Early involvement of other organs occurs with general symptoms of lethargy, anorexia, and weight loss.

- ***Gastrointestinal involvement*** includes heartburn, reflux, or dysphagia due to oesophageal involvement (see Chapter 22), which is almost invariable. Anal incontinence occurs in many patients. Malabsorption from bacterial overgrowth due to dilatation and atony of the small bowel is not infrequent; more rarely, dilatation and atony of the colon occurs. Pseudo-obstruction is a known complication.
- ***Renal involvement*** is acute or chronic (see Chapter 26). Acute hypertensive renal crisis used to be the most common cause of death in systemic sclerosis. Angiotensin-converting enzyme (ACE) inhibitors and better care, along with dialysis and renal transplantation, have changed this.
- ***Lung disease***, both fibrosis (in 41% of cases) and pulmonary hypertension (17% of cases), contributes significantly to mortality in SSc. Pulmonary hypertension can be isolated or secondary to fibrosis, and high plasma levels of endothelin-1 are seen.
- ***Myocardial fibrosis*** leads to arrhythmias and conduction defects. Pericarditis is found occasionally.

Sometimes, these systemic features occur without skin involvement (SSc *sine* scleroderma).

Investigations

- ***Full blood count.*** A normochromic, normocytic anaemia occurs and a microangiopathic haemolytic anaemia with thrombocytopaenia is seen in some people with hypertension and renal disease.
- ***Serum creatinine and electrolytes.*** Urea and creatinine rise in acute kidney injury.
- ***Autoantibodies*** (see **Box 32.38**):
  - ***In LcSSc:*** speckled, nucleolar, or anti-centromere antibodies (ACAs) occur in 70% of cases. Anti-Th/To antibodies target small nuclear ribonucleoproteins and are found in up to 7% of cases of LcSSC.
  - ***In DcSSc:*** there are anti-topoisomerase-1 antibodies (called anti–Scl-70) in 30% of cases, and anti-RNA polymerase (I, II, and III) antibodies in 20%–25%. Anti–Scl-70 is highly specific for DcSSC. Anti-RNA polymerase positivity is associated with an increased risk of renal involvement. Anti-U3RNP antibodies bind a small nuclear ribonucleoprotein called fibrillarin and are present in up to 6% of patients with SSc—more commonly in DcSSc than LcSSc.
  - ***RF*** is positive in 30%.
  - ***ANA*** is positive in 95%.
  - ***Urine*** microscopy and, if there is proteinuria, the urine albumin/creatinine ratio should be measured.
- Imaging:
  - ***Chest X-ray:*** to exclude other pathology, for changes in cardiac size and established lung disease.
  - ***Hands:*** deposits of calcium around the fingers (in severe cases, erosion and absorption of the tufts of the distal phalanges, termed 'acro-osteolysis').
  - ***Barium swallow:*** generally confirms impaired oesophageal motility. Scintigraphy, manometry, impedance, and upper gastrointestinal endoscopy are also valuable.
  - ***Echocardiogram and lung function tests*** should be carried out annually to monitor for development of pulmonary fibrosis or pulmonary hypertension.
  - ***High-resolution CT:*** to demonstrate fibrotic lung involvement.
- ***Other investigations*** of gastrointestinal tract, lung, renal, and cardiac as appropriate.

Management

Treatment should be organ-based in order to try to control the disease. Currently, there is no cure. In contrast to many other ARDs, corticosteroids and immunosuppressants are rarely used in SSc, with the exception of SSc-related pulmonary fibrosis.

- ***Education***, counselling, and family support are essential.
- ***Regular exercises*** and skin lubricants may limit contractures and methotrexate is given in early diffuse SSc to improve skin fibrosis.
- ***Raynaud's phenomenon*** may be improved by hand warmers and oral vasodilators (calcium-channel blockers, ACE inhibitors, angiotensin receptor blockers). In severe cases, especially where there are digital ulcers, oral phosphodiesterase inhibitors such as sildenafil are used as second-line therapy and parenteral vasodilators (prostacyclin analogues and calcitonin gene-related peptide) as third-line. Lumbar sympathectomy can help foot symptoms. Radical micro-arteriolysis (digital sympathectomy) can be used where individual fingers or toes are severely ischaemic, and thoracic sympathectomy under video-assisted thoracic surgery is now performed.
- ***Oesophageal symptoms*** can almost always be improved by proton pump inhibitors but prokinetic drugs are rarely helpful.
- ***Symptomatic malabsorption*** requires nutritional supplements and rotational antibiotics to treat small intestinal bacterial overgrowth.
- ***Renal involvement*** requires intensive control of hypertension. The first drug of choice is an ACE inhibitor. Vigilance for hypertensive scleroderma renal crisis (SRC) is critical, especially in early-stage DcSSc with rapidly progressive skin and tendon friction rubs. High-dose corticosteroids (above 10 mg prednisolone daily) may increase the risk of SRC.
- ***Pulmonary hypertension*** is treated with oral vasodilators, oxygen, and warfarin. Advanced cases should receive prostacyclin therapy (inhaled, subcutaneous, or intravenous) or the oral endothelin-receptor antagonist, bosentan. Right heart failure is treated conventionally and transplantation (heart–lung or single lung) is used in eligible cases.
- ***Pulmonary fibrosis*** is currently treated with immunosuppression, most often with cyclophosphamide or azathioprine combined with low-dose oral prednisolone.

Prognosis

In limited cutaneous scleroderma, the disease is often milder, with much less severe internal organ involvement and a 70% 10-year survival. Pulmonary hypertension is a significant later cause of death. Lung fibrosis and severe gut involvement also determine mortality. In diffuse disease, where organ involvement is often severe at an earlier stage, many patients die of pulmonary, cardiac, or renal involvement. Overall, pulmonary involvement (vascular or interstitial) accounts for around 50% of scleroderma-related deaths.

Localized forms of scleroderma occur either in patches (morphea; see Chapter 45) or linear forms. These are more commonly seen in children and adolescents, and do not convert into systemic forms, although ANA may occur in localized scleroderma and, very occasionally, localized and systemic forms coexist.

Polymyositis and dermatomyositis

Polymyositis (PM) is a rare disorder of unknown cause, in which the clinical picture is dominated by inflammation of striated muscle, causing proximal muscle weakness. When the skin is involved, it is called 'dermatomyositis' (DM). The incidence is about 2–10/million population *per annum* and it occurs in all races and at all ages. The aetiology is unknown, although viruses (e.g., *Coxsackie*, rubella, influenza) have been implicated and persons with HLA-B8/DR3 appear to be genetically predisposed.

Clinical features

Adult polymyositis

Females are affected three times more commonly than males.

The onset can be insidious, over months, or acute. General malaise, weight loss, and fever can develop during the acute phase, but the cardinal symptom is proximal muscle weakness. The shoulder and pelvic girdle muscles become wasted but are not usually tender. Face and distal limb muscles are not usually affected. Movements such as squatting and climbing stairs become difficult. As the disease progresses, involvement of pharyngeal, laryngeal, and respiratory muscles can lead to dysphonia and respiratory failure. These severe complications are rare if the disease is treated early.

Adult dermatomyositis

This condition is also more common in females. Apart from muscle weakness, these patients often suffer from myalgia, polyarthritis, and Raynaud's phenomenon, but DM is primarily distinguished from PM by the characteristic rash. This typically affects the eyelids, where heliotrope (purple) discoloration is accompanied by periorbital oedema, and the fingers, where purple–red, raised vasculitic patches are seen. These patches occur over the knuckles (Gottron's papules; **Fig. 32.38**) in 70% of patients, and this appearance is highly specific for DM. Ulcerative vasculitis and calcinosis of the subcutaneous tissue are seen in 25% of cases. In the long term, muscle fibrosis and contractures of joints occur.

Antisynthetase syndrome

Some 20%–30% of people with PM or DM have antibodies to tRNA synthetase enzymes, of which anti-Jo1 is the most common. These people are more likely to develop pulmonary interstitial fibrosis, Raynaud's phenomenon, arthritis, and hardening and fissuring of skin over the pulp surface of the fingers (mechanic's hands). This variant of PM/DM is sometimes called antisynthetase syndrome and often has a poor outcome. Respiratory muscles are affected in PM/DM and this compounds the effects of interstitial fibrosis. Dysphagia is seen in about 50% of patients owing to oesophageal muscle involvement.

Association with other ARDs

There is an association with other ARDs (e.g., SLE, RA, and SSc) and their associated clinical features, such as deforming arthritis, malar rash, and skin sclerosis.

Association with malignancies

The relative risk of cancer is 2.4 for male and 3.4 for female patients, and a wide variety of cancers have been reported. The onset and clinical picture do not differ in malignancy-associated disease from those of typical DM/PM. The associated cancer may not become apparent for 2–3 years, and recurrent, refractory, or ANA-negative DM should prompt a search for occult malignancy.

Malignancy (e.g., lung, ovary, breast, stomach) can also predate the onset of myositis, particularly in males with DM.

Fig. 32.38 Purplish Gottron's papules over the knuckles (dermatomyositis). *(Courtesy of David Paige, London.)*

Childhood dermatomyositis

This variant most commonly affects children between the ages of 4 and 10 years. The typical rash of DM is usually accompanied by muscle weakness. Muscle atrophy, subcutaneous calcification, and contractures may be widespread and severe. Ulcerative skin vasculitis is common and recurrent abdominal pain due to vasculitis is also a feature. Adults who have suffered from childhood DM may have chronic damage (e.g., contractures) or show effects of long-term corticosteroid or immunosuppressive therapy.

Investigations

- ***Serum creatine kinase (CK)***, aminotransferases, lactate dehydrogenase (LDH), and aldolase are usually raised and are useful guides to muscle damage, but may not reflect activity.
- ***ESR*** and ***CRP*** may be raised.
- ***Serum autoantibody studies*** include antinuclear antibody testing, which is usually positive in people with DM. RF is present in up to 50%, and many ***myositis-specific antibodies*** (MSAs) have been recognized and correlate with certain subsets. Antisynthetase antibodies have been described above.
- ***Electromyography (EMG)*** shows a typical triad of changes with myositis: spontaneous fibrillation potentials at rest, polyphasic or short-duration potentials on voluntary contraction, and salvos of repetitive potentials on mechanical stimulation of the nerve.
- ***MRI*** can be used to detect abnormally inflamed muscle.
- ***Needle muscle biopsy*** shows fibre necrosis and regeneration in association with an inflammatory cell infiltrate with lymphocytes around the blood vessels and between muscle fibres. Open biopsy allows more thorough assessment.
- ***Screening for malignancy*** is usually limited to relatively non-invasive investigations such as chest X-ray, mammography, pelvic/abdominal ultrasound and CT scanning, urine microscopy, and a search for circulating tumour markers.
- ***PET*** scanning can be considered as enhanced screening for malignancy in selected patients deemed to be at high risk (e.g., older age, resistance to immunosuppressive therapy, negative myositis-specific autoantibody panel).

Management

Bed rest may be helpful but must be combined with an exercise programme. Prednisolone is the mainstay of treatment: 0.5–1.0 mg/kg body weight as initial therapy, continued until at least 1 month after myositis has become clinically and enzymatically inactive. Tapering of steroids must be slow. Early intervention with steroid-sparing agents, such as methotrexate, azathioprine, ciclosporin, cyclophosphamide, and mycophenolate mofetil, is common, especially where there is clinical relapse or a rise in CK as the dose of steroids is reduced. Intravenous immunoglobulin therapy (IVIG) is helpful in some recalcitrant cases. Treatment of childhood DM tends to be more intensive, with earlier use of immunosuppressive agents. Use of biological agents such as rituximab has been described. Rituximab is more likely to be effective in autoantibody-positive cases.

Inclusion body myositis

Inclusion body myositis is an idiopathic inflammatory myopathy usually occurring in males over 50 years. Weakness of the pharyngeal muscles causes difficulty in swallowing in over 50%. It is a slowly progressive weakness of mainly distal muscles. In contrast to polymyositis, the CK is only slightly elevated; the EMG shows both myopathic and neuropathic changes. On MRI, the changes are often more distal but can be similar to those of polymyositis. A muscle biopsy shows inflammation and basophilic rimmed vacuoles with diagnostic filamentous inclusions and vacuoles on electron microscopy. A trial of corticosteroids is worthwhile but generally the response is poor.

Sjögren's syndrome

The syndrome of dry eyes (keratoconjunctivitis sicca) in the absence of RA or any of the autoimmune diseases is known as 'primary Sjögren's syndrome'. There is an association with HLA-B8/DR3. Dryness of the mouth, skin, or vagina may also be a problem. Salivary and parotid gland enlargement is seen. In the majority of cases, dryness and fatigue are the only symptoms, and Sjögren's syndrome is irritating and inconvenient rather than dangerous. However, in a minority there may be systemic symptoms such as:

- arthralgia and occasional non-progressive polyarthritis, like that seen in SLE (but much less common)
- Raynaud's phenomenon
- dysphagia and abnormal oesophageal motility, as seen in systemic sclerosis (but less common)
- other organ-specific autoimmune disease, including thyroid disease, myasthenia gravis, primary biliary cirrhosis, autoimmune hepatitis, tubulo-interstitial nephritis and pancreatitis
- renal tubular defects (uncommon) causing AVP-resistance (also referred to as 'nephrogenic diabetes insipidus') and renal tubular acidosis
- pulmonary diffusion defects and fibrosis
- polyneuropathy, fits, and depression
- vasculitis
- increased incidence of non-Hodgkin's B-cell lymphoma.

Pathology and investigations

Biopsies of the salivary gland or of the lip show a focal infiltration of lymphocytes and plasma cells.

- ***Schirmer tear test.*** A standard strip of filter paper is placed on the inside of the lower eyelid; wetting of less than 10 mm in 5 minutes indicates defective tear production.
- ***Rose Bengal staining.*** Staining of the eyes shows punctate or filamentary keratitis.
- ***Laboratory abnormalities.*** These include raised immunoglobulin levels, circulating immune complexes and autoantibodies. RF is usually positive. Antinuclear antibodies are found in 80% of cases and anti-mitochondrial antibodies are found in 10%. Anti-Ro (SSA) antibodies are found in 60%–90%, compared with 10% of cases of RA and secondary Sjögren's syndrome.

Management

Symptomatic treatment is with artificial tears and saliva replacement solutions. Hydroxychloroquine may help fatigue and arthralgia. Corticosteroids are rarely needed but are used to treat persistent salivary gland swelling or neuropathy. Pilocarpine can be used to stimulate secretion of saliva and tears. A trial of rituximab in Sjögren's syndrome did not demonstrate efficacy and this drug is not used.

'Overlap' syndromes and undifferentiated autoimmune rheumatic disease

An ***overlap syndrome*** is one where the patient shows the characteristic clinical features of more than one ARD. Treatment of each ARD is usually the same as when they occur separately.

Undifferentiated ARD is a term used for the condition of patients who have evidence of autoimmunity (e.g., positive autoantibody test) and some clinical features of such diseases (commonly, Raynaud's phenomenon and/or arthralgia) but not enough to make a clear diagnosis of any individual ARD. These patients sometimes develop a clearer ARD over time, but some always remain undifferentiated and tend to have relatively mild disease without major organ problems.

Further reading

Bakshi J, Tejera Segura B, Wincup C et al. Unmet needs in the pathogenesis and treatment of systemic lupus erythematosus. *Clin Rev Allergy Immunol* 2018; 55:352–367.
Giannokopoulos B, Krilis S. The pathogenesis of the antiphospholipid syndrome. *N Engl J Med* 2013; 368:1033–1044.
Kowal-Bielecka O, Fransen J, Avouac J et al. Update of EULAR recommendations for the treatment of systemic sclerosis. *Ann Rheum Dis* 2017; 76:1327–1339.
Rahman A, Isenberg DA. Systemic lupus erythematosus. *N Engl J Med* 2008; 358:929–939.
Tektonidou MG, Andreoli L, Limper Z et al. EULAR recommendations for the management of antiphospholipid syndrome in adults. *Ann Rheum Dis* 2019; 78:1296–1304.
Tsokos G. Systemic lupus erythematosus. *N Engl J Med* 2011; 378:1949–1961.

SYSTEMIC INFLAMMATORY VASCULITIS

Vasculitis is a histological term describing inflammation of the vessel wall. Vasculitis can be seen in many diseases (**Boxes 32.39 and 32.40**). The group of diseases described in this section (systemic inflammatory vasculitides) is characterized by widespread vasculitis leading to systemic symptoms and signs, generally requiring treatment with corticosteroids and/or immunosuppressive drugs. Two main features are helpful in classifying these vasculitides: the ***size of the blood vessels involved*** and the ***presence or absence of anti-neutrophil cytoplasmic antibodies (ANCAs)*** in the blood. Revisions in the commonly used terms for various vasculitides have been proposed to reflect increased pathophysiological understanding of these conditions (**Fig. 32.39** and see **Box 32.39**).

- ***Large-vessel vasculitis*** refers to the aorta and its major tributaries.
- ***Medium-vessel vasculitis*** refers to medium- and small-sized arteries and arterioles.

Box 32.39 Types of systemic vasculitis

Large vessel
- Giant cell arteritis/polymyalgia rheumatica
- Takayasu's arteritis

Medium vessel
- Classical polyarteritis nodosa (PAN)
- Kawasaki's disease

Small vessel
- Microscopic polyangiitis
- Granulomatosis with polyangiitis
- Eosinophilic granulomatosis with polyangiitis (30%–50% ANCA-positive)

ANCA-associated

- Immunoglobulin A (Henoch–Schönlein) vasculitis
- Cutaneous leucocytoclastic vasculitis
- Essential cryoglobulinaemia

ANCA-negative

Box 32.40 Other conditions associated with vasculitis[a]

Infective
- E.g., subacute infective endocarditis

Non-infective
- Vasculitis with rheumatoid arthritis
- Systemic lupus erythematosus
- Scleroderma
- Polymyositis/dermatomyositis
- Drug-induced Behçet's disease
- Goodpasture's syndrome
- Hypocomplementaemia
- Serum sickness
- Paraneoplastic syndromes
- Inflammatory bowel disease

[a]See also **Box 32.39**.

- ***Small-vessel vasculitis*** refers to small arteries, arterioles, venules, and capillaries.

Large-vessel vasculitis

PMR and giant cell (temporal) arteritis are systemic illnesses of the elderly. Both are associated with the finding of a giant cell arteritis (GCA) on temporal artery biopsy.

Polymyalgia rheumatica

Polymyalgia rheumatica (PMR) causes a sudden onset of severe pain and stiffness of the shoulders and neck, and of the hips and lumbar spine, which is a limb-girdle pattern. These symptoms are worse in the morning, lasting from 30 minutes to several hours. The clinical history is usually diagnostic and the patient is always over 50 years old. Up to 25% of patients may have some inflammation of peripheral joints.

Approximately one-third of patients develop systemic features of tiredness, fever, weight loss, depression, and occasionally nocturnal sweats, especially if PMR is not diagnosed and treated early. A differential diagnosis is shown in **Box 32.41**.

Investigation of PMR

- ***A raised ESR and/or CRP*** is a hallmark of this condition. It is rare to see PMR without an acute-phase response. If it is absent, the diagnosis should be questioned and the tests repeated a few weeks later before treatment is started.
- ***Serum alkaline phosphatase and γ-glutamyl-transpeptidase*** may be raised as markers of the acute inflammation.
- ***Anaemia of inflammation*** (mild normochromic, normocytic) is often present.
- ***Temporal artery biopsy*** shows GCA in 10%–30% of cases, but is rarely performed unless GCA is also suspected.

Giant cell arteritis

Giant cell arteritis (GCA) is inflammatory granulomatous arteritis of large cerebral arteries, which occurs in association with PMR. The patient may have current PMR or a history of recent PMR, or may be on treatment for PMR. It is extremely rare under 50 years of age. Presenting symptoms of GCA include severe headaches, tenderness of the scalp (combing the hair may be painful) or of the temple, claudication of the jaw when eating, and tenderness and swelling of one or more temporal or occipital arteries. The most feared manifestation is sudden, painless, temporary, or permanent loss of vision in one eye due to involvement of the ophthalmic artery (see Chapter 28). Systemic manifestations of severe malaise, tiredness, and fever occur.

Investigations in GCA

- Normochromic, normocytic anaemia.
- ESR is usually raised (in the region of 50–120 mm/h) and the CRP is very high.
- Liver biochemistry may be abnormal, as in PMR. The albumin may be low.
- A temporal artery biopsy from the affected side is the definitive diagnostic test. This should be taken before, or within 7 days of starting, high doses of corticosteroids. The lesions are patchy and the whole length of the biopsy (>1 cm long) must be examined; even so, negative biopsies occur.
- Ultrasound scanning of the temporal arteries is increasingly used—if diagnostic features are seen, biopsy can be avoided.

The ***histological features*** of GCA are:
- Cellular infiltrates of $CD4^+$ T lymphocytes, macrophages, and giant cells in the vessel wall. Note that giant cells are not visible in all cases.
- Granulomatous inflammation of the intima and media.
- Breaking up of the internal elastic lamina.
- Giant cells, lymphocytes, and plasma cells in the internal elastic lamina.

Management of PMR or GCA

Corticosteroids produce a dramatic reduction of symptoms of PMR within 24–48 hours of starting treatment, provided the dose is adequate. If this improvement does not occur, the diagnosis should be questioned and an alternative cause sought, such as RA, vasculitis, infection, or malignancy. Corticosteroid treatment should reduce the risk of patients who have PMR developing GCA. NSAIDs are less effective and should not be used.

In GCA, corticosteroids are obligatory because they significantly reduce the risk of irreversible visual loss and other focal ischaemic lesions, but much higher doses are needed than in PMR. If GCA is suspected, it may not be possible to arrange an ultrasound scan or temporal artery biopsy rapidly. In these circumstances, treatment should not be delayed, especially if there have already been episodes of visual loss or stroke.

Starting daily doses of prednisolone are:
- ***PMR:*** 15 mg prednisolone as a single dose in the morning
- ***GCA:*** 60–100 mg prednisolone, usually in divided doses.

The dose should then be reduced gradually in weekly or monthly steps. While the dose is above 20 mg, the step reductions are 5 mg, reducing the evening doses first. Between 20 mg and 10 mg, the reduction can be in 2.5-mg steps, but below 10 mg, the rate should be slower and the steps of 1 mg each. Most patients will eventually be able to stop corticosteroids after 12–18 months but up to 25% may need low doses long term. Steroid-sparing immunosuppressive

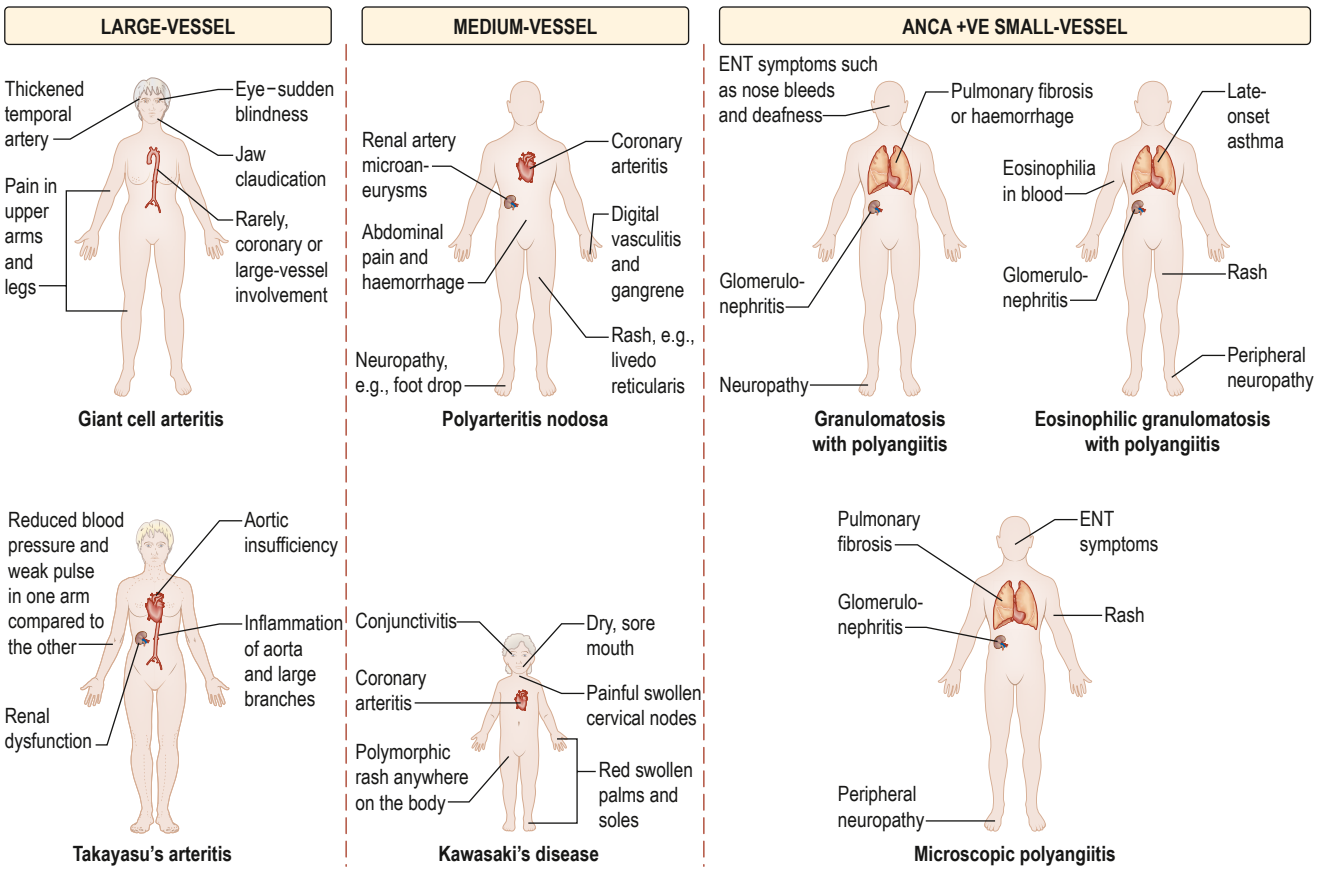


Fig. 32.39 Clinical features of systemic inflammatory vasculitides. These illnesses all frequently present with joint pain, as well as systemic symptoms such as fatigue, malaise, and weight loss. The most common organ-specific manifestations are shown. Giant cell arteritis is usually seen in people over 50, whereas Takayasu's disease is seen in people under 50. Kawasaki's disease is usually seen in children under 5. *ANCA,* Anti-neutrophil cytoplasmic antibodies; *ENT,* ear, nose and throat.

? Box 32.41 Symptom patterns in some muscle disorders

- Polymyositis—proximal muscle ache and weakness
- Polymyalgia rheumatica—proximal morning stiffness and pain
- Myopathy—weakness, but no pain or stiffness

agents are used in refractory cases where it is hard to reduce the corticosteroid dose without causing a flare of disease or a rise in ESR or CRP. A clinical trial showed that a combination of the anti-interleukin receptor 6 receptor agent tocilizumab with corticosteroid gave higher rates of remission than corticosteroid alone and this biologic is now used in refractory cases of GCA.

Calcium and vitamin D supplements, and sometimes bisphosphonates, are necessary to prevent osteoporosis while high-dose steroids are being used (see Chapter 33).

Takayasu's arteritis

Takayasu's arteritis is a granulomatous inflammation of the aorta and its major branches (see Chapter 17).

Medium-sized vessel vasculitis

Polyarteritis nodosa

Classical polyarteritis nodosa (PAN) is a rare condition that usually occurs in middle-aged males. It is accompanied by severe systemic manifestations, and its occasional association with hepatitis B antigenaemia suggests a vasculitis secondary to the deposition of immune complexes. Pathologically, there is fibrinoid necrosis of vessel walls with microaneurysm formation, thrombosis, and infarction.

Clinical features

These include fever, malaise, weight loss, and myalgia. These initial symptoms are followed by dramatic acute features that are due to organ infarction.

- ***Neurological.*** Mononeuritis multiplex is due to arteritis of the vasa nervorum.
- ***Abdominal.*** Pain is due to arterial involvement of the abdominal viscera, mimicking acute cholecystitis, pancreatitis, or appendicitis. Gastrointestinal haemorrhage occurs because of mucosal ulceration.
- ***Renal.*** Presentation is with haematuria and proteinuria. Hypertension and acute/chronic kidney disease occur.
- ***Cardiac.*** Coronary arteritis causes myocardial infarction and heart failure. Pericarditis also occurs.
- ***Skin.*** Subcutaneous haemorrhage and gangrene occur. A persistent livedo reticularis is seen in chronic cases. Cutaneous and subcutaneous palpable nodules occur but are uncommon.
- ***Lung.*** Involvement is rare.

Investigations and management

- ***Blood count.*** Anaemia, leucocytosis, and a raised ESR occur.
- ***Biopsy.*** Material from an affected organ shows features listed above.

- ***Angiography.*** Microaneurysms in hepatic, intestinal, or renal vessels can be demonstrated if necessary.
- ***Other investigations.*** These are performed as appropriate (e.g., electrocardiogram and abdominal ultrasound), depending on the clinical problem. ANCA is positive but rarely in classic PAN.

Treatment is with corticosteroids, usually in combination with immunosuppressive drugs such as azathioprine.

Kawasaki's disease

Kawasaki's disease is an acute systemic vasculitis involving medium-sized vessels, affecting mainly children under 5 years of age. It is very frequent in Japan and an infective trigger is suspected. Fewer than 10% of cases occur in adults.

Clinical features and management

The characteristic clinical features are fever lasting longer than 4 days, bilateral conjunctival congestion, dry red lips and oral cavity, cervical lymphadenopathy, rash and redness of the palms and soles.

Cardiovascular changes in the acute stage include pancarditis and coronary arteritis, leading to aneurysms or dilatation visible on echocardiography, MRI, or angiography.

Treatment is with a single dose of high-dose intravenous immunoglobulin (2 g/kg), which prevents the coronary artery disease, followed after the acute phase by aspirin 200–300 mg daily. There is no evidence that steroid treatment improves the outcome.

Small-vessel vasculitis

The two revised categories of small-vessel vasculitis (SVV) are:

- ANCA-associated vasculitis (AAV), characterized by a ***paucity of vessel wall immunoglobulin*** and presence (in many but not all patients) of ANCA, including:
  - granulomatosis with polyangiitis (GPA; see Chapter 20)
  - eosinophilic granulomatosis with polyangiitis (EGPA; see Chapter 20)
  - microscopic polyangiitis (MPA; see Chapter 20).
- Immune complex (ANCA-negative) SVV, with a ***prominence of vessel wall immunoglobulin*** including:
  - anti-glomerular basement membrane (anti-GBM) disease (see Chapter 20)
  - IgA vasculitis (IgAV, Henoch–Schönlein) (see Chapter 26)
  - cryoglobulinaemic vasculitis (CV) (see Chapter 26)
  - hypocomplementaemic urticarial vasculitis (HUV, anti-C1q vasculitis)
  - cutaneous leucocytoclastic vasculitis (see Chapter 45).

Cutaneous leucocytoclastic vasculitis is the characteristic acute purpuric lesion that histologically involves the dermal postcapillary venules. This lesion affects only the skin (a form of 'single-organ vasculitis') and should be differentiated from similar lesions produced in systemic vasculitis. The purpura may be accompanied by arthralgia, although the appearance of additional organ disease (e.g., glomerulonephritis) suggests evolution towards a systemic vasculitis. Hepatitis C infection is common and may be an aetiological agent. The condition can also be caused by drugs such as sulphonamides and penicillin.

Primary CNS vasculitis is a very rare condition (incidence 2–4 cases per million person-years) characterized by clinical features confined to the CNS and characteristic appearances on angiography or cerebral biopsy. The most common clinical features are headache, altered cognition, focal weakness, and stroke. Blood tests, including inflammatory markers and autoantibodies, are usually normal but a normal cerebral MRI essentially excludes the diagnosis.

Management of small-cell vasculitis

The treatment depends on the organs involved. Vasculitis confined to the skin may not require systemic treatment, whereas involvement of major organs (e.g., lungs or kidneys in GPA) requires high-dose corticosteroids, immunosuppression, and, sometimes, plasma exchange. Clinical trials have shown that depletion of B cells with rituximab is as effective as cyclophosphamide in treating AAV and can help reduce the dose of corticosteroids required to maintain remission. In EGPA, the anti-interleukin 5 agent mepolizumab has been found to be effective in a clinical trial.

Behçet's disease

Behçet's disease is an inflammatory disorder of unknown cause. There is a striking geographical distribution, it being most common in Turkey, Iran, and Japan. The prevalence per 100 000 is 10–15 in Japan and 80–300 in Turkey. There is a link to the HLA-B51 allele, with a relative risk of 5–10; this association is not seen in patients in the USA and Europe.

Clinical features

The cardinal clinical feature is recurrent oral ulceration. The ***international criteria for diagnosis*** require oral ulceration and any two of the following:

- genital ulcers
- defined eye lesions, including an anterior or posterior uveitis or retinal vascular lesions
- defined skin lesions—erythema nodosum, pseudofolliculitis, and papulopustular lesions
- positive skin pathergy test (see below)
- oral ulcers—aphthous or herpetiform.

Other manifestations include a self-limiting peripheral mono- or oligoarthritis affecting knees, ankles, wrists, and elbows; gastrointestinal symptoms of diarrhoea, abdominal pain, and anorexia; pulmonary and renal lesions; thrombophlebitis (especially in the legs); vasculitis; and a brainstem syndrome, organic confusional states and a meningoencephalitis. All the common manifestations are self-limiting except for the ocular attacks. Repeated attacks of uveitis can cause blindness.

The ***pathergy reaction*** is highly specific to Behçet's disease. Skin injury—by a needle prick, for example—leads to papule or pustule formation within 24–48 hours. Blood tests usually show raised ESR and CRP but not autoantibodies.

Management

Corticosteroids, immunosuppressive agents, and ciclosporin are used for chronic uveitis and the rare neurological complications. Colchicine helps erythema nodosum and joint pain. Thalidomide may be useful in some cases, although side-effects of drowsiness and peripheral neuropathy are common. It should not be used in pregnant women because of phocomelia (limb abnormalities) and females of childbearing potential must use effective contraception for at least 4 weeks before, during, and for at least 4 weeks after stopping treatment (the combined oral contraceptive pill is not recommended due to increased risk of thromboembolism). Anti-TNF agents can be used to control severe uveitis and serious manifestations such as neurological and gastrointestinal Behçet's disease.

Further reading

Jennette JC, Falk RJ, Bacon PA et al. 2012 revised International Chapel Hill Consensus Conference nomenclature of vasculitides. *Arthritis Rheum* 2013; 65:1–11.
Kermani TA, Warrington KJ. Polymyalgia rheumatica. *Lancet* 2013; 381:63–72.
Stone JH, Tuckwell K, Dimonaco S et al. Trial of tocilizumab in giant cell arteritis. *N Engl J Med* 2017; 377:317–328.
Warrell DA, Cox TM, Firth JD (eds). *Oxford Textbook of Medicine*, 5th edn. Oxford: Oxford University Press; 2010. Section 19.11 Autoimmune rheumatic disorders and vasculitides; Section 21.10.2 The kidney in systemic vasculitis.
Wechsler ME, Akuthota P, Jayne D et al. Mepolizumab or placebo for eosinophilic granulomatosis with polyangiitis. *N Engl J Med* 2017; 376:1921–1932.

ARTHRITIS IN CHILDREN

Joint and limb pains are common in children but arthritis is, fortunately, rare. Discussion here is limited to juvenile idiopathic arthritis, as a disease which may persist into adulthood (and so falling under the care of adult rheumatologists).

Juvenile idiopathic arthritis

A full description of juvenile idiopathic arthritis (JIA) is beyond the scope of this textbook. A brief account is provided below.

JIA is classified according to the extent and type of joint involvement. A particularly severe form, ***systemic onset JIA (Still's disease)*** presents with high fever (>39°C), rash, lymphadenopathy, hepatosplenomegaly, and/or serositis. Joint involvement is often not the most prominent symptom.

In contrast, patients with ***oligoarticular*** (<5 joints involved) and ***polyarticular*** (≥5 joints) JIA experience mainly of joint pain and swelling.

The oligoarticular form is more common. In childhood, these patients can develop uveitis and require regular screening by slit lamp examination. Outlook is usually good but approximately 25% develop an extended form where more joints can be involved over time leading to destructive arthritis persisting into adulthood.

Polyarticular JIA is divided into RF-positive and RF-negative subtypes, of which the RF-positive form is more aggressive.

Enthesitis-related arthritis tends to affect teenage and younger boys, mainly involves lower limb joints and entheses and is associated with HLA-B27 positivity.

Children with JIA are generally managed by specialist paediatric rheumatologists but come under the care of adult rheumatologists later in life. Treatment recommendations are based on clinical phenotypes involving cs/b/tsDMARDs for which management and monitoring are as for adult-onset inflammatory arthritis. The goal of therapy is to rapidly reduce disease activity to inactive disease or clinical remission, minimize drug side-effects, and achieve quality of life comparable to peers. Joint replacement surgery in patients who have had JIA may be required earlier in life than in other groups.

RHEUMATOLOGICAL PROBLEMS SEEN IN OTHER DISEASES

Gastrointestinal and liver disease

- Enteropathic synovitis (see 'Enteropathic arthritis associated with inflammatory bowel disease' section).
- Autoimmune hepatitis (see Chapter 23) may be accompanied by an arthralgia similar to that seen in SLE. Joint pain occurs in a bilateral, symmetrical distribution, predominantly affecting the small joints of the hands. Joints usually look normal but sometimes there is a slight soft tissue swelling. These patients often have positive tests for antinuclear antibodies.
- Primary biliary cholangitis patients occasionally have a symmetrical arthropathy.
- Hereditary haemochromatosis is associated with arthritis in 50% of cases, which is often the first sign of the disease and CC is common.
- Whipple's disease (see Chapter 22) is accompanied by fever and arthralgia.

Respiratory disease

Sarcoidosis

- Sarcoidosis (see Chapter 20) is a multisystem granulomatous disease associated with erythema nodosum in 20% of cases at or soon after the onset of the disease. A chest X-ray shows hilar lymphadenopathy in 80% of cases. The serum ACE may be raised.
- Several patterns of arthritis occur later in the disease. These include a transient rheumatoid-like polyarthritis and an acute monoarthritis that can be mistaken for gout. Bone cysts can also develop.
- ***Treatment*** is with NSAIDs; if these fail to control the symptoms, corticosteroids are usually very effective.

Malignant disease

It is not uncommon for malignant diseases to present with musculoskeletal symptoms. Gout occurs in conditions such as chronic myeloid leukaemia. Neoplastic disease of bone is described in Chapter 33.

Hypertrophic pulmonary osteoarthropathy

Hypertrophic pulmonary osteoarthropathy (HPO) is a paraneoplastic, non-metastatic complication, frequently associated with carcinoma of the bronchus. It may be the presenting feature of the disease. It rarely occurs with other conditions that also cause clubbing, presenting most often in middle-aged men with pain and swelling of the wrists and ankles. Other joints are occasionally involved. Primary HPO is a hereditary condition involving a mutation in the *HPGD* gene that degrades prostaglandin E_2 (PGE_2), thus allowing over-production of PGE_2, which may cause clubbing.

The diagnosis is made on the presence of clubbing of the fingers (usually gross in primary HPO) and periosteal new bone formation along the shafts of the distal ends of the radius, ulna, tibia, and fibula on X-ray. A chest X-ray usually shows the malignancy.

Management should be directed at the underlying carcinoma; if this can be removed, the arthropathy disappears. NSAIDs relieve the symptoms.

Paraneoplastic polyarthritis

This condition is seen with carcinoma of the breast in women and of the lung in men, and also with renal cell carcinoma. The neoplasm may be occult at onset and the diagnosis is then difficult to make.

Skin disease

Psoriatic arthritis

This disease is discussed earlier (see 'Psoriatic arthritis' section).

Erythema nodosum

Erythema nodosum (see Chapter 45) is accompanied by arthritis in over 50% of cases. The knees and ankles are particularly affected,

being swollen, red, and tender. The arthritis subsides, along with the skin lesions, within a few months. Treatment is with NSAIDs or occasionally steroids.

Neurological disease

Neuropathic (Charcot's) joints

These are damaged by trauma as a result of the loss of the protective pain sensation. The site of the neuropathic joint depends on the localization of the pain loss:

- In tabes dorsalis (spinal cord degeneration resulting from neurosyphilis), the knees and ankles are most often affected.
- In diabetes mellitus, the joints of the tarsus are involved.
- In syringomyelia, the shoulder is involved.

Neuropathic joints are not painful, although there may be painful episodes associated with crystal deposition. Presentation is usually with swelling and instability. Eventually, severe deformities develop.

The characteristic finding is a swollen joint with abnormal but painless movement, in association with neurological findings that depend on the underlying disease (e.g., dissociated sensory loss in syringomyelia or polyneuropathy in diabetes). X-ray changes are characteristic, with gross joint disorganization and bony distortion.

Treatment is symptomatic. Surgery may be required in advanced cases.

Blood disease (see also Chapter 38)

Arthritis due to haemarthrosis is a common presenting feature of people with ***haemophilia***. Attacks begin in early childhood in most cases and are recurrent. The knee is the most commonly affected joint but the elbows and ankles are sometimes involved. The arthritis can lead to bone destruction and disorganization of joints. Apart from replacement of factor VIII, affected joints require initial immobilization followed by physiotherapy to restore movement and measures to prevent and correct deformities.

Sickle cell crises are often accompanied by joint pain that particularly affects the hands and feet in a bilateral, symmetrical distribution. Affected joints usually look normal but are occasionally swollen. This condition may also be complicated by avascular necrosis and by osteomyelitis.

Arthritis can also occur in ***acute leukaemia*** (see Chapter 39); it may be the presenting feature in childhood. The knee is particularly affected and is very painful, warm, and swollen. Treatment is directed at the underlying leukaemia. Arthritis may also occur in chronic leukaemia, with leukaemic deposits in and around the joints.

Individuals with ***transfusion-dependent thalassaemia*** are living longer and are presenting with back pain due to premature disc degeneration, secondary spondylosis, and crush fractures caused by osteoporosis. There is marked discal calcification.

Endocrine and metabolic disorders

Hypothyroid patients may complain of pain and stiffness of proximal muscles, resembling PMR. They may also have carpal tunnel syndrome. Less often, there is an arthritis accompanied by joint effusions, particularly in the knees, wrist, and small joints of the hands and feet. These problems respond rapidly to thyroxine.

Hyperparathyroidism may be complicated by CC and acute CPPD.

In ***acromegaly***, arthralgia occurs in about 50% of patients, particularly in the small joints of the hands and knees. There may be a carpal tunnel syndrome.

In ***Cushing's disease***, back pain is common.

Joint disorders related to ***diabetes mellitus*** are described in Chapter 35.

Familial hypercholesterolaemia is associated with oligo- or polyarthritis, usually with tendon xanthomata. Arthritis also occurs in combined hyperlipidaemia.

MISCELLANEOUS ARTHROPATHIES

Familial Mediterranean fever

Periodic fever syndromes are a group of autoinflammatory disorders that present with recurrent attacks of fever in childhood and adult life.

Familial Mediterranean fever (FMF) is the most common inherited autoinflammatory disease. It is an autosomal recessive condition recognized worldwide but most prevalent in populations originating from the eastern Mediterranean. Mutations in the *MEFV* gene on chromosome 16 that encodes for pyrin, a suppressor of the activation of caspase 1, lead to increased biosynthesis of IL-1β and inflammation in FMF.

Symptoms of FMF are recurrent episodes of fever, arthritis, and serositis. Peritonitic abdominal pain occurs in 80% of attacks. Pleuritic chest pain occurs in 15% of cases and is unilateral. Arthritis is rare and usually monoarticular, lasting up to 1 week. The CRP is markedly raised during the attacks. The condition may be mistaken for palindromic rheumatism (see 'Other presentations' section), but such attacks are not usually accompanied by fever.

The diagnosis can be made by PCR, if available, but is based on the clinical picture and exclusion of other conditions.

Treatment involves lifelong colchicine 1000–1500 μg daily, which can usually prevent the attacks. In resistant patients, IL-1 inhibitors (anakinra and canakinumab) are licensed. In general, the disorder is benign, but in 25% of cases, renal amyloidosis develops (see Chapter 26).

Other periodic fevers are being increasingly recognized. These include tumour necrosis factor receptor-associated periodic syndrome (TRAPS), mevalonate kinase deficiency (MKD), cryopyrin-associated periodic syndrome (CAPS), and periodic fever, aphthous stomatitis, pharyngitis, and adenitis (PFAPA).

SAPHO (synovitis, acne, palmoplantar pustulosis, hyperostosis, osteitis)

This rare autoinflammatory disorder, linked to various HLA associations, dysregulation of IL-1, and *Cutibacterium acnes* infection, is considered on a spectrum with SpA, autoinflammatory disorders, and pustular dermatoses. It produces chronic multifocal osteitis with anterior chest wall pain and peripheral synovitis. There is inflammatory cytokine release and global neutrophil activation. NSAIDs, methotrexate, or TNF inhibitors (see 'Rheumatoid arthritis, Drug therapy' section) may help.

Osteochondromatosis

In this condition, foci of cartilage form within the synovial membrane. These foci become calcified and then ossified (osteochondromas). They may give rise to loose bodies within the joint. The condition occurs in a single joint of a young adult and X-rays are usually diagnostic.

Management involves removal of loose bodies and synovectomy.

Pigmented villonodular synovitis

This is characterized by exuberant synovial proliferation that occurs either in joints or in tendon sheaths. The main manifestation in joints is recurrent haemarthrosis. It may produce progressive local bone destruction. A malignant form is seen occasionally.

Management is synovectomy or radiotherapy for more widespread cases or those that have failed surgery.

Relapsing polychondritis

Relapsing polychondritis is a rare inflammatory condition of cartilage. It occurs equally in males and females, usually the elderly. Tenderness, inflammation, and eventual destruction of cartilage occur, mainly in the ear, nose, larynx, or trachea. A seronegative polyarthritis occurs, as well as episcleritis and evidence of a vasculitis (e.g., glomerulonephritis). The diagnosis is clinical with laboratory evidence of acute inflammation.

Treatment involves corticosteroids and immunosuppressive agents.

Further reading

Lachmann HJ. Periodic fever syndromes. *Best Pract Res Clin Rheumatol* 2017; 31:596–609.

Questions for this chapter can be found on Elsevier eBooks+.

33 Bone disease

Richard Conway

Clinical skills 1121
History 1121
Examination 1121
Biochemical tests 1122
Anatomy and physiology of bone 1123
Bone structure 1123
Bone growth and remodelling 1124
Calcium homeostasis and its regulation 1124
Investigation of bone and calcium disorders 1126
Osteoporosis 1128
Osteonecrosis 1132
Paget's disease of bone 1133
Rickets and osteomalacia 1134
Bone infections 1135
Neoplastic disease of bone 1135
Scheuermann's disease 1136

CORE SKILLS AND KNOWLEDGE

Bone disease is common in the general population, although frequently undiagnosed and often asymptomatic. Some 50% of females and 25% of males will suffer an osteoporotic fracture in their lifetime. Bone diseases are even more prevalent in those with coexisting medical conditions and the elderly. Bone disease specialists work in a variety of disciplines including general practice, rheumatology, endocrinology, and medicine for the elderly, dividing their time between outpatient bone health clinics and inpatient fracture liaison services. Specialist nurses, radiographers, orthopaedic surgeons, and allied health professionals all have important roles in helping to care for patients with bone disease.

Key learning objectives in bone disease include:

- Understanding osteoporosis—its causes, diagnosis, and treatment
- Knowing when to investigate for other bone disorders, including bone infections and malignancy
- Understanding 'bone profile' blood tests, dual-energy X-ray absorptiometry (DXA) scanning, and fracture risk assessment tools.

Learning opportunities in bone health include attending specialist bone health clinics, and attending fracture liaison service ward rounds where the underlying causes of fragility fractures are addressed.

CLINICAL SKILLS

History

Clinical features

- Bone disease such as osteoporosis is often asymptomatic.
- ***Severe, acute-onset pain*** suggests a fracture (Box 33.1), which can occur with minimal trauma if the bone is fragile due to osteoporosis or malignancy.
- ***Progressive localized bone pain*** may suggest primary or secondary malignancy, or localized Paget's disease.

Risk factors for osteoporosis

- Increased age.
- Early menopause without use of hormone replacement therapy.
- Hypogonadism in males.
- Corticosteroid use.
- Intestinal malabsorption.
- Obesity.
- Immobility.
- Other medical conditions (Box 33.4).

Examination

Bone disease often presents with no physical signs, at least until secondary consequences ensue. This is particularly true of osteoporosis. The commonest anatomical sites for various diseases of bone to occur are shown in the diagram; physical signs will vary from severe acute deformity and pain (if fracture occurs), through varying levels of tenderness, swelling, and deformity in slowly progressive disease, to no observable changes in clinically silent disease.

Biochemical tests

A number of biochemical tests can be used to investigate the presence of bone disease. These are best described by their use in different clinical settings: in a patient with a fracture, or in patients with bony pain, or no symptoms at all (when these abnormalities may be picked up incidentally during routine or screening blood tests) (Boxes 33.2 and 33.3).

Box 33.1 Evaluation of a patient with a fracture

Why has fracture occurred?
- Traumatic—high impact trauma
- Fragility—force equivalent to fall from less than standing height—implies osteoporosis or other bone disorder
- Pathological—consider when multiple fractures, systemic symptoms, unusual sites.

Investigations of fragility fracture
- Calcium/phosphate/magnesium
- 1,25 dihydroxyvitamin D/parathyroid hormone/alkaline phosphatase
- Myeloma screening—serum electrophoresis, serum, and urine-free light chains, immunoglobulins
- Thyroid function tests, coeliac testing
- Dual-energy X-ray absorptiometry (DXA) scanning.

Treatment of fragility fracture
- Fracture risk assessment
- Most patients with fragility fracture have osteoporosis and warrant treatment
- Calcium/vitamin D if required
- Bisphosphonate
  - Denosumab is alternative
  - Consider recombinant parathyroid hormone (PTH) if very high risk.

Box 33.2 Evaluation of a patient with a fragility fracture

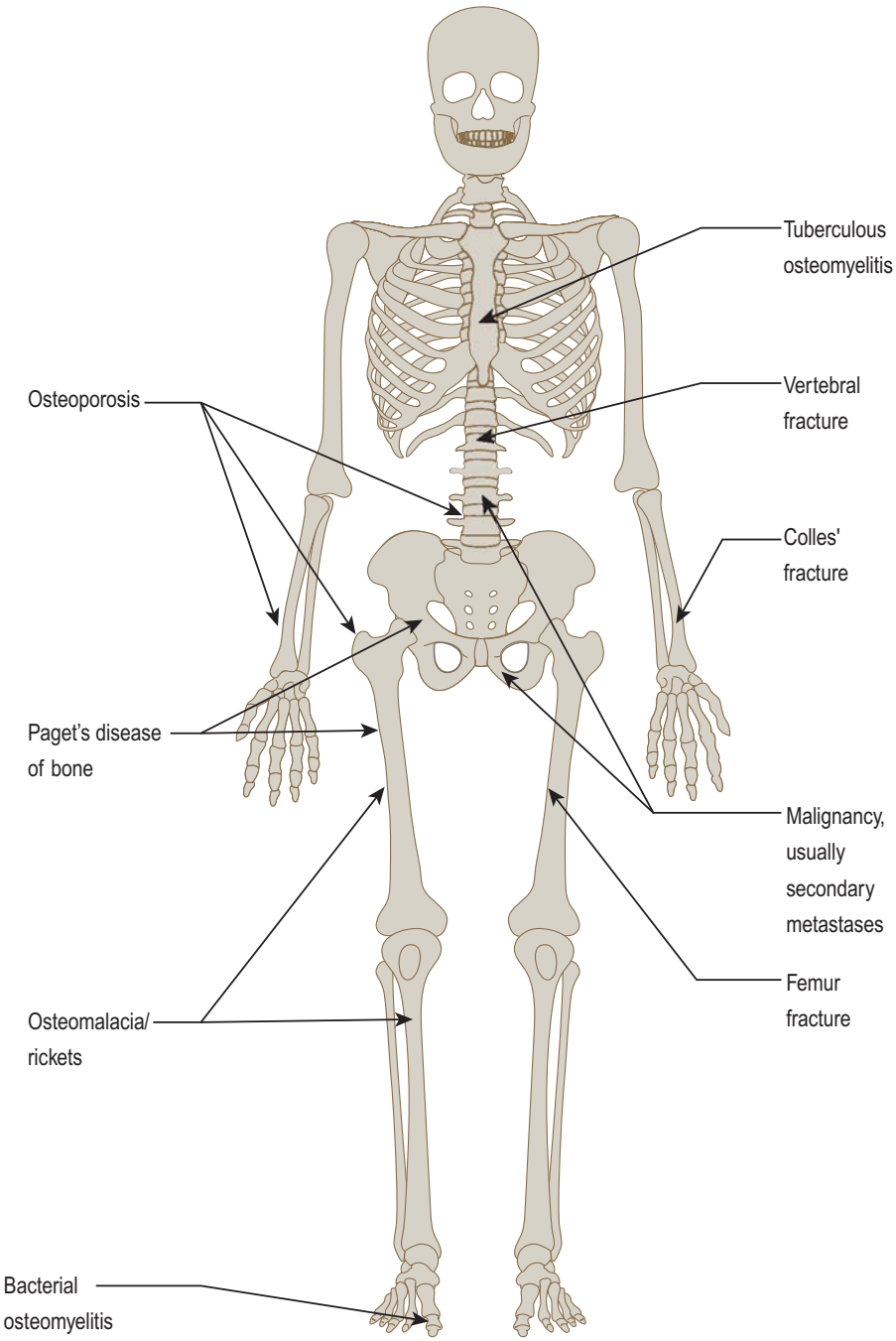
Ca	ALP	PTH	Probable diagnosis	Confirmatory test
N	N/H	N	**Osteoporosis**	DXA scanning
H	H	L	**Malignancy**	Nuclear medicine bone scan, investigation for likely site of primary (e.g., CT chest/abdomen/pelvis, mammogram)
N/H	H	H	**Primary hyperparathyroidism**	MIBI nuclear medicine parathyroid scan (see Chapter 34)
L	H	H	**Secondary hyperparathyroidism**	Check renal function for evidence of chronic kidney disease mineral and bone disorder (CKD-MBD, Chapter 26)

ALP, Alkaline phosphatase; *Ca,* calcium; *CT,* computed tomography; *DXA,* dual-energy X-ray absorptiometry; *H,* high; *L,* low; *MIBI,* sestamibi; *N,* normal; *PTH,* parathyroid hormone.

Box 33.3 Evaluation of a patient with bone pain or no symptoms

Ca	Phos	ALP	Probable diagnosis	Confirmatory test
L	L	H	**Osteomalacia**	Low vitamin D
N	N	H	**Paget's disease of bone**	Plain X-ray, isotope bone scan
L/N/H	H	H	**CKD-MBD**	Check renal function

ALP, Alkaline phosphatase; *Ca,* calcium; *CKD-MBD,* chronic kidney disease mineral and bone disorder; *H,* high; *L,* low; *N,* normal; *phos,* phosphate.



ANATOMY AND PHYSIOLOGY OF BONE

Bone is a specialized connective tissue serving three major functions:

- ***mechanical***—supplying structure and muscle attachment for movement
- ***metabolic***—providing the body's primary store of calcium and phosphate
- ***protective***—enclosing the marrow and other vital organs.

This chapter covers the connective tissue functions of bone; bone marrow function is discussed in Chapter 38.

Bone structure

Bone comprises cells and a matrix of organic protein and inorganic mineral. Long bones (femur, tibia, humerus) and flat bones (skull, scapula) have different embryological templates, with varying proportions of cortical and trabecular bone.

- ***Cortical (compact or lamellar) bone*** forms the shaft of long bones and the outer shell of flat bones. Formed of concentric rings of bone, it is particularly adapted to withstand bending strain.
- ***Trabecular (cancellous) bone*** is found at the ends of long bones and inside flat bones. Comprising a network of interconnecting rods and plates of bone, it offers resistance to compressive loads. It is also the main site of bone turnover for mineral homeostasis.
- ***Woven bone*** lacks an organized structure. It appears in the first few years of life, at sites of fracture repair and in high-turnover bone disorders such as Paget's disease.

Matrix components

- ***Type I collagen*** is the main protein, forming parallel lamellae of differing density (which impairs spreading of cracks). In cortical

bone, concentric lamellae form around a central blood supply (Haversian system), which communicates via transverse (Volkmann's) canals.

- ***Non-collagen proteins*** include osteopontin, osteocalcin and fibronectin.
- ***Bone mineral*** largely consists of calcium and phosphate in the form of hydroxyapatite.

Bone cells

Osteoblasts

Derived from local mesenchymal stem cells, these cells synthesize matrix (osteoid) and regulate its mineralization. After bone formation, the majority of osteoblasts are removed by apoptosis, others remaining at the bone/marrow interface as lining cells or within the bone as osteocytes. Osteoblasts regulate bone resorption through the balance in expression of the stimulatory receptor activator of nuclear factor kappa B ligand (RANKL) and its antagonist, osteoprotegerin (OPG). Osteoblasts are rich in alkaline phosphatase and express receptors for parathyroid hormone (PTH), oestrogen, glucocorticoids, vitamin D, inflammatory cytokines, and the transforming growth factor-beta (TGF-β) family, all of which may therefore influence bone remodelling.

Osteocytes

These small cells, derived from osteoblasts, are embedded in bone and interconnected with each other and with bone-lining cells through cytoplasmic processes. They respond to mechanical strain by undergoing apoptosis or through altered cell signalling, which in turn activates bone formation with or without prior resorption. As osteocytes also express RANKL and OPG, the relative importance of osteocytes and osteoblasts in bone resorption function continues to be explored.

Osteoclasts

These cells have the unique capacity to resorb bone and are derived from haemopoietic precursors of the macrophage lineage. In response to RANKL, macrophage colony-stimulating factor (M-CSF), and local adhesion factors (integrins), osteoclasts attach to bone, creating a ruffled border that forms a number of extracellular lysosomal compartments. Hydrogen ions are actively secreted into these spaces and the acid environment removes the mineral phase before specialized cysteine proteases (e.g., cathepsin K) resorb the collagen matrix.

Bone growth and remodelling

Longitudinal growth occurs at the epiphyseal growth plate, a cartilage structure between the epiphysis and metaphysis (**Fig. 33.1**). Cartilage production is tightly regulated, with subsequent mineralization and growth finally arrested at 18 to 21 years, when the epiphysis and metaphysis fuse.

In adults, bone is regularly remodelled to ensure repair of microdamage and turnover of calcium and phosphate for homeostasis. This remodelling cycle is carried out by the basic multicellular unit (BMU; **Fig. 33.2**). Signals initiating resorption include osteocyte apoptosis and altered signalling (sclerostin, prostaglandins, RANKL, and other molecules), resulting in localized retraction of bone-lining cells and binding of multinucleate osteoclasts to the bone surface, followed by bone resorption. Bone formation involves reciprocal effects of *wnt* versus *dickkopf (Dkk)* and sclerostin on the LRP5/6-β-catenin pathway. The switch from resorption to formation may rely on osteocyte signalling or on release of signals from the bone matrix, such as TGF-β. Bone remodelling is said to be coupled when formation follows resorption, but may be unbalanced when the amount of bone removed is not replaced with an equal amount.

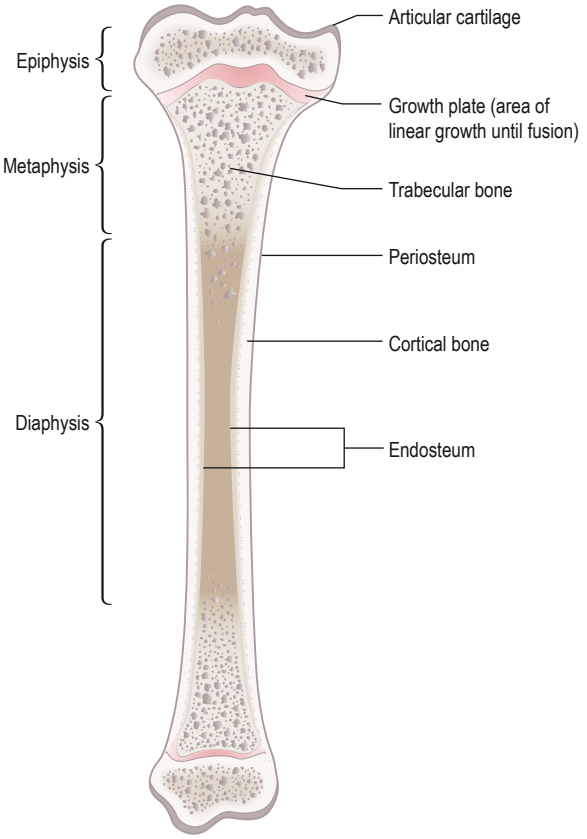


Fig. 33.1 **A longitudinal section of a growing long bone.**

Examples of bone remodelling include

- Myeloma cells have dual lytic effects, with enhanced expression of RANKL and *Dkk.*
- In rheumatoid arthritis, RANKL and *Dkk* are increased.
- In spondyloarthritis (characterized by new bone formation alongside erosion), *Dkk* is inhibited, with the increased *wnt* activity also increasing OPG relative to RANKL.
- Corticosteroids may increase osteocyte SOST expression, and stimulate expression of *Dkk.*

Calcium homeostasis and its regulation

Calcium homeostasis is regulated by the effects of PTH and 1,25-dihydroxyvitamin D_3 on gut, kidney, and bone. Calcium-sensing receptors are present in the parathyroid glands, kidney, and brain.

Calcium absorption and distribution

Daily calcium consumption (**Fig. 33.3**), primarily from dairy foods, is 20 to 25 mmol (800–1000 mg). The combined effect of calcium and vitamin D deficiency contributes to the bone fragility seen in some older people. Intestinal absorption of calcium is reduced by vitamin D deficiency and in malabsorption states (see Chapter 25).

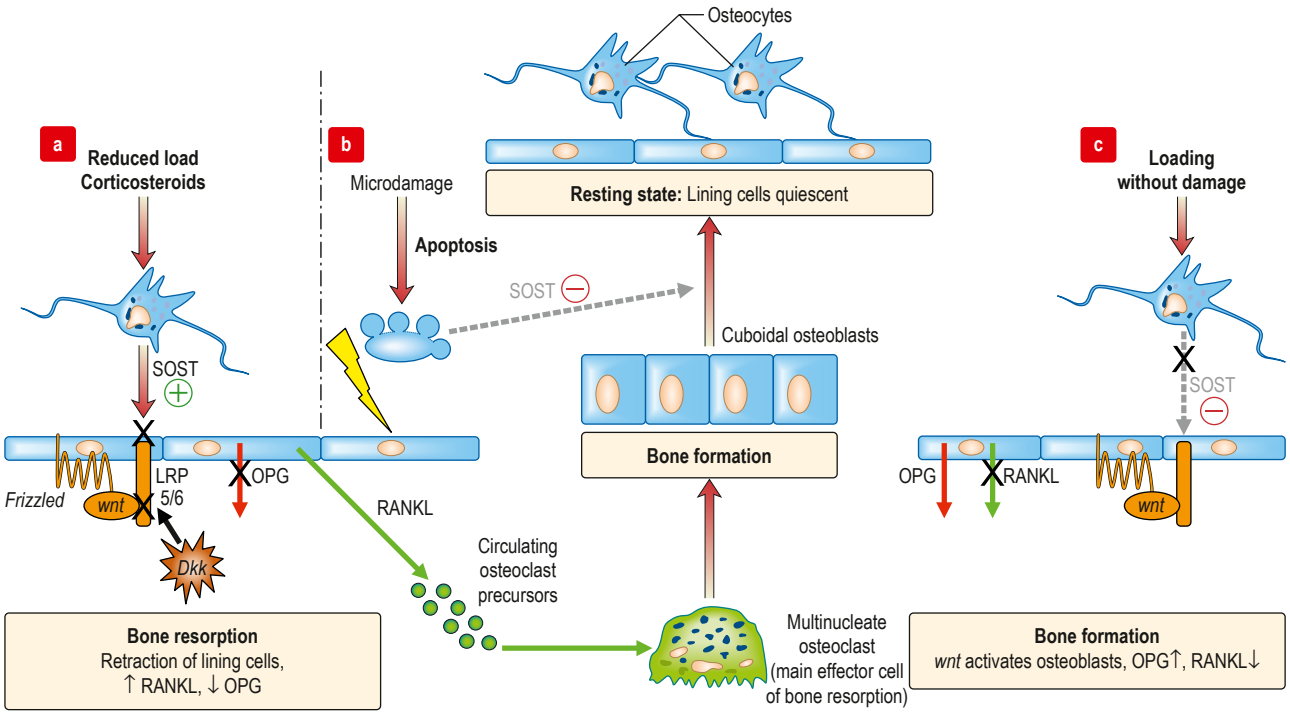


Fig. 33.2 Bone is remodelled in response to alterations in two reciprocal systems. (a) ***Quiescent bone*** may experience *reduced load*, in response the osteocyte increases expression of ***sclerostin (SOST+)***, inhibiting the response to *wnt* via the ***LRP5/frizzled co-receptor complex*** (orange). **(b)** ***Microdamage*** causes osteocyte apoptosis with direct osteoclast-generating effects via receptor activator of nuclear factor kappa B ligand (RANKL), and loss of the inhibitory effect of SOST on later bone formation (dashed line, SOST–). The osteoblastic lining cells retract from an area of bone, forming a bone remodelling unit, while increased expression of RANKL and reduced osteoprotegerin (OPG) activates the formation of bone-resorbing multinucleate osteoclasts from circulating precursors; the resorbed area is then replaced by new osteoid, formed by cuboidal osteoblasts. As new osteocytes are formed, SOST levels rise and the osteoblasts cease formation, the majority undergoing apoptosis. **(c)** If bone experiences *loading without damage*, ***sclerostin expression*** is reduced (SOST–), increasing signal response to *wnt*, activation of osteoblast bone formation, and increased OPG expression.

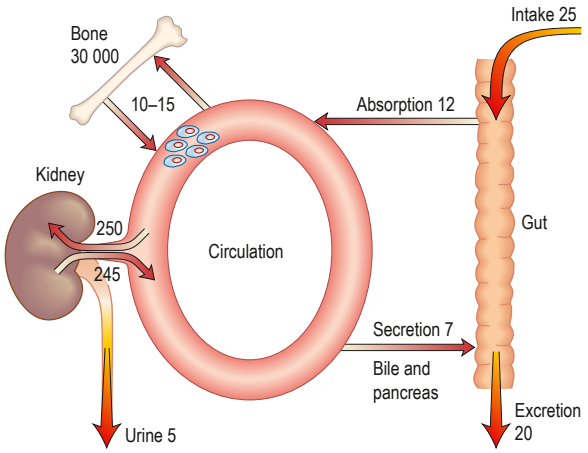


Fig. 33.3 Calcium exchange in the normal human. The amounts are shown in mmol per day.

Vitamin D metabolism

The primary source of vitamin D (Fig. 33.4) in humans is photoactivation in the skin of 7-dehydrocholesterol to cholecalciferol, which is then converted first in the liver to 25-hydroxyvitamin D (25(OH)D_3) and subsequently in the kidney (by the enzyme 1α-hydroxylase) to 1,25-dihydroxyvitamin D_3. (This step can occur in lymphomatous and sarcoid tissue, resulting in hypercalcaemia.) Regulation of the latter step is by PTH, phosphate, and feedback inhibition by 1,25-dihydroxyvitamin D_3.

Parathyroid hormone

Parathyroid hormone (PTH), an 84-amino-acid hormone, is secreted from the chief cells of the parathyroid gland, which have calcium-sensing and vitamin D receptors. PTH increases renal phosphate excretion and increases plasma calcium by

- increasing osteoclastic activity in bone (a rapid response)
- increasing intestinal absorption of calcium (a slower response)
- increasing 1α-hydroxylation of vitamin D (the rate-limiting step)
- increasing renal tubular reabsorption of calcium.

Hypomagnesaemia can suppress the normal PTH response to hypocalcaemia.

Calcitonin

Calcitonin is produced by thyroid C cells. Although calcitonin inhibits osteoclastic bone resorption and increases the renal excretion of calcium and phosphate, neither excess calcitonin (seen in medullary carcinoma of the thyroid) nor its deficiency following thyroidectomy has significant skeletal effects in humans.

Further reading

Rosen CJ. Vitamin D insufficiency. *N Engl J Med.* 2011;364:248–254.
Xiong J, Onal M, Jilka RL et al. Matrix-embedded cells control osteoclast formation. *Nat Med.* 2011;17:1235–1241.
http://courses.washington.edu/bonephys/. *Bone physiology*.

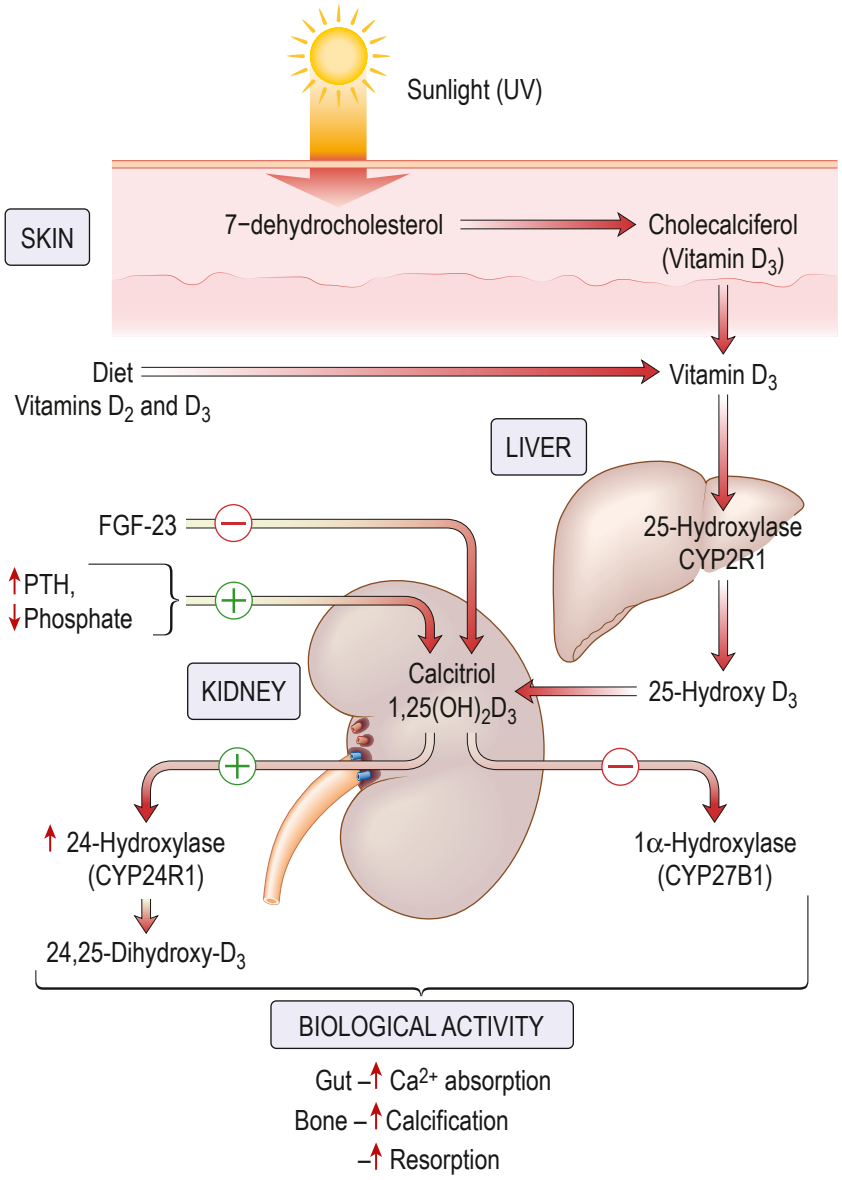


Fig. 33.4 The metabolism and actions of vitamin D. The biologically active agent is 1,25-dihydroxycholecalciferol (calcitriol). It inhibits 1α-hydroxylase and stimulates 24-hydroxylase to produce 24,25-dihydroxy-D_3, which is not biologically active. *CYP,* Cytochrome P enzyme; *FGF23,* fibroblast growth factor 23 secreted mainly by osteocytes (see Chapter 27).

INVESTIGATION OF BONE AND CALCIUM DISORDERS

See **Box 33.1**.

Total plasma calcium

Normal range is 2.2 to 2.6 mmol/L. About 40% of circulating calcium is ionized and physiologically active; the remainder is complexed or protein bound. As ionized calcium is difficult to measure, normal practice is to measure total calcium, correcting the value to allow for protein binding according to the following formula: add or subtract 0.02 mmol/L for each g/L of a simultaneous albumin level below or above 40 g/L. For critical measurements, samples should be taken in the fasting state and without a tourniquet (the latter may increase local plasma calcium concentration).

Plasma phosphate

Normal range is 0.8 to 1.4 mmol/L. Phosphate is essential to most biological systems. High levels are found in chronic kidney disease (CKD) and hypoparathyroidism, while low levels are associated with primary hyperparathyroidism, hypophosphataemic rickets and osteomalacia, and other disorders associated with reduced renal tubular phosphate reabsorption.

Plasma PTH

Normal range is 10 to 65 ng/L (1.3–7.6 pmol/L). The PTH assay measures the intact hormone. In hypercalcaemia not due to hyperparathyroidism, serum PTH levels are suppressed. Lithium toxicity may be associated with raised PTH levels; in familial hypocalciuric hypercalcaemia (FHH), serum PTH may be normal or marginally elevated.

Serum 25-hydroxyvitamin D

Vitamin D status is best assessed using serum 25-hydroxyvitamin D_3, as 1,25-dihydroxyvitamin D_3 has a short half-life and does not accurately reflect true vitamin D status. ***Vitamin D deficiency*** is defined as <25 nmol/L (10 ng/mL) and ***vitamin D insufficiency*** as less than 75 nmol/L (30 ng/mL). Rickets and osteomalacia occur with prolonged vitamin D deficiency.

The extra-skeletal significance of vitamin D insufficiency is uncertain but it has been linked to a wide range of conditions, including ischaemic heart disease, multiple sclerosis, and a variety of cancers. Recent evidence suggests that racial differences are present in the assessment of vitamin D status. Black Americans have consistently lower levels of 25-hydroxyvitamin D_3 than white Americans but also lower levels of ***vitamin D-binding protein***, resulting in equivalent bioavailable 25-hydroxyvitamin D_3.

24-hour urinary calcium

Normal range 2.5 to 7.5 mmol/24 h (100–300 mg/24 h). This is increased where renal tubular reabsorption of calcium is decreased, and in hypercalcaemia. One exception is FHH, where the genetic defect leads to inappropriately reduced calcium excretion. Measurement of 24-hour urinary calcium excretion should be performed in the assessment of hypercalcaemic patients.

Biochemical markers of bone formation and resorption

The clinical use of these biochemical markers is limited by large biovariability and measurement variance. Serial measurements at the same time of day in individual patients are useful in assessing response to treatment of metabolic bone diseases.

- ***Bone-specific alkaline phosphatase.*** Circulating alkaline phosphatase is derived from bone, liver, neutrophils, and placenta. The bone-specific isoenzyme can be measured as a marker of formation, although there is some overlap with the liver isoenzyme. Elevated serum levels occur during bone growth: for example, in adolescents, fracture repair and high-bone-turnover states.
- ***Type 1 collagen pro-peptides.*** These are by-products of collagen synthesis. Serum levels of both the carboxyterminal (P1CP) and aminoterminal (P1NP) pro-peptides reflect bone formation.
- ***Serum osteocalcin.*** This is another bone formation marker.
- ***Serum or urine levels of N-terminal (NTX) and C-terminal (CTX) cross-linked telopeptides.*** These reflect bone resorption. They may change rapidly in response to anti-resorptive drugs or in disease states, and have been used to assess fracture risk.

Diagnostic imaging

- ***Plain radiographs.*** These identify fractures, tumours, and infections. Other specific features may be seen (see following sections).
- ***Radionucleotide imaging.*** The uptake of a ^{99m}technetium-labelled bisphosphonate in bone reflects bone turnover and blood flow. Increased uptake is therefore seen in fractures, tumour and metastatic deposits, infection, and Paget's disease of bone.
- ***Magnetic resonance imaging (MRI).*** This is the most sensitive and specific test for the diagnosis of osteomyelitis. It is also useful in the detection of stress fractures, which may not be demonstrated on plain radiographs. A technique to suppress the high signal associated with bone marrow (such as STIR sequences; see Chapter 8) allows highly sensitive recognition of 'bone marrow oedema', a non-specific feature of a number of bone disorders, including osteonecrosis.
- ***Bone biopsy*** (Fig. 33.5). A core of bone is removed, including both cortices of the iliac crest, using a trephine. The non-decalcified specimen is examined for static and dynamic (bone turnover) indices. An oral tetracycline is given to the patient prior to the biopsy, for 2 days on two occasions 10 days apart, allowing assessment of the rate of bone turnover and mineralization. Biopsy is rarely performed but is useful in the assessment of renal bone disease and osteomalacia.
- ***Bone densitometry measurements*** (see 'Bone density' section).

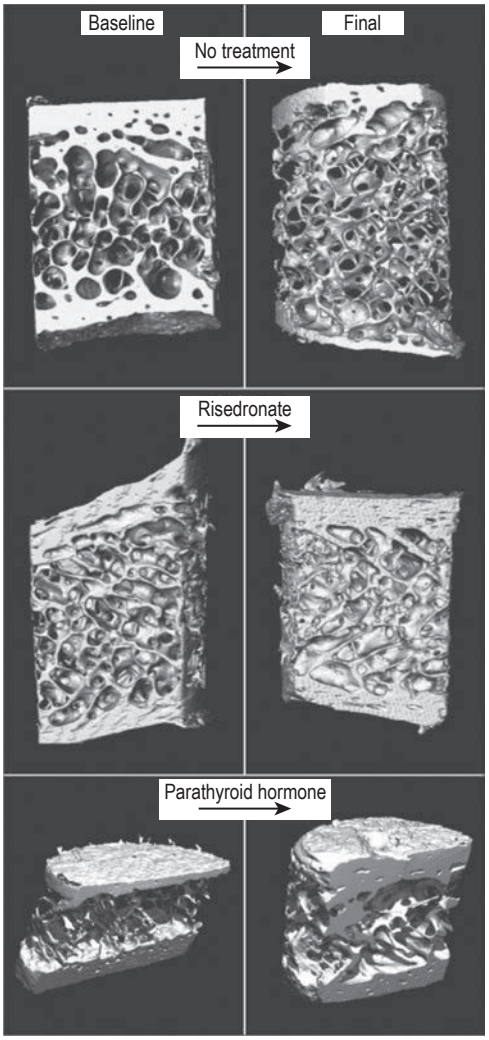


Fig. 33.5 Computerized images of bone biopsies in osteoporosis. These images demonstrate changes in bone architecture over time. ***Upper pair:*** Without treatment, the bone cortex becomes thinner and numerous trabecular rods and plates are lost. ***Centre pair:*** In response to bisphosphonate treatment, resorption is reduced, with preservation of bone mineral and structure. ***Lower pair:*** In response to anabolic therapy, bone is increased at both trabecular and cortical sites. *(From Jiang Y, Zhao JJ, Mitlak BH et al. Recombinant human parathyroid hormone [1–34] improves both cortical and cancellous bone structure.* J Bone Miner Res *2003; 18:1932–1941, with permission of the American Society for Bone and Mineral Research.)*

OSTEOPOROSIS

Osteoporosis is defined as 'a disease characterized by low bone mass and micro-architectural deterioration of bone tissue, leading to enhanced bone fragility and an increase in fracture risk'.

Using bone densitometry at the hip or spine measured by dual-energy X-ray absorptiometry (DXA), the World Health Organization (WHO) also defines osteoporosis as a bone density of 2.5 standard deviations (SDs) below the young healthy adult mean value (T-score ≤–2.5) or lower. Values between –1 and –2.5 SDs below the young adult mean are termed 'osteopenia'. The rationale for this definition is the inverse relationship between bone mineral density (BMD) and fracture risk in postmenopausal females and older males. However, this definition should not be applied to younger populations.

Fractures due to osteoporosis are a major cause of morbidity and mortality in elderly populations, with osteoporotic fractures of the spine causing acute pain or deformity and postural back pain. One in two females and one in five males aged 50 years will have an osteoporotic fracture during their remaining lifetime. As the risk of fracture increases with age, changing population demographics will increase the burden of disease.

Pathogenesis

Osteoporosis results from increased bone breakdown by osteoclasts and decreased bone formation by osteoblasts, leading to loss of bone mass.

Bone mass decreases with age (Fig. 33.6) but will depend on the 'peak' mass attained in adult life and on the rate of loss in later life. Genetic factors are the single most significant influence on peak bone mass. Multiple genes are involved, including collagen type 1A1, vitamin D receptor, and oestrogen-receptor genes. Nutritional factors, sex hormone status, and physical activity also affect peak mass. Not all causes of osteoporosis affect bone remodelling and architecture in the same way.

Oestrogen deficiency results in increased numbers of remodelling units, premature arrest of osteoblastic synthetic activity, and perforation of trabeculae, with a loss of resistance to fracture that is not fully reflected in the bone density measurement.

Glucocorticoids induce a high-turnover state initially, with increased fracture risk evident within 3 months of starting therapy. More prolonged use leads to a reduced-turnover state but with a net loss due to reduced synthesis (through inhibition of the *wnt*-LRP5/6 axis).

Ageing results in increased turnover at the bone/vascular interface within cortical bone, resulting in structural weakness (trabecularization of cortical bone).

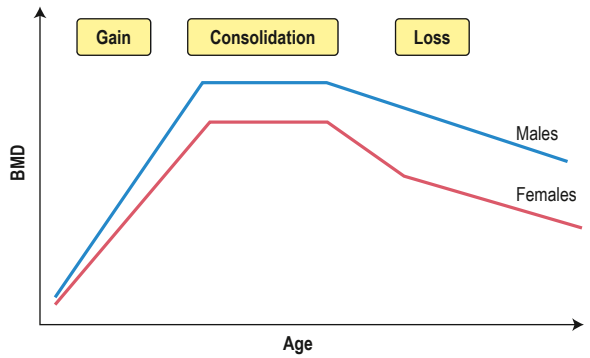


Fig. 33.6 Lifetime changes in bone mineral density (BMD). Peak bone mass is achieved between 20 and 30 years of age, and consolidated up to around the age of 40 years. Then, age-related bone loss occurs in both males and females, with an accelerated loss in females starting around the time of menopause, which lasts between 5 and 10 years.

Risk factors

Risk factors for fracture may exert their effect through reducing BMD, or they may increase risk over that attributable to BMD (BMD independent; Box 33.4). Oestrogen deficiency is a major factor in the pathogenesis of accelerated bone loss due to a normal or premature menopause, or amenorrhoea in patients with anorexia nervosa and in athletes. In the elderly, vitamin D insufficiency and consequent hyperparathyroidism reduce BMD. However, previous fracture, increasing age, glucocorticoid therapy, smoking, and falls increase the risk of fracture at any given BMD. For instance, 10% of females who are 65 years old and have a T-score of –2 at the hip would be expected to sustain a fracture over the next 10 years; if similar females had a Colles' fracture, smoked, and had prolonged exposure to steroids, their risk would be closer to 26% in the same period.

Box 33.4 Risk factors for fragility fractures

Bone mineral density-dependent
- Female sex
- White/Asian ethnicity
- Gastrointestinal disease
- Hypogonadism
- Immobilization
- Chronic liver disease
- Chronic kidney disease
- Low dietary calcium intake
- Vitamin D insufficiency
- Chronic obstructive pulmonary disease
- Cushing's syndrome
- Hyperthyroidism
- Hyperparathyroidism
- Diabetes mellitus
- Mastocytosis
- Multiple myeloma
- Osteogenesis imperfecta

Drugs
- Heparin
- Calcineurin inhibitors, e.g., ciclosporin
- Anticonvulsants
- Thiazolidinediones
- Aromatase inhibitors
- Anti-androgens
- Gonadotrophin-releasing hormone (GnRH) analogues
- Proton pump inhibitors
- ?Selective serotonin reuptake inhibitors

Bone mineral density-independent
- Increasing age
- Previous fragility fracture
- Family history of hip fracture
- Low body mass index
- Smoking
- Excess alcohol use
- Glucocorticoid therapy
- High bone turnover
- Increased risk of falling
- Rheumatoid arthritis

BMD-dependent risk factors respond to bone-directed treatment. BMD-independent risk factors benefit from additional targeted intervention (e.g., reduction of falls risk—see Chapter 49).

Clinical features

Fracture is the only cause of symptoms in osteoporosis. ***Vertebral fracture*** is suggested by the sudden onset of severe back pain. However, two in every three vertebral fractures are asymptomatic. Pain from mechanical derangement, increasing kyphosis, height loss, and abdominal protuberance may follow vertebral fracture. ***Colles' fractures*** typically follow a fall on an outstretched arm. ***Fractures of the proximal femur*** usually occur in older individuals falling on their side or back.

Other causes of low-trauma fractures must not be overlooked, including metastatic disease and myeloma.

Investigations

Plain X-rays (Fig. 33.7) may show fractures including previously asymptomatic vertebral fractures. Such clinically silent fractures may also be detected during DXA scanning with an additional analysis (called ***vertebral fracture assessment***, Box 33.5) carried out with a much lower radiation dose than conventional imaging. Increased bone radiolucency suggests low BMD but requires confirmation with DXA.

Bone density

DXA measures areal bone density (mineral per surface area rather than a true volumetric density), usually of the lumbar spine and proximal femur. It is precise and accurate, uses low doses of radiation and is the 'gold standard' in diagnosis of osteoporosis (Fig. 33.8). Because of osteophytes, spinal deformity, and vertebral fractures, spinal values may be artefactually elevated and should be interpreted with caution in the elderly.

Associated disease and risk factors

Investigations to exclude other diseases or identify contributory factors associated with osteoporosis should be performed and they are particularly necessary in males, in whom secondary causes are more common (see Box 33.4).

Selection of individuals for treatment: risk assessment

The goal of osteoporosis treatment is to reduce fracture risk (Box 33.6). Thus, assessment of absolute fracture risk should be made in every case. A previous fracture is the strongest risk factor for further fractures. DXA is useful in guiding treatment decisions (Box 33.7; Fig. 33.8). Patients with a previous fragility fracture or a BMD T-score less than or equal to –2.5 should be considered for treatment. It is vital to recognize, however, that the majority of fragility fractures occur in females with a T-score better than –2.5 (Fig. 33.9). In these patients, an assessment for future fracture risk should be made; this can be performed using clinical judgement by an experienced physician or using formal risk calculators, for example FRAX (Box 33.8). The threshold for recommending treatment in this setting will be determined by the cost-effectiveness of treatment in a particular healthcare setting and by clinical judgement. For example,

Box 33.5 Vertebral fracture assessment (VFA)

- Radiographic image of lateral spine generated using DXA
- Lower radiation than formal X-ray
- Can be done at same time as DXA
- Used to identify previously unsuspected vertebral fractures
- If vertebral fragility fracture identified, treatment for osteoporosis indicated irrespective of T-score

DXA, Dual-energy X-ray absorptiometry.

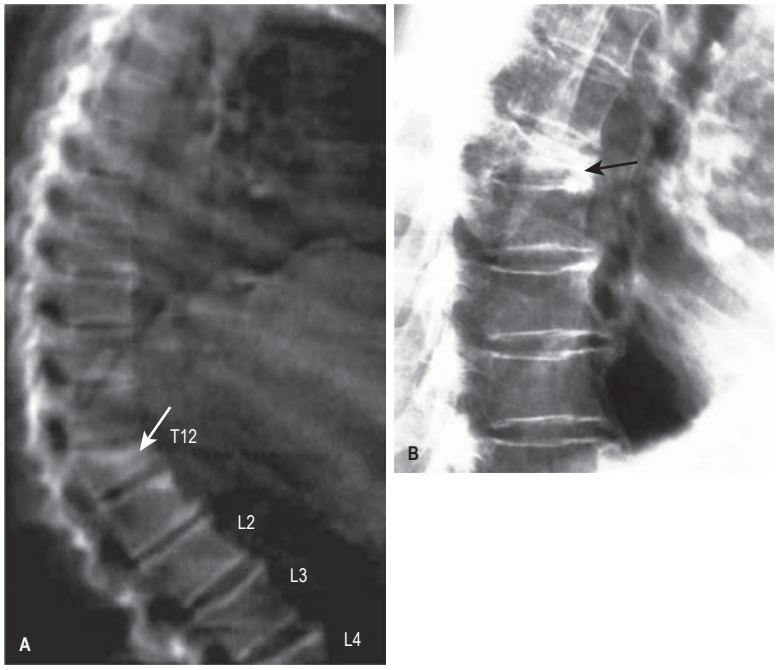


Fig. 33.7 Vertebral fracture as seen on DXA vertebral fracture assessment (VFA) and plain radiograph. (A) The VFA image shows a vertebra with reduced height at T12 *(arrow)*. **(B)** The radiograph shows a typical wedge compression fracture.

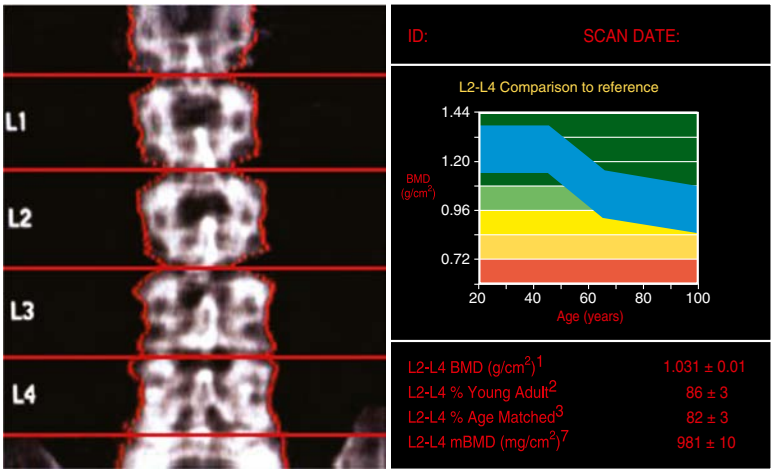


Fig. 33.8 Dual-energy X-ray absorptiometry (DXA) of the lumbar spine. **(A)** Anteroposterior image of the lumbar spine. **(B)** The patient's value is expressed in relation to the reference range.

i Box 33.6 Management of osteoporosis: summary

Treatment is guided by risk of fracture, not BMD alone.
Fracture risk assessment should be performed, either using clinical judgement or a formal risk assessment calculator (e.g., http://www.shef.ac.uk/FRAX and many others; see **Box 33.8**).

Do not under-estimate the risk from steroids or previous fracture.

Oral bisphosphonates (alendronate, risedronate) are first-line treatment in many cases. Other options include:

- ***denosumab:***
  - difficulty taking bisphosphonate
  - poor adherence to treatment
  - falling BMD on bisphosphonate
  - new fracture on a bisphosphonate
  - after 5–10 years on bisphosphonate
- ***teriparatide*** if there are multiple vertebral fractures or very low BMD
- ***romosozumab*** in very low BMD
- ***i.v. zoledronate*** if difficulty tolerating oral bisphosphonate

***BMD monitoring* is recommended in:**

- all treated patients to monitor therapy
- all untreated cases if a BMD reduction would lead to treatment.

BMD, Bone mineral density.

i Box 33.7 DXA scanning

Indications for DXA scanning

- Females aged ≥65
- Males aged ≥70
- Fragility fracture
- Females <65 with risk factors (see **Box 33.4**)
- Males <70 with risk factors (see **Box 33.4**)
- Monitoring of BMD in those on treatment

Features of DXA scanning

- Low-dose radiation X-ray
- Measures a real BMD
- Gives BMD in g/cm^2, which is difficult to interpret so translated to T- or Z-score
- T-score generated by comparing result with female, white, aged 20–29 years taken from the NHANES III study database
- Z-score generated by comparing result with person of same age and ethnicity and used to report BMD in premenopausal females and males aged <50

Interpretation of DXA

BMD measured at lumbar spine, total hip, femoral next (± radius)

In postmenopausal females and males aged ≥50:

- T-score ≤–2.5 = osteoporosis
- T-score ≤–1.0 to <–2.5 = osteopenia/low BMD
- T-score >–1.0 = normal BMD
- Z-scores ≤–2.0 is 'low BMD for age'

Interpretation of Z-scores is complex and beyond the scope of this text

BMD, Bone mineral density; *DXA*, dual-energy X-ray absorptiometry.

in the USA, the Bone Health and Osteoporosis Foundation recommends treatment if the 10-year probability of hip fracture is more than or equal to 3% or if major osteoporotic fracture is more than or equal to 20%. In the UK, the National Osteoporosis Guideline Group (NOGG) models a FRAX-based age-dependent intervention threshold that ranges from 7.5% to 30% for 10-year probability of major osteoporotic fracture.

Prevention and management

- ***Symptomatic management.*** New vertebral fractures may require bed rest for 1 to 2 weeks with strong analgesia and gradual physiotherapy to restore confident mobilization (see also Chapter 32). Pain following vertebral fracture takes an average of 3 months to resolve. Non-spinal fractures should be treated by conventional orthopaedic means. In fractured neck of femur, early surgical intervention, and early mobilization post surgery, are associated with improved outcomes.
- ***Calcium and vitamin D.*** Daily intakes of 800 to 1200 mg of calcium and 400 to 800 IU of vitamin D are recommended throughout life for optimum bone health. Dietary intake of calcium, and vitamin D from sunshine and diet, are preferable. For those not meeting calcium intake targets, or those with low serum 25-hydroxyvitamin D_3 levels, supplements are recommended (**Box 33.9**).
- ***Lifestyle measures.*** Weight-bearing exercise for 30 minutes three times a week may increase BMD, while gentle exercise in the elderly may reduce the risk of falls and improve the protective

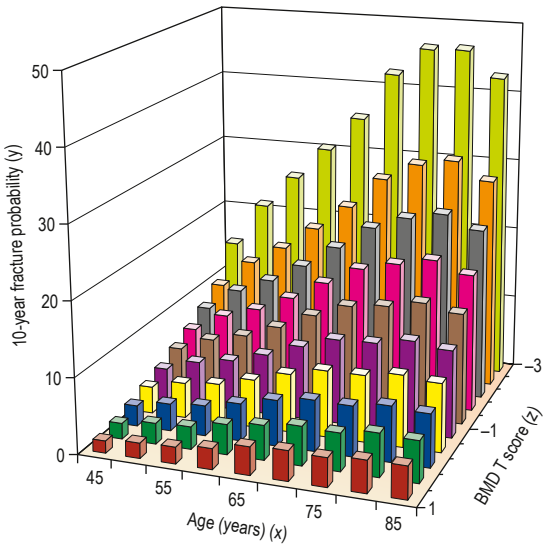


Fig. 33.9 Graph illustrating the combined effects of age (x-axis) and reduced bone mass (expressed as T-scores on the z-axis) on the 10-year probability of fracture (y-axis) in a population of females. Reduced bone mass osteopenia, T-scores between –1 and –2.5. Osteoporotic bone (T-scores <–2.5) is shown in pale green. Note that the risk of fracture may be greater in older females with osteopenia than in young females with osteoporosis. *(Data from Kanis JA. Diagnosis of osteoporosis and assessment of fracture risk.* Lancet *359:1929–1936.)*

Box 33.8 Fracture risk assessment

Fracture risk assessment:
- Should be performed in all patients
- Can be performed clinically or using risk assessment calculators.

FRAX is one of a number of fracture risk assessment tools available:
- Treatment is generally indicated if a 10-year major fracture risk ≥20% or hip fracture risk ≥3%
- These tools enable quantification of fracture risk and guide treatment decisions
- They provide estimates of fracture risk based on a set number of variables
- In certain situations they will not accurately estimate fracture risk so must be used in combination with clinical judgement

Box 33.9 Management of vitamin D deficiency

Treatment phase
- Vitamin D 50 000 IU weekly PO × 6–8 weeks

Maintenance phase
- Vitamin D 800–1000 IU daily PO × lifelong

responses to falling (see Chapter 49). Smoking and excess alcohol use should be avoided.
- ***Reduction of falls.*** Physiotherapy and assessment of home safety are helpful. Hip protectors do reduce fractures in the elderly in residential care when worn correctly, but compliance is poor.

Pharmacological intervention

Most interventions (see **Box 33.6**) act by inhibiting bone resorption (anti-resorptives), the exception being PTH peptides and romosozumab, which stimulate bone formation.

Box 33.10 Evidence for fracture risk reduction of current treatment options at specific sites in postmenopausal females

Intervention	Vertebral fracture	Non-vertebral fracture	Hip fracture
Alendronate	+	+	+
Risedronate	+	+	+
Ibandronate	+	+[a]	ND
Zoledronate	+	+	+
Bazedoxifene	+	ND	ND
Raloxifene	+	ND	ND
Denosumab	+	+	+
Teriparatide	+	+	ND
Abaloparatide	+	+	ND
Romosozumab	+	+	+

[a]Demonstrated only in high-risk subgroup.
ND, Not demonstrated.

The evidence base for the ***anti-fracture efficacy*** of interventions varies. Treatments with proven efficacy at all major fracture sites (particularly spine and hip) are preferable (**Box 33.10**). The bisphosphonates are generally regarded as the first-line option in the majority of postmenopausal females with osteoporosis.

Bisphosphonates

As synthetic analogues of bone pyrophosphate, bisphosphonates adhere to hydroxyapatite and inhibit osteoclasts. Alendronate and risedronate are given as once-weekly doses, zoledronate as a once-yearly infusion, and ibandronate usually as a once-monthly oral therapy (a 3-monthly intravenous injection is rarely used).

Oral bisphosphonates should be taken fasting, with a large drink of water, while standing or sitting upright. The patient should then remain upright and avoid food and drink for at least 30 minutes.

Bisphosphonates are generally well tolerated but may be associated with upper gastrointestinal side-effects such as oesophagitis, particularly if the dosing instructions are not closely followed. Bisphosphonates should be used with careful monitoring in patients who have CKD (stage 4 or 5). Osteonecrosis of the jaw is a rare complication of high-dose intravenous bisphosphonates in patients who have malignant disease. It is associated with poor dental hygiene. As prolonged suppression of bone turnover is linked with atypical femoral fractures, it is currently advised to reassess bisphosphonate treatment after 5 years for oral agents and 3 years for zoledronate. For those estimated to remain at high risk of fracture, a further 5 years of oral treatment or 3 years of intravenous treatment is appropriate.

Denosumab

Denosumab is a fully human monoclonal antibody to RANKL administered as a single subcutaneous injection every 6 months. It is an anti-resorptive agent that increases BMD and reduces fractures at the spine, hip, and other non-vertebral sites. Fracture risk reduction is equivalent to that of bisphosphonates. Unlike bisphosphonates, the benefit of denosumab is rapidly lost on cessation. This results in a precipitous drop in BMD and often multiple vertebral fractures. Therefore, it should generally not be stopped and should continue as a lifelong treatment. Strategies to substitute other agents such as bisphosphonates to allow denosumab cessation appear partially effective at best. Adverse effects are infrequent: most commonly

dysuria, rarely cellulitis. Osteonecrosis of the jaw and atypical femoral fractures have also been reported.

Selective oestrogen-receptor modulators

Selective oestrogen-receptor modulators (SERMs) include raloxifene and bazedoxifene. They are rarely used to treat osteoporosis. They have no stimulatory effect on the endometrium but activate oestrogen receptors in bone. Both prevent BMD loss at the spine and hip in postmenopausal females but have been found to reduce only vertebral fracture rates. Leg cramps and flushing may occur, and the risk of thromboembolic complications is also increased to a degree similar to that seen with hormone replacement therapy (HRT; see Chapter 47 and Box 47.16). The use of SERMs is associated with a small increase in the risk of stroke.

Recombinant human parathyroid hormone

Recombinant human PTH peptide 1–34 (teriparatide) and recombinant human PTH 1–84 are anabolic agents that stimulate bone formation. Teriparatide reduces vertebral and non-vertebral fractures in postmenopausal females with established osteoporosis, although specific data on hip fracture are not available. It is given by daily subcutaneous injection for 24 months. Recombinant human PTH 1–84 is also administered by once-daily subcutaneous injection but has been shown to reduce only vertebral fractures. An anti-resorptive drug should be given on completion of PTH peptide therapy to maintain the increase in BMD. Non-osteoporotic bone diseases, such as osteomalacia, should be excluded prior to treatment. PTH peptide therapy is difficult for patients to take and associated with more adverse events than other agents and therefore is mainly indicated in severe cases of vertebral osteoporosis or in those who fail to respond to other therapies. Teriparatide may cause mild transient hypercalcaemia but routine monitoring is not required. Nausea and headache may occur. Recombinant human PTH 1–84 is associated with a higher incidence of hypercalcaemia and hypercalciuria, and routine monitoring is advised. Neither agent should be used in people with skeletal metastases or osteosarcoma.

Romosozumab

Romosozumab is a monoclonal antibody that binds sclerostin. It is an anabolic agent that has been demonstrated to reduce vertebral, hip, and non-vertebral fracture. Treatment is by monthly subcutaneous injection for 1 year and should be followed by an anti-resorptive agent. There is a concern over increased adverse cardiovascular events from one study, and therefore romosozumab is best avoided in patients at increased risk of cardiovascular events. Romosozumab has a similar role in treatment to PTH agents, and is best reserved for those with severe osteoporosis or who fail to respond to other agents.

Hormone replacement therapy

Because of its adverse effects profile and the availability of effective bone-specific agents, HRT is generally not indicated solely for osteoporosis, but may benefit bone health early postmenopausal females who commence HRT for perimenopausal symptoms.

Calcitriol (1,25-dihydroxyvitamin D_3) and calcitonin

These may reduce vertebral fracture rate, although the data are inconsistent.

Combination therapies

Combination therapies, either with two anti-resorptive agents, or an anti-resorptive and an anabolic agent, often produce larger increases in BMD than monotherapy but have not been shown to result in greater fracture reduction. Therefore, they are generally not advised. An exception to this may be in non-responding denosumab-treated patients, who may benefit from the addition of recombinant PTH.

Surgery

This is very rarely indicated in osteoporotic vertebral fractures. Percutaneous vertebroplasty and balloon kyphoplasty are discussed in Chapter 32. Hip fractures are dealt with by hip replacements or stabilization with pins.

Treatment of specific conditions

Glucocorticoid-induced osteoporosis

Individuals requiring continuous oral glucocorticoid therapy for 3 months or more (at any dose) should be assessed for coexisting risk factors (age, previous fracture, hormone status). Postmenopausal females, males aged over 50 years and any individuals who have sustained a fragility fracture should receive treatment without waiting for DXA scanning. DXA results and fracture risk assessment guide treatment for other patients (see Boxes 33.5, 33.7, and 33.8). Bisphosphonates, denosumab, and teriparatide are approved in this setting. Bisphosphonates are generally preferred due to the cost and adverse event profile of teriparatide, and the often temporary nature of treatment making denosumab less preferable.

Osteoporosis in males

Alendronate, risedronate, zoledronate, denosumab, and teriparatide are approved for use in males. In males with osteoporosis who have clinical and biochemical evidence of hypogonadism, testosterone replacement is also used.

Further reading

Black DM, Rosen CJ. Clinical Practice. Postmenopausal Osteoporosis. *N Engl J Med.* 2016;374(3):254–62.
Buckley L, Humphrey MB. Glucocorticoid-Induced Osteoporosis. *N Engl J Med.* 2018;379(26):2547–2556.
Compston JE, McClung MR, Leslie WD. Osteoporosis. *Lancet.* 2019;393(10169):364–376.
Fuggle NR, Curtis B, Clynes M et al. The treatment gap: The missed opportunities for osteoporosis therapy. *Bone*. 2021;144:115833.
Reid IR, Billington EO. Drug therapy for osteoporosis in older adults. *Lancet.* 2022;399(10329):1080–1092.
Vilaca T, Eastell R, Schini M. Osteoporosis in men. *Lancet Diabetes Endocrinol.* 2022;10(4):273–283.
https://www.bonehealthandosteoporosis.org/. *Bone Health and Osteoporosis Foundation clinical guidelines*.
https://www.nogg.org.uk/. *National Osteoporosis Guideline Group clinical guidelines*.
http://www.shef.ac.uk/FRAX. *FRAX tool for 10-year probability of hip fracture and 10-year probability of major osteoporotic fracture*.

OSTEONECROSIS

This is also known as aseptic, avascular, or ischaemic necrosis of bone. Over 80% of cases are attributed to glucocorticoid treatment or alcohol excess. Less frequent causes include sickle cell disease, systemic lupus erythematosus (SLE), deep-sea diving (Caisson's disease), endocrine disorders (e.g., Cushing's, diabetes mellitus), trauma, human immunodeficiency virus (HIV) infection, and irradiation.

Osteonecrosis usually presents with joint pain, the shoulder or hip being most commonly affected, but can be asymptomatic, particularly on the opposite side in the same person. It may only be recognized when it results in collapse of the articular bone.

MRI best confirms the diagnosis by showing bone marrow oedema. If advanced, it can be seen on plain X-rays.

Management is mainly symptomatic. Surgical options include drilling through the bone cortex (decompression), vascularized bone grafts, or rotation of the affected bone away from the load-bearing area; however, joint replacement is often required. Bisphosphonates may reduce pain and progression but the evidence is mixed.

PAGET'S DISEASE OF BONE

Paget's disease of bone is a focal disorder of bone remodelling. Increased osteoclastic bone resorption is followed by a compensatory increase in new bone formation, increased local bone blood flow, and fibrous tissue in adjacent bone marrow. Ultimately, formation exceeds resorption but the new woven bone is weaker than normal bone, which leads to deformity and increased fracture risk. Paget's disease does not spread but can become symptomatic at previously silent sites.

Epidemiological studies are difficult because most affected individuals are asymptomatic. Paget's disease is most often seen in Europe and particularly in northern England. It affects males and females (2:3) over the age of 40 years. The incidence approximately doubles per decade thereafter, with up to 10% of individuals radiologically affected by the age of 90. For unknown reasons, the incidence and severity of Paget's disease have decreased in recent years.

Aetiology and pathogenesis

Genetic factors are implicated in Paget's disease. A positive family history is noted in about 15%. Mutations in *SQSTM1*, which encodes the osteoclast mediator protein p62, have been reported in up to 10% of cases. Intracellular inclusions in the osteoclasts in pagetic lesions are believed to be paramyxovirus nucleocapsid (e.g., canine distemper virus, measles, or respiratory syncytial virus). However, similar inclusions are seen in other bone disorders, and theories of a viral aetiology in Paget's remain contentious. Altered expression of *c-fos* (an oncogene) is one suggested mechanism linking viral infection with the pathogenic changes in osteoclasts, which are more numerous and contain an increased number of nuclei (up to 100).

Clinical features

Clinical features are summarized in Fig. 33.10A. Between 60% and 80% of people with radiologically identified Paget's disease are asymptomatic. Diagnosis often follows the finding of an asymptomatic elevation of serum alkaline phosphatase, or a plain X-ray performed for other indications. The disease may involve one bone (monostotic, in 15%) or many (polyostotic). The most common sites, in order of frequency, are pelvis, femur, lumbar spine, skull, and tibia (see Fig. 33.10C). Small bones of the feet and hands are rarely involved.

Symptoms and complications include:

- bone pain
- joint pain when an involved bone is close to a joint, leading to cartilage damage and osteoarthritis
- deformities, in particular bowed tibia and skull changes
- neurological complications—nerve compression (deafness from VIIIth cranial nerve involvement; cranial nerves II, V, and VII may also be involved); spinal stenosis; hydrocephalus due to blockage of the aqueduct of Sylvius
- high-output cardiac failure and myocardial hypertrophy due to increased bone blood flow (rarely seen with modern management)
- hypercalcaemia—rarely seen outside the setting of fracture
- pathological fractures
- osteosarcoma—occurs in fewer than 1% of cases of Paget's and may be heralded by an increase in bone pain or swelling in a previous pagetic bone area.

Investigations

- ***Increased serum alkaline phosphatase*** with normal serum calcium and phosphate reflects increased bone turnover.

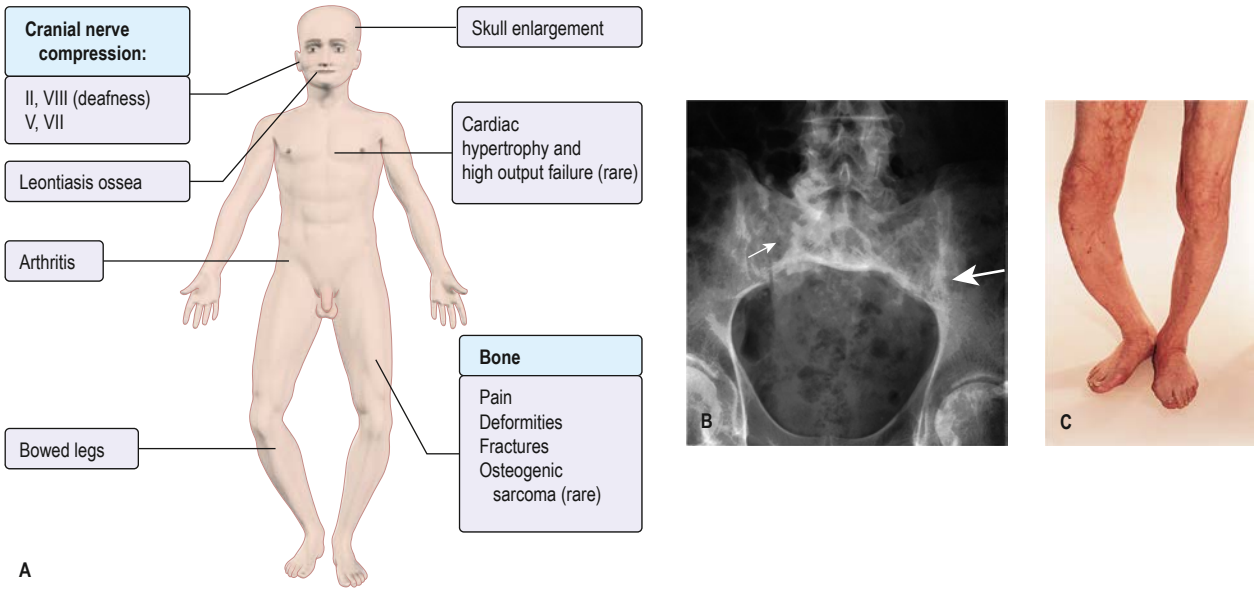


Fig. 33.10 Paget's disease. **(A)** Clinical features. **(B)** X-ray appearance of the pelvis, showing osteolytic *(thin arrow)* and osteosclerotic *(thick arrow)* lesions. **(C)** Legs showing bowing of the tibia caused by increased bone growth. Note the erythema ab igne on the medial aspect of the thigh.

Levels may be normal with limited or monostotic Paget's disease. Levels are reduced with treatment and increased during relapse.

- ***Vitamin D*** should be measured, as deficiency is frequent in the age group affected by Paget's disease and should be corrected (see **Box 33.9**) to avoid hypocalcaemia following bisphosphonate treatment.
- ***X-ray*** features (see **Fig. 33.10B**) vary from predominantly lytic lesions (osteoporosis circumscripta in the skull is characteristic), through a mixed phase, to a mainly sclerotic phase of bone expansion, cortical thickening and coarsening of the trabecular pattern.
- ***Isotope bone scans*** are useful to determine the extent of skeletal involvement but are unable to distinguish between Paget's disease and sclerotic metastatic carcinoma (especially breast and prostate).

Management

Bisphosphonates are the mainstay of treatment. New bone formed after treatment is lamellar, not woven (reflecting normalization of bone turnover rather than a direct effect on osteoblasts). Repeat treatment courses are guided by recurrence of symptoms and by rising alkaline phosphatase. In addition to treatment of symptomatic patients, treatment of asymptomatic lesions is appropriate if there is a significant risk of potential complications, such as fracture in weight-bearing long bones or the spine, nerve entrapment, or deafness with skull involvement, and before orthopaedic procedures in involved bone (to reduce vascularity).

Intravenous bisphosphonates

Zoledronate is the most commonly used agent for Paget's disease, administered as a single infusion over 15 minutes. ***Pamidronate*** is an alternative but takes longer to infuse and is less potent; some patients develop drug resistance for unknown reasons. Both drugs can be associated with a first-dose reaction characterized by 'flu-like' symptoms, including transient pyrexia over 24 to 48 hours, which can be ameliorated with paracetamol.

Oral bisphosphonates

Oral bisphosphonates are used at doses higher than those for osteoporosis (e.g., 30 mg risedronate daily for 2 months, or 40 mg alendronate daily for 6 months) and are less effective than intravenous zoledronic acid.

Surgery

Joint replacement or osteotomy is sometimes necessary to correct deformity or pain due to associated degenerative joint disease. Neurosurgery may be needed where there is spinal disease. Osteosarcoma usually requires amputation, though wide excision and limb salvage can be successful at distal sites.

Further reading

Ralston SH. Paget's disease of bone. *N Engl J Med.* 2013; 368:644–650.

RICKETS AND OSTEOMALACIA

Osteomalacia is defective mineralization of newly formed bone matrix or osteoid. Rickets is defective mineralization at the epiphyseal growth plate and is found in association with osteomalacia in children.

Aetiology

Many factors can result in defective mineralization of the osteoid. For normal mineralization, adequate levels of vitamin D, calcium, and phosphate, adequate activity of alkaline phosphatase, a normal pH at the osteoid surface, and normal osteoid composition are all necessary (**Box 33.11**).

The most common cause of osteomalacia is hypophosphataemia due to hyperparathyroidism secondary to vitamin D deficiency. The most common cause of vitamin D deficiency worldwide is dietary deficiency. Bread, milk, and cereals in high-income countries are now fortified with vitamin D. This has led to a much-reduced incidence of osteomalacia and rickets.

Vitamin D is produced in the skin through the action of sunlight on 7-dehydrocholesterol (see **Fig. 33.4**). Lack of sun exposure can lead to vitamin D deficiency, especially in individuals living in temperate regions who keep large parts of the skin covered throughout the year.

Vitamin D is a fat-soluble vitamin, so gastrointestinal disease can result in malabsorption. Gastrectomy, various forms of bariatric surgery, cystic fibrosis, coeliac disease, Crohn's disease, and primary biliary cirrhosis are well-recognized causes.

Due to the intimate involvement of the kidney in phosphate balance, a number of causes of osteomalacia are mediated by the kidney (see Chapter 26). Primary renal phosphate wasting occurs in tumour-induced osteomalacia, multiple myeloma, and Fanconi's syndrome. Proximal (type 2) renal tubular acidosis can cause osteomalacia due to both renal phosphate wasting and abnormal osteoid pH secondary to metabolic acidosis.

Clinical features

Osteomalacia may be asymptomatic and identified incidentally on routine investigations following a fragility fracture. When symptomatic, it characteristically causes muscle weakness and widespread bone pain. Muscle weakness is due to a multifactorial proximal myopathy, with low vitamin D, hypophosphataemia, and high PTH levels all contributing. It results in a characteristic waddling gait with difficulty climbing stairs and getting out of a chair. Generalized bone pain and tenderness are thought to be caused by hydration of the

? Box 33.11 Causes of rickets/osteomalacia

Deficient intake or absorption of vitamin D
- Inadequate sun exposure or deficient synthesis in skin
- Low dietary intake
- Malabsorption: coeliac disease, Crohn's disease, gastrectomy, bariatric surgery, primary biliary cirrhosis

Defective 1α-hydroxylation
- Chronic kidney disease
- Vitamin D-dependent rickets type I—due to deficiency of 1α-hydroxylase

Primary renal phosphate wasting
- Renal tubular acidosis (also by causing metabolic acidosis)
- Hereditary hypophosphataemic rickets (vitamin D-resistant rickets)
- Dent's disease
- Fanconi's syndrome
- Multiple myeloma
- Tumour-induced osteomalacia

Inhibitors of mineralization
- Fluoride, aluminium, bisphosphonates
- Metabolic acidosis

Defective vitamin D receptors
- Hereditary vitamin D-resistant rickets (previously known as vitamin D-dependent rickets type II)

demineralized matrix, resulting in periosteal distension. The pain is typically a dull ache that is worse on weight-bearing and walking. It can be reproduced by pressure on the sternum or tibia. Insufficiency fractures can occur when the quality of the bone is insufficient to handle the stress of weight-bearing.

At birth, neonatal rickets may present as craniotabes (a thin, deformed skull). In the first few years of life, there may be widened epiphyses at the wrists and beading at the costochondral junctions, producing the 'rickety rosary', or a groove in the rib cage (Harrison's sulcus). In older children, lower limb deformities are seen. A myopathy may also occur. Hypocalcaemic tetany may occur in severe cases.

Investigations

- ***Serum alkaline phosphatase*** is elevated in 90% of cases.
- ***Low serum calcium***, ***low phosphate***, and ***elevated PTH*** are each present in approximately half of the cases.
- ***Serum*** 25-hydroxyvitamin D_3 is low, usually less than 25 nmol/L (10 ng/mL).
- ***Serum FGF-23*** is elevated in many people with tumour-induced osteomalacia and in hypophosphataemic rickets.
- ***Plain radiographs*** demonstrate decreased bone mineralization. The characteristic finding in osteomalacia is Looser's pseudofractures. These are narrow radiolucent lines with sclerotic borders running perpendicular to the cortex. They can be found at any site but are most commonly seen in the femur and pelvis.
- ***Tetracycline-labelled bone biopsy*** is the gold standard diagnostic test. This is not practical in most clinical settings and is used mainly in research studies.

Management

Vitamin D replacement is the cornerstone of treatment. Treatment involves two stages: an initial loading stage to replenish body stores of vitamin D, and a subsequent maintenance phase to avoid repeat deficiency (see **Box 33.9**). All patients should also receive supplementary calcium of 1000 to 1200 mg/day. In nutritional deficiency, recommended initial replacement is with vitamin D 50 000 units per week orally. The initial replacement dose should be continued for 6 to 8 weeks. Adequacy of vitamin D replacement should be confirmed by repeat measuring of vitamin D levels. This should be followed by maintenance supplementation with 800 to 1000 units of vitamin D per day.

Doses for children are lower and are age-dependent. People with gastrointestinal disease and vitamin D deficiency due to malabsorption need higher doses of 10 000 to 50 000 units per day of vitamin D.

Tumour-induced osteomalacia is best treated by removal of the causative neoplasm, which is usually occult and frequently benign. This leads to rapid resolution of symptoms.

Further reading

Elder CJ, Bishop NJ. Rickets. *Lancet.* 2014; 383:1665–1676.
Pearce SH, Cheetham TD. Diagnosis and management of vitamin D deficiency. *BMJ.* 2010; 340:b5664.

BONE INFECTIONS

Acute and chronic osteomyelitis

Osteomyelitis is more common in children and arises due to haematogenous spread to the vascular metaphysis. In high-income countries, it occurs in around 8 per 100 000 children per year, but it is more common in low-income countries, with males affected twice as often as females. In adults, it generally occurs in the setting of other medical conditions, such as type 2 diabetes mellitus and chronic skin ulceration where it infiltrates the bone through direct spreading.

Staphylococcus is the organism responsible for 90% of cases of acute osteomyelitis. Other organisms include *Haemophilus influenzae* and *Salmonella*; infection with the latter may occur as a complication of sickle cell disease. The classic presentation is with fever and localized bone pain with overlying tenderness and erythema.

Diagnosis

- Imaging:
  - Plain X-ray is not sensitive in early infection but demonstrates chronic damage.
  - MRI is highly sensitive in early disease.
  - Nuclear medicine bone scans are also helpful.
- ***Blood cultures*** may be positive.
- ***Bone biopsy and culture*** identifies the organism and sensitivities.

Management

Treatment of osteomyelitis is with intravenous antibiotics. Switching to oral antibiotics can be considered after 2 weeks with continuation for a further 4 weeks. Choice of antibiotic is determined by the causative agent and local guidelines. Surgical drainage and removal of dead bone (sequestrum) may be necessary but recurrence is common.

Delayed treatment leads to chronic osteomyelitis. In chronic osteomyelitis, sinus formation is common. Subacute osteomyelitis is associated with a chronic abscess within the bone (Brodie's abscess).

Tuberculous osteomyelitis

This is usually due to haematogenous spread from a reactivated primary focus in the lungs or gastrointestinal tract. The disease starts in intra-articular bone. The spine is commonly involved ***(Pott's disease)***, with damage to the bodies of two neighbouring vertebrae leading to vertebral collapse and acute angulation of the spine (gibbus). Later, an abscess forms ('cold abscess'). Pus can track along tissue planes and discharge at a point far from the affected vertebrae. Symptoms consist of local pain and later swelling if pus has collected. Systemic symptoms of malaise, fever, and night sweats occur.

Management is as for pulmonary tuberculosis but extended to 9 months (see Chapter 20), together with initial immobilization.

NEOPLASTIC DISEASE OF BONE

Bone pain may be due to multiple myeloma, lymphoma, a primary tumour of bone, or secondary deposits. The pain is typically unremitting and worse at night, and there are other clinical clues such as weight loss or ill-health.

Malignant tumours of bone are shown in **Box 33.12**. The most common tumours are metastases from the bronchus, breast and prostate. Metastases from kidney and thyroid are less common. Primary bone tumours are rare and most commonly seen in children and young adults.

Symptoms are usually related to the anatomical position of the tumour, with local bone pain. Systemic symptoms (e.g., malaise and

Box 33.12 Malignant neoplasms of bone

- Metastases (osteolytic):
  - Lung
  - Breast
  - Prostate (often osteosclerotic as well)
  - Thyroid
  - Kidney
- Multiple myeloma
- Primary bone tumours (rare; seen in the young), e.g.:
  - Osteosarcomas
  - Fibrosarcomas
  - Chondromas
  - Ewing's tumour

pyrexia) and aches and pains occur and may be related to hypercalcaemia. The diagnosis of metastases may be evident from the history and examination, particularly in the setting of a known primary tumour. Symptoms from bony metastases may also be the first presenting feature of a malignancy, however.

Investigations

- ***Skeletal isotope scans*** show bony metastases as 'hot' areas.
- ***X-rays*** may show metastases as osteolytic areas with bony destruction. Osteosclerotic metastases are characteristic of prostatic carcinoma.
- ***MRI*** is used particularly for vertebral lesions.
- ***CT and computed tomography-positron emission tomography (CT-PET)*** are useful.
- ***Serum alkaline phosphatase*** may be raised.
- ***Hypercalcaemia*** occurs in 10% to 20% of patients who have metastatic malignancies, or can be due to ectopic parathormone or PTH-related protein secretion.
- ***Prostate-specific antigen*** (PSA) and serum acid phosphatase are raised in the presence of prostatic metastases.

Management

Treatment is usually with analgesics and anti-inflammatory drugs. Local radiotherapy to bone metastases relieves pain and reduces the risk of pathological fracture. Some tumours respond to chemotherapy; others are hormone-dependent and respond to hormonal therapy. Bisphosphonates (see 'Bisphosphonates' section) can help symptomatically. Occasionally, pathological fractures require internal fixation.

SCHEUERMANN'S DISEASE

This disease predominantly occurs in adolescent males. The main feature is a progressive dorsal kyphosis of the thoracic spine. Pain may or may not be present. The cause is unknown. A genetic predisposition, exacerbated by excessive exercise prior to epiphyseal fusion, is one suggested explanation. Older patients with kyphosis may be referred with suspected osteoporotic fractures but found to have long-standing kyphosis due to Scheuermann's. Management is focused on postural exercises and avoidance of precipitants. Surgery may be undertaken to correct kyphosis in severe cases.

Bibliography

Rosen CJ, ed. *Primer on the Metabolic Bone Diseases and Disorders of Mineral Metabolism*, 8th edn. Hoboken, New Jersey: Wiley-Blackwell, American Society for Bone and Mineral Research; 2013.

Significant websites

https://www.bonehealthandosteoporosis.org/. *Bone Health and Osteoporosis Foundation—useful clinicians' guide.*
http://www.iscd.org. *International Society for Clinical Densitometry—guidelines on DXA scanning.*
https://theros.org.uk/. *Royal Osteoporosis Society—useful information and reviews of ongoing research.*

Questions for this chapter can be found on Elsevier eBooks+.

34 Endocrinology

Helena Gleeson, Narendra L. Reddy, and Miles J. Levy

Clinical skills 1138
History and examination 1138
Dynamic endocrine tests 1138
Background 1140
Hormonal activity 1140
Aetiology and pathology 1143
Endocrine investigations 1144
The pituitary gland and hypothalamus 1145
Presentation of pituitary and hypothalamic disease 1146
Hypothalamo–pituitary–adrenal axis 1151
The adrenal cortex 1152
The adrenal medulla 1161
The thyroid axis 1163
Anatomy and physiology 1163
Hypothyroidism 1165
Hyperthyroidism 1167
Graves' orbitopathy (ophthalmopathy) 1171
Goitre (thyroid enlargement) 1173
Thyroid cancer 1174
Hypothalamo–pituitary–gonadal axis 1175
Anatomy and physiology 1175
Disorders in the male 1178
Disorders in the female 1180
Disorders of sex development 1185
The thirst axis 1190
Disorders of calcium metabolism 1193
Hypoglycaemia in the non-diabetic patient 1196
Other endocrine disorders 1198
Diseases of many glands 1198
Ectopic hormone secretion 1199

CORE SKILLS AND KNOWLEDGE

Many endocrine conditions are common and seen outside the hospital setting. They often affect young people and can usually be cured or completely controlled with appropriate therapy. For example, the UK prevalence of hypothyroidism is 2%, and it is usually managed in primary care. Excluding obesity (Chapter 33) and diabetes mellitus (Chapter 35), the commonest endocrine disorders are thyroid disorders; menstrual disorders; disorders affecting reproduction; hyperparathyroidism; and increasingly incidental findings in endocrine organs found on scans being done for other reasons—termed 'incidentalomas'.

Most endocrinologists specialize in both diabetes and endocrinology, seeing patients in outpatient clinics and often looking after general medical inpatients. Joint clinics that include specialist endocrine surgeons, endocrine oncologists, obstetricians, gynaecologists, and paediatricians combine the expertise patients require in one clinical environment. Specialist nurses oversee dynamic testing of endocrine function and coordinate education and management of patients with endocrine conditions.

Key skills required in endocrinology include:

- Mastering key principles of hormone action (such as negative feedback) in order to distinguish primary from secondary gland failure
- Understanding the concept of dynamic endocrine tests and interpreting their results
- Becoming familiar with the hormones most commonly used in endocrine replacement including levothyroxine, hydrocortisone, testosterone, and oestrogen/progestogen.

The best way to learn about endocrinology is to attend specialist and general outpatient clinics, where patients presenting with endocrine syndromes are evaluated and complex dynamic tests are planned and interpreted. Attend specialist multidisciplinary team meetings (MDTs) to learn about the radiology, histology, and different management approaches to endocrine disease.

CLINICAL SKILLS

History and examination

Specific features of the history relevant to endocrine disease are described in **Box 34.1**. On examination, weight, height, body mass index (BMI), blood pressure, and general habitus should be documented, together with the presence or absence of specific signs of deficiency or excess of individual hormone axes indicated in the diagram opposite. Emergency presentations are listed in **Box 34.2**. In addition:

- In people with suspected pituitary disease, visual fields and adjacent cranial nerves should be assessed.
- In thyroid disease, the presence of goitre or orbitopathy should be documented.
- In hypogonadism both sexes require assessment of other secondary sexual characteristics, and testicular volumes in males should be documented as well as presence or absence of gynaecomastia.

Dynamic endocrine tests

Because of intra-individual and diurnal variations in the levels of many clinically relevant hormones, a single blood test often does not provide the information necessary to make a diagnosis. Dynamic endocrine tests assess the response of an endocrine axis to a specific challenge and are performed according to detailed protocols, which must be followed precisely. There are two types of dynamic tests: stimulation and suppression tests (**Box 34.3**).

Box 34.1 Taking a history in endocrine assessment

General symptoms of a variety of endocrine disorders

- Tiredness, weakness, lack of energy, or drive

Other common 'hormonal' symptoms

- Changes in body size and shape
- Changes in appetite or thirst
- Problems with libido and sexual function
- Changes in menstrual cycle
- Changes in the skin (dryness, greasiness, acne, bruising, thinning or thickening)
- Abnormal hair distribution (loss or excess)
- Subfertility or pregnancy loss

Role of hormones in physical development

- Growth and development in childhood
- Age of menarche
- Age of puberty
- Problems of sexual development

Risk factors common to a range of endocrine disorders

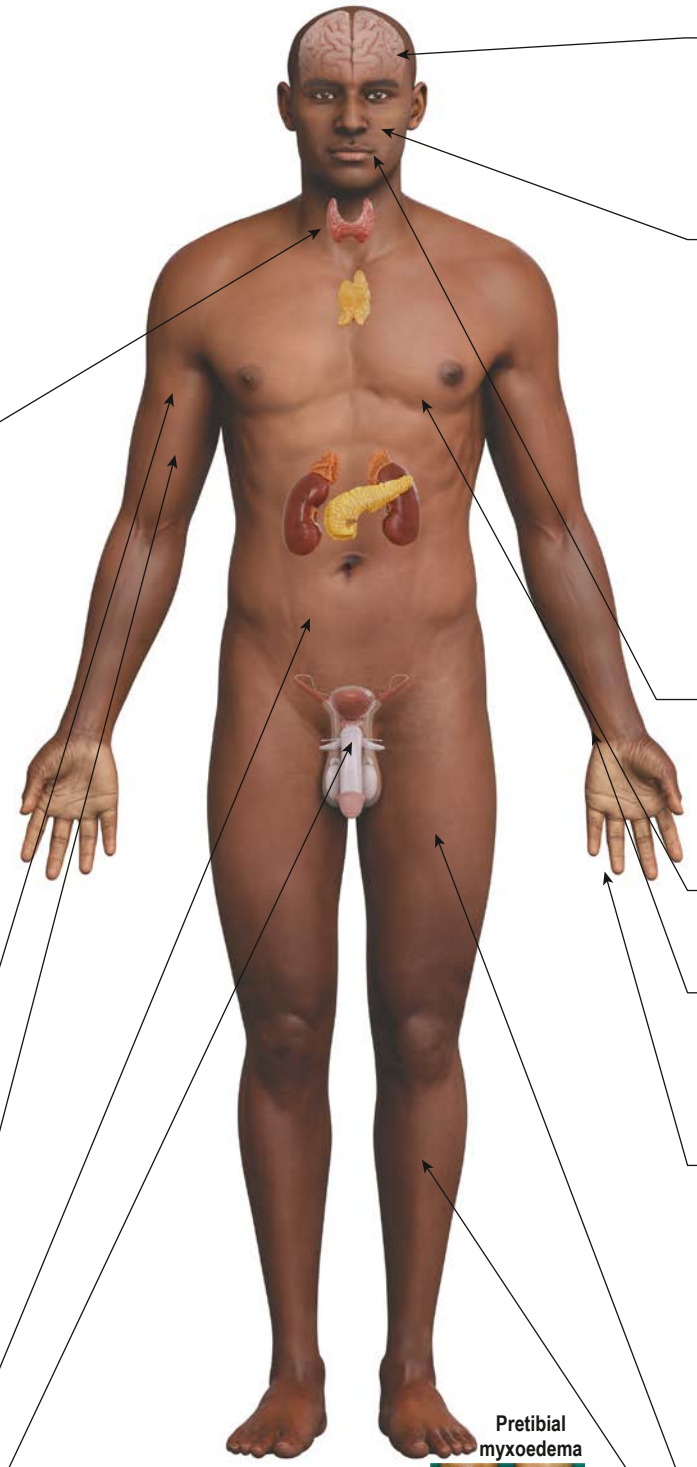
- Previous surgery or radiation involving endocrine glands
- Current or previous medication with endocrine side-effects
- A previous or family history of autoimmune disease
- A previous or family history of atypical tumours
- A family history of obesity, early-onset cardiovascular disease, or any symptoms similar to those the patient is experiencing

Box 34.2 The 'top 10' endocrine emergencies

Diagnosis	Presentation	Location in chapter/book
Severe hyponatraemia	Non-specific neurological symptoms and signs including confusion, drowsiness, coma, and seizures if severe	Box 27.13
Severe hypercalcaemia	Thirst, dehydration, polyuria	Box 34.54
Hypoadrenal crisis	Collapse, postural hypotension, hypoglycaemia, history of chronic steroid use	Box 34.18
Rapid dehydration in diabetes insipidus	Severe dehydration with ongoing polyuria; thirst; drowsiness, coma, or seizures; history of diabetes insipidus or chronic lithium use	'Diabetes insipidus/History and examination' section
Pituitary apoplexy	Sudden severe headache (may be mistaken for subarachnoid haemorrhage) in a patient with known or undiagnosed pituitary tumour	'Pituitary apoplexy' section
Thyrotoxicosis	Fever, tachycardia, anxiety, tremor, dysrhythmia, e.g., fast atrial fibrillation	'Thyroid crisis or 'thyroid storm" section
Severe hypothyroidism	Bradycardia, hypotension, drowsiness, or coma	'Myxoedema coma' section
Hypoglycaemia	Sweating, confusion, drowsiness, or coma. Regular user of insulin or sulphonylurea	Chapter 11, 'Hypoglycaemia' section
Diabetic ketoacidosis	Vomiting, fever, dehydration, acidosis with Kussmaul respiration, hyperglycaemia. History of type 1 diabetes	Chapter 11, 'Diabetic ketoacidosis' section
Hyperglycaemic hyperosmolar state	Drowsiness or coma. Severe dehydration and hyperglycaemia. Signs of intercurrent illness. History of type 2 diabetes	Chapter 11, 'Hyperosmolar hyperglycaemic state' section

Box 34.3 Dynamic endocrine tests

	Stimulation test	Suppression test
Assess for	**Hormone deficiency**	**Hormone excess**
Abnormal result	Inadequate rise in hormone level after stimulation, suggesting damage to target gland limiting maximum secretory output	Inadequate suppression of hormone level, indicating failure of negative feedback control and likely autonomous hormone secretion
Example	Short ACTH (synacthen) test for adrenal insufficiency	Dexamethasone suppression test for Cushing's disease



General
Mood
- depressed/lethargic
- restlessness
- memory
Voice – hoarse
Weight – loss or gain
Sweating
Hirsutism

Neck
Goitre
Thyroid nodules

Goitre

Vitiligo

Skin
Vitiligo (autoimmune disease)
Pigmentation
Hair distribution

Blood pressure
↑ Cushing's
Conn's
Phaeochromocytoma
Hypertension
Acromegaly
Thyroid disease
↓ Adrenal insufficiency

Central
Obesity
Striae (Cushing's)

Genitalia
Testicular volume
Pubertal development
Virilization

Hair
Alopecia
Coarse hair
Frontal balding

Eyes
- Signs of Graves' orbitopathy - proptosis, erythema, swelling
- Diplopia
- Visual fields defects
- Ophthalmoscopy
  - Papilloedema
  - Pale optic discs

Exophthalmos

Facial features
Hypothyroid
- coarse features
- 'peaches and cream' complexion
- puffy eyes
- dry skin
Hyperthyroid
- signs of Grave's orbitopathy

Breasts
Gynaecomastia
Galactorrhoea

Pulse
Atrial fibrillation

Hands
Sweating, tremor, palmar erythema
Carpal tunnel
Large, spade-like hands (acromegaly)
Pigmentation of creases (Addison's)

Acromegaly

Pretibial myxoedema

Proximal myopathy

Pretibial myxoedema

(From Rojanametin K, Masaru T. J Am Coll Dermatol 2015; 73:e195-e196, with permission)

BACKGROUND

The endocrine system consists of glands that exert their actions at distant parts of the body via the production of biologically active hormones secreted into the bloodstream. Unlike the neurological system, which produces an immediate response, the endocrine system typically has a slower and longer-lasting effect on the body. The main endocrine glands are the pituitary, thyroid, adrenals, gonads, parathyroids, and pancreas, and the common endocrine problems seen in clinical practice are shown in Fig. 34.1. The pituitary gland, a pea-sized structure situated at the base of the brain, plays a key role in the control and feedback mechanisms of the endocrine system and has been termed the 'conductor of the endocrine orchestra'.

Hormonal activity

Synthesis, storage, and release of hormones

Hormones have several chemical structures: polypeptide, glycoprotein, steroid, or amine. Hormone release is the end-product of a cascade of intracellular events. In the case of polypeptide hormones neural or endocrine stimulation of the cell causes increased transcription from DNA to a specific messenger RNA (mRNA), which is, in turn, translated to the peptide product. This is often a precursor molecule that may itself be biologically inactive. This 'prohormone' is further processed before being packaged into granules in the Golgi apparatus. These granules are transported to the plasma membrane before release, which is itself regulated by a complex combination of intracellular regulators. Hormone release may be in a brief spurt caused by the sudden stimulation of granules, often induced by an intracellular Ca^{2+}-dependent process, or 'constitutive' (immediate and continuous secretion).

Plasma transport

Most classical hormones are secreted into the systemic circulation. In contrast, hypothalamic releasing hormones are released into the pituitary portal system so that higher concentrations of the releasing hormones reach the pituitary than occur in the systemic circulation.

Many hormones are bound to proteins within the circulation. In most cases only the free (unbound) hormone is available and thus biologically active. This binding serves to buffer against rapid changes in plasma levels of the hormone, and some binding protein interactions are also involved in the active regulation of hormone action. Many tests of endocrine function measure total rather than free hormone, which can give rise to difficulties in interpretation when binding proteins are altered in disease states or by drugs.

Binding proteins comprise both specific, high-affinity proteins of limited capacity, such as thyroxine-binding globulin (TBG), cortisol-binding globulin (CBG), sex-hormone-binding globulin (SHBG), and insulin-like growth factor (IGF)–binding proteins (e.g., IGF-BP3), and other less specific, low-affinity ones, such as prealbumin and albumin.

Hormone action and receptors

Hormones act by binding to specific receptors in the target cell. Most hormone receptors are proteins with complex tertiary structures. The structure of the hormone-binding domain of the receptor complements the tertiary structure of the hormone, while changes in other parts of the receptor in response to hormone binding are responsible for the effects of the activated receptor within the cell. The structure of common hormones and their receptors is described under individual hormone axes.

Hormone receptors are broadly divided into:

- ***cell surface or membrane receptors:*** transmembrane receptors that contain hydrophobic sections spanning the lipid-rich plasma membrane and trigger internal cellular messengers

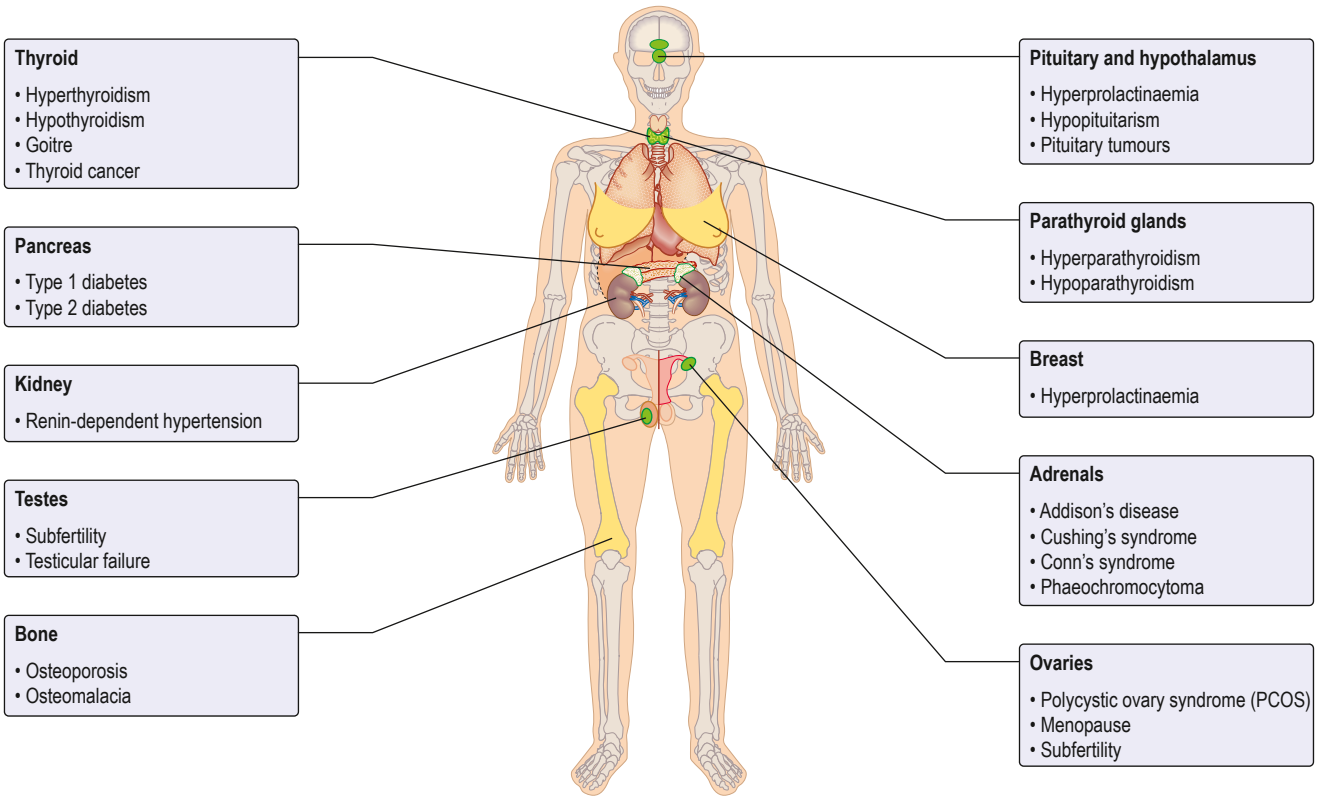


Fig. 34.1 The major endocrine organs and common endocrine problems.

- ***nuclear receptors:*** they bind hormones and translocate them to the nucleus, where they bind hormone response elements of nuclear DNA via characteristic amino acid sequences (e.g., so-called 'zinc fingers').

Abnormal receptors are an occasional cause of endocrine disease.

Mechanisms of hormone–receptor action

Common structural mechanisms of hormone–receptor action include:

- ***G-protein coupled receptors (7-transmembrane or serpentine receptors).*** These bind hormones on their extracellular domain and activate the membrane G-protein complex with their intracellular domain. The activated complex may then:
  - Stimulate cyclic AMP (cAMP) generation by adenylate cyclase—activating further intracellular kinases and leading to phosphorylation
  - Activate phospholipase C (PLC), leading to generation of inositol 1,4,5-triphosphate (IP_3) and release of intracellular calcium—in turn, leading to calmodulin-dependent kinase activity and phosphorylation
  - Lead to diacylglycerol (DAG) activation of C-kinase and subsequent protein phosphorylation
  - Most peptide hormones act via G-protein coupled receptors.
- ***Dimeric transmembrane receptors.*** These receptors, from several receptor superfamilies, bind hormone in their extracellular components (sometimes causing dimerization of the receptor monomer) and directly phosphorylate intracellular messengers via intracellular components, leading to a variety of intracellular activation cascades. Growth hormone (GH), prolactin, and IGF-1 act via this type of receptor.
- ***Lipid-soluble molecules.*** These pass through the cell membrane and bind with their ***nuclear receptors*** in the cell cytoplasm before translocation of the activated hormone–receptor complex to the nucleus; here, it binds to nuclear DNA, often in combination with a multi-component complex of promoters, inhibitors, and transcription factors. This interaction leads to increased transcription of the relevant gene product. Steroid and thyroid hormones act via this type of receptor.

Effects of hormone binding on target tissues

The activation of intracellular kinases, phosphorylation, release of intracellular calcium and other 'second messenger' pathways, and the direct stimulation of DNA transcription may result in the following:

- stimulation/release of pre-formed secondary hormones from storage granules
- stimulation/synthesis of secondary hormones and other cellular components
- opening/closing of ion or water channels in the cell membrane (e.g., calcium channels or aquaporin water channels)
- activation/deactivation of other DNA binding proteins, leading to stimulation or inhibition of DNA transcription.

In each case, binding of the hormone to its receptor is the first step in a cascade of intracellular events, which eventually lead to the overall effects of that hormone on cellular function.

The sensitivity and/or number of receptors for a hormone are often decreased after prolonged exposure to a high hormone concentration, the receptors thus becoming less sensitive ('downregulation', e.g., angiotensin II receptors, β-adrenoceptors). The reverse is true when stimulation is absent or minimal, the receptors showing increased numbers or sensitivity ('upregulation').

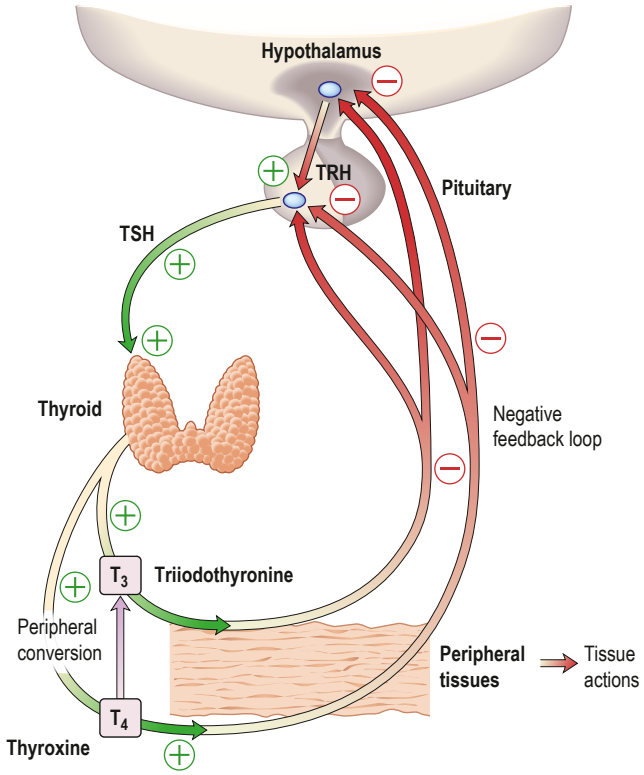


Fig. 34.2 The hypothalamic–pituitary–thyroid axis. Pituitary thyroid-stimulating hormone (TSH) is secreted in response to hypothalamic thyrotrophin-releasing hormone (TRH) and stimulates secretion of T_4 and T_3 from the thyroid. T_4 and T_3 have actions in peripheral tissues and exert negative feedback on the pituitary and hypothalamus.

Control and feedback

Most hormone systems are under tight regulatory control (typically by the hypothalamic–pituitary [HP] axis) by a system known as negative feedback. An example of the negative feedback system in the hypothalamic–pituitary–thyroid axis is demonstrated in Fig. 34.2 and described here:

- Thyrotrophin-releasing hormone (TRH) is secreted in the hypothalamus and travels via the portal system to the pituitary, where it stimulates the thyrotrophs to produce thyroid-stimulating hormone (TSH).
- TSH is secreted into the systemic circulation, where it stimulates increased iodine uptake by the thyroid, and synthesis and release of thyroxine (T_4) and triiodothyronine (T_3).
- Serum levels of T_3 and T_4 are increased by TSH; in addition, the conversion of T_4 to T_3 (the more active hormone) in peripheral tissues is stimulated by TSH.
- T_3 and T_4 enter cells, where they bind to nuclear receptors and promote increased metabolic and cellular activity.
- Levels of T_3, from the blood and from local conversion of T_4, are sensed by receptors in the pituitary and the hypothalamus. If they rise above normal, TRH and TSH production is suppressed, leading to reduced T_3 and T_4 secretion.
- Peripheral T_3 and T_4 levels fall to normal.
- If T_3 and T_4 levels are low—for example, after thyroidectomy, increased amounts of TRH and TSH are secreted, stimulating the

remaining thyroid to produce more T_3 and T_4; blood levels of T_3 and T_4 may be restored to normal, at the expense of increased TSH drive, reflected by a high TSH level: 'compensated euthyroidism'.

- Conversely, in thyrotoxicosis, when factors other than TSH itself are maintaining high T_3 and T_4 levels, the same mechanisms lead to suppression of TSH secretion.

Primary and secondary gland failure

It is useful in clinical endocrinology to distinguish between 'primary' disease of the end-organ gland (e.g., due to autoimmune destruction, atrophic change, infiltration or surgical removal of the gland) and 'secondary' disorders of the same axis, caused by disease of the pituitary gland. An understanding of the negative feedback system is key to interpreting endocrine blood results and diagnosing the site of the disease process in clinical practice. In general terms:

- ***'Primary' gland failure*** due to a disease process in the endocrine end-organ (thyroid, adrenal or gonad) will lead to deficiency of the hormone produced by that gland and a loss of negative feedback and subsequent elevation in the corresponding anterior pituitary hormone.
- In ***'secondary gland failure'***, there are low or 'inappropriately normal' levels of the pituitary trophic hormone in the face of a low end-organ hormone level. For example, if a patient has low circulating free T_3 (fT_3) and T_4 levels in the context of a low TSH, pituitary disease should be suspected.

Hormone excess

Abnormal hormone excess due to a disease process in the primary endocrine gland, or excess amount of exogenous hormone, will lead to increased negative feedback and suppression of the corresponding pituitary hormones. For example, in autoimmune hyperthyroidism, free T_3 and T_4 levels are elevated in the context of a suppressed TSH. Equally, the presence of a non-suppressed plasma pituitary hormone in the context of excess primary hormone implies that the pituitary, rather than the primary gland itself, is the cause. For example, in pituitary-driven Cushing's disease, cortisol levels are elevated but adrenocorticotrophin (ACTH) is not suppressed.

Hormone resistance

In certain situations, receptor abnormalities can give rise to abnormal negative feedback due to hormone resistance, which can lead to an unusual pattern of blood results. For example, thyroid hormone resistance, due to mutations in the thyroid hormone receptor, is characterized by an elevation in thyroid hormones with a non-suppressed TSH. With this pattern of thyroid results, the clinician should also consider the rare diagnosis of a TSH-secreting pituitary tumour, as well as the possibility of assay issues.

Measurement of hormones

Hormones are routinely measured by biochemical assays in the laboratory. It is possible to measure pituitary trophic hormones and the hormones produced by the end-organ glands, but hypothalamic hormones are not measured in practice because of their low concentration and local action within the HP axis. Circulating levels of most hormones are very low (10^{-9}–10^{-12} mol/L) and cannot be measured by simple chemical techniques. Hormones are therefore usually measured by immunoassays, which rely on highly specific polyclonal or monoclonal antibodies, which bind to the hormone being measured. This hormone–antibody interaction is measured by use of labelled hormone after separation of bound and free fractions (Fig. 34.3).

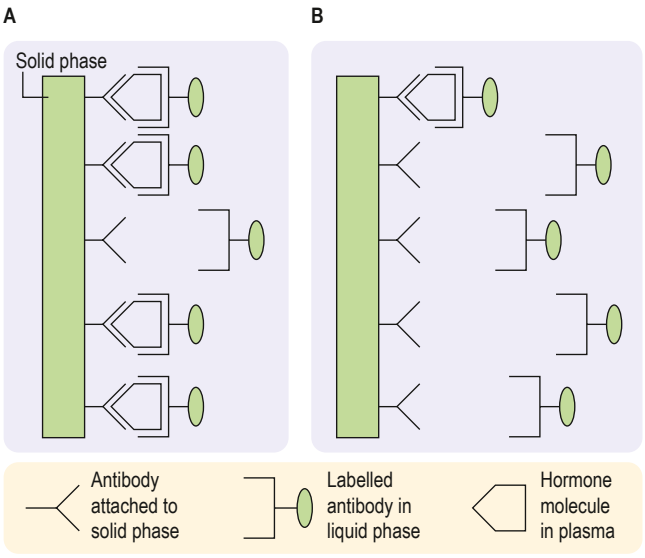


Fig. 34.3 Principles of measurement of hormone levels in plasma by immunoassay. (Precise details vary with different assays and manufacturers.) Immunoassays use two antibodies specific to the hormone being measured: one typically attached to a solid phase and one labelled antibody in the liquid phase. **(A)** High hormone levels in plasma: a large amount of hormone binds to antibody in solid phase; a large amount of labelled antibody is linked to solid phase via molecules of the hormone. **(B)** Low hormone levels in plasma: less hormone, and therefore less labelled antibody, is linked to the solid phase. Label (radioactive, chemiluminescent, enzymatic or fluorescent) can be measured in either solid or liquid phase after separation of phases; levels of label will be proportional to the amount of hormone in the sample.

Immunoassays are sensitive but have limitations. In particular, the immunological activity of a hormone, as used in developing the antibody, may not necessarily correspond to biological activity and there may be false-positive and negative results. The patient's blood may also contain heterophile antibodies, which interact with the animal antibodies used in the assay and produce falsely low or high values. When there is a discrepancy between endocrine blood results and the clinical presentation, the clinician must question the validity of an endocrine result and liaise with chemical pathology. It may be necessary for the sample to be measured in a different laboratory using an alternative antibody, or to measure hormones in ways other than immunoassay. Examples of alternative techniques include equilibrium dialysis, high-pressure liquid chromatography (HPLC), and mass spectroscopy.

Hormone-binding proteins

Many hormones are transported in the bloodstream by a specific binding protein (see 'Hormonal activity' section). It is more helpful to measure the free hormone rather than total bound hormone level, as this is the part that is biologically active. Some modern assays attempt to measure the free hormone level directly (e.g., free T_4) and are a more accurate reflection of biological activity, although many assays still measure total hormone level.

Cortisol, which is bound to CBG, and testosterone, which is bound to SHBG, are still usually measured in their total form and can be affected by alterations in binding protein levels. In women who are pregnant or on the combined oral contraceptive pill, high oestrogen levels may lead to an elevation in CBG, which can overestimate cortisol and give the false impression of hypercortisolaemia. In people with diabetes mellitus or other insulin-resistant states

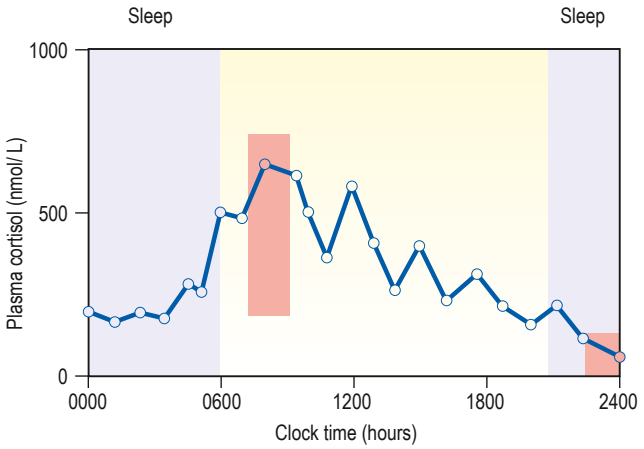


Fig. 34.4 Plasma cortisol levels during a 24-hour period. Note both the pulsatility and the shifting baseline. Normal ranges for 09:00 hours (180–700 nmol/L) and 24:00 hours (<100 nmol/L; must be taken when asleep) are shown in the *red* boxes. *Grey shading* shows sleep.

that may lower SHBG levels, low total testosterone levels may give the false impression of androgen deficiency. Conversely, hyperthyroidism or oestrogen excess can cause an elevation in SHBG, leading to apparently high total testosterone levels.

Patterns of hormonal secretion

Hormone secretion can be continuous or intermittent, for example:

- ***Continuous secretion*** is shown by the thyroid hormones, with a half-life of 7–10 days for T_4 and 6–10 h for T_3, and with little variation in levels over the day, month, and year.
- ***Pulsatile secretion*** is the normal pattern for the gonadotrophins, luteinizing hormone (LH), and follicle-stimulating hormone (FSH), with major pulses released every 1–2 h, depending on the phase of the menstrual cycle. GH is also secreted in a pulsatile fashion, with undetectable levels in between pulses. A single measurement is therefore not helpful to diagnose GH deficiency or excess.

Biological rhythms

Circadian means changes over the 24 hours of the day–night cycle and is best shown for the pituitary–adrenal axis. Fig. 34.4 shows plasma cortisol measured over 24 hours; levels are highest in the early morning and lowest overnight. Additionally, cortisol release is pulsatile, following the pulsatility of pituitary ACTH. Thus 'normal' cortisol levels vary during the day and great variations can be seen in samples taken only 30 minutes apart.

The menstrual cycle is an example of a longer and more complex (28-day) biological rhythm (see 'Hypothalamo–pituitary–gonadal axis' section).

Other regulatory factors

- ***Stress.*** Physiological 'stress' and acute illness produce rapid increases in ACTH and cortisol, GH, prolactin, adrenaline (epinephrine), and noradrenaline (norepinephrine). These can occur within seconds or minutes.
- ***Sleep.*** Secretion of GH and prolactin is increased during sleep, especially the rapid eye movement (REM) phase.
- ***Feeding and fasting.*** Many hormones regulate the body's control of energy intake and expenditure and are therefore influenced by feeding and fasting. Secretion of insulin increases, while testosterone and GH decrease after ingestion of food, and secretion of a number of hormones is altered during prolonged food deprivation.

Box 34.4 Types of autoimmune disease affecting endocrine organs

Organ and frequency, if known[a]	Antibody	Antigen, if known	Clinical syndrome
Stimulating			
Thyroid (1 in 100)	Thyroid-stimulating immunoglobulin (TSI, TSAb)	Thyroid-stimulating hormone (TSH) receptor	Graves' disease, neonatal thyrotoxicosis
Destructive			
Thyroid (1 in 100)	Thyroid microsomal	Thyroid peroxidase enzyme (TPO)	Primary hypothyroidism (myxoedema)
	Thyroglobulin	Thyroglobulin	
	Immunoglobulin (Ig) G4		Riedel's thyroiditis
Adrenal (1 in 20 000)	Adrenal cortex	21-Hydroxylase enzyme	Primary hypoadrenalism (Addison's disease)
Pancreas (1 in 500)	Islet cell	Glutamic acid decarboxylase (GAD; see Fig. 35.8)	Type 1 diabetes
Stomach	Gastric parietal cell	Gastric parietal cell	Pernicious anaemia
	Intrinsic factor	Intrinsic factor	
Skin	Melanocyte		Vitiligo
Ovary (1 in 500)	Ovary		Primary ovarian failure
Testis	Testis		Primary testicular failure
Parathyroid	Parathyroid chief cell		Primary hypoparathyroidism
Pituitary	Pituitary-specific cells		Autoimmune hypophysitis
			Selective hypopituitarism (e.g., growth hormone [GH] deficiency, diabetes insipidus)

[a]Frequencies are approximate and refer to the population in Northern Europe.
Note: Other related diseases include myasthenia gravis and autoimmune liver diseases.

Aetiology and pathology

Aetiological mechanisms common to many endocrine disorders are described as follows.

Autoimmune disease

Organ-specific autoimmune diseases can affect every major endocrine organ (Box 34.4). They are characterized by the presence of specific antibodies in the serum, often present years before clinical symptoms are evident, usually more common in women, and have a strong genetic component, often with an identical-twin concordance rate of 50% and with human leucocyte antigen (HLA) associations (see individual diseases).

Endocrine tumours

Most endocrine tumours are benign, although a cytological or histological diagnosis may be needed if there is suspicion of malignancy. Clinical presentation depends on whether the tumour is functional or non-functional, the latter presenting only as a mass clinically or on imaging. Palpable thyroid nodules are common, and mass effects are a frequent presentation of pituitary adenomas, but the increased use of high-resolution ultrasound and detailed cross-sectional imaging has revealed a very high prevalence of asymptomatic, incidentally discovered thyroid, adrenal and pituitary lesions, commonly termed 'incidentalomas'.

Functional tumours cause their effects via excess secretion of the relevant hormone. While often considered to be 'autonomous'—that is, independent of the physiological control mechanisms—many functional tumours do show evidence of feedback occurring at a higher 'set-point' than normal (e.g., adrenocorticotrophic hormone [ACTH] secretion from a corticotroph adenoma). This is relevant in the dynamic assessment of endocrine diseases, such as in the differential diagnosis of Cushing's syndrome.

Endocrine adenomas typically present in a single gland, although rarer multiple endocrine neoplasia (MEN) syndromes exist that are due to very specific mutations of a single gene, such as the mutations of the *RET* proto-oncogene in MEN 2 or the *MEN1* gene mutation in MEN 1 (see 'Diseases of many glands' section). The diagnosis of some endocrine tumours, particularly at a young age, or the identification of a family history should prompt genetic screening.

Enzyme defects

The biosynthesis of most hormones involves many stages. Deficient or abnormal enzymes can lead to absent or reduced production of the secreted hormone. In general, severe deficiencies present early in life with obvious signs; partial deficiencies usually present later with mild signs or are only evident under stress. An example of an enzyme deficiency is congenital adrenal hyperplasia (CAH), whose molecular basis has also been identified as mutations or deletions of the gene encoding the relevant enzymes (see 'Hypothalamo–pituitary–adrenal axis' section).

Receptor abnormalities

There are rare conditions in which hormone secretion and control are normal but the receptors are defective; thus, if androgen receptors are defective, normal levels of androgen will not produce masculinization (e.g., androgen insensitivity syndrome). There are also a number of rare syndromes of diabetes and insulin resistance from receptor abnormalities (see **Box 23.4**); other examples include AVP-resistance (also referred to as 'nephrogenic diabetes insipidus'), pseudohypoparathyroidism, and thyroid hormone resistance.

Endocrine investigations

Endocrine function is assessed by measurement of hormone levels in blood (or, more precisely, in plasma or serum) and sometimes in other body fluids on samples obtained basally and in response to stimulation and suppression tests.

Basal blood levels

Assays for all clinically relevant pituitary and end-organ hormones are available.

The time, day, and conditions of measurement make great differences to hormone levels, and the method and timing of samples therefore depend upon the characteristics of the endocrine system involved. There are also gender, pubertal, and age differences.

Basal levels are especially useful for systems with long half-lives (e.g., T_4 and T_3, IGF-1, androstenedione, SHBG). These vary little over the short term and random samples are therefore satisfactory.

Basal samples for many hormones need to be interpreted with respect to normal ranges for the time of day/month, diet, or posture concerned. Hormones with a marked circadian rhythm (e.g., testosterone and cortisol) must be measured at an appropriate time of day. Testosterone should be measured before 11 a.m. in the fasting state; cortisol should be checked between 8 a.m. and 10 a.m. to exclude hypoadrenalism, but at midnight to demonstrate normal low levels (to exclude Cushing's). LH/FSH, oestrogen, and progesterone vary with time of menstrual cycle, and renin/aldosterone may vary with sodium intake, posture, and age. For these hormones, all relevant details must be recorded or the results may prove uninterpretable.

Stress-related hormones

Measurement of stress-related hormones may be problematic either because the patient is stressed by hospital attendance or venepuncture, leading to falsely high levels (e.g., catecholamines, prolactin; sampling via an indwelling needle may be required, some time after initial venepuncture), or because low levels in a non-stressed individual are unable to confirm an adequate reserve required for normal physiological stress (cortisol and GH).

Urine collections

Collections over 24 hours have the advantage of providing an 'integrated mean' of a day's secretion but, in practice, are often incomplete or wrongly timed. They also vary with sex and body size or age. Written instructions should be provided for the patient to ensure accurate collection. Examples of hormones measured in this way are catecholamines and urinary free cortisol levels.

Saliva

Saliva is sometimes used for steroid estimations, especially in children or for samples taken at home. Midnight salivary cortisol levels are increasingly used for the diagnosis of Cushing's syndrome due to the practical difficulties in obtaining a midnight blood sample.

Stimulation and suppression tests

These tests are used when basal levels give equivocal information. In general, stimulation tests are used to confirm suspected deficiency of hormone secretion, and suppression tests to confirm suspected excess (see **Box 34.3**). These tests are valuable in many instances. For example, where the secretory capacity of a gland is damaged, maximal stimulation by the trophic hormone will give a diminished output. Thus, in the short ACTH stimulation test for adrenal reserve (**Box 34.5**; **Fig. 34.5A**), the healthy subject shows a normal response, while the subject with primary hypoadrenalism (Addison's disease) demonstrates an impaired cortisol response to tetracosactide (an ACTH analogue).

A patient with a hormone-producing tumour usually fails to show normal negative feedback. A patient with Cushing's disease (excess pituitary ACTH) will thus fail to suppress ACTH and cortisol production when given a dose of synthetic steroid, in contrast to normal subjects. **Fig. 34.5B** shows the response of a normal subject given dexamethasone 1 mg at midnight; cortisol is suppressed the following morning. The subject with Cushing's disease shows inadequate suppression.

The detailed protocol for each test must be followed exactly, since differences in technique will produce variations in results.

Box 34.5 Short ACTH (tetracosactide) stimulation test

Indications
- Diagnosis of Addison's disease
- Screening test for ACTH deficiency

Procedure
- I.V. cannula for sampling
- Any time of day, but best at 09:00 h; non-fasting
- Tetracosactide 250 μg, i.v. or i.m. at time 0
- Measure serum cortisol at time 0 and time +30 min

Normal response
- 30 min cortisol >600 nmol/L[a]
  - (400–600 nmol/L borderline and may indicate deficiency)

[a]Precise cortisol normal ranges are variable between laboratories and assays; appropriate local reference ranges must be used.

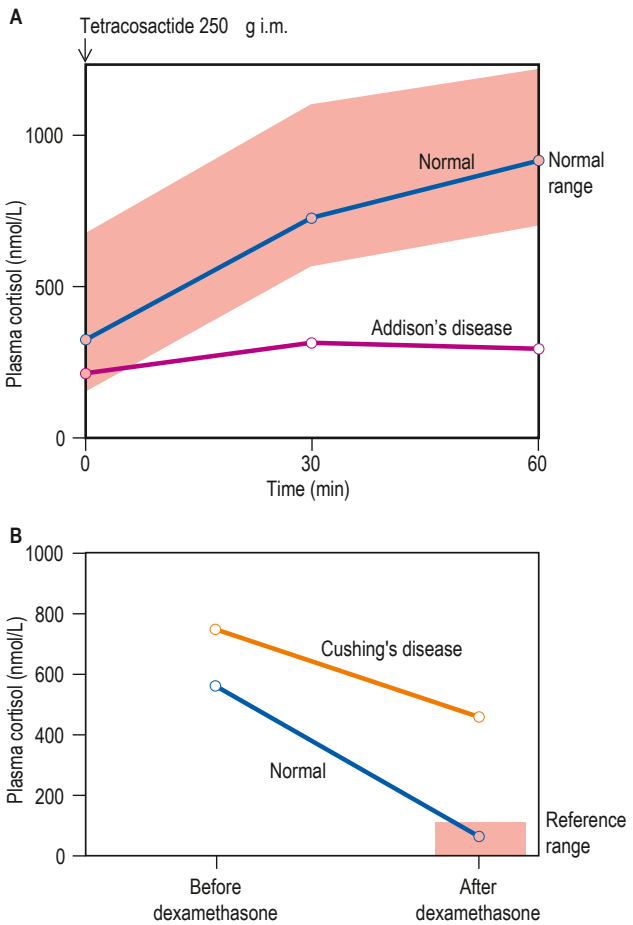


Fig. 34.5 Short ACTH stimulation and dexamethasone suppression tests. **(A)** The short ACTH stimulation test shows a normal response in a healthy subject and a decreased response in a patient with Addison's disease. **(B)** Dexamethasone suppression tests in a normal subject and in a patient with Cushing's disease, showing inadequate suppression.

Radiological investigations

Magnetic resonance imaging (MRI), computed tomography (CT) and ultrasound give detailed images of endocrine glands but should be interpreted in clinical and biochemical context. Incidental endocrine tumours are common ('incidentalomas') and may have no clinical significance. Conversely, functional endocrine tumours may be too small to visualize on routine imaging. There is an increasing role for functional imaging, for example, positron emission tomography (PET), to detect endocrine tumours. Selective venous sampling is a useful interventional technique to localize hormone excess and is discussed in the relevant section.

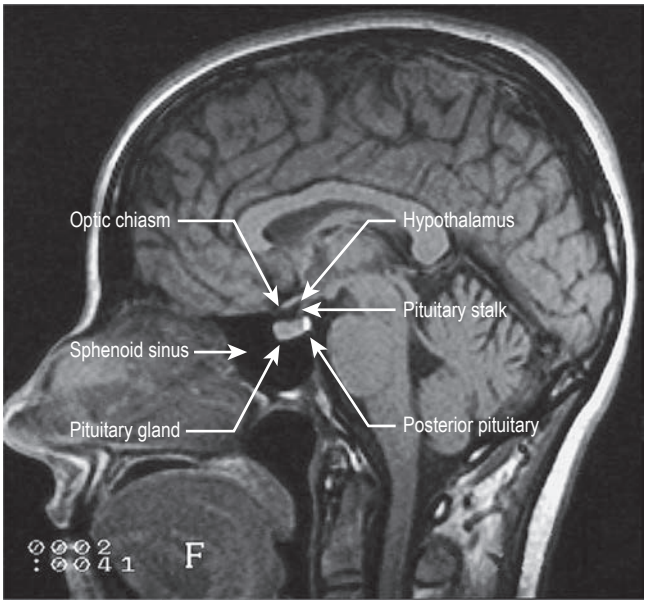


Fig. 34.6 Magnetic resonance imaging (MRI) of a sagittal section of the brain, showing the pituitary fossa and adjacent structures. *(By kind permission of Dr Martin Jeffree.)*

THE PITUITARY GLAND AND HYPOTHALAMUS

Anatomy

Most peripheral hormone systems are controlled by the hypothalamus and pituitary. The hypothalamus is sited at the base of the brain around the third ventricle and above the pituitary stalk, which leads down to the pituitary itself, carrying the hypophyseal–pituitary portal blood supply.

The anatomical relations of the hypothalamus and pituitary (Fig. 34.6) include the optic chiasm above the pituitary fossa; any expanding lesion from the pituitary or hypothalamus can thus produce visual field defects by pressure on the chiasm. Such upward expansion of the gland through the diaphragma sellae is termed 'suprasellar extension'. Lateral extension of pituitary lesions may involve the vascular and nervous structures in the cavernous sinus and may rarely reach the temporal lobe of the brain. The pituitary is itself encased in a bony box; therefore any lateral, anterior, or posterior expansion must cause bony erosion.

Embryologically, the anterior pituitary is formed from an upgrowth of Rathke's pouch (ectoderm); this meets an outpouching of the third ventricular floor, which becomes the posterior pituitary. This combination of primitive gut and neural tissue provides an essential link between the rapidly responsive central nervous system and the longer-acting endocrine system. Several transcription factors—LHX3, HESX1, PROP1, POU1F1—are responsible for the differentiation and development of the pituitary cells. Mutation of these produces pituitary disease.

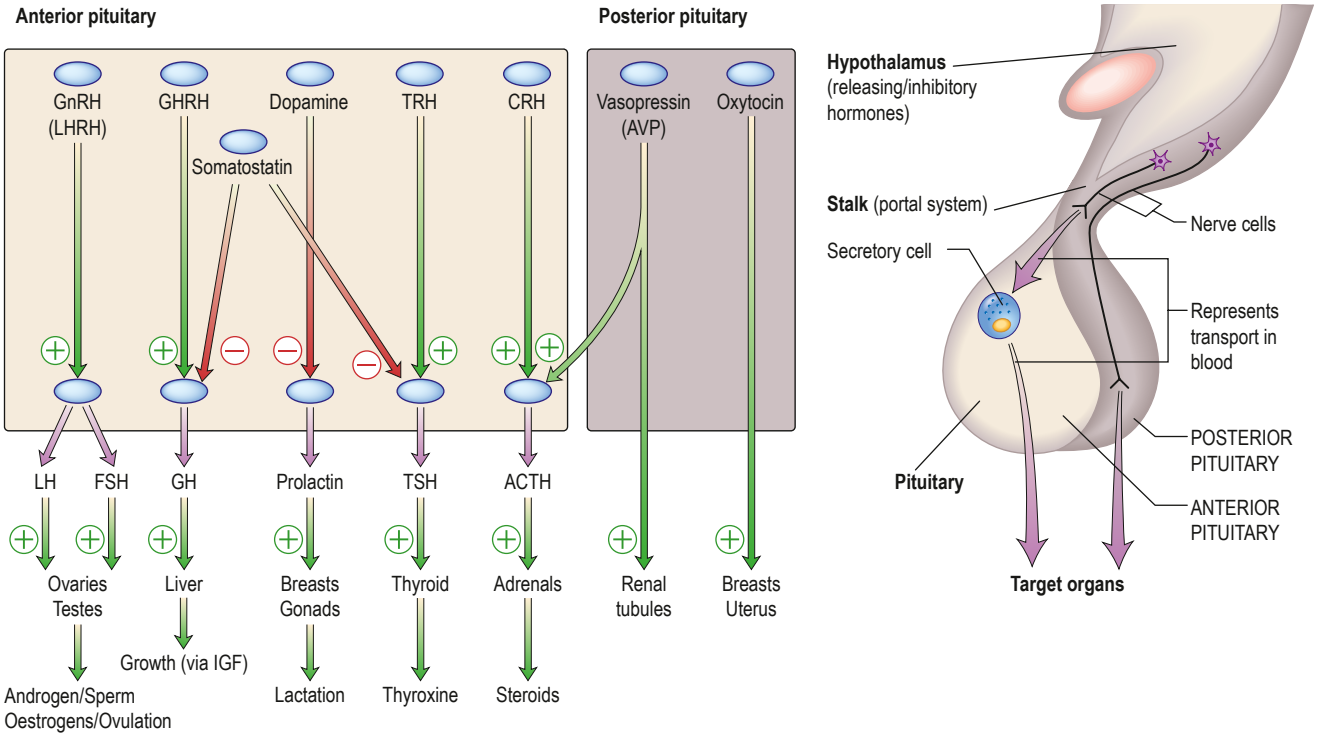


Fig. 34.7 Hypothalamic releasing hormones and the pituitary trophic hormones. *AVP,* Arginine vasopressin; *CRH,* corticotrophin-releasing hormone; *FSH,* follicle-stimulating hormone; *GH,* growth hormone; *GHRH,* growth hormone-releasing hormone; *GnRH,* gonadotrophin-releasing hormone; *IGF,* insulin-like growth factor; *LH,* luteinizing hormone; *TRH,* thyrotrophin-releasing hormone.

Physiology

Hypothalamus

This contains many vital centres for such functions as appetite, thirst, thermal regulation, and sleeping/waking. It acts as an integrator of many neural and endocrine inputs to control the release of pituitary hormone-releasing factors. It plays a role in circadian rhythm, menstrual cyclicity, and responses to stress, exercise, and mood.

Hypothalamic neurones secrete pituitary hormone releasing and inhibiting factors and hormones into the portal system; these run down the stalk to the pituitary. As well as the classical hormones (Fig. 34.7), the hypothalamus also contains large amounts of other neuropeptides and neurotransmitters such as neuropeptide Y, vasoactive intestinal peptide (VIP), and nitric oxide, which can also alter pituitary hormone secretion.

Synthetic hypothalamic hormones and their antagonists are available for the testing of many aspects of endocrine function and for treatment.

Anterior pituitary

The majority of anterior pituitary hormones are under predominantly positive control by the hypothalamic releasing hormones; the exception is prolactin, which is under tonic inhibition by dopamine. Pathological conditions interrupt the flow of hormones between the hypothalamus and pituitary gland and therefore cause deficiency of most hormones but oversecretion of prolactin. There are five major anterior pituitary axes: the gonadotrophin axis, the growth axis, prolactin, the thyroid axis, and the adrenal axis.

Posterior pituitary

The posterior pituitary is neuroanatomically connected to specific hypothalamic nuclei and acts merely as a storage organ. Arginine vasopressin (AVP, also called vasopressin) and oxytocin, both nonapeptides, are synthesized in the supraoptic and paraventricular nuclei in the anterior hypothalamus. They are then transported along the axon and stored in the posterior pituitary (see Fig. 34.7). This means that damage to the stalk or pituitary alone does not prevent synthesis and release of AVP and oxytocin. AVP is discussed in 'The thirst axis' section; oxytocin produces milk ejection and uterine myometrial contraction.

Presentation of pituitary and hypothalamic disease

Diseases of the pituitary can cause under- or overactivity of each hypothalamic–pituitary–end-organ axis. The clinical features of the syndromes associated with such altered pituitary function, such as Cushing's syndrome, can be the presenting symptom of pituitary disease or of end-organ disease and are discussed later. First, however, we look at the clinical features of pituitary disease that are common to all hormonal axes.

Pituitary space-occupying lesions and tumours

Pituitary tumours (Box 34.6) are the most common cause of pituitary disease, and the great majority are benign pituitary adenomas, usually monoclonal in origin. Problems are caused by:

- local effects of a tumour
- excess hormone secretion
- inadequate production of hormone by the remaining normal pituitary, i.e., hypopituitarism.

Approach for a possible or proven mass

Is there a tumour?

If there is, how big is it and what ***local anatomical effects*** is it exerting? Pituitary and hypothalamic space-occupying lesions, hormonally active or not, can cause symptoms by pressure on, or infiltration of:

- the visual pathways, with field defects and visual loss (most common)
- the cavernous sinus, with III, IV, and VI cranial nerve lesions
- bony structures and the meninges surrounding the fossa, causing headache
- hypothalamic centres: altered appetite, obesity, thirst, somnolence/wakefulness, or precocious puberty
- the ventricles, causing interruption of cerebrospinal fluid (CSF) flow and hydrocephalus
- the sphenoid sinus, causing CSF rhinorrhoea.

Box 34.6 Characteristics of common pituitary and related tumours

Tumour or condition	Usual size	Most common clinical presentation
Prolactinoma	Most <10 mm (microprolactinoma)	Galactorrhoea, amenorrhoea, hypogonadism, erectile dysfunction
	Some >10 mm (macroprolactinoma)	As above plus headaches, visual field defects, and hypopituitarism
Acromegaly	Few mm to several cm	Change in appearance, visual field defects, and hypopituitarism
Cushing's disease	Most small: few mm (some cases are hyperplasia)	Central obesity, cushingoid appearance (local symptoms rare)
Nelson's syndrome	Often large: >10 mm	Post-adrenalectomy, pigmentation, sometimes local symptoms
Non-functioning tumours	Usually large: >10 mm	Visual field defects; hypopituitarism (microadenomas may be incidental finding)
Craniopharyngioma	Often very large and cystic (skull X-ray abnormal in >50%; calcification common)	Headaches, visual field defects, growth failure (50% occur below age 20; about 15% arise from within sella)

Investigations

- ***MRI of the pituitary.*** MRI is superior to CT (Fig. 34.8) and readily shows any significant pituitary mass. Small lesions within the pituitary fossa, 'microadenomas', are very common on MRI (10% of normal individuals in some studies). Such small lesions are sometimes detected during MRI of the head for other reasons—so-called 'pituitary incidentalomas'.
- ***Visual fields.*** These should be plotted formally by automated computer perimetry or Goldmann perimetry, but clinical assessment by confrontation using a small red pin as the target is also sensitive and valuable. Common defects are upper temporal quadrantanopia and bitemporal hemianopia (see Fig. 28.6).

Is there a hormonal excess?

There are three major conditions that are usually caused by excess secretion from pituitary adenomas and which will show positive immunostaining for the relevant hormone:

- prolactin excess ***(prolactinoma or hyperprolactinaemia):*** histologically, prolactinomas are 'chromophobe' adenomas (a description of their appearance on classical histological staining)
- GH excess ***(acromegaly or gigantism):*** somatotroph adenomas, usually 'acidophil', and sometimes due to specific G-protein mutations (see 'Presentation of pituitary and hypothalamic disease' section)
- excess ACTH secretion ***(Cushing's disease and Nelson's syndrome):*** corticotroph adenomas, usually 'basophil'.

Many tumours are able to synthesize several pituitary hormones, and occasionally more than one hormone is secreted in clinically significant excess (e.g., both GH and prolactin).

The clinical features of acromegaly, Cushing's disease, or hyperprolactinaemia are usually (but not always) obvious (see 'Presentation of pituitary and hypothalamic disease' section). Tumours producing LH, FSH, or TSH are well described but very rare.

Some common pituitary tumours cause no clinically apparent hormone excess and are referred to as 'non-functioning' tumours.

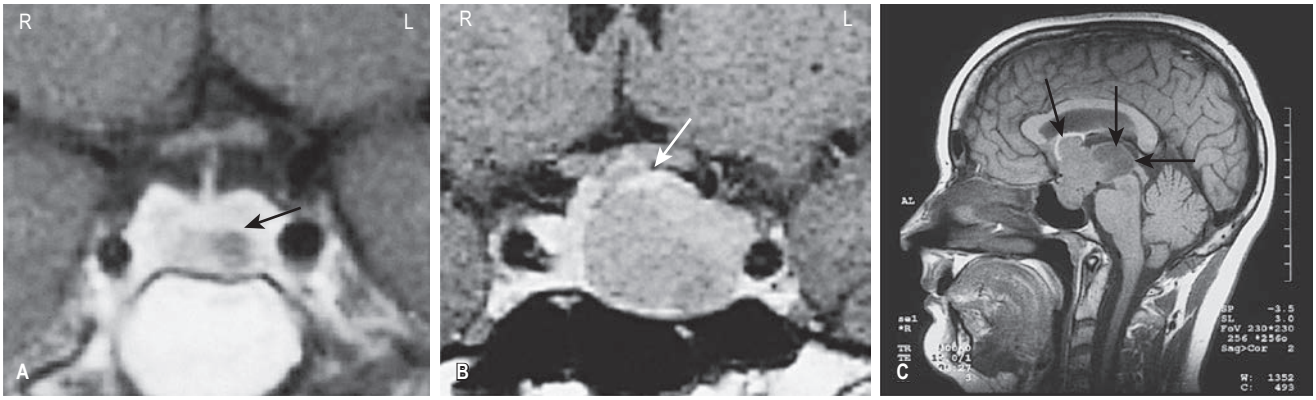


Fig. 34.8 Investigation of the pituitary gland. **(A)** Left-sided lucent intrasellar microadenoma *(arrowed)* (coronal MRI). The pituitary stalk is deviated slightly to the right. **(B)** Macroadenoma with moderate suprasellar extension and lateral extension compressing the left cavernous sinus (coronal MRI). The top of the adenoma is compressing the optic chiasm *(arrowed)*. **(C)** Pituitary macroadenoma with massive suprasellar extension *(arrowed)* (sagittal MRI).

Box 34.7 Tests of hypothalamic–pituitary (HP) function

All hormone levels are measured in plasma unless otherwise stated.
Tests **shown in bold** are those normally measured on a single basal 09:00 h sample in the initial assessment of pituitary function.

Axis	Basal investigations		Common dynamic tests	Other tests
	Pituitary hormone	End-organ product/function		
Anterior pituitary				
HP–ovarian	**LH** **FSH**	**Oestradiol** Progesterone (day 21 of cycle)		Ovarian ultrasound LHRH test[a]
HP–testicular	**LH** **FSH**	**Testosterone**		Sperm count LHRH test[a]
Growth hormone	GH	**IGF-1** IGF-BP3	Insulin tolerance test Glucagon test Macimorelin test	Other tests available but lack sufficient evidence base
Prolactin	**Prolactin**		–	–
HP–thyroid	**TSH**	**Free T_4, T_3**		TRH test[a]
HP–adrenal	ACTH	**Cortisol**	Insulin tolerance test Short ACTH (tetracosactide) stimulation test	Glucagon test CRH test[a] Metyrapone test
Posterior pituitary				
Thirst and osmoregulation		Plasma/urine osmolality	Water deprivation test	Hypertonic saline infusion

[a]Releasing hormone tests were a traditional part of pituitary function testing but have been largely replaced by the advent of more reliable assays for basal hormones. They test only the 'readily releasable pool' of pituitary hormones, and normal responses may be seen in hypopituitarism. Bold text indicates commonly performed investigations.
ACTH, Adrenocorticotropic hormone; *CRH,* corticotrophin-releasing hormone; *FSH,* follicle-stimulating hormone; *GH,* growth hormone; *GHRH,* growth hormone-releasing hormone; *IGF,* insulin-like growth factor; *LH,* luteinizing hormone; *LHRH,* luteinizing hormone-releasing hormone; *TRH,* thyroid-releasing hormone; *TSH,* thyroid-stimulating hormone.

Laboratory studies such as immunocytochemistry or *in situ* hybridization show that these tumours may often produce small amounts of LH and FSH or the α-subunit of LH, FSH, and TSH, and occasionally ACTH.

Is there a deficiency of any hormone?

Clinical examination may give clues; thus, short stature in a child with a hypothalamic or pituitary mass is likely to be due to GH deficiency. An adult complaining of lethargy with pale skin is likely to be deficient in TSH and/or ACTH. Milder deficiencies may not be obvious and require specific testing (**Box 34.7**).

Management

Management depends on the type and size of tumour (**Box 34.8**). Decisions about pituitary tumour management are made in an MDT setting, which typically comprises an endocrinologist, pituitary surgeon, and radiologist. In general, therapy has three aims: removal/control of the tumour, reduction of excess hormone secretion, and replacement of hormone deficiencies.

Removal/control of tumour

This is only required if the tumour is large enough to cause, or is likely to cause, anatomical effects, or if it is secreting excess hormones. Small tumours producing no significant symptoms, pressure, or endocrine effects may be observed with appropriate clinical, visual field, imaging, and endocrine assessments.

- ***Surgery*** via the trans-sphenoidal route is usually the treatment of choice. Very large tumours are occasionally removed via the open transcranial (usually transfrontal) route.
- ***Radiotherapy***—by conventional linear accelerator or newer stereotactic techniques—is usually employed when surgery is impracticable or incomplete, as it controls but rarely abolishes tumour mass. The conventional regimen involves a dose of 45 Gy, given as 20–25 fractions via three fields. Stereotactic techniques use either a linear accelerator or multiple cobalt sources ('gamma-knife'; see Chapter 41).

Box 34.8 Comparisons of primary treatments for pituitary tumours

Treatment method	Advantages	Disadvantages
Surgical		
Trans-sphenoidal adenomectomy or hypophysectomy	Relatively minor procedure Potentially curative for microadenomas and smaller macroadenomas	Some extrasellar extensions may not be accessible Risk of CSF leakage and meningitis
Transcranial (usually transfrontal) route	Good access to suprasellar region	Major procedure; danger of frontal lobe damage High chance of subsequent hypopituitarism
Radiotherapy		
External (40–50 Gy)	Non-invasive Reduces recurrence rate after surgery	Slow action, often over many years Not always effective Possible late risk of tumour induction
Stereotactic	Precise administration of high dose to lesion	Long-term follow-up data limited
Medical		
Dopamine agonist therapy (e.g., bromocriptine, cabergoline)	Non-invasive; reversible	Usually not curative; significant side-effects in minority Concerns about fibrotic reactions
Somatostatin analogue therapy (octreotide, lanreotide)	Non-invasive; reversible	Usually not curative; causes gallstones; expensive
Growth hormone receptor antagonist (pegvisomant)	Highly selective	Usually not curative; very expensive

Box 34.9 Replacement therapy for hypopituitarism

Axis	Usual replacement therapies
Adrenal	Hydrocortisone 15–25 mg daily (starting dose 10 mg on rising/5 mg lunch-time/5 mg evening) (Normally no need for mineralocorticoid replacement)
Thyroid	Levothyroxine 100–150 μg daily
Gonadal	
Male	Testosterone intramuscularly, orally, transdermally, or implant
Female	Cyclical oestrogen/progestogen orally or as patch
Fertility	HCG plus FSH (purified or recombinant) or pulsatile GnRH to produce testicular development, spermatogenesis or ovulation
Growth	Recombinant human GH used routinely to achieve normal growth in children
	Also advocated for replacement therapy in adults, where GH has effects on muscle mass and wellbeing
Thirst	Desmopressin 10–20 μg one to three times daily by nasal spray or orally 100–200 μg three times daily Carbamazepine, thiazides, and chlorpropamide are very occasionally used in mild diabetes insipidus
Breast (prolactin inhibition)	Dopamine agonist (e.g., cabergoline 500 μg weekly)

FSH, Follicle-stimulating hormone; *GH*, growth hormone; *HCG*, human chorionic gonadotrophin.

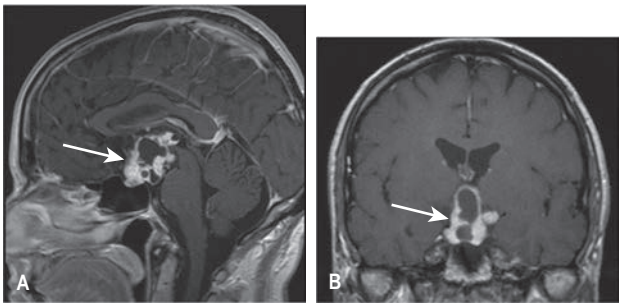


Fig. 34.9 Craniopharyngioma *(arrowed)*: a partially cystic pituitary and suprasellar mass. **(A)** Sagittal MRI. **(B)** Coronal MRI.

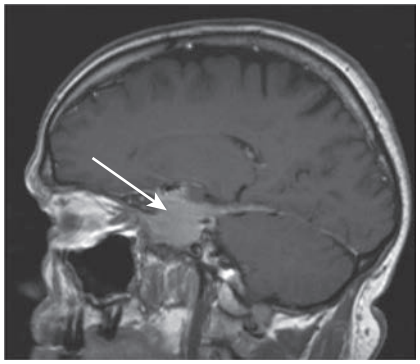


Fig. 34.10 Meningioma *(arrowed)* involving the pituitary fossa. Note the typical 'dural tail'.

- ***Medical therapy*** with somatostatin analogues and/or dopamine agonists sometimes causes shrinkage of specific tumour types and, if successful, can be used as primary therapy, particularly in the case of prolactinomas.

Reduction of excess hormone secretion

Reduction is usually obtained by surgical removal but sometimes by medical treatment. Useful control can be achieved with dopamine agonists for prolactinomas or somatostatin analogues for acromegaly, but ACTH secretion usually cannot be controlled by medical means. GH antagonists are also available for acromegaly (see 'Presentation of pituitary and hypothalamic disease' section).

Replacement of hormone deficiencies

See **Box 34.9**. Replacement of hormone deficiencies, i.e., hypopituitarism, is discussed further in the 'Presentation of pituitary and hypothalamic disease' section.

Differential diagnosis of pituitary or hypothalamic masses

Although pituitary adenomas are the most common mass lesion of the pituitary (90%), a variety of other conditions may also present as a pituitary or hypothalamic mass and form part of the differential diagnosis.

Other tumours

- ***Craniopharyngioma*** (1%–2%), a usually cystic hypothalamic tumour that is often calcified and arises from Rathke's pouch. It is the most common pituitary tumour in children but may present at any age (**Fig. 34.9**).
- ***Uncommon tumours*** include meningiomas (**Fig. 34.10**), gliomas, chondromas, germinomas, and pinealomas. Primary pituitary carcinomas are very rare, but occasionally prolactin- and ACTH-secreting tumours present in an aggressive manner, which may require chemotherapy in addition to conventional treatment. Metastases occasionally present as apparent pituitary tumours, typically accompanied by headache and DI.

Hypophysitis and other inflammatory masses

A variety of inflammatory masses occur in the pituitary or hypothalamus. These include rare pituitary-specific conditions (e.g., autoimmune [lymphocytic] hypophysitis, giant cell hypophysitis, postpartum hypophysitis), or pituitary manifestations of more generalized disease processes (sarcoidosis, Langerhans' cell histiocytosis, granulomatosis with polyangiitis). These lesions may be associated with DI and/or an unusual pattern of hypopituitarism.

Other lesions

Carotid artery aneurysms may masquerade as pituitary tumours and must be diagnosed before surgery. Cystic lesions may also present as a pituitary mass, including arachnoid and Rathke cleft cysts.

Hypopituitarism

Pathophysiology

Deficiency of hypothalamic releasing hormones or of pituitary trophic hormones can be selective or multiple. Thus isolated deficiencies of GH, LH/FSH, ACTH, TSH, and vasopressin (AVP) are all seen, some cases of which are genetic and congenital, and others sporadic and autoimmune or idiopathic in nature.

Multiple deficiencies usually result from tumour growth or other destructive lesions. There is generally a progressive loss of anterior pituitary function. GH and gonadotrophins are usually first affected. Hyperprolactinaemia, rather than prolactin deficiency, occurs relatively early because of loss of tonic inhibitory control by dopamine. TSH and ACTH are usually last to be affected.

Box 34.10 Causes of hypopituitarism

Congenital
- Isolated deficiency of pituitary hormones (e.g., Kallmann's syndrome)
- *POU1F1 (Pit-1)*, *Prop1*, *HESX1* mutations

Infective
- Basal meningitis (e.g., tuberculosis)
- Encephalitis
- Syphilis

Vascular
- Pituitary apoplexy
- Sheehan's syndrome (postpartum necrosis)
- Carotid artery aneurysm

Immunological
- Autoimmune (lymphocytic) hypophysitis
- Pituitary antibodies

Neoplastic
- Pituitary or hypothalamic tumour
- Craniopharyngioma
- Meningioma
- Gliomas
- Pinealoma
- Secondary deposits, especially breast
- Lymphoma

Traumatic
- Skull fracture through base
- Surgery, especially transfrontal
- Perinatal trauma

Infiltrations
- Sarcoidosis
- Langerhans' cell histiocytosis
- Hereditary haemochromatosis
- Hypophysitis:
  - Postpartum
  - Giant cell

Others
- Radiation damage
- Fibrosis
- Chemotherapy
- Immunotherapy (ipilimumab, pembrolizumab, nivolumab)
- Empty sella syndrome

'Functional'
- Anorexia nervosa
- Starvation
- Emotional deprivation

Panhypopituitarism refers to deficiency of all anterior pituitary hormones; it is most commonly caused by pituitary tumours, surgery, or radiotherapy. Vasopressin (AVP) will only be significantly affected if the hypothalamus is involved by a hypothalamic tumour or major suprasellar extension of a pituitary lesion, or if there is an infiltrative/inflammatory process. Posterior pituitary deficiency with DI is rare in an uncomplicated pituitary adenoma.

Genetics

Specific genes are responsible for the development of the anterior pituitary, involving interaction between signalling molecules and transcription factors. For example, mutations in *PROP1* and *POU1F1* (previously *PIT-1*) prevent the differentiation of anterior pituitary cells (precursors to somatotroph, lactotroph, thyrotroph and gonadotroph cells), leading to deficiencies of GH, prolactin, TSH, and gonadotrophin-releasing hormone (GnRH). In addition, novel mutations within GH and growth hormone-releasing hormone (GHRH) receptor genes have been identified, which explain the pathogenesis of isolated GH deficiency in some children. Despite these advances, most cases of hypopituitarism do not have specific identifiable genetic causes.

Aetiology

Disorders that cause hypopituitarism are listed in **Box 34.10**. Pituitary and hypothalamic tumours, and surgical or radiotherapy treatment, are the most common.

Clinical features

Symptoms and signs depend upon the extent of hypothalamic and/or pituitary deficiencies, and mild deficiencies may not lead to any complaint. In general, symptoms of deficiency of a pituitary-stimulating hormone are the same as those of primary deficiency of the peripheral endocrine gland (e.g., TSH deficiency and primary hypothyroidism cause similar symptoms due to lack of thyroid hormone secretion).

- ***Secondary hypothyroidism, hypoadrenalism, hypogonadism***, and ***GH deficiency*** lead to tiredness and general malaise and reduced quality of life.
- ***Hypothyroidism*** causes weight gain, slowness of thought and action, dry skin, cold intolerance, constipation, and potentially bradycardia and hypothermia.
- ***Hypoadrenalism*** causes mild hypotension, hyponatraemia, and, ultimately, cardiovascular collapse during severe intercurrent stressful illness.
- ***Hypogonadism*** leads to loss of libido, loss of secondary sexual hair, amenorrhoea and erectile dysfunction, and, eventually, osteoporosis.
- ***Hyperprolactinaemia*** may cause galactorrhoea and hypogonadism, including amenorrhoea.
- ***GH deficiency*** causes growth failure in children and impaired wellbeing in some adults.
- ***Weight gain*** (due to hypothyroidism; see above), or ***weight loss*** in severe combined deficiency.
- Longstanding ***panhypopituitarism*** gives the classic picture of pallor with hairlessness ('alabaster skin').

Particular syndromes related to hypopituitarism are described below.

Kallmann's syndrome

This is isolated gonadotrophin (GnRH) deficiency (see 'Hypothalamo–pituitary–gonadal axis' section) caused by mutations in the *KAL1* gene, which is located on the short (p) arm of the X chromosome. Kallmann's classically causes anosmia because the *KAL1* gene provides instructions to make anosmin, which has a role both in development of the olfactory system and in migration of GnRH-secreting neurones.

Septo-optic dysplasia

This rare congenital syndrome (rarely associated with mutations in the *HESX1* gene) presents in childhood with two out of the clinical triad of midline forebrain abnormalities, optic nerve hypoplasia, and hypopituitarism.

Sheehan's syndrome

This is due to pituitary infarction following postpartum haemorrhage and is rare in high-income countries but not uncommon in countries where there are not established obstetric services.

Pituitary apoplexy

A pituitary tumour occasionally enlarges rapidly owing to infarction or haemorrhage. This may produce severe headache, double vision, and sudden severe visual loss, sometimes followed by acute life-threatening hypopituitarism. Often, pituitary apoplexy can be managed conservatively with replacement of hormones and close monitoring of vision, although if there is a rapid deterioration in visual acuity and fields, surgical decompression of the optic chiasm may be necessary (**Fig. 34.11**).

'Empty sella' syndrome

An 'empty sella' is sometimes reported on pituitary imaging. This is sometimes due to a defect in the diaphragma and extension of the subarachnoid space (cisternal herniation) or may follow spontaneous infarction or regression of a pituitary tumour. All or most of the sella turcica is devoid of apparent pituitary tissue but,

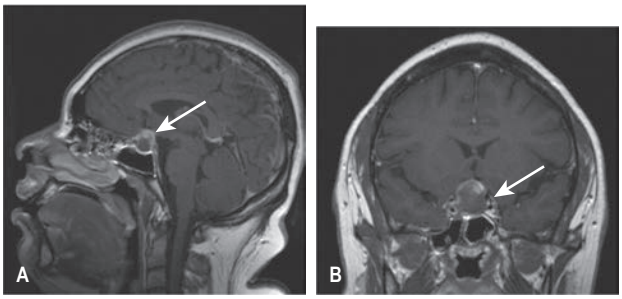


Fig. 34.11 Pituitary apoplexy. A bright area of haemorrhage at the top of a pituitary adenoma *(arrowed)* is shown. **(A)** Sagittal MRI. **(B)** Coronal MRI.

despite this, pituitary function is usually normal, with the pituitary being eccentrically placed and flattened against the floor or roof of the fossa.

Investigations

Each axis of the HP system requires separate investigation. However, the presence of normal gonadal function in women as evidenced by the menstrual cycle obviates the need for investigation.

Investigations range from measurement of simple basal levels (e.g., TSH and free T_4 for the thyroid axis) to stimulatory tests for the pituitary and tests of feedback for the hypothalamus (see **Box 34.7**). Assessment of the hypothalamic–pituitary–adrenal (HPA) axis is complex: depending on the assay basal 9 a.m. cortisol levels above 400 nmol/L usually indicate adequate reserve, while levels below 100 nmol/L predict an inadequate stress response. In many cases, basal levels are equivocal and a dynamic test is essential. The insulin tolerance test (**Box 34.11**) is widely regarded as the 'gold standard' but the short ACTH stimulation test (see **Box 34.5**), though an indirect measure, is used by many as a routine test of HPA status as long as it is performed 6 weeks after any pituitary insult. Occasionally, the difference between ACTH deficiency and normal HPA axis can be subtle, and interpretation requires experience.

Management

- Glucocorticoid and thyroid hormones are essential for life. Both are given as oral replacement drugs, hydrocortisone and levothyroxine, respectively, as in primary thyroid and adrenal failure, with the aim of restoring the patient to clinical and biochemical normality (see **Box 34.9**); for levothyroxine replacement, levels are monitored by routine hormone assays, whereas for hydrocortisone replacement monitoring is limited to symptoms. ***Note:*** Thyroid replacement therapy should not commence until normal adrenal axis function has been demonstrated or glucocorticoid replacement therapy initiated, as an adrenal 'crisis' may otherwise be precipitated.
- Sex hormone replacement is with testosterone in males and a combination of oestrogen and progestogen in females, both for symptomatic control and for prevention of long-term problems related to deficiency (e.g., osteoporosis).
- When fertility is desired, gonadal function is stimulated directly by human chorionic gonadotrophin (HCG, mainly acting as LH) or by purified or biosynthetic gonadotrophins, or indirectly by pulsatile GnRH (also known as luteinizing hormone-releasing hormone, LHRH) if gonadotrophs are undamaged; their use should be restricted to specialist units. HCG and biosynthetic gonadotrophins combination therapy is a recognized novel way of inducing male puberty.
- GH replacement therapy is given for growth in a child, under the care of a paediatric endocrinologist. If the adolescent remains deficient at the achievement of adult height, GH replacement therapy can be offered until patients reach their mid-20s to maximize muscle and bone mass. In adult GH deficiency, GH replacement therapy also produces improvements in body composition, work capacity, and psychological wellbeing, together with reversal of lipid abnormalities associated with high cardiovascular risk, and often results in significant symptomatic benefit. The National Institute for Health and Care Excellence (NICE) recommends GH replacement for children and adolescents with GH deficiency, and in adults with severe GH deficiency and significant quality-of-life impairment. It is expensive, and in the UK it costs £2500–6000 per annum.
- Glucocorticoid deficiency may mask impaired urine concentrating ability, with DI only becoming apparent after glucocorticoid replacement therapy because glucocorticoids are required for excretion of free water.

i Box 34.11 Insulin tolerance test

Indications

- Diagnosis or exclusion of ACTH and GH deficiency

Procedure

- Test explained to patient and consent obtained
- Should only be performed in experienced, specialist units
- Exclude cardiovascular disease (ECG), epilepsy, or unexplained blackouts; exclude severe untreated hypopituitarism (basal cortisol must be >100 nmol/L; normal free T_4)
- I.V. hydrocortisone and glucose available for emergency
- Overnight fast, begin at 08:00–09:00 h
- Soluble insulin, 0.15 U/kg, i.v. at time 0
- Glucose, cortisol and GH levels at 0, 30, 45, 60, 90, 120 min

Normal response

- Cortisol rises above 550 nmol/L[a]
- GH rises above 7 μg/L (severe deficiency ≤ 3 μg/L (<9 mU/L))
- Glucose must be <2.2 mmol/L to achieve adequate stress response

[a]Precise cortisol normal ranges are variable between laboratories and assays; appropriate local reference ranges must be used.

Further reading

Cerbone M, Visser J, Bulwer C et al. Management of children and young people with idiopathic pituitary stalk thickening, central diabetes insipidus, or both: a national clinical practice consensus guideline. *Lancet Child Adolesc Health* 2021; 5(9):662–676.
Fleseriu M, Biller BMK, Freda PU, et al. A Pituitary Society update to acromegaly management guidelines. *Pituitary* 2021; 24(1):1–13.
National Institute for Health and Clinical Excellence. *NICE Technology Appraisal 64: Human Growth Hormone (Somatotropin) in Adults with Growth Hormone Deficiency*. NICE 2003. http://www.nice.org.uk/guidance/ta64.
Rajasekaran S, Vanderpump M, Baldeweg S et al. UK guidelines for the management of pituitary apoplexy. *Clin Endocrinol* 2011; 74:9–20.

HYPOTHALAMO–PITUITARY–ADRENAL AXIS

Anatomy and physiology

The human adrenals (**Box 34.12**) together weigh 8–10 g and comprise an outer cortex and inner medulla. The cortex has three zones: the zona glomerulosa, which secretes aldosterone under the control of

Box 34.12 Hormones and receptors of the hypothalamic–pituitary–adrenal axis

Hormone	Source	Site of action	Hormone structure	Receptor	Post-receptor activity
Corticotrophin-releasing hormone (CRH)	Hypothalamus	Pituitary	Peptide—41 AA	Membrane—7TM CRF1	G-proteins cAMP
Adrenocorticotrophic hormone (ACTH)	Pituitary	Adrenal	Peptide—39 AA	Membrane—7TM ACTHR (MCR2)	G-proteins cAMP
Cortisol	Adrenal	All tissues	Steroid ring	Nuclear GRα	Transcription GRE

7TM, 7 transmembrane (G-protein coupled receptor); *AA*, amino acids; *GRE*, glucocorticoid response element.

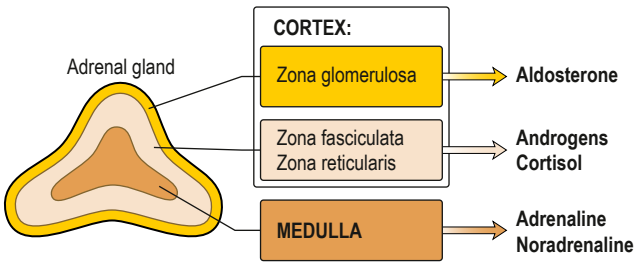


Fig. 34.12 **The adrenal gland and its hormones.**

Box 34.13 The major actions of glucocorticoids

Increased or stimulated:
- Gluconeogenesis
- Glycogen deposition
- Protein catabolism
- Fat deposition
- Sodium retention
- Potassium loss
- Free water clearance
- Uric acid production
- Circulating neutrophils

Decreased or inhibited:
- Protein synthesis
- Host response to infection
- Lymphocyte transformation
- Delayed hypersensitivity
- Circulating lymphocytes
- Circulating eosinophils

the renin–angiotensin system, and the zona fasciculata and zona reticularis, which produce glucocorticoids, cortisol, and sex steroids and androgens under feedback control of the HPA axis. The inner medulla synthesizes, stores, and secretes catecholamines (see later and Fig. 34.12).

The adrenal cortex

The steroids produced by the adrenal cortex are grouped into three classes, based on their predominant physiological effects: glucocorticoids, mineralocorticoids, and androgens.

Glucocorticoids

These are named after their effects on carbohydrate metabolism. Major actions are listed in Box 34.13. They act on intracellular corticosteroid receptors and combine with coactivating proteins to bind the 'glucocorticoid response element' (GRE) in specific regions of DNA to cause gene transcription. Glucocorticoid action is modified locally by the action of 11β-hydroxysteroid dehydrogenase (11βHSD). 11βHSD type 1 converts inactive cortisone into active cortisol, hence amplifying the hormone signal, while 11βHSD type 2 does the opposite.

The relative potency of common steroids is shown in Box 34.14.

Mineralocorticoids

The predominant effect of mineralocorticoids is on the extracellular balance of sodium and potassium in the distal tubule of the kidney. Aldosterone, produced solely in the zona glomerulosa, is the predominant mineralocorticoid in humans (about 50%); corticosterone makes a small contribution to overall mineralocorticoid activity. Mineralocorticoids act on type 1 corticosteroid receptors, while glucocorticoids act on type 2 receptors, both having a very similar structure. The mineralocorticoid activity of cortisol is weak but cortisol is present in considerable excess. The mineralocorticoid receptor in the kidney is protected from this excess by the intrarenal conversion ('shuttle') of cortisol to cortisone by 11βHSD type 2.

Box 34.14 The relative potency of equal amounts of common natural and synthetic steroids

Steroid	Glucocorticoid effect	Mineralocorticoid effect
Cortisol (hydrocortisone)[a]	1	1
Prednisolone	4	0.7
Dexamethasone	40	2
Aldosterone	0.1	400
Fludrocortisone	10	400

[a]Cortisol is arbitrarily defined as 1.

Androgens

Although secreted in considerable quantities, most androgens have relatively weak intrinsic androgenic activity until metabolized peripherally to testosterone or dihydrotestosterone. Dihydrotestosterone is metabolized from testosterone by 5α-reductase and is a potent androgen receptor agonist. The androgen receptor has been well characterized and mutations within this gene may cause androgen insensitivity syndromes.

Biochemistry

All steroids have the same basic skeleton (Fig. 34.13B) and the chemical differences between them are slight. The major biosynthetic pathways are shown in Fig. 34.13A.

Physiology

Glucocorticoid production by the adrenal is under HP control (Fig. 34.14).

Corticotrophin-releasing hormone (CRH) is secreted in the hypothalamus in response to circadian rhythm, stress, and other stimuli. CRH travels down the portal system to stimulate ***adrenocorticotrophin (ACTH)*** release from the anterior pituitary. Hypothalamic vasopressin (AVP) also stimulates ACTH secretion and acts synergistically. ACTH is derived from the prohormone

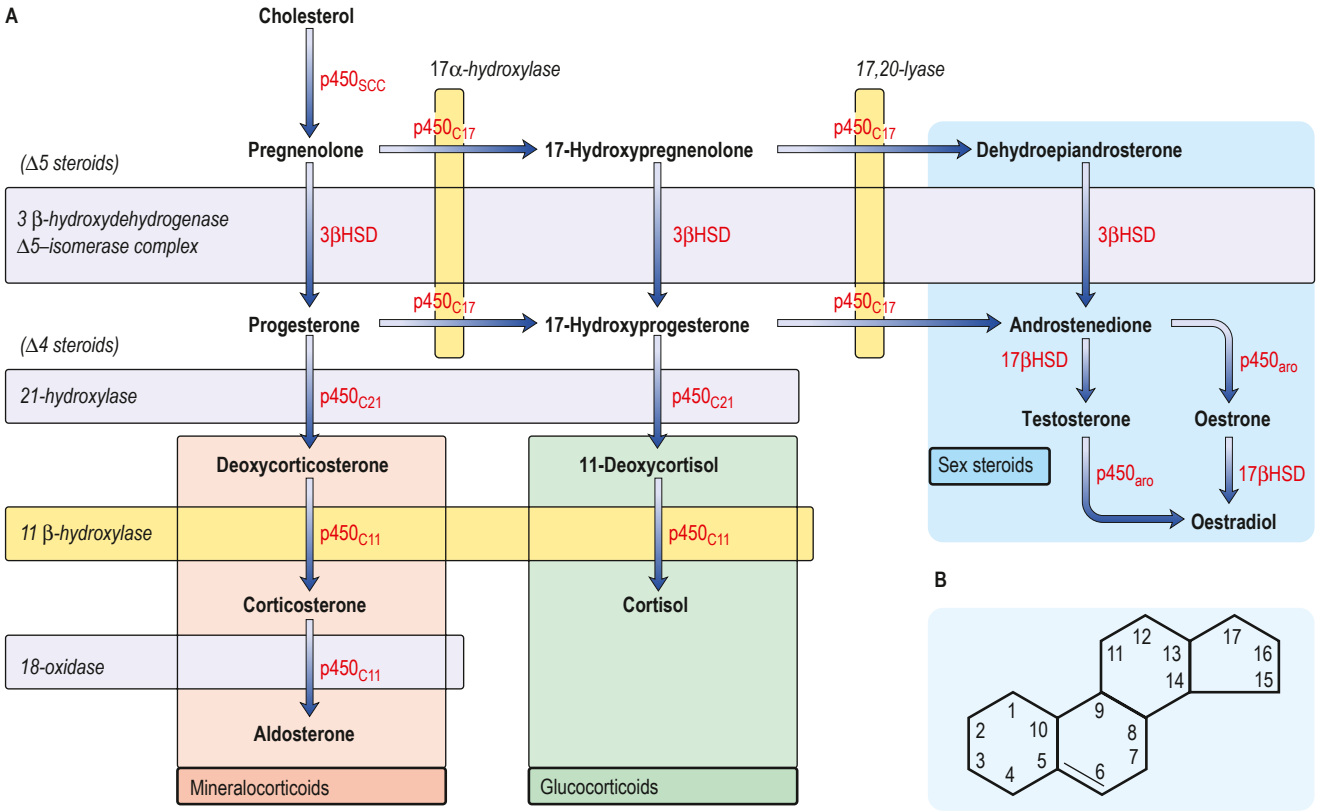


Fig. 34.13 Steroid pathways and molecule. **(A)** The major steroid biosynthetic pathways. Steroid hormones are synthesized in the adrenal glands and gonads from the initial substrate via a series of interconnected enzymatic steps. The reactions catalyzed are shown by shaded boxes with italic labels. The molecular identity of the enzymes is shown in *red*; some catalyze more than one reaction. The p450 enzymes are in mitochondria, 3β-hydroxysteroid dehydrogenase (3βHSD) in cytoplasm; 17βHSD and $p450_{aro}$ (aromatase) are found mainly in gonads. **(B)** The steroid molecule.

pro-opiomelanocortin (POMC), which undergoes processing within the pituitary to produce ACTH and a number of other peptides, including beta-lipotrophin and beta-endorphin. Many of these peptides, including ACTH, contain melanocyte-stimulating hormone (MSH)–like sequences, which cause pigmentation when levels of ACTH are markedly raised as in Addison's disease.

Circulating ACTH stimulates cortisol production in the adrenal. The secreted cortisol (or any other synthetic corticosteroid administered to the patient) causes negative feedback on the hypothalamus and pituitary to inhibit further CRH/ACTH release. The set-point of this system varies through the day according to the circadian rhythm and is usually overridden by severe stress. Unlike cortisol, mineralocorticoids and sex steroids do not cause negative feedback on the CRH/ACTH axis.

Following adrenalectomy or other adrenal damage (e.g., Addison's disease), cortisol secretion is absent or reduced; ACTH levels will therefore rise.

Mineralocorticoid secretion is mainly controlled by the renin–angiotensin system (see 'The adrenal cortex' section).

Investigation of glucocorticoid abnormalities

Basal levels

Cortisol has a circadian rhythm and increases in response to stress. When taking a blood sample, remember:

- Sampling time should be recorded.
- Basal cortisol levels should be taken between 8 a.m. and 10 a.m., near the peak of the circadian variation, if investigating for deficiency, and at midnight, near the peak of variation, if investigating for excess.
- Stress should be minimized.
- Appropriate reference ranges (for time and assay method) should be used.

The measurement of ACTH levels is only really required once an abnormality in cortisol levels has been identified. A plasma sample is required and has to be transported on ice to the lab immediately.

Suppression tests are used if excess cortisol is suspected, and stimulation tests are used if cortisol deficiency is suspected.

Dexamethasone suppression tests

Administration of a synthetic glucocorticoid (dexamethasone) to a normal subject produces suppression of CRH and ACTH levels and thus of endogenous cortisol secretion (dexamethasone is not measured by most cortisol assays). Three forms of the test, used in the diagnosis and differential diagnosis of Cushing's syndrome, are available (**Box 34.15**).

ACTH stimulation tests

Synthetic ACTH (tetracosactide, which consists of the first 24 amino acids of human ACTH) is given to stimulate adrenal cortisol production (see **Box 34.5** and **Fig. 34.5**).

Cushing's syndrome

Cushing's syndrome describes the clinical state of increased free circulating glucocorticoid. It is most often iatrogenic following the

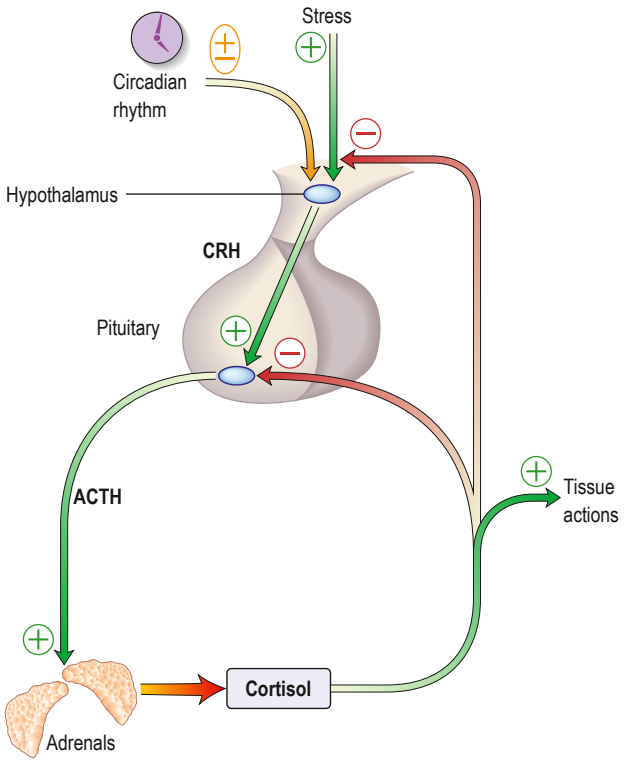


Fig. 34.14 Control of the hypothalamic–pituitary–adrenal axis. Pituitary adrenocorticotrophic hormone (ACTH) is secreted in response to hypothalamic corticotrophin-releasing hormone (CRH), triggered by circadian rhythm, stress, and other factors, and stimulates secretion of cortisol from the adrenal. Cortisol has multiple actions in peripheral tissues and exerts negative feedback on the pituitary and hypothalamus.

therapeutic administration of synthetic steroids. The next most common cause is excess endogenous secretion of ACTH from a pituitary adenoma, when it is called Cushing's disease (see below).

Pathophysiology and aetiology

Spontaneous Cushing's syndrome is rare, with an incidence of fewer than 5/million per year.

Causes of Cushing's syndrome are usually subdivided into two groups (Box 34.16):

- Increased circulating ACTH from the pituitary (65% of cases), known as Cushing's disease, or from an 'ectopic', non-pituitary, ACTH-producing tumour elsewhere in the body (10%) with consequent glucocorticoid excess ('ACTH-dependent' Cushing's).
- A primary excess of endogenous cortisol secretion (25%) by an adrenal tumour or, more rarely, by bilateral primary pigmented nodular hyperplasia (PPNH) with suppression of ACTH ('non-ACTH-dependent' Cushing's). Rare cases are due to aberrant expression of receptors for other hormones (e.g., glucose-dependent insulinotropic peptide [GIP], LH, or catecholamines) in adrenal cortical cells. Germline changes in protein kinase A (PKA) can also result in adrenal hyperplasia and adenomas.

History and examination

The clinical features of Cushing's syndrome are illustrated in Fig. 34.15.

- ***Pigmentation*** occurs only with ACTH-dependent causes (most frequently in ectopic ACTH syndrome).

Box 34.15 Dexamethasone suppression test in the diagnosis of Cushing's syndrome

Test and protocol	Measurement[a]	Normal test result or positive suppression	Use and explanation
Dexamethasone (for Cushing's): overnight			
Take 1 mg on going to bed at 23:00 h	Plasma cortisol at 09:00 h next morning	Plasma cortisol <100 nmol/L	Outpatient screening test Some 'false-positives'
'Low-dose'			
0.5 mg 6-hourly Eight doses from 09:00 h on day 0	Plasma cortisol at 09:00 h on days 0 and +2	Plasma cortisol <50 nmol/L on second sample	For diagnosis of Cushing's syndrome
'High-dose' used in differential diagnosis			
2 mg 6-hourly Eight doses from 09:00 h on day 0	Plasma cortisol at 09:00 h on days 0 and +2	Plasma cortisol on day +2 <50% of that on day 0 suggests pituitary-dependent disease	Differential diagnosis of Cushing's syndrome Pituitary-dependent disease suppresses in about 90% of cases

[a]Plasma cortisol values are very dependent upon the assay used; local reference ranges must be consulted.

Box 34.16 Causes of Cushing's syndrome

ACTH-dependent disease
- Pituitary-dependent (Cushing's disease)
- Ectopic ACTH-producing tumours

Non-ACTH-dependent disease
- Adrenal adenomas
- Adrenal carcinomas
- Exogenous steroids

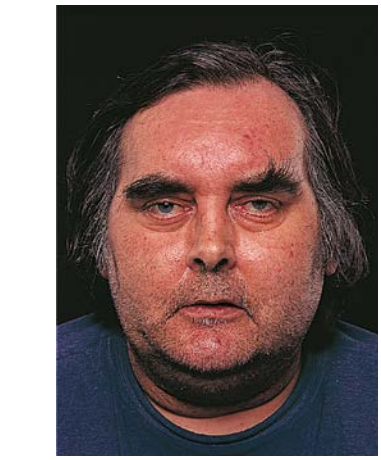
- ***A cushingoid appearance*** can also be caused by excess alcohol consumption (pseudo-Cushing's syndrome); the pathophysiology is poorly understood.
- ***Impaired glucose tolerance*** or frank diabetes is common, especially with ectopic ACTH.
- ***Hypokalaemia*** due to the mineralocorticoid activity of cortisol is common with ectopic ACTH secretion.
- ***Hypertension*** is common in all causes of Cushing's syndrome.

Investigations

There are two phases to investigation.

Confirmation

Most patients with obesity and hypertension or women with obesity and hirsutism do not have Cushing's syndrome, but some cases of genuine Cushing's can also have relatively subtle clinical signs. Confirmation rests on demonstrating inappropriate cortisol secretion that is not suppressed by exogenous glucocorticoids; difficulties occur with obesity and depression, where cortisol dynamics are often abnormal. Random cortisol measurements are of no value.



Symptoms

General
Change of appearance
Weight gain (central)
Hair growth/acne
Thin skin/easy bruising

Mental changes
Depression
Psychosis
Insomnia

Muscular weakness
Back pain

Amenorrhoea/oligomenorrhoea

Poor libido

Growth arrest in children

Polyuria/polydipsia

Old photographs may be useful

Signs

General
Plethora
Moon face
Hypertension
'Buffalo hump'
Central obesity
Depression/psychosis
Glycosuria
Oedema

Skin
Thin skin
Hirsutism
Acne
Bruising
Poor wound healing
Skin infections
Striae (purple or red)
Pigmentation

Musculoskeletal
Osteoporosis
Pathological fractures (esp. vertebrae and ribs)
Kyphosis
Rib fractures

Proximal myopathy
Proximal muscle wasting

Fig. 34.15 Cushing's syndrome: symptoms and signs. ***Bold italic*** type indicates signs of most value in discriminating Cushing's syndrome from simple obesity and hirsutism.

Occasional patients are seen with 'cyclical Cushing's', in which the abnormalities come and go.

Investigations to confirm the diagnosis include:

- ***48-h low-dose dexamethasone test*** (see Box 34.15). Normal individuals suppress plasma cortisol to less than 50 nmol/L. People with Cushing's syndrome fail to show complete suppression of plasma cortisol levels (although levels may fall substantially in a few cases). This test is highly sensitive (>97%). The overnight dexamethasone test is simpler but has a higher false-positive rate, so often used as an initial screening test, before progressing to the longer test.
- ***24-hour urinary free cortisol measurements.*** This is another screening test. It is simple but less reliable; repeatedly normal values render the diagnosis unlikely, but some people with Cushing's syndrome have normal values on some collections (approximately 10%).
- ***Circadian rhythm.*** After 48 h in hospital, cortisol samples are taken at 09:00 and 24:00 h (without warning the patient and ideally when they are asleep). Normal subjects show a pronounced circadian variation (see Fig. 34.4); those with Cushing's syndrome have high midnight cortisol levels (>100 nmol/L), though the 09:00 h value may be normal. Increasingly, loss of circadian rhythm recorded using ***midnight or late-night salivary cortisol*** collected at home can be used for the diagnosis and surveillance of Cushing's, removing the need for a hospital stay.
- ***Other tests.*** There are frequent exceptions to the classic responses to diagnostic tests in Cushing's syndrome. If any clinical suspicion of Cushing's remains after preliminary tests, then specialist investigations are still indicated. These ***may include the insulin stress test, desmopressin stimulation test*** (see Box 34.53), and ***corticotrophin-releasing hormone (CRH)*** tests.

Differential diagnosis of the cause

This can be extremely difficult on symptoms and signs alone since all causes can result in clinically identical Cushing's syndrome. The classical ectopic ACTH syndrome is distinguished by a short history, pigmentation, weight loss, hypokalaemia, diabetes, and plasma ACTH levels above 200 ng/L, but many ectopic tumours are benign and mimic pituitary disease closely, both clinically and biochemically. Severe hirsutism/virilization suggests an adrenal tumour.

Biochemical and radiological procedures for diagnosis include:

- If ***plasma ACTH levels*** are low or undetectable (<10 ng/L) on two or more occasions, this is a reliable indicator of non-ACTH-dependent disease and adrenal imaging should be planned, using ***adrenal CT or MRI scan.*** Adrenal adenomas and carcinomas causing Cushing's syndrome are relatively large and always detectable by CT. Carcinomas are distinguished by their large size and irregular outline, as well as signs of infiltration or metastases. Bilateral adrenal hyperplasia may be seen in ACTH-dependent causes or in ACTH-independent nodular hyperplasia.
- If ***plasma ACTH levels*** are normal or high on two or more occasions, this is a reliable indicator of ACTH-dependent disease and pituitary imaging should be planned in the first instance and biochemical tests to distinguish between pituitary and ectopic sources of ACTH.
- ***Pituitary MRI.*** A pituitary adenoma may be seen but the adenoma is often small and not visible in a significant proportion of cases.
- ***High-dose dexamethasone suppression test*** (see Box 34.15). Failure of significant plasma cortisol suppression suggests an ectopic source of ACTH (or an adrenal tumour).
- ***Desmopressin stimulation test.*** An exaggerated ACTH and cortisol response to exogenous desmopressin suggests

pituitary-dependent Cushing's disease, as ectopic sources rarely respond. This has replaced the CRH test.

- ***Chest, abdominal, and pelvic imaging.*** A carcinoma of the bronchus or a bronchial carcinoid is sought. Carcinoid lesions may be very small, and whole-body CT or PET imaging may be indicated.

Further investigations may involve:

- ***Selective catheterization*** of the inferior petrosal sinus to measure ACTH for pituitary lesions, or blood samples taken throughout the body to look for ectopic sources.
- ***Radiolabelled octreotide*** (^{111}In octreotide) and/or **Gallium-68 DOTATATE PET-CT** are occasionally helpful in locating ectopic ACTH sites.

Management

Cushing's syndrome

Successful treatment with a normal biochemical profile should lead to reversal of clinical features. However, untreated Cushing's syndrome has a very poor prognosis, with death from venous thromboembolism, hypertension, myocardial infarction, infection, and heart failure. Whatever the underlying cause, cortisol hypersecretion should be controlled prior to surgery or radiotherapy. Considerable morbidity and mortality are otherwise associated with operating on unprepared patients, especially when abdominal surgery is required.

The usual drug is ***metyrapone***, an 11β-hydroxylase blocker, which is given in doses of 750 mg to 4 g daily in 3–4 divided doses. ***Osilodrostat*** has a similar mechanism of action to metyrapone and appears more effective. ***Ketoconazole*** (200 mg three times daily) is also used and is synergistic with metyrapone. Plasma cortisol should be monitored, aiming to reduce the mean level during the day to 150–300 nmol/L, equivalent to normal production rates.

Aminoglutethimide, ***trilostane*** (which reversibly inhibits 3-hydroxysteroid dehydrogenase/5-5,4 isomers), and ***etomidate*** infusion (in severe cases) are occasionally used.

Choice of further treatment depends on the cause.

Cushing's disease (pituitary-dependent hypercortisolism)

- ***Trans-sphenoidal removal of the tumour*** is the treatment of choice. Selective adenomectomy nearly always leaves the patient ACTH-deficient immediately postoperatively, and this is a good prognostic sign. Overall, pituitary surgery results in remission in 75%–80% of cases, but results vary considerably and an experienced surgeon is essential.
- ***External pituitary irradiation*** alone is slow-acting, only effective in 50%–60% even after prolonged follow-up, and mainly used after failed pituitary surgery. Children respond much better to radiotherapy, with 80% being cured. Stereotactic radiotherapy can be useful in selected cases.
- ***Medical therapy*** to reduce ACTH (e.g., bromocriptine, cabergoline and cyproheptadine) is rarely effective. The somatostatin analogue pasireotide may provide medical control of Cushing's in some patients but is associated with hyperglycaemia as a common side-effect. Aggressive corticotroph adenomas may respond to temozolomide chemotherapy.
- ***Bilateral adrenalectomy*** is an effective last resort if other measures fail to control the disease (see 'Nelson's syndrome' later). This can be performed laparoscopically.

Cushing's syndrome due to other causes

Adrenal adenomas should be resected laparoscopically. Contralateral adrenal suppression may last for a year or more.

Adrenal carcinomas are highly aggressive and the prognosis is poor. In general, if there are no widespread metastases, tumour bulk should be reduced surgically. The adrenolytic drug mitotane may inhibit growth of the tumour and prolong survival, though it can cause nausea and ataxia. Sometimes radiotherapy is directed to the tumour bed after surgery.

Tumours secreting ACTH ectopically should be removed if possible. Otherwise chemotherapy/radiotherapy may be used, depending on the tumour type and location. Control of Cushing's syndrome with metyrapone or ketoconazole is beneficial for symptoms, and bilateral adrenalectomy may be appropriate to give complete control of Cushing's syndrome if prognosis from the tumour itself is reasonable.

If the source of ACTH is not clear, cortisol hypersecretion should be controlled with medical therapy until a diagnosis can be made.

Nelson's syndrome

Nelson's syndrome occurs in about 20% of cases after bilateral adrenalectomy for Cushing's disease and is characterized by increased pigmentation (because of high levels of ACTH), associated with an enlarging pituitary tumour. The syndrome is rare now that adrenalectomy is an uncommon primary treatment, and its incidence may be reduced by pituitary radiotherapy soon after adrenalectomy. The Nelson's adenoma may be treated by pituitary surgery and/or radiotherapy (unless given previously).

Further reading

Colao A, Petersenn S, Newell-Price J et al. A 12-month phase 3 study of pasireotide in Cushing's disease. *N Engl J Med* 2012; 366:914–924.
Feelders RA, Hofland LJ. Medical treatment of Cushing's disease. *J Clin Endocrinol Metab* 2013; 98:425–438.
Fleseriu M, Auchus R, Bancos I et al. Consensus on diagnosis and management of Cushing's disease: a guideline update. *Lancet Diabetes Endocrinol* 2021; 9(12):847–875.
Howlett TA, Willis D, Walker G et al. Control of growth hormone and IGF1 in patients with acromegaly in the UK: responses to medical treatment with somatostatin analogues and dopamine agonists. *Clin Endocrinol* 2013; 79:689–699.

Addison's disease

Pathophysiology and aetiology

In this condition, there is destruction of the entire adrenal cortex. Glucocorticoid, mineralocorticoid, and sex steroid production are therefore all reduced. This differs from adrenal insufficiency caused by HP disease, in which mineralocorticoid secretion remains largely intact, being predominantly stimulated by angiotensin II. Adrenal sex steroid production is also largely independent of pituitary action. In Addison's disease, reduced cortisol levels lead, through feedback, to increased CRH and ACTH production, the latter being directly responsible for the hyperpigmentation.

Addison's disease is rare, with an incidence of 3–4/million per year and prevalence of 40–60/million. Primary adrenal insufficiency shows a marked female preponderance and is usually caused by autoimmune disease (>90% in the UK), but in countries with a high prevalence of HIV/AIDS, tuberculosis is a common cause. Autoimmune adrenalitis results from the destruction of the adrenal cortex by organ-specific autoantibodies, with 21-hydroxylase as the common antigen. There are associations with other autoimmune conditions in polyglandular autoimmune syndromes types I and II (e.g., type 1 diabetes mellitus, pernicious anaemia, thyroiditis, hypoparathyroidism, premature ovarian insufficiency; see 'Diseases of many glands' section).

All other causes are rare (**Box 34.17**). In patients with degenerative neurological symptoms, a diagnosis of adrenoleukodystrophy should be excluded. In childhood the most common cause of primary adrenal insufficiency is CAH.

History and examination

See Fig. 34.16. The symptomatology of Addison's disease is often non-specific. These symptoms may be the prelude to an addisonian crisis with severe hypotension and dehydration precipitated by intercurrent illness, accident, or operation.

Pigmentation (dull, slaty, grey–brown) is the predominant sign in over 90% of cases.

Postural systolic hypotension, due to hypovolaemia and sodium loss, is present in 80%–90% of cases, even if supine blood pressure is normal. Mineralocorticoid deficiency is the cause of the hypotension.

Investigations

Once Addison's disease is suspected, investigation is urgent. If the patient is seriously ill or hypotensive, hydrocortisone 100 mg should be given intravenously or intramuscularly, together with intravenous 0.9% saline. Ideally, this should be done immediately after a blood sample is taken for later measurement of plasma cortisol. Full investigation should be delayed until emergency treatment (see below) has improved the patient's condition. Otherwise, tests are as follows:

- ***Single cortisol measurements*** are of little value, although a random cortisol below 100 nmol/L during the day is highly suggestive, and a random cortisol higher than 550 nmol/L makes the diagnosis unlikely.
- ***The short ACTH stimulation test*** should be performed (see Box 34.5 and Fig. 34.5). Note that an absent or impaired cortisol response confirms the presence of hypoadrenalism but does not differentiate Addison's disease from ACTH deficiency or iatrogenic suppression by steroid medication. The long ACTH test is no longer used because of the availability of accurate ACTH assays.
- ***A 09:00-hours plasma ACTH level*** is measured, a high level (>80 ng/L) with low cortisol confirming primary adrenal insufficiency.
- ***Adrenal antibodies*** are present in approximately 90% of cases of autoimmune adrenalitis.
- ***Abdominal imaging*** (plain radiographs, CT, or MRI) may show evidence of tuberculosis and/or calcified adrenals (Fig. 34.17).
- ***Plasma renin activity*** is high due to low serum aldosterone.
- ***Hyponatraemia, hyperkalaemia, and acute kidney injury*** may be present, but electrolytes can be normal.
- ***Blood glucose*** may be low.
- ***Hypercalcaemia and anaemia*** (after rehydration) are sometimes seen.

? Box 34.17 Causes of primary hypoadrenalism

- Autoimmune disease
- Tuberculosis (<10% in UK)
- Surgical removal
- Haemorrhage/infarction:
  - Meningococcal septicaemia
  - Venography
- Infiltration:
  - Malignant destruction
  - Amyloid
- Schilder's disease (adrenal leukodystrophy)

Management

Acute hypoadrenalism needs urgent treatment (Box 34.18).

Long-term treatment is with replacement glucocorticoid and mineralocorticoid; tuberculosis must be treated if present or suspected. Replacement dosage details are shown in Box 34.19. Dual-release, oral, once-daily hydrocortisone preparations are becoming increasingly available. Once-daily prednisolone may also be effective at doses of 3–4 mg. Dehydroepiandrosterone (DHEA) replacement has also been advocated by some for quality of life; some studies suggest that this may cause symptomatic improvements, although others show no clear benefit and acne and hirsutism occur.

Adequacy of glucocorticoid dose is judged by:

- clinical wellbeing and restoration of normal, but not excessive, weight; ideally aiming for a low dose, for example, total dose of hydrocortisone between 15 and 25 mg split in two or three divided doses throughout the day

Symptoms

Weight loss
Malaise
Weakness
Fever

Anorexia
Nausea/vomiting
Diarrhoea
Abdominal pain
Constipation
Depression
Confusion
Myalgia
Joint or back pain

Impotence/amenorrhoea

Syncope from postural hypotension

Signs

General
Loss of weight
General wasting
Pigmentation, especially of new scars and palmar creases
Buccal pigmentation
Postural hypotension
Loss of body hair (vitiligo)
Dehydration

Fig. 34.16 Primary hypoadrenalism (Addison's disease): symptoms and signs. *Bold italic* type indicates signs of greater discriminant value.

Fig. 34.17 CT scan showing bilateral adrenal calcification.

Box 34.18 Management of acute hypoadrenalism

Clinical context

Hypotension, hyponatraemia, hyperkalaemia, hypoglycaemia, dehydration, and pigmentation often with precipitating infection, infarction, trauma, or operation. The major deficiencies are of salt, steroid, and glucose.

Requirements

Assuming normal cardiovascular function, the following are required:

- 1 L of 0.9% saline should be given over 30–60 min with 100 mg of i.v. bolus hydrocortisone
- Subsequent requirements are several litres of saline within 24 h (assessing with central venous pressure line if necessary) plus hydrocortisone, 100 mg i.m., 6-hourly, until the patient is clinically stable
- Glucose should be infused if there is hypoglycaemia
- Oral replacement medication is then started, unless the patient is unable to take oral medication: initially, hydrocortisone 20 mg, 8-hourly, reducing to 20–30 mg in divided doses over a few days (see **Box 34.19**)
- Fludrocortisone is unnecessary acutely, as the high cortisol doses provide sufficient mineralocorticoid activity—it should be introduced before discharge

Box 34.19 Average replacement steroid dosages for adults with primary hypoadrenalism

Drug	Dose
Glucocorticoid	
Hydrocortisone	20–30 mg daily (e.g., 10 mg on waking, 5 mg at 12:00 h, 5 mg at 18:00 h)
rarely	
Prednisolone	5 mg daily on waking
very rarely	
Dexamethasone	0.75 mg daily (0.5 mg on waking, 0.25 mg at 18:00 h)
Mineralocorticoid	
Fludrocortisone	50–300 μg daily

- measuring cortisol levels during the day while on replacement hydrocortisone (cortisol levels cannot be used for synthetic steroids) is not recommended as an assessment of dose adequacy. Adequacy of fludrocortisone replacement is assessed by:
- restoration of serum electrolytes to normal
- no evidence of hypertension or postural hypotension (it should not fall >10 mmHg systolic after 2 minutes' standing)
- suppression of plasma renin activity to high normal.

Patient advice

All patients requiring replacement steroids should:

- know how to increase steroid replacement by doubling the dose for intercurrent illness
- carry a 'steroid card'
- wear a MedicAlert bracelet (or similar), which gives details of their condition so that emergency replacement therapy can be given if found unconscious
- keep an (up-to-date) ampoule of hydrocortisone at home in case oral therapy is impossible, for administration by self, family, ambulance, or doctor.

Secondary hypoadrenalism

Pituitary disease

Most people with HP disease have panhypopituitarism (see 'Presentation of pituitary and hypothalamic disease' section) and need T_4 replacement, as well as cortisol; in this case, hydrocortisone must be started before T_4.

Adrenal suppression due to long-term corticosteroid therapy

Long-term corticosteroid medication for non-endocrine disease is the most common cause of secondary hypoadrenalism. The HP axis and the adrenal may both be suppressed and the patient may have vague symptoms of feeling unwell. ACTH levels are low in secondary hypoadrenalism. Weaning off steroids is often a long and difficult process.

Congenital adrenal hyperplasia

Pathophysiology

Congenital adrenal hyperplasia (CAH) is an autosomal recessive disorder resulting in deficiency of an enzyme in the cortisol synthetic pathways. There are six major types; most common is 21-hydroxylase deficiency (CYP21A2), which occurs in about 1 in 15 000 births and is due to defects on chromosome 6 near the HLA region, affecting one of the cytochrome p450 enzymes (p450C21).

As a result, cortisol secretion is reduced and feedback leads to increased ACTH secretion, causing adrenal hyperplasia. Diversion of the steroid precursors into the androgenic steroid pathways occurs (see **Fig. 34.13A**). Thus, 17-hydroxyprogesterone, androstenedione, and testosterone levels are increased, leading to virilization. Aldosterone synthesis may be impaired with resultant salt wasting.

The other forms affect 11β-hydroxylase, 17α-hydroxylase, 3β-hydroxysteroid dehydrogenase, and a cholesterol side-chain cleavage enzyme (p450scc) (see **Fig. 34.13A**).

History and examination

Classical CAH presents at birth with:

- Ambiguous genitalia in females (clitoral hypertrophy, urogenital abnormalities and labioscrotal fusion are common) *or*
- Adrenal failure (collapse, hypotension, hypoglycaemia), sometimes with a salt-losing state (hypotension, hyponatraemia). The syndrome may be unrecognized in the male until a salt-losing crisis occurs, usually within 10 days of birth.

Non-classical disease presents later as precocious puberty with hirsutism; rare, milder cases present in adult life, usually accompanied by primary amenorrhoea. Hirsutism developing before puberty is suggestive of CAH.

Investigations

In many high-income countries neonatal screening for CAH takes place by assessing 17OHP in dried blood spots soon after birth. This approach would prevent undiagnosed children presenting with adrenal failure and identify those that may present later with precocious puberty.

Expert advice is essential in the confirmation and differential diagnosis of 21-hydroxylase deficiency; with ambiguous genitalia, advice must be sought urgently before any assignment of sex is made.

Unless 17OHP levels are very elevated (>30 nmol/L) a synacthen test should be performed with measurement of 17OHP and cortisol at 0 and 60 minutes. If 17OHP levels at baseline or after synacthen are above 30 nmol/L 21-hydroxylase deficiency is confirmed.

Assessment of mineralocorticoid activity is also required.

Management

Glucocorticoid activity must be replaced, as must mineralocorticoid activity if deficient. In ***adults***, clinical features and biochemistry (plasma renin, androstenedione and 17-hydroxyprogesterone) are used to modify treatment. The use of modified-release hydrocortisone may have a role in the management of CAH.

Prenatal diagnosis

Prenatal genetic counselling is essential (see Chapter 2). There is a 1:4 chance of an affected child if one child in the family is already affected. In a parent with CAH, genetic testing of the partner provides important information about the risk of having an affected child. Dexamethasone can be offered to the mother as it may prevent virilization of an affected female fetus by suppressing ACTH levels. There are, however, concerns about the potential adverse consequences of this practice.

Problems of therapeutic steroid therapy

Apart from their use as therapeutic replacement for endocrine deficiency states, synthetic glucocorticoids are widely used for many non-endocrine conditions. Short-term use (e.g., for acute asthma) carries only small risks of significant side-effects. In general, therapy for 3 weeks or less, or a dose of prednisolone of less than 5 mg per day, will not result in significant long-term suppression of the normal adrenal axis.

Long-term therapy with synthetic or natural steroids will, in most respects, mimic endogenous Cushing's syndrome. Exceptions are the relative absence of hirsutism, acne, hypertension, and severe sodium retention, as the common synthetic steroids have low androgenic and mineralocorticoid activity.

Excessive doses of steroids may also be absorbed from skin when strong dermatological preparations are used, and along with inhaled steroids may rarely cause Cushing's syndrome; they may, however, quite commonly cause adrenal suppression.

The major hazards are detailed in **Box 34.20**. In the long term, many are of such severity that the clinical need for high-dose steroids should be continually and critically assessed. Steroid-sparing agents should always be considered, and screening and prophylactic therapy for osteoporosis introduced (see Chapter 33, 'Osteoporosis' section). New targeted biological therapies for inflammatory conditions may reduce the incidence of steroid-induced adrenal suppression.

Supervision of steroid therapy

All patients receiving steroids should carry a 'steroid card'. They should be made aware of the following points:

- Long-term steroid therapy must never be stopped suddenly.
- Doses should be reduced gradually, with most being given in the morning at the time of withdrawal; this minimizes adrenal suppression. 'Alternate-day therapy' may produce less suppression.
- Depending on current dose, doses need to be increased at times of serious intercurrent illness, accident, and stress. Double doses should be taken during these periods.
- Other physicians, anaesthetists, and dentists must be told about steroid therapy.

Patients should also be informed of potential side-effects, and all this information should be documented in the clinical record. If prophylactic use of bisphosphonate therapy is required to prevent the development of osteoporosis (following NICE guidance), patients should be informed of the rationale.

Box 34.20 Major adverse effects of corticosteroid therapy

Physiological
- Adrenal suppression (secondary)

Pathological

Cardiovascular
- Increased blood pressure

Gastrointestinal
- Pancreatitis

Renal
- Polyuria
- Nocturia

Central nervous
- Depression
- Euphoria
- Psychosis
- Insomnia

Endocrine
- Weight gain
- Glycosuria/hyperglycaemia/diabetes
- Impaired growth
- Amenorrhoea

Bone and muscle
- Osteoporosis
- Proximal myopathy and wasting
- Avascular necrosis of the hip
- Pathological fractures

Skin
- Thinning
- Easy bruising

Eyes
- Cataracts

Increased susceptibility to infection

(Signs and fever are frequently masked)
- Septicaemia
- Fungal infections
- Reactivation of tuberculosis
- Skin infections (e.g., fungi)

Steroids and surgery

Any patient receiving steroids, or who has recently received a 4-week course of steroids (within the last 6 months) and may still have adrenal suppression, requires steroid cover around the time of surgery (**Box 34.21**).

Incidental adrenal tumours ('incidentalomas')

With the advent of abdominal CT, MRI, and high-resolution ultrasound scanning, unsuspected adrenal masses have been discovered in 3%–10% of scans (increasing with age). The two issues with an incidental adrenal mass are:

- whether the lesion is functional or non-functional
- whether it is benign or malignant.

Most incidentalomas are asymptomatic and benign, but direct questioning may reveal symptoms of endocrine hypersecretion such as cushingoid features, catecholamine excess, virilization in women, or evidence of endocrine hypertension (see **Box 31.2**). Even in the absence of symptoms, biochemical tests to exclude secretory activity should be performed, as adrenal adenomas often secrete cortisol at a low level—'autonomous cortisol secretion', which may confer increased cardiovascular risk. If no endocrine activity is found, then most guidelines recommend removal only of large adrenal tumours (>4–5 cm) because of the risk of malignancy. Smaller, hormonally inactive lesions are usually left alone. Adrenal biopsy

Box 34.21 Steroid cover for operative procedures[a]

Procedure	Premedication	Intra- and postoperative	Resumption of normal maintenance
Simple procedures (e.g., gastroscopy, simple dental extractions)	Hydrocortisone 100 mg i.m.	–	Immediately if no complications and eating normally
Minor surgery (e.g., laparoscopic surgery, veins, hernias)	Hydrocortisone 100 mg i.m.	Hydrocortisone 20 mg orally 6-hourly or 50 mg i.m. 6-hourly for 24 h if not eating	After 24 h if no complications
Major surgery (e.g., hip replacement, vascular surgery)	Hydrocortisone 100 mg i.m.	Hydrocortisone 50–100 mg i.m. 6-hourly for 72 h	After 72 h if normal progress and no complications Perhaps double normal dose for 2–3 days
GI tract surgery or major thoracic surgery (not eating or ventilated)	Hydrocortisone 100 mg i.m.	Hydrocortisone 100 mg i.m. 6-hourly for 72 h or longer if still unwell	When patient eating normally again Until then, higher doses (to 50 mg 6-hourly) may be needed

[a]A useful summary of surgical steroid guidelines can be found at: http://www.addisons.org.uk/.

is usually contraindicated in evaluation of adrenal tumours, as the procedure can precipitate potentially fatal phaeochromocytoma crisis from sudden catecholamine surge or can breach the capsule of an adrenal malignancy, facilitating metastasis due to the procedure.

Phaeochromocytoma must be excluded before surgery due to the risk of perioperative hypertensive or hypotensive crises (see 'The adrenal medulla' section).

Primary hyperaldosteronism

Increased mineralocorticoid secretion from the adrenal cortex, termed primary hyperaldosteronism, is one of the endocrine causes of 5%–10% of all hypertension, which is termed 'secondary' (see Chapter 18). Other endocrine causes of hypertension should also be considered if there is clinical suspicion (**Box 34.22**). It is impracticable to screen all hypertensive patients for secondary endocrine causes. The highest chances of detecting such causes are in patients:

- under 35 years, especially those without a family history of hypertension
- with accelerated (malignant) hypertension
- with hypokalaemia
- resistant to conventional antihypertensive therapy (e.g., more than three drugs) *or*
- with unusual symptoms (e.g., sweating attacks or weakness).

Pathophysiology

Primary hyperaldosteronism is characterized by excess aldosterone production leading to sodium retention, potassium loss, and the combination of hypokalaemia and hypertension. This must be distinguished from secondary hyperaldosteronism (see Chapter 27), which arises when there is excess renin (and hence angiotensin II) stimulation of the zona glomerulosa. Common causes of secondary hyperaldosteronism include:

- renal artery stenosis or renin-secreting renal tumours, when the patient will also be hypertensive
- heart failure and cirrhosis, when the patient is typically normotensive. Excess aldosterone production contributes to the sodium retention and excess extracellular fluid accumulation seen in these conditions.

Box 34.22 Endocrine causes of hypertension

Excessive renin, and thus angiotensin II, production
- Renal artery stenosis
- Other local renal disease
- Renin-secreting tumours

Excessive production of catecholamines
- Phaeochromocytoma

Excessive growth hormone (GH) production
- Acromegaly

Excessive aldosterone production
- Adrenal adenoma (Conn's syndrome)
- Idiopathic adrenal hyperplasia
- Dexamethasone-suppressible hyperaldosteronism

Excessive production of other mineralocorticoids
- Cushing's syndrome (massive excess of cortisol, a weak mineralocorticoid)
- Congenital adrenal hyperplasia (in rare cases)
- Tumours producing other mineralocorticoids, e.g., corticosterone

Exogenous 'mineralocorticoids' or enzyme inhibitors
- Liquorice ingestion (inhibits 11β-hydroxylase)
- Misuse of mineralocorticoid preparations

Aetiology

Adrenal adenomas (Conn's syndrome; see **Box 34.22**) originally accounted for 60% of cases of primary hyperaldosteronism but represented a rare cause of hypertension. The use of the aldosterone:renin ratio (ARR) in the routine investigation of hypertension now suggests that hyperaldosteronism due to bilateral adrenal hyperplasia (idiopathic hyperaldosteronism) is much more common than the classical Conn's adenoma. There is an increasing recognition that primary hyperaldosteronism may have a genetic basis in certain individuals and families.

History and examination

The usual presentation is simply hypertension; hypokalaemia (<3.5 mmol/L) is frequently not present. The few symptoms are nonspecific; rarely, muscle weakness, nocturia, and tetany are seen. Hypertension may be severe and associated with renal, cardiac, and retinal damage.

Adenomas, often very small, are more common in young females, while bilateral hyperplasia rarely occurs before the age of 40 years and is more common in males.

Investigations

Beta-blockers and other drugs may interfere with renin activity, and spironolactone, angiotensin-converting enzyme (ACE) inhibitors, and angiotensin II receptor antagonists will all affect results; all should be discontinued, if possible. The characteristic features of investigation are as follows:

- ***High plasma aldosterone:renin ratio (ARR).*** This is now the most frequently used screening test for the condition, but raised ARR alone does not confirm the diagnosis (if the renin is low

enough, ARR will always be high). Note that normal ranges are highly assay-dependent.

- ***Elevated plasma aldosterone levels*** which are not suppressed with 0.9% saline infusion (2 L over 4 h) or fludrocortisone administration. Between 30% and 50% of people with a raised ARR on screening will suppress normally, excluding the diagnosis.
- ***Suppressed plasma renin activity*** or immunoreactivity is seen.
- ***Hypokalaemia*** is often present but a normal potassium does not exclude the diagnosis.
- ***Urinary potassium loss*** greater than 30 mmol daily during hypokalaemia is inappropriate.

Once a diagnosis of hyperaldosteronism is established, differentiation of adenoma from hyperplasia involves adrenal CT or MRI, but small adenomas may be missed and non-functioning incidentalomas also occur. Further information is obtained from diurnal/postural changes in plasma aldosterone levels (which tend to rise with adenomas between 09:00 hours supine and 13:00 hours erect samples; in contrast, they fall with hyperplasia), measurement of 18-OH cortisol levels (raised in adenoma), and venous catheterization for aldosterone levels. All of these tests have pitfalls and exceptions.

Management

An adenoma can be removed surgically, usually laparoscopically; blood pressure falls in 70% of patients. Those with hyperplasia should be treated with the aldosterone antagonist, spironolactone (100–400 mg daily); frequent side-effects include nausea, rashes, and gynaecomastia. The pure aldosterone receptor antagonist, eplerenone, is a useful alternative if side-effects preclude the use of spironolactone (see Chapter 27, 'Potassium-sparing diuretics' section). Amiloride and calcium-channel blockers are moderately effective in controlling hypertension but do not correct the hyperaldosteronism.

Glucocorticoid (or dexamethasone)-suppressible hyperaldosteronism

This rare condition is caused by a chimeric gene on chromosome 8. A fusion gene resulting from an unusual crossover at meiosis between the genes encoding aldosterone synthase and adrenal 11β-hydroxylase produces aldosterone, which is under ACTH control. Treatment with glucocorticoid resolves the problem.

Syndrome of apparent mineralocorticoid excess

This causes the clinical syndrome of primary hyperaldosteronism but with low renin and aldosterone levels. Reduced activity of the enzyme 11β-hydroxysteroid dehydrogenase type 2 (11βHSD2) prevents the normal conversion in the kidney of cortisol (which is active at the mineralocorticoid receptor) to cortisone (which is not) and therefore 'exposes' the mineralocorticoid receptor in the kidney to the usual molar excess of cortisol over aldosterone in the blood. While the inherited syndrome is rare, the same clinical syndrome can occur with excessive ingestion of liquorice, which inhibits the 11βHSD2 enzyme.

The adrenal medulla

The adrenal medulla is the innermost part of the adrenal gland, consisting of cells that secrete the major catecholamines, noradrenaline (norepinephrine), and adrenaline (epinephrine), which produce the sympathetic nervous response. The catecholamines are interconverted in the adrenal medulla, and an increase in levels of their metabolites in the urine is a marker of abnormal hypersecretion (Fig. 34.18).

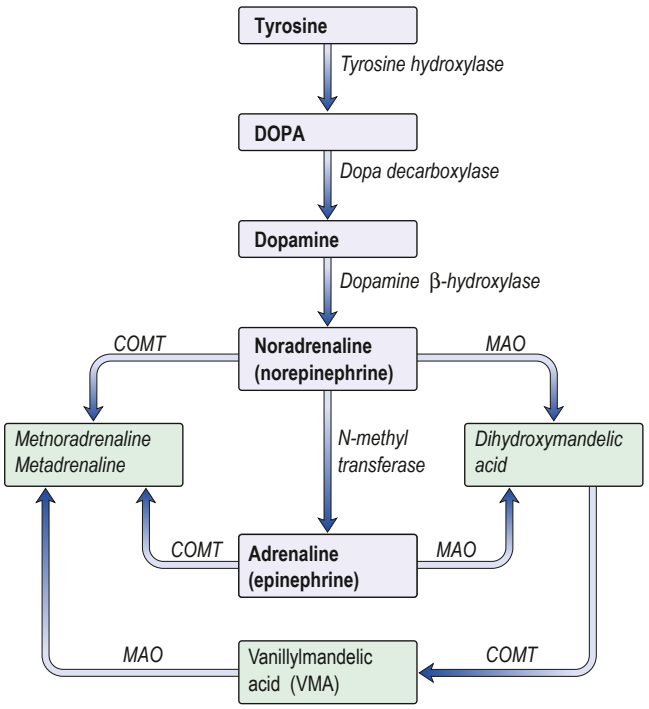


Fig. 34.18 The synthesis and metabolism of catecholamines.
COMT, Catechol-O-methyl transferase; *MAO,* monoamine oxidase.

Phaeochromocytoma and paraganglioma

These are very rare tumours of the sympathetic nervous system (less than 1 in 1000 cases of hypertension) that secrete catecholamines, noradrenaline (norepinephrine), adrenaline (epinephrine), and their metabolites (see Fig. 34.18):

- 90% arise in the adrenal medulla ***(phaeochromocytomas)***
- 10% occur elsewhere in the sympathetic chain ***(paragangliomas).***

Some are associated with MEN 2 (see below), succinate dehydrogenase (*SDHB* and *SDHD*) mutations, von Hippel–Lindau's (VHL) syndrome (see 'Diseases of many glands' section), and less commonly neurofibromatosis type 1. Most tumours release both noradrenaline (norepinephrine) and adrenaline (epinephrine), but large tumours and extra-adrenal tumours produce almost entirely noradrenaline.

Paragangliomas typically occur in the head and neck but are also found in the thorax, pelvis, and bladder. They are more closely associated with other genetic associations than phaeochromocytoma. The association of paraganglioma, bilateral adrenal phaeochromocytomas, positive family history, or young age at presentation is seen in MENs (see 'The adrenal medulla' section). Mutations in the *SDHB* and *SDHD* genes have been shown to be strongly associated with the development of paraganglioma.

Pathology

Oval groups of cells occur in clusters and stain for chromogranin A; 25% are multiple and 10% malignant, the latter being more frequent in extra-adrenal tumours. Malignancy cannot be determined on simple histological examination alone.

History and examination

The clinical features are those of catecholamine excess and are frequently, but not necessarily, intermittent (Box 34.23). All people with

Box 34.23 Symptoms and signs of phaeochromocytoma

Symptoms
- Anxiety or panic attacks
- Palpitations
- Tremor
- Sweating
- Headache
- Flushing
- Nausea and/or vomiting
- Weight loss
- Constipation or diarrhoea
- Raynaud's phenomenon
- Chest pain
- Polyuria/nocturia

Signs
- Hypertension
- Tachycardia/arrhythmias
- Bradycardia
- Orthostatic hypotension
- Pallor or flushing
- Glycosuria
- Fever
- (Signs of hypertensive damage)

suspected phaeochromocytomas must be investigated because phaeochromocytomas may cause acute cardiovascular compromise during routine medical procedures and can also present with sudden death if the diagnosis is missed.

Investigations

Specific tests are:

- ***Measurement of urinary catecholamines and metabolites*** is a useful screening test (metanephrines are most sensitive and specific; see Fig. 34.18); normal levels on three 24-hour collections of metanephrines virtually exclude the diagnosis. Many drugs, e.g., tricyclics, and dietary vanilla interfere with these tests.
- ***Resting plasma metanephrines*** are raised.
- ***Plasma chromogranin A*** (a storage vesicle protein) is raised.
- ***Clonidine suppression test*** may be appropriate but should only be performed in specialist centres.
- ***CT/MRI scans***, initially of the abdomen, are helpful to localize the tumours, which are often large (Fig. 34.19).
- ***Scanning with [^{131}I]meta-iodobenzylguanidine (MIBG)*** produces specific uptake in sites of sympathetic activity with about 90% success. It is particularly useful with extra-adrenal tumours. ^{18}F-deoxyglucose PET is also used by some centres (Fig. 34.20).
- ***68-Gallium PET scans*** may be useful in non-MIG avid phaeochromocytomas
- ***Genetic testing*** for *MEN2*, *VHL*, *SDHB*, and *SDHD* mutations should be performed in all people with confirmed phaeochromocytoma or paraganglioma.

Management

Tumours should be removed if possible; 5-year survival is about 95% for non-malignant tumours. Medical preoperative and perioperative treatment is vital and includes complete alpha- and beta-blockade with phenoxybenzamine (20–80 mg daily in divided doses), then propranolol (120–240 mg daily), plus intravenous hydration to re-expand the contracted plasma volume. Alpha-blockade must precede the beta-blockade, as worsened hypertension may otherwise result. Labetalol is not recommended. Surgery in the unprepared patient is fraught with dangers of both hypertension and hypotension; expert anaesthesia and an experienced surgeon are vital, and sodium nitroprusside, phentolamine (a rapid-acting α-blocker), and magnesium should be available in case sudden severe hypertension develops.

When operation is not possible, combined alpha- and beta-blockade can be used long term. Radionucleotide treatment with MIBG has been employed but has had limited success in malignant phaeochromocytoma.

Patients should be kept under clinical and biochemical review after tumour resection, as over 10% of tumours recur or a further tumour develops. Catecholamine excretion measurements should be performed at least annually.

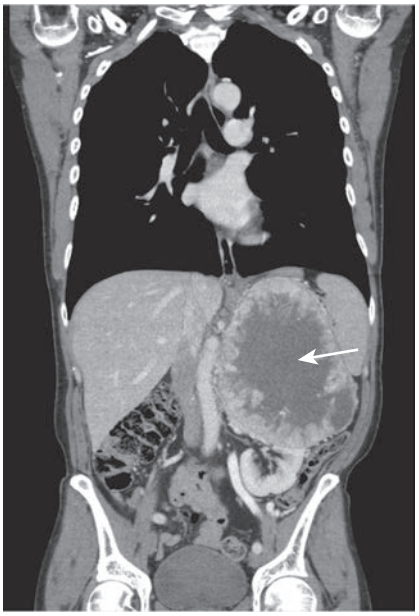


Fig. 34.19 MRI scan showing a massive left-sided phaeochromocytoma.

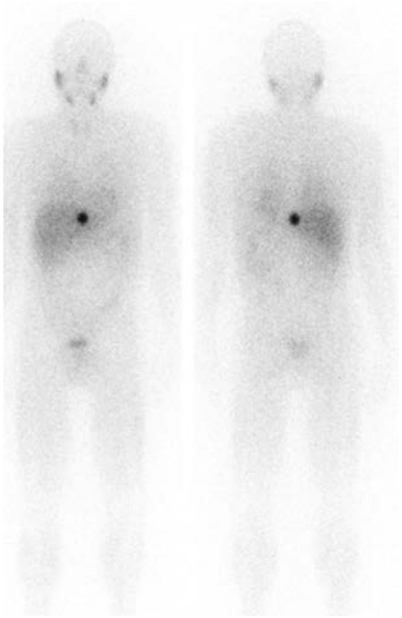


Fig. 34.20 MIBG scan showing a solitary area of increased uptake in the liver consistent with metastasis from a phaeochromocytoma. *MIBG,* Meta-iodobenzylguanidine.

Further reading

Auchus RJ, Arlt W. Approach to the patient: the adult with congenital adrenal hyperplasia. *J Clin Endocrinol Metab* 2013; 98:2645–2655.
Fassnacht M, Tsagarakis S, Terzolo et al. European Society of Endocrinology clinical practice guidelines on the management of adrenal incidentalomas, in collaboration with the European Network for the Study of Adrenal Tumors. *Eur J Endocrinol* 2023; 189(1):G1–G42.
Lenders JW, Duh QY, Eisenhofer G et al. and the Endocrine Society. Pheochromocytoma and paraganglioma: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab* 2014; 99:1915–1942.
Muth A, Crona J, Gimm O et al. Genetic testing and surveillance guidelines in hereditary pheochromocytoma and paraganglioma. *J Intern Med* 2019; 285(2):187–204.
Speiser PW, Arlt W, Auchus RJ et al. Congenital adrenal hyperplasia due to steroid 21-hydroxylase deficiency: an Endocrine Society clinical practice guideline. *Clin Endocrinol Metab* 2018; 103(11):4043–4088.

THE THYROID AXIS

The metabolism of virtually all nucleated cells of many tissues is controlled by the thyroid hormones. Overactivity or underactivity of the gland is the most common of all endocrine problems.

Anatomy and physiology

Anatomy

The thyroid gland consists of two lateral lobes connected by an isthmus. It is closely attached to the thyroid cartilage and to the upper end of the trachea and thus moves on swallowing. It is often palpable in healthy women.

Embryologically, it originates from the base of the tongue and descends to the middle of the neck. Remnants of thyroid tissue can sometimes be found at the base of the tongue (lingual thyroid) and along the line of descent. The gland has a rich blood supply from superior and inferior thyroid arteries.

The thyroid gland consists of follicles lined by cuboidal epithelioid cells. Inside is colloid (the iodinated glycoprotein thyroglobulin), which is synthesized by the follicular cells. Each follicle is surrounded by basement membrane, and between them are parafollicular cells containing calcitonin-secreting C cells.

Physiology

Synthesis

The thyroid synthesizes two hormones:

- triiodothyronine (T_3), which acts at the cellular level
- L-thyroxine (T_4), which is the prohormone.

Inorganic iodide is trapped by the gland via an enzyme-dependent system, oxidized, and incorporated into the glycoprotein thyroglobulin to form mono- and diiodotyrosine and then T_4 and T_3 (Fig. 34.21).

More T_4 than T_3 is produced, but T_4 is converted in some peripheral tissues (liver, kidney and muscle) to the more active T_3 by 5′-monodeiodination; an alternative 3′-monodeiodination yields the inactive reverse T_3 (rT_3). The latter step occurs particularly in severe non-thyroidal illness (see later).

In plasma, more than 99% of all T_4 and T_3 is bound to hormone-binding proteins (thyroxine-binding globulin, TBG; thyroid-binding prealbumin, TBPA; and albumin). Only free hormone is available for action in the target tissues, where T_3 binds to specific nuclear receptors within target cells. Many drugs and other factors affect TBG; all may result in confusing total T_4 levels in blood, and most laboratories therefore now measure free T_4 levels.

Control of the hypothalamic–pituitary–thyroid axis

Thyrotrophin-releasing hormone (TRH), a peptide produced in the hypothalamus, stimulates the pituitary to secrete thyroid-stimulating

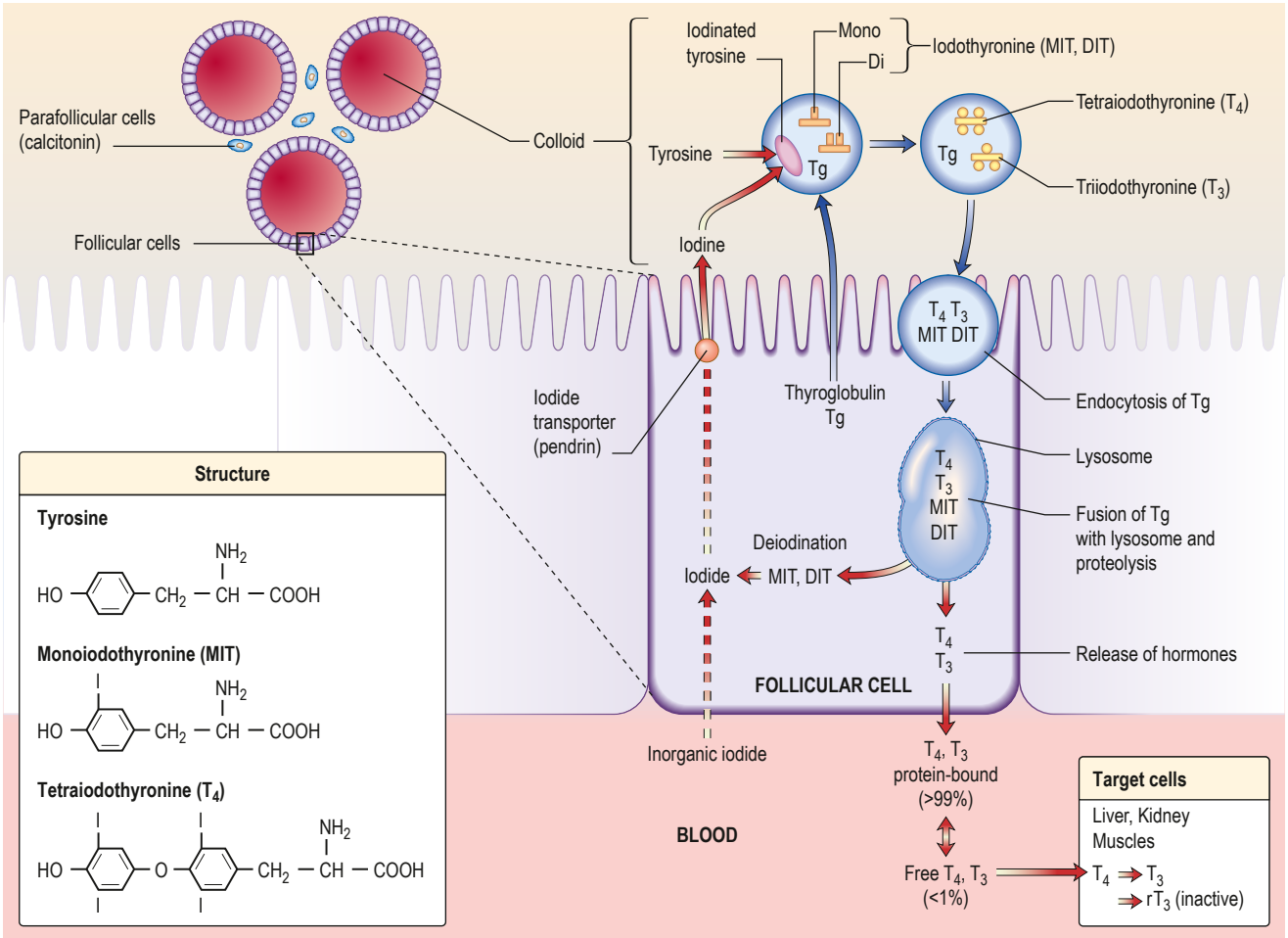


Fig. 34.21 Synthesis and metabolism of the thyroid hormones. *Tg,* Thyroglobulin.

Box 34.24 Hormones and receptors of the hypothalamic–pituitary–thyroid axis

Hormone	Source	Site of action	Hormone structure	Receptor	Post-receptor activity
Thyrotrophin-releasing hormone (TRH)	Hypothalamus	Pituitary	Peptide—3 AA	Membrane—7TM TRHR	G-proteins PLC/IP_3
Thyroid-stimulating hormone (TSH)	Pituitary	Thyroid	Glycoprotein: α and β subunits	Membrane—7TM TSHR	G-proteins cAMP/ PLC/IP_3
Thyroxine and triiodothyronine (T_4 and T_3)	Thyroid	All tissues	Thyronines—4/3 iodine atoms	Nuclear TR-α and β	Transcription TRE/RXR

7TM, 7 transmembrane (G-protein coupled receptor); *AA*, amino acids; *TRE/RXR*, thyroid hormone response element/response silencing region.

hormone (TSH) (see **Fig. 34.2**). TSH, in turn, stimulates growth and activity of the thyroid follicular cells via the G-protein-coupled TSH membrane receptor (**Box 34.24**). The T_3 and T_4 subsequently secreted into the circulation by follicular cells exert negative feedback on the hypothalamus, as described in the 'Dynamic endocrine tests' section.

Circulating T_4 is peripherally deiodinated to T_3, which binds to the thyroid hormone nuclear receptor (TR) on target organ cells to cause modified gene transcription. There are two TR receptors (TR-α and TR-β) and the tissue-specific effects of T_3 are dependent on the local expression of these TR receptors. TR-α knockout mice show poor growth, bradycardia, and hypothermia, while TR-β knockout mice show thyroid hyperplasia and high T_4 levels in the presence of inappropriately normal circulating TSH, suggesting a role for the latter receptors in thyroid hormone resistance (see 'Thyroid hormone resistance' section).

Physiological effects of thyroid hormones

The physiological effects of thyroid hormones are summarized in **Box 34.25**.

Box 34.25 Physiological effects of thyroid hormone

Target	Effect
Cardiovascular system	Increases heart rate and cardiac output
Bone	Increases bone turnover and resorption
Respiratory system	Maintains normal hypoxic and hypercapnic drive in respiratory centre
Gastrointestinal system	Increases gut motility
Blood	Increases red blood cell 2,3-BPG[a], facilitating oxygen release to tissues
Neuromuscular function	Increases speed of muscle contraction/relaxation and muscle protein turnover
Carbohydrate metabolism	Increases hepatic gluconeogenesis/ glycolysis and intestinal glucose absorption
Lipid metabolism	Increases lipolysis and cholesterol synthesis and degradation
Sympathetic nervous system	Increases catecholamine sensitivity and β-adrenergic receptor numbers in heart, skeletal muscle, adipose cells, and lymphocytes Decreases cardiac α-adrenergic receptors

[a]2,3-BPG, 2,3-bisphosphoglyceric acid.

Dietary iodine requirement

Globally, dietary iodine deficiency is a major cause of thyroid disease, as iodine is an essential requirement for thyroid hormone synthesis. The recommended daily intake of iodine should be at least 140 μg, and dietary supplementation of salt and bread has reduced the number of areas where 'endemic goitre' still occurs (see later).

Investigations: thyroid function tests

Immunoassays for free T_4, free T_3, and TSH are widely available. There are only minor circadian rhythms and measurements may be made at any time. Particular uses of the tests are summarized in **Box 34.26**, with typical findings in common disorders.

TSH measurement

In most circumstances, TSH levels can discriminate between hyperthyroidism, hypothyroidism, and euthyroidism (normal thyroid gland function). Exceptions are hypopituitarism, and the 'sick euthyroid' syndrome, where low levels (which normally imply hyperthyroidism) occur in the presence of low or normal T_4 and T_3 levels. As a single test of thyroid function, TSH measurement is the most sensitive in most circumstances, but accurate diagnosis requires at least two tests: for example, TSH plus free T_4 or free T_3 where hyperthyroidism is suspected; TSH plus serum free T_4 where hypothyroidism is likely.

TRH test

This has been rendered almost obsolete by modern sensitive TSH assays, except for investigation of HP dysfunction. TRH (protirelin) is occasionally used to differentiate between thyroid hormone resistance and a pituitary TSH-secreting tumour (TSHoma) in the context of raised fT_4 and TSH levels. Typically, after TRH administration, there is a rise in TSH in thyroid hormone resistance, while in TSHoma there is a flat response due to continued autonomous TSH secretion, which does not respond to TRH.

Problems in interpretation of thyroid function tests

There are three major areas of difficulty: serious acute or chronic illness, pregnancy and use of oral contraceptives, and certain drugs.

Serious acute or chronic illness

Thyroid function is affected in several ways:

- reduced concentration and affinity of binding proteins
- decreased peripheral conversion of T_4 to T_3 with more rT_3
- reduced HP TSH production.

Systemically ill patients can therefore have an apparently low total and free T_4 and T_3 with a normal or low basal TSH (the 'sick euthyroid' syndrome). Levels are usually only mildly below normal and are thought to be mediated by interleukins IL-1 and IL-6; the tests should be repeated after resolution of the underlying illness.

Box 34.26 Characteristics of blood thyroid function tests in common thyroid disorders[a]

	TSH (0.3–3.5 mU/L)	Free T_4 (10–25 pmol/L)	Free T_3 (3.5–7.5 pmol/L)
Thyrotoxicosis	**Suppressed (<0.05 mU/L)**	**Increased**	**Increased**
Primary hypothyroidism	Increased (>10 mU/L)	Low/low-normal	Normal or low
TSH deficiency	Low-normal or subnormal	**Low/low-normal**	Normal or low
T_3 toxicosis	Suppressed (<0.05 mU/L)	Normal	**Increased**
Compensated euthyroidism	**Slightly increased (5–10 mU/L)**	**Normal**	Normal

TSH, Thyroid-stimulating hormone.
[a]The clinically most informative tests in each situation are shown in **bold**.

Pregnancy and oral contraceptives

These lead to greatly increased TBG levels and thus to high or high-normal total T_4. Free T_4 is usually normal. Normal ranges for free T_4 and TSH alter with the normal physiological changes during pregnancy and TSH is often slightly suppressed in the first trimester, but this rarely causes clinical problems. See Chapter 48 for management of thyroid disease in pregnancy.

Drugs

Amiodarone decreases T_4 to T_3 conversion and free T_4 levels may therefore be above normal in a euthyroid patient; conversely, amiodarone may induce both hyper- and hypothyroidism—the TSH level is usually reliable.

Many drugs affect thyroid function tests by interfering with protein binding but this now rarely causes a problem with free T_4 assays.

Anti-thyroid antibodies

Serum antibodies to the thyroid are common.

- ***Destructive antibodies*** are directed against the microsomes or against thyroglobulin; the antigen for thyroid microsomal antibodies is the thyroid peroxidase (TPO) enzyme. TPO antibodies are found in up to 20% of the normal population, especially older women, but only 10%–20% of these develop overt hypothyroidism.
- ***TSH receptor IgG antibodies (TRAb)*** typically stimulate, but occasionally block, the receptor. They are specific for Graves' disease.

Hypothyroidism

Pathophysiology

Underactivity of the thyroid is usually primary, caused by disease of the thyroid, but may be secondary to HP disease (reduced TSH drive) (**Box 34.27**). Primary hypothyroidism is one of the most common endocrine conditions, with an overall UK prevalence of over 2% in women but under 0.1% in men; lifetime prevalence for an individual is higher—perhaps as high as 9% for women and 1% for men, with mean age at diagnosis around 60 years. The worldwide prevalence of subclinical hypothyroidism varies from 1% to 10%.

Box 34.27 Causes of hypothyroidism

Primary disease of thyroid

Congenital
- Agenesis
- Ectopic thyroid remnants

Defects of hormone synthesis
- Iodine deficiency
- Dyshormonogenesis
- Anti-thyroid drugs
- Other drugs (e.g., lithium, amiodarone, interferon)

Autoimmune
- Atrophic thyroiditis
- Hashimoto's thyroiditis
- Postpartum thyroiditis

Infective
- Post-subacute thyroiditis

Post-surgery

Post-irradiation
- Radioactive iodine therapy
- External neck irradiation

Infiltration
- Tumour

Secondary (to hypothalamic–pituitary disease)

Hypopituitarism
- Isolated TSH deficiency

Peripheral resistance to thyroid hormone

Aetiology of primary hypothyroidism

See **Box 34.27**.

Autoimmune

Atrophic (autoimmune) hypothyroidism

This is the most common cause of hypothyroidism and is associated with anti-thyroid autoantibodies, leading to lymphoid infiltration of the gland and eventual atrophy and fibrosis. It is six times more common in females and the incidence increases with age. The condition is associated with other autoimmune diseases, such as pernicious anaemia, vitiligo, and other endocrine deficiencies. Occasionally, intermittent hypothyroidism occurs with subsequent recovery; antibodies that block the TSH receptor may sometimes be involved in the aetiology.

Hashimoto's thyroiditis

This form of autoimmune thyroiditis, again more common in women and most common in late middle age, produces atrophic changes with regeneration, leading to goitre formation. The gland is usually firm and rubbery but may range from soft to hard. TPO antibodies are present, often in very high titres (>1000 IU/L). Patients may be hypothyroid or euthyroid, though they may go through an initial toxic phase, 'Hashi-toxicity'. Levothyroxine therapy may shrink the goitre, even when the patient is not hypothyroid.

Postpartum thyroiditis

This is usually a transient phenomenon observed following pregnancy. It may cause hyperthyroidism, hypothyroidism, or the two sequentially. It is believed to result from the modifications to the immune system necessary in pregnancy, and histologically is a lymphocytic thyroiditis. The process is normally self-limiting, but when conventional antibodies are found there is a high chance of this proceeding to permanent hypothyroidism. Postpartum thyroiditis may be misdiagnosed as postnatal depression, emphasizing the need for thyroid function tests in this situation. See Chapter 48.

Defects of hormone synthesis

Iodine deficiency

Dietary iodine deficiency still exists as 'endemic goitre' in some areas where goitre, occasionally massive, is common. The patients may be euthyroid or hypothyroid, depending on the severity of iodine deficiency. The mechanism is thought to be borderline hypothyroidism, leading to TSH stimulation and thyroid enlargement in the face of continuing iodine deficiency. Iodine deficiency is still a problem in the

Netherlands, Western Pacific, India, South-east Asia, Russia, and parts of Africa. Efforts to prevent deficiency by providing iodine in salt continue worldwide but often with incomplete success, and iodine deficiency continues to cause a high prevalence of congenital hypothyroidism (cretinism) in some parts of India.

Dyshormonogenesis

This rare condition is due to genetic defects in the synthesis of thyroid hormones; patients develop hypothyroidism with goitre. One particular familial form is associated with sensorineural deafness due to a deletion mutation in chromosome 7, causing a defect of the transporter pendrin (Pendred's syndrome) (see Fig. 34.21).

History and examination

Hypothyroidism produces many symptoms (Fig. 34.22). The alternative term 'myxoedema' refers to the accumulation of mucopolysaccharide in subcutaneous tissues. The classic picture of the slow, dry-haired, thick-skinned, deep-voiced patient with weight gain, cold intolerance, bradycardia, and constipation makes the diagnosis easy. Milder symptoms are, however, more common and hard to distinguish from other causes of non-specific tiredness. Many cases are detected on biochemical screening.

Special difficulties in diagnosis may arise in certain circumstances:

- ***Children with hypothyroidism*** may not show classic features but often have a slow growth velocity, poor school performance, and sometimes arrest of pubertal development.
- ***Young women with hypothyroidism*** may not show obvious signs. Hypothyroidism should be excluded in all people with oligomenorrhoea/amenorrhoea, menorrhagia, infertility, or hyperprolactinaemia.
- ***The elderly*** show many clinical features that are difficult to differentiate from normal ageing. Hypothyroidism should be excluded in elderly patients with cognitive impairment.

Investigation of primary hypothyroidism

Serum TSH is the investigation of choice; a high TSH level confirms primary hypothyroidism. A low free T_4 level confirms the hypothyroid state (and is also essential to exclude TSH deficiency if clinical hypothyroidism is strongly suspected and TSH is normal or low).

Thyroid and other organ-specific antibodies may be present. Other abnormalities include the following:

- ***anaemia***, which is usually normochromic and normocytic but, where co-morbidity exists, may be macrocytic (sometimes this is due to associated pernicious anaemia) or microcytic (in women, due to menorrhagia or undiagnosed coeliac disease)
- ***increased serum aspartate transferase levels***, from muscle and/or liver
- ***increased serum creatine kinase levels***, with associated myopathy
- ***hypercholesterolaemia*** and ***hypertriglyceridaemia***
- ***hyponatraemia*** due to an increase in AVP and impaired free water clearance.

Management

Replacement therapy

Replacement therapy with levothyroxine (thyroxine, i.e., T_4) is given for life. The starting dose will depend upon the severity of the deficiency and on the age and fitness of the patient, especially their cardiac performance: 100 μg daily for the young and fit, 50 μg (increasing to 100 μg after 2–4 weeks) for the small, old, or frail. People with ischaemic heart disease require even lower initial doses, especially if the hypothyroidism is severe and longstanding. Most physicians would then begin with 25 μg daily and perform serial electrocardiograms (ECGs), increasing the dose at 3- to 4-week intervals if angina does not occur or worsen, and the ECG does not deteriorate. Occasional patients develop 'thyrotoxic' (hyperthyroid) symptoms despite normal fT_4 levels if the dose is increased too rapidly. It is unclear whether replacement with liothyronine (T_3) as monotherapy or with levothyroxine is beneficial to patients.

Symptoms

General
Tiredness/malaise
Weight gain
Cold intolerance
Change in appearance
Goitre

Depression
Psychosis
Coma
Poor memory

Hair – dry, brittle, unmanageable
Skin – dry, coarse

Arthralgia
Myalgia
Muscle weakness/stiffness
Poor libido
Puffy eyes
Deafness
Constipation
Anorexia

Signs

Mental slowness
Psychosis/dementia
Ataxia
Poverty of movement
Deafness

'Peaches and cream' complexion
Dry thin hair
Loss of eyebrows

Hypertension
Hypothermia
Heart failure
Bradycardia
Pericardial effusion

Cold peripheries
Carpal tunnel syndrome
Oedema

Periorbital oedema
Deep voice
(Goitre)

Dry skin
Mild obesity

Myotonia
Muscular hypertrophy
Proximal myopathy
Slow-relaxing reflexes

Anaemia

Fig. 34.22 Hypothyroidism: symptoms and signs. ***Bold italic*** type indicates symptoms or signs of greater discriminant value. A history from a relative is often revealing. Symptoms of other autoimmune disease may be present.

Monitoring

The aim is to restore T_4 and TSH to well within the normal range. Adequacy of replacement is assessed clinically and by thyroid function tests after at least 6 weeks on a steady dose. If serum TSH remains high, the dose of T_4 should be increased in increments of 25–50 μg, and the tests repeated at 6–8-week intervals until TSH becomes normal, ideally in the lower third of the normal range. Complete suppression of TSH should be avoided because of the risk of atrial fibrillation and osteoporosis. The usual maintenance dose is 100–150 μg given as a single daily dose. An annual thyroid function test is recommended—this is usually performed in the primary care setting, often assisted and prompted by district 'thyroid registers'.

Clinical improvement on T_4 may not begin for 2 weeks or more, and full resolution of symptoms may take 6 months. The necessity for lifelong therapy must be emphasized and the possibility of other autoimmune endocrine disease developing, especially Addison's disease or pernicious anaemia, should be considered. During pregnancy, an increase in T_4 dosage of about 25–50 μg is often needed to maintain relatively stricter TSH range of 0.3–2.5 mU/L, and the necessity for optimal replacement during pregnancy is emphasized by the finding of reductions in cognitive function in children of mothers with elevated TSH during pregnancy.

In cases where there is difficulty normalizing the TSH, compliance issues, concurrent medication that can interfere with thyroxine absorption (such as iron, calcium compounds and proton pump inhibitors), and undiagnosed coeliac disease should be considered.

A few people with primary hypothyroidism complain of incomplete symptomatic response to T_4 replacement. Combination T_4 and T_3 replacement has been advocated in this context, but randomized clinical trials show no consistent benefit in quality-of-life symptoms.

Borderline hypothyroidism or 'compensated euthyroidism'

Patients are frequently seen with low-normal serum T_4 levels and slightly raised TSH levels. Treatment with levothyroxine is normally recommended where the TSH is consistently above 10 mU/L, or when possible symptoms, high-titre thyroid antibodies, or lipid abnormalities are present. Where the TSH is only marginally raised, the tests should be repeated 3–6 months later, as a significant proportion will be normal on repeat testing. Conversion to overt hypothyroidism is more common in men or when TPO antibodies are present. In practice, vague symptoms in people with marginally elevated TSH (below 10 mU/L) rarely respond to treatment, but a 'therapeutic trial' of replacement may be needed to confirm that symptoms are unrelated to the thyroid. It is also considered best to normalize TSH during (and ideally before) pregnancy to avoid potential fetal adverse effects.

Myxoedema coma

Severe hypothyroidism, especially in the elderly, may present with confusion or even coma. Myxoedema coma is very rare; hypothermia is often present and the patient may have severe cardiac failure, pericardial effusions, hypoventilation, hypoglycaemia, and hyponatraemia. The mortality was previously at least 50% and patients require full intensive care. Optimal treatment is controversial and data are lacking; most physicians would advise T_3 orally or intravenously in doses of 2.5–5 μg every 8 hours, then increasing as above. Large intravenous doses should not be used. Additional measures, though unproven, should include:

- oxygen (by ventilation if necessary)
- monitoring of cardiac output and pressures
- gradual rewarming
- hydrocortisone 100 mg i.v. 8-hourly
- glucose infusion to prevent hypoglycaemia.

'Myxoedema madness'

Depression is common in hypothyroidism. Rarely, with severe hypothyroidism in the elderly, the patient may become frankly demented or psychotic, sometimes with striking delusions. This may occur shortly after starting T_4 replacement.

Screening for hypothyroidism

The incidence of ***congenital*** hypothyroidism is approximately 1 in 3500 births. Untreated, severe hypothyroidism produces permanent neurological and intellectual damage ('cretinism'). Routine screening of the newborn using a blood spot, as in the Guthrie test, to detect a high TSH level as an indicator of primary hypothyroidism is efficient and cost-effective; cretinism is prevented if T_4 is started within the first few months of life.

Screening of ***elderly*** patients for thyroid dysfunction has a low pick-up rate, is controversial, and is not currently recommended unless as part of a confusion screen. However, patients who have undergone thyroid surgery or received radio-iodine should have regular thyroid function tests, as should those receiving lithium or amiodarone therapy.

Hyperthyroidism

Hyperthyroidism (thyroid overactivity, thyrotoxicosis) is common, affecting perhaps 2%–5% of all females at some time and having a sex ratio of 5:1; it most often occurs between the ages of 20 and 40 years. Nearly all cases (>99%) are caused by intrinsic thyroid disease; a pituitary cause is extremely rare (**Box 34.28**). Most common patterns of hyperthyroidism and their distinctive features are outlined in **Box 34.29**.

Graves' disease

This is the most common cause of hyperthyroidism and is due to an autoimmune process. Serum IgG antibodies bind to TSH receptors in the thyroid, stimulating thyroid hormone production; that is, they behave like TSH. These TSH receptor antibodies (TSHR-Ab) are specific for Graves' disease. Persistent high levels predict a relapse when drug treatment is stopped. There is an association with HLA-B8, DR3, and DR2; the concordance rate among monozygotic twins is 50% and that in dizygotic twins is 5%. There is an association with cytotoxic T lymphocyte-associated antigen 4

? Box 34.28 Causes of hyperthyroidism

Common
- Graves' disease (autoimmune)
- Toxic multinodular goitre
- Solitary toxic nodule/adenoma

Uncommon
- Acute thyroiditis
  - Viral (e.g., de Quervain's)
  - Autoimmune
  - Post-irradiation
  - Postpartum
- Gestational thyrotoxicosis (HCG-stimulated)
- Neonatal thyrotoxicosis (maternal thyroid antibodies)
- Exogenous iodine
- Drugs—amiodarone, immunotherapy (ipilimumab, pembrolizumab, nivolumab)
- Thyrotoxicosis factitia (secret T_4 consumption)

Rare
- TSH-secreting pituitary tumours
- Metastatic differentiated thyroid carcinoma
- HCG-producing tumours
- Hyperfunctioning ovarian teratoma (struma ovarii)

Box 34.29 Differential diagnosis of thyrotoxicosis

	Graves' disease	Solitary or multinodular thyrotoxicosis	Thyroiditis
History	Symptoms of thyrotoxicosis	Symptoms of thyrotoxicosis	Symptoms of thyrotoxicosis Recent viral infection Recent delivery of child Painful neck swelling
Clinical examination	Signs of thyrotoxicosis Diffuse goitre Extrathyroidal manifestations including ophthalmopathy	Signs of thyrotoxicosis Palpable solitary nodule or multinodular goitre	Signs of thyrotoxicosis Diffuse goitre Tender on palpation
Biochemical tests	TSH suppressed fT_4 and fT_3 elevated	TSH suppressed fT_4 and fT_3 elevated (only fT_3 elevated in T3 toxicosis)	TSH suppressed fT_4 and fT_3 elevated Raised inflammatory markers (ESR, CRP)
Immunological tests	TSH receptor antibodies raised TPO antibodies raised		TPO antibodies may be raised
Thyroid isotope uptake scan (Scintiscan ^{99}Tm)	Diffuse uptake of tracer	Focal uptake of tracer in nodules with reduced uptake in rest of thyroid gland	Minimal uptake of tracer

(CTLA-4), HLA-DRBP*08, and DRB3*0202 on chromosome 6. The mechanism of immune damage is illustrated in Fig. 3.16.

Yersinia enterocolitica, *Escherichia coli,* and other Gram-negative organisms contain TSH-binding sites. This raises the possibility that the initiating event in the pathogenesis may be an infection with possible 'molecular mimicry' in a genetically susceptible individual, but the precise initiating mechanisms remain unproven in most cases.

Thyroid eye disease is a feature of Graves' disease. It may accompany the hyperthyroidism in many cases but patients may also be euthyroid or hypothyroid. Other components of Graves' disease, such as Graves' dermopathy, are very uncommon. Rarely, lymphadenopathy and splenomegaly may occur. Graves' disease is also associated with other autoimmune disorders such as pernicious anaemia, vitiligo, and myasthenia gravis.

The natural history is one of fluctuation, with many patients showing a pattern of alternating relapse and remission; perhaps only 40% of subjects have a single episode. Many patients eventually become hypothyroid.

Other causes of hyperthyroidism/thyrotoxicosis

Solitary toxic adenoma/nodule

This is the cause of about 5% of cases of hyperthyroidism. While the hyperthyroidism will be controlled by anti-thyroid drugs, it does not usually remit after a course of anti-thyroid drugs.

Toxic multinodular goitre

This commonly occurs in older women. Again, anti-thyroid drugs are rarely successful in inducing a remission, although they can control the hyperthyroidism.

de Quervain's thyroiditis

This is transient hyperthyroidism from an acute inflammatory process, probably viral in origin. Apart from the toxicosis, there is usually fever, malaise, and pain in the neck with tachycardia and local thyroid tenderness. Thyroid function tests show initial hyperthyroidism, the erythrocyte sedimentation rate (ESR) and plasma viscosity are raised, and thyroid uptake scans show suppression of uptake in the acute phase. Hypothyroidism, usually transient, may then follow after a few weeks. Treatment of the acute phase is with aspirin, using short-term prednisolone in severely symptomatic cases.

Postpartum thyroiditis

This is described in Chapter 48.

Amiodarone-induced thyrotoxicosis

Amiodarone, a class III antiarrhythmic drug (see Chapter 17), causes two types of hyperthyroidism.

- ***Type I amiodarone-induced thyrotoxicosis (AIT)*** is associated with pre-existing Graves' disease or multinodular goitre. In this situation, hyperthyroidism is probably triggered by the high iodine content of amiodarone.
- ***Type II AIT*** is not associated with previous thyroid disease and is thought to be due to a direct effect of the drug on thyroid follicular cells, leading to a destructive thyroiditis with release of T_4 and T_3. Type II AIT may be associated with a hypothyroid phase several months after presentation. Because amiodarone inhibits the deiodination of T_4 to T_3, biochemical presentation of both types of AIT may be associated with higher T_4:T_3 ratios than usual.

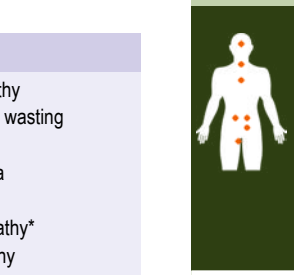
Immunotherapy-induced thyrotoxicosis

Cancer immunotherapy drugs that block negative regulators on T cell, called immune checkpoint inhibitors (ipilimumab, pembrolizumab, nivolumab), have been found to be effective in a wide range of cancers. These drugs cause immune-related adverse events (irAEs) in various organs, including in endocrine glands, resulting in thyroiditis (15%), hypophysitis (9%), adrenalitis (1%), and type 1 diabetes mellitus (<1%). Thyroiditis is the most common of endocrine irAEs and occurs between 3 and 12 weeks after initiation of the drug. Treatment is with beta-blockers, analgesia and, rarely, with anti-thyroid medication if thyrotoxicosis persists, especially with symptoms.

History and examination

The symptoms and signs of hyperthyroidism affect many systems (Fig. 34.23). Symptomatology and signs vary with age and with the underlying aetiology.

- ***The eye signs—lid lag and 'stare'***—may occur with hyperthyroidism of any cause but other features of thyroid eye disease (see below) occur only in Graves' disease.
- ***Graves' dermopathy*** is rare and can occur on any extensor surface. ***Pretibial myxoedema*** is the most commonly described and is an infiltration of the skin on the shin. ***Thyroid acropachy*** is very rare and consists of clubbing, swollen fingers, and periosteal new bone formation.
- ***In the elderly***, a frequent presentation is with atrial fibrillation, other tachycardias, and/or heart failure, often with few other



Symptoms
General
Weight loss
Irritability/behaviour change
Restlessness
Malaise
Itching
Sweating
Tall stature (in children)
Breathlessness
Palpitations
Heat intolerance
Stiffness
Muscle weakness
Tremor
Choreoathetosis
Thirst
Vomiting
Diarrhoea
Eye complaints*
Oligomenorrhoea
Loss of libido
Gynaecomastia
Onycholysis
Goitre
*Only in Graves' disease

Signs	
Tremor	Proximal myopathy
Hyperkinesis	Proximal muscle wasting
Psychosis	Onycholysis
	Palmar erythema
Tachycardia or atrial fibrillation	Graves' dermopathy*
Full pulse	Thyroid acropachy
Warm vasodilated peripheries	Pretibial myxoedema
Systolic hypertension	
Cardiac failure	
Exophthalmos*	
Lid lag and 'stare'	
Conjunctival oedema	
Ophthalmoplegia*	
Periorbital oedema	
Goitre, bruit	
Weight loss	
*Only in Graves' disease	

Fig. 34.23 Hyperthyroidism: symptoms and signs. ***Bold italic*** type indicates symptoms or signs of greater discriminant value.

signs. Thyroid function tests are mandatory in any patient with atrial fibrillation.

- ***Children*** frequently present with excessive height or excessive growth rate, or with behavioural problems such as hyperactivity. They may also show weight gain rather than loss.
- ***So-called 'apathetic thyrotoxicosis'*** in some elderly patients presents with a clinical picture more like that of hypothyroidism. There may be very few signs and a high degree of clinical suspicion is essential.

Differential diagnosis

Clinical differentiation from anxiety states may be difficult; useful positive clinical markers are weight loss, eye signs, a diffuse goitre, proximal myopathy with wasting, and evidence of a hyperdynamic circulation.

Investigations

- ***Serum TSH*** is suppressed in hyperthyroidism (<0.05 mU/L), except for the very rare instances of TSH hypersecretion.
- ***A raised free T_4 or T_3*** confirms the diagnosis; T_4 is almost always raised but T_3 is more sensitive, as there are occasional cases of isolated 'T_3 toxicosis'.
- ***TSH receptor stimulating antibodies (TSHR-Ab)*** are now measured routinely and the third-generation tests are 97%–99% specific for Graves' disease.
- ***Thyroid peroxidase (TPO)*** and thyroglobulin antibodies are present in 80% of cases of Graves' disease but are also found in normal individuals.
- ***Scintiscan ^{99}Tm*** is used in patients who are antibody negative to look for toxic nodular disease.

Management

Three possibilities are available: anti-thyroid drugs, radio-iodine, and surgery. Practices and beliefs differ widely within and between countries.

Anti-thyroid drugs

Carbimazole is most often used in the UK, and propylthiouracil (PTU) is also an option. ***Thiamazole*** (methimazole), the active metabolite of carbimazole, is used in the USA. These drugs inhibit the formation of thyroid hormones and also have other minor actions; carbimazole/thiamazole is also an immunosuppressive agent. Initial doses and side-effects are detailed in **Box 34.30**.

Although thyroid hormone synthesis is reduced very quickly, the long half-life of T_4 (7 days) means that clinical benefit is not apparent for 10–20 days.

As many of the manifestations of hyperthyroidism are mediated via the sympathetic system, ***beta-blockers*** are used to provide rapid partial symptomatic control; they also decrease peripheral conversion of T_4 to T_3. Preferred drugs are those without intrinsic sympathomimetic activity, e.g., propranolol (see **Box 34.30**). They should not be used alone for hyperthyroidism except when the condition is self-limiting, as in subacute thyroiditis.

Subsequent management is either by gradual dose titration or by a 'block and replace' regimen. Neither regimen has been shown to be unequivocally superior. TSH often remains suppressed for

Box 34.30 Drugs used in the treatment of hyperthyroidism

Drug	Usual starting dose	Side-effects	Remarks
Anti-thyroid drugs			
Carbimazole	20–40 mg daily, 8-hourly or in single dose	Rash, nausea, vomiting, arthralgia, agranulocytosis (0.1%), jaundice	Active metabolite is thiamazole (methimazole) Mild immunosuppressive activity
Propylthiouracil (PTU)	100–200 mg 8-hourly	Rash, nausea, vomiting, agranulocytosis, hepatotoxicity	Additionally blocks conversion of T_4 to T_3
Beta-blocker for symptomatic control			
May need higher doses than normal			
Propranolol	40–80 mg every 6–8 h		Avoid in asthma Use agents without intrinsic sympathomimetic activity as receptors are highly sensitive

many months after clinical improvement and normalization of T_4 and T_3.

Dosage regimen

Gradual dose titration

1. Start carbimazole 20–40 mg daily alongside propranolol for symptomatic control.
2. Review after 4–6 weeks and reduce the dose of carbimazole, depending on clinical state and fT_4/fT_3 levels. TSH levels may remain suppressed for several months and are unhelpful at this stage.
3. When the patient is clinically and biochemically euthyroid, stop beta-blockers.
4. Review thyroid function regularly during the planned course of treatment (typically 18 months—but some use courses between 6 and 24 months).
5. Reduce carbimazole if fT_4 falls below or TSH rises above normal, and when approaching the end of the planned course.
6. Increase carbimazole if fT_4 or fT_3 is above normal (and consider if TSH remains suppressed after several months with a normal fT_4).
7. Stop treatment at the end of the course if the patient is euthyroid on 5 mg daily carbimazole. PTU is used in a similar fashion (see **Box 34.30**).

'Block and replace' regimen

With this policy, full doses of anti-thyroid drugs, usually carbimazole 40 mg daily, are given to suppress the thyroid completely while replacing thyroid activity with 100 μg of levothyroxine daily once euthyroidism has been achieved. This is continued usually for 18 months, the claimed advantages being the avoidance of over- or undertreatment and the better use of the immunosuppressive action of carbimazole. This regimen is contraindicated in pregnancy, as T_4 crosses the placenta less well than carbimazole.

Relapse

About 50% of patients will relapse after a course of carbimazole or PTU, mostly within the following 2 years but occasionally much later. Long-term anti-thyroid therapy is then used, or surgery or radiotherapy is considered (see below). Most patients (90%) with hyperthyroidism have a diffuse goitre, but those with large single or multinodular goitres are unlikely to remit after a course of anti-thyroid drugs and will usually require definitive treatment. Severe biochemical hyperthyroidism is also less likely to remain in remission.

Toxicity

The major side-effect of drug therapy is agranulocytosis, which occurs in approximately 1 in 1000 patients, usually within 3 months of treatment. All patients must be warned to seek immediate medical attention and to have a check of their white blood cell count if they develop unexplained fever or sore throat; written information is essential. Rashes are more frequent and usually require a change of drug. If toxicity occurs on carbimazole, PTU may be used, and vice versa; side-effects are only occasionally repeated on the other drug.

Radioactive iodine

Radioactive iodine (RAI) is given to patients of all ages, although it is contraindicated in pregnancy and while breast-feeding. RAI is the most common treatment modality in the USA, whereas anti-thyroid drugs tend to be favoured in Europe.

^{131}Iodine is given in an empirical dose (usually 400–550 MBq) because of variable uptake and radiosensitivity of the gland. It accumulates in the thyroid and destroys the gland by local radiation, although it takes several months to be fully effective.

Patients must be rendered euthyroid before treatment. They should stop anti-thyroid drugs at least 4 days before radio-iodine, and not recommence until 3 days after radio-iodine. Patients on PTU should stop anti-thyroid medication before RAI earlier than those on carbimazole because it has a radioprotective action. Many patients do not need to restart anti-thyroid medication after treatment.

Early discomfort in the neck and immediate worsening of hyperthyroidism are sometimes seen. If worsening occurs, the patient should receive propranolol (see **Box 34.30**); if necessary, carbimazole can be restarted. Euthyroidism normally returns in 2–3 months. People with dysthyroid eye disease are more likely to show worsening of eye problems after radio-iodine than after anti-thyroid drugs; this represents a partial contraindication to RAI, although worsening can usually be prevented by steroid administration.

Long-term surveillance

Hypothyroidism affects the majority of subjects over the following 20 years. Some 75% of patients are rendered euthyroid in the short term, but a small proportion remain hyperthyroid and may require a second dose of radio-iodine. Long-term surveillance of thyroid function is necessary, with frequent tests in the first year after therapy and at least annually thereafter.

There is no increased risk of malignancy after RAI.

Surgery

Thyroidectomy should be performed only in patients who have previously been rendered euthyroid. Conventional practice is to stop

the anti-thyroid drug 10–14 days before operation and to give potassium iodide (60 mg three times daily), which reduces the vascularity of the gland and reduces thyroid hormone synthesis by inhibiting organification of iodine (Wolff–Chaikoff effect).

The operation should be performed only by experienced surgeons to reduce the chance of complications:

- Early postoperative bleeding causing tracheal compression and asphyxia is a rare emergency that mandates immediate removal of all clips/sutures to allow escape of the blood/haematoma.
- Laryngeal nerve palsy occurs in 1%. Vocal cord movement should be checked preoperatively. Mild hoarseness is more common and thyroidectomy is best avoided in professional singers.
- Transient hypocalcaemia occurs in up to 10% but with permanent hypoparathyroidism in fewer than 1%.
- Ongoing thyroid function depends on the operation performed. With a single toxic nodule, excision of the lesion is curative. With Graves' disease or multinodular goitre, the traditional 'subtotal' thyroidectomy, aiming for euthyroidism on no treatment, results in recurrent hyperthyroidism in 1%–3% within 1 year, then 1% per year and hypothyroidism in about 10% of patients within 1 year, and then increasing with time. 'Near-total' thyroidectomy is therefore now preferred, with inevitable hypothyroidism but a much-reduced risk of recurrence.

  Indications for either surgery or radio-iodine are given in **Box 34.31**.

Special situations in hyperthyroidism

Thyroid crisis or 'thyroid storm'

This rare condition, with a mortality of 10%, is a rapid deterioration of hyperthyroidism with hyperpyrexia, severe tachycardia, extreme restlessness, cardiac failure, and liver dysfunction. It is usually precipitated by stress, infection, or surgery in an unprepared patient, or by radio-iodine therapy. With careful management, it should no longer occur and most cases referred to as a 'crisis' are simply severe but uncomplicated thyrotoxicosis.

Treatment is urgent. Propranolol in full doses is started immediately together with potassium iodide, anti-thyroid drugs, corticosteroids (which suppress many of the manifestations of hyperthyroidism) and full supportive measures. Control of cardiac failure and tachycardia is also necessary.

Occasionally, hyperthyroidism can lead to a thyrotoxic cardiomyopathy, which causes ischaemic changes on a 12-lead ECG; these reverse after euthyroidism is achieved (**Fig. 34.24**).

Hyperthyroidism in pregnancy and neonatal life

This is covered in Chapter 48.

Thyroid hormone resistance

Thyroid hormone resistance is an inherited condition, caused by an abnormality of the thyroid hormone receptor. Mutations to the receptor (TR β) result in the need for higher levels of thyroid hormones to achieve the same intracellular effect. As a result, the normal feedback control mechanisms (see **Fig. 34.2**) result in high blood levels of T_4 with a normal TSH in order to maintain a euthyroid state. This has two consequences:

- Thyroid function tests appear abnormal, even when the patient is euthyroid and requires no treatment. Specialist review is needed to differentiate this condition from hyperthyroidism due to inappropriate TSH secretion.
- Different tissues contain different thyroid hormone receptors and, in some families, receptors in certain tissues may have normal activity. In this case, the level of thyroid hormones to maintain euthyroidism at pituitary and hypothalamic levels (which controls secretion of TSH) may be higher than that required in other tissues such as heart and bone, so these tissues may exhibit 'thyrotoxic' effects in spite of a normal serum TSH. This 'partial thyroid hormone resistance' can be very difficult to manage effectively.

Thyroid hormone resistance is diagnosed on the basis of a raised T_4/T_3 in the context of a non-elevated TSH. The differential diagnosis of this pattern of blood results includes TSHoma and laboratory assay interference. There is evidence that there may be increased cardiovascular mortality in patients with thyroid hormone resistance, similar to those with hyperthyroidism.

Long-term consequences of hyperthyroidism

Long-term follow-up studies of hyperthyroidism show a slight increase in overall mortality, which affects all age groups, is not fully explained, and tends to occur in the first year after diagnosis. Thereafter, the only long-term risk of adequately treated hyperthyroidism appears to be an increased risk of osteoporosis. People with persistently suppressed TSH levels have an increased likelihood of developing atrial fibrillation, which may predispose to thromboembolic disease.

Box 34.31 Choice of surgery or radio-iodine therapy for hyperthyroidism

Indications for surgery or radio-iodine

- Patient choice
- Persistent drug side-effects
- Poor compliance with drug therapy
- Recurrent hyperthyroidism after drugs

Indications for surgery

- A large goitre, which is unlikely to remit after anti-thyroid medication

Graves' orbitopathy (ophthalmopathy)

Pathophysiology

Graves' orbitopathy is due to a specific immune response that causes retro-orbital inflammation (**Fig. 34.25**). Swelling and oedema of the extraocular muscles lead to limitation of movement and to proptosis, which is usually bilateral but can sometimes be unilateral. Ultimately, increased pressure on the optic nerve may cause optic atrophy. Histology of the extraocular muscles shows focal oedema and glycosaminoglycan deposition followed by fibrosis. The precise autoantigen that leads to the immune response remains to be

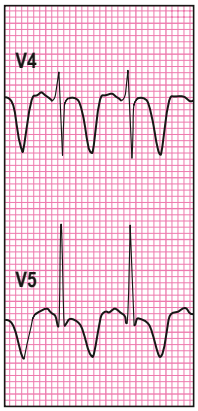


Fig. 34.24 Thyrotoxic cardiomyopathies. ECG showing marked lateral T-wave inversion.

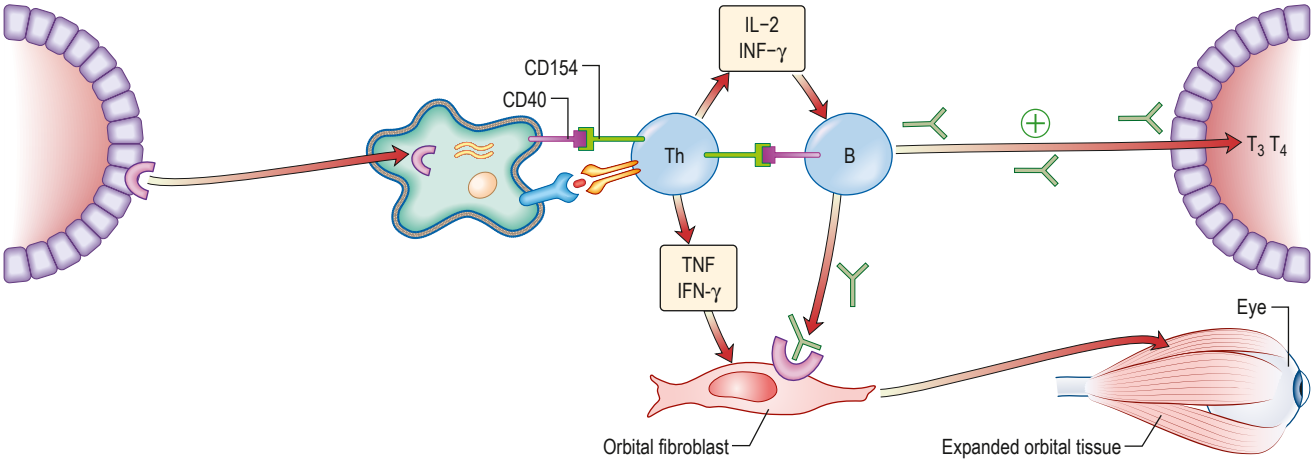


Fig. 34.25 Model of the initiation of thyrotrophin-receptor autoimmunity in Graves' ophthalmopathy and its consequences. Thyrotrophin receptor is internalized and degraded by antigen-presenting cells that present thyrotrophin-receptor peptides, in association with major histocompatibility complex (MHC) class II antigens, to helper T (Th) cells. These cells become activated, interact with autoreactive B cells through CD154–CD40 bridges, and secrete interleukin 2 (IL-2) and interferon-gamma (IFN-γ). These cytokines induce the differentiation of B cells into plasma cells that secrete anti-thyrotrophin-receptor antibodies. Anti-thyrotrophin-receptor antibodies *(green)* recognize the thyrotrophin receptor on orbital fibroblasts and, in conjunction with the secreted type 1 helper T cytokines, IFN-γ, and tumour necrosis factor (TNF), initiate the tissue changes characteristic of Graves' ophthalmopathy. These antibodies stimulate the thyrotrophin receptor on thyroid follicular epithelial cells, leading to hyperplasia and increased production of the thyroid hormones triiodothyronine (T_3) and thyroxine (T_4). *(Adapted from Bahn RS. Graves' ophthalmology.* N Engl J Med *2010; 362:726–738.)*

identified, but it appears to be an antigen in retro-orbital tissue with similar immunoreactivity to the TSH receptor.

Eye disease is a manifestation of Graves' disease and can occur in patients who may be hyperthyroid, euthyroid, or hypothyroid. Thyroid dysfunction and orbitopathy usually occur within 2 years of each other, although sometimes a gap of many years is seen. TSH receptor antibodies are almost invariably found in the serum; their role in the pathogenesis is becoming clearer (see Fig. 34.25). Orbitopathy is more common and more severe in smokers.

Clinical features

The clinical appearances are characteristic (see Fig. 34.23) but thyroid eye disease demonstrates a wide range of severity. A high proportion of people with Graves' disease notice some soreness, painful watering, or prominence of the eyes, and the 'stare' of lid retraction is relatively common. More severe proptosis occurs in a minority of cases, and limitation and discomfort of eye movement and visual impairment due to optic nerve compression are relatively uncommon. Proptosis and lid retraction may limit the ability to close the eyes completely so that corneal damage may occur. There is periorbital oedema, and conjunctival oedema and inflammation.

Eye manifestations do not parallel the degree of biochemical thyrotoxicosis or the need for anti-thyroid therapy, but exacerbation of eye disease is more common after radio-iodine treatment (15% versus 3% on anti-thyroid drugs). Sight is threatened in only 5%–10% of cases, but the discomfort and cosmetic problems cause great patient anxiety.

Investigations

If the appearance is characteristic, consideration is needed as to whether the patient requires review by an ophthalmologist. Investigation includes:

- TSH, fT_3, and fT_4 measurement.
- MRI or CT of the orbits to exclude retro-orbital space-occupying lesions. In Graves' orbitopathy, there may be enlarged muscles and oedema, with a taut optic nerve due to raised intra-orbital pressure.
- Assessment of vision, particularly optic nerve function.

Management

If the patient is thyrotoxic, this should be treated, but treatment will not directly result in an improvement of the ophthalmopathy; hypothyroidism must be avoided, as it may exacerbate the eye problem. Smoking should be stopped. Treatment of the eyes may be either local or systemic and always requires close liaison between specialist endocrinologist and ophthalmologist:

- ***Methylcellulose or hypromellose eyedrops*** or ointment are given to aid lubrication and improve comfort.
- ***Sleeping upright*** affords some patients relief.
- ***Taping the eyelids*** ensures closure at night.
- ***Systemic steroids*** (prednisolone 30–120 mg daily) usually reduce inflammation if more severe symptoms are present. Pulse intravenous methylprednisolone may be used initially and is more rapidly effective in severe cases.
- ***Surgical decompression of the orbit(s)*** may be required, particularly if pressure of orbital contents on the optic nerve threatens vision (often called a 'hot' decompression) and at a later, stable stage for cosmetic reasons ('cold' decompression).
- ***Lid surgery*** will protect the cornea if lids cannot be closed and can be useful later for cosmetic reasons.
- ***Corrective eye muscle surgery*** may improve diplopia due to muscle changes; it should be deferred until the situation has been stable for 6 months and should follow any orbital decompression.
- ***Irradiation of the orbits*** (20 Gy in divided doses) is used in some centres. This reduces inflammation and ocular motility but has little effect on proptosis; its precise role is debated.

- ***Immunomodulatory agents*** may produce a response in some patients when conventional treatments fail, although clinical trial evidence is inconsistent.
- ***Selenium*** may have a beneficial effect on mild orbitopathy in some patients.

Goitre (thyroid enlargement)

Goitre is more common in women than in men and may be either physiological or pathological.

Clinical features

Goitres are present on examination in up to 9% of the population. Most commonly, a goitre is noticed as a cosmetic defect by the patient or by friends or relatives. The majority are painless, but pain or discomfort can occur in acute varieties. Large goitres can produce dysphagia and difficulty in breathing, implying oesophageal or tracheal compression.

A small goitre may be more easily visible (on swallowing) than palpable. Clinical examination should record the size, shape, consistency, and mobility of the gland, as well as whether its lower margin can be demarcated (thus implying the absence of retrosternal extension). A bruit may be present. Associated lymph nodes should be sought and the tracheal position determined if possible. Examination should never omit an assessment of the patient's clinical thyroid status.

Specific enquiry should be made about any medication, especially iodine-containing preparations, and possible exposure to radiation.

Particular points of note are as follows:

- Puberty and pregnancy may produce a diffuse increase in size of the thyroid.
- Pain in a goitre may be caused by thyroiditis, bleeding into a cyst, or (rarely) a thyroid tumour.
- Excessive doses of carbimazole or PTU will induce goitre.
- Iodine deficiency and dyshormonogenesis (see earlier) can also cause goitre.

Diagnosis

There are four major aspects to any goitre: its pathological nature, whether it is causing any compressive symptoms, the patient's thyroid status, and whether it is of cosmetic concern. The nature can often be judged clinically. Goitres (**Box 34.32**) are usually separable into diffuse and nodular types, the causes of which differ.

Box 34.32 Goitre: causes and types

Diffuse
- Simple:
  - Physiological (puberty, pregnancy)
- Autoimmune:
  - Graves' disease
  - Hashimoto's disease
- Thyroiditis:
  - Acute (de Quervain's thyroiditis)
- Iodine deficiency (endemic goitre)
- Dyshormonogenesis
- Goitrogens (e.g., sulphonylureas)

Nodular
- Multinodular goitre
- Solitary nodular
- Fibrotic (Riedel's thyroiditis)
- Cysts

Tumours
- Adenomas
- Carcinoma
- Lymphomas

Miscellaneous
- Sarcoidosis
- Tuberculosis

Diffuse goitre

Simple goitre

In this instance, no clear cause is found for enlargement of the thyroid, which is usually smooth and soft. It may be associated with thyroid growth-stimulating antibodies.

Autoimmune thyroid disease

Hashimoto's thyroiditis and Grave's disease thyrotoxicosis are both associated with firm diffuse goitre of variable size. A bruit is often present in thyrotoxicosis.

Thyroiditis

Acute tenderness in a diffuse swelling, sometimes with severe pain, is suggestive of an acute viral thyroiditis (de Quervain's). It may produce transient clinical hyperthyroidism with an increase in serum T_4.

Nodular goitres

Multinodular goitre

Most common is the multinodular goitre, especially in older patients. The patient is usually euthyroid but may have clinical hyperthyroidism or subclinical hyperthyroidism (suppressed or low TSH levels but normal free T_4 and T_3). Multinodular goitre is the most common cause of tracheal and/or oesophageal compression and can lead to laryngeal nerve palsy. It may also extend retrosternally (**Fig. 34.26**).

The classical 'multinodular goitre' is usually readily apparent clinically, but it should be noted that modern, high-resolution ultrasound frequently reports multiple small nodules in glands, which are clinically diffusely enlarged and associated with autoimmune thyroid disease. These nodules are also found in up to 40% of the normal population.

Solitary nodular goitre

Such a goitre presents a difficult problem of diagnosis. Malignancy should be of concern in any solitary nodule; however, the majority of such nodules are cystic or benign and, indeed, may simply be the largest nodule of a multinodular goitre. The diagnostic challenge is to differentiate the small minority of malignant nodules, which require surgery, from the majority of benign nodules, which do not. Sometimes, a history of rapid enlargement, associated lymph nodes, or pain may suggest an aggressive malignancy, but most thyroid cancers are painless and slow-growing, so investigations are paramount. Risk factors for malignancy include previous irradiation, longstanding iodine deficiency, and occasional familial cases.

Solitary toxic nodules are quite uncommon and may be associated with T_3 toxicosis.

Fibrotic goitre (Riedel's thyroiditis)

Fibrotic goitre is a rare condition, usually producing a 'woody' gland. It is associated with other midline fibrosis and is often difficult to distinguish from carcinoma, being irregular and hard. Clinical clues include systemic symptoms of inflammation and elevation in inflammatory markers; it has been shown to be an IgG4-related disease (see Chapter 6).

Malignancy

In addition to thyroid carcinomas (see later), the thyroid is rarely the site of a metastatic deposit or the site of origin of a lymphoma.

Investigations

Clinical findings will dictate the appropriate initial tests:

- ***Thyroid function tests*** should include measurement of TSH plus free T_4 or T_3 (see **Box 34.7**).

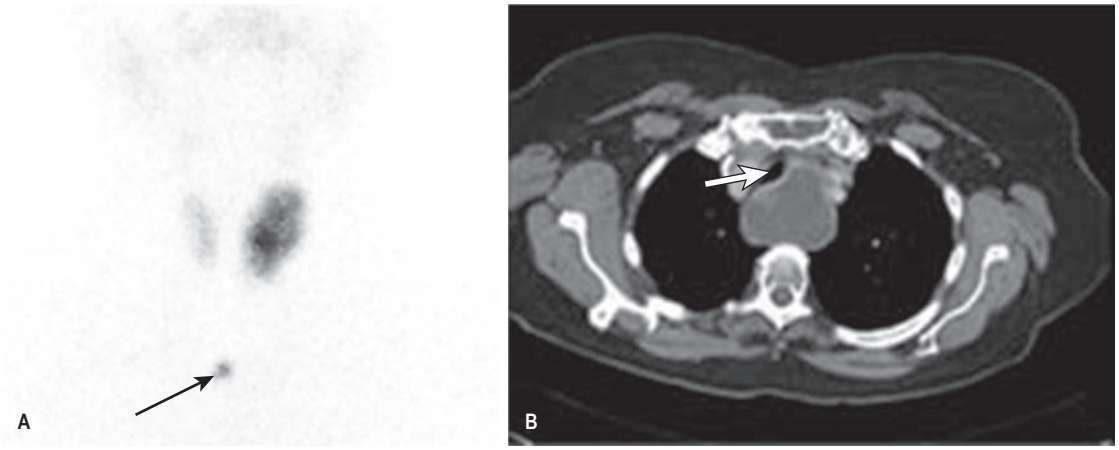


Fig. 34.26 Imaging of goitre. **(A)** Tc scan showing a multinodular goitre, with predominant involvement of the left lobe of the thyroid (anterior view with *arrowed* anatomical marker at the sternal notch). **(B)** Retrosternal goitre showing deviation and compression of the trachea *(arrowed)* on CT.

- ***Thyroid antibodies*** are assessed to exclude an autoimmune aetiology.
- ***Ultrasound*** with high resolution is a sensitive method for delineating nodules and can demonstrate whether they are cystic or solid. In addition, a multinodular goitre may be demonstrated when only a single nodule is palpable. Unfortunately, even cystic lesions can be malignant and thyroid tumours may arise within a multinodular goitre; therefore fine-needle aspiration (see below) is often required and is performed under ultrasound control at the same time as the scan.
- ***Chest and thoracic inlet X-rays or CT scan*** may detect tracheal compression and large retrosternal extensions in people with a very large goitre or clinical symptoms.
- ***Fine-needle aspiration (FNA)*** should be performed in the outpatient clinic or during ultrasound if the appearance is suggestive of potential malignancy, based on defined ultrasound criteria, as in people with a solitary nodule or a dominant nodule in a multinodular goitre, there is a 5% chance of malignancy. Cytology in expert hands can usually differentiate the suspicious or definitely malignant nodule. FNA reduces the necessity for surgery but there is a 5% false-negative rate, which must be borne in mind (and the patient appropriately counselled). Continued observation is required when an isolated thyroid nodule is assumed to be benign without excision.
- ***Thyroid scan*** (^{99m}Tc) can be useful to distinguish between functioning (hot) or non-functioning (cold) nodules. A hot nodule is only rarely malignant; however, a cold nodule is malignant in 10% of cases, and FNA has largely replaced isotope scans in the diagnosis of thyroid nodules.

Management

Euthyroid goitre

Many goitres are small, cause no symptoms, and can be observed (including self-monitoring by the patient in the long term). In particular, during puberty and pregnancy, a goitre associated with euthyroidism rarely requires intervention and the patient can be reassured that spontaneous resolution is likely. Indications for surgical intervention are:

- ***The possibility of malignancy.*** A positive or suspicious FNA makes surgery mandatory, and surgery may be necessary if doubt persists, even in the presence of a negative FNA (especially if the patient is concerned by the false-negative rate).
- ***Pressure symptoms on the trachea or, more rarely, the oesophagus.*** The possibility of retrosternal extension should be excluded.
- ***Cosmetic reasons.*** A large goitre often causes considerable anxiety to the patient, even though it is functionally and anatomically benign.

RAI has also been advocated for the treatment of euthyroid goitre, particularly when surgery is an unattractive option.

Toxic nodule

This is initially treated with anti-thyroid drugs but surgery or radioiodine is often required.

Thyroid cancer

Types of thyroid carcinoma, with their characteristics and treatment, are listed in **Box 34.33**. While not common, these tumours are responsible for 400 deaths annually in the UK and there is an annual incidence of 30 000 cases in the USA. Over 75% occur in women. They present in 90% of cases as thyroid nodules (see above), but occasionally with cervical lymphadenopathy (about 5%), or with lung, cerebral, hepatic, or bone metastases.

Carcinomas derived from thyroid epithelium may be papillary or follicular (differentiated) or anaplastic (undifferentiated). Medullary carcinomas (about 5% of all thyroid cancers) arise from the calcitonin-producing C cells. The pathogenesis of thyroid epithelial carcinomas is not understood, except for occasional familial papillary carcinoma, and those cases related to previous head and neck irradiation or ingestion of RAI (e.g., post-Chernobyl). These tumours are minimally active hormonally and are extremely rarely associated with hyperthyroidism; over 90%, however, secrete thyroglobulin, which can therefore act as a tumour marker after thyroid ablation.

Papillary and follicular carcinomas

The primary treatment is surgical: normally total or near-total thyroidectomy for local disease. Regional or more extensive neck dissection is needed where there is local nodal spread or involvement of local structures.

Most tumours will take up iodine, and UK and other guidelines currently recommend RAI ablation of residual thyroid tissue postoperatively for most people with differentiated thyroid cancer. After ablation of normal thyroid in this way, RAI may be used to localize

residual disease (scanning using low doses) or to treat it (using high doses: 5.5–7.5 GBq). When disease does recur, local invasion and lymph node involvement is most common, and lungs and bone are the most common sites of distant metastases.

To minimize risk of recurrence patients are treated with suppressive doses of levothyroxine (sufficient to suppress TSH levels below the normal range). Patient progress is monitored, both clinically and biochemically, using serum thyroglobulin levels as a tumour marker. The measurement of thyroglobulin is most sensitive when TSH is high but this requires the withdrawal of levothyroxine therapy. Recombinant TSH (thyrotropin alpha, rhTSH) 900 μg (2 doses over 48 h) is used to stimulate thyroglobulin without stopping levothyroxine therapy. Detectable thyroglobulin suggests recurrence, in which case whole-body ^{131}I scanning is required. Unfortunately, the presence of anti-thyroglobulin antibodies can make the assay unreliable.

The ***prognosis*** is extremely good when differentiated thyroid cancer is excised while confined to the thyroid gland, and the specific therapies available lead to a relatively good prognosis even in the presence of metastases at diagnosis. Accepted markers of high risk include greater age (>40 years), larger primary tumour size (>4 cm), and macroscopic invasion of capsule and surrounding tissues. Recently, oral sorafenib has shown a good short-term response in locally advanced and metastatic differentiated thyroid cancers that were refractory to RAI.

Medullary carcinoma

Medullary thyroid carcinoma (MTC) is a neuroendocrine tumour of the calcitonin-producing C cells of the thyroid. This condition is often associated with MEN 2 (see 'Diseases of many glands' section). Approximately 25% of patients diagnosed with MTC have a mutation of the *RET* proto-oncogene, although the other manifestations of MEN 2 may be absent, hence the importance of genetic counselling and family screening. People with MEN 2 mutations are advised to have a prophylactic thyroidectomy as early as 5 years of age to prevent the development of MTC.

Total thyroidectomy and wide lymph node clearance are usually indicated in MTC. Local invasion or metastasis is frequent, and the tumour responds poorly to treatment, although progression is often slow. Biological therapy with vandetanib and cabozantinib has shown benefit in advanced medullary thyroid carcinoma.

Box 34.33 Types of thyroid malignancy

Cell type	Frequency	Behaviour	Spread	Prognosis
Papillary	70%	Occurs in young people	Local, sometimes lung/bone secondaries	Good, especially in young
Follicular	20%	More common in females	Metastases to lung/bone	Good if resectable
Medullary cell	5%	Often familial	Local and metastases	Poor, but indolent course
Anaplastic	<5%	Aggressive	Locally invasive	Very poor
Lymphoma	2%	Variable		Sometimes responsive to radiotherapy

Anaplastic carcinomas and lymphoma

These typically do not respond to RAI, and external radiotherapy produces only a brief respite, although molecular-based treatments are proving promising.

Further reading

Burman KD, Wartofsky L. Thyroid nodules. *N Engl J Med* 2015; 373:2347–2356.
Franklyn JA, Boelaert K. Subclinical thyroid disease: where is the evidence? *Lancet Diabetes Endocrinol* 2013; 1:172–173.
Haugen BR, Wartofsky L. American Thyroid Association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer. *Thyroid* 2016; 26:1–133.
Ross DS. Radioiodine therapy for hyperthyroidism. *N Engl J Med* 2011; 364:542–550.
Thyroid Cancer: Assessment and Management. *NICE guideline* [NG230]. Published: 19 December 2022.
Vasileiou M, Gilbert J, Fishburn S et al. Guideline Committee. Thyroid disease assessment and management: summary of NICE guidance. *BMJ* 2020; 368:m41.
Wiersinga WM, Poppe KG, Effraimidis G. Hyperthyroidism: aetiology, pathogenesis, diagnosis, management, complications, and prognosis. *Lancet Diabetes Endocrinol* 2023; 11(4):282–298.

HYPOTHALAMO–PITUITARY–GONADAL AXIS

The hypothalamo–pituitary–gonadal axis controls progress through puberty and the capacity for reproduction. Terminology in reproductive medicine is shown in **Box 34.34**.

Anatomy and physiology

Embryology

Up to 8 weeks of gestation, the sexes share a common development, with a primitive genital tract including the Wolffian and

Box 34.34 Definitions in reproductive medicine

Term	Definition
Erectile dysfunction	Inability of the male to achieve or sustain an erection adequate for satisfactory intercourse
Azoospermia	Absence of sperm in the ejaculate
Oligospermia	Reduced numbers of sperm in the ejaculate
Libido	Sexual interest or desire; often difficult to assess and greatly affected by stress, tiredness and psychological factors
Menarche	Age at first period
Primary amenorrhoea	Failure to begin spontaneous menstruation by age 16
Secondary amenorrhoea	Absence of menstruation for 3 months in a woman who has previously had cycles
Oligomenorrhoea	Irregular long cycles; often used for any length of cycle above 35 days
Dyspareunia	Pain or discomfort in the female during intercourse
Menstruation	Onset of spontaneous (usually regular) uterine bleeding in the female
Virilization	Occurrence of male secondary sexual characteristics in the female

Müllerian ducts. Additionally, there are a primitive perineum and primitive gonads.

- In the presence of a Y chromosome, the potential testis develops while the ovary regresses.
- In the absence of a Y chromosome, the potential ovary develops and related ducts form a uterus and the upper vagina.

Production of Müllerian inhibitory factor from the early 'testis' produces atrophy of the Müllerian duct, while, under the influence of testosterone and dihydrotestosterone, the Wolffian duct differentiates into an epididymis, vas deferens, seminal vesicles, and prostate. Androgens induce transformation of the perineum to include a penis, penile urethra, and scrotum containing the testes, which descend in response to androgenic stimulation. At birth, testicular volume is 0.5–1 mL. The number of oocytes in the ovaries of a female fetus has reached a maximum by the end of the second trimester of *in utero* development. At birth, the ovaries contain about 3 million oocytes, of which only 400 000 will remain by the time puberty occurs. With the onset of menarche, these oocytes will activate and grow, leading to ovulation and then menstruation.

Physiology

An outline of the hypothalamic–pituitary–gonadal axis is shown in Fig. 34.27.

Pulses of GnRH are released from the hypothalamus and stimulate LH and FSH release from the pituitary. LH and FSH are composed of two glycoprotein chains (α and β subunits). The α subunits are identical and are shared with TSH, while the β subunit confers specific biological activity (Box 34.35)

The male

1. LH stimulates testosterone production from Leydig cells of the testis.
2. Testosterone acts via nuclear androgen receptors, which interact with co-regulatory proteins to produce the appropriate tissue responses: progress through puberty, anabolism, and the maintenance of libido. It also acts locally within the testis to aid spermatogenesis. Testosterone circulates largely bound to sex-hormone-binding globulin (SHBG). Testosterone feeds back on the hypothalamus/pituitary to inhibit GnRH secretion.
3. FSH stimulates the Sertoli cells in the seminiferous tubules to produce mature sperm and the inhibins A and B.
4. Inhibin feeds back to the pituitary to decrease FSH secretion. Activin, a related peptide, counteracts inhibin.

The female

Female physiology is more complex (see Figs. 34.27 and 34.28).

1. During puberty, in response to pulses of hypothalamic GnRH, the pituitary secretes LH and FSH, which stimulate androgen and oestrogen production from the ovaries resulting in progress through puberty followed by menarche and the establishment of the menstrual cycle.
2. The menstrual cycle depends upon the activity of hypothalamic GnRH. Pulses of GnRH, at about 2-hour intervals, stimulate release of pituitary LH and FSH. The KNDy neurons in the hypothalamus control GnRH secretion. Kisspeptin and neurokinin B are stimulatory and dynorphin A is inhibitory.
3. Menstrual cycles are often irregular during adolescence, particularly the interval from the first cycle to the second cycle as immaturity of the hypothalamic–pituitary–ovarian axis during the early years after menarche often results in anovulation and cycles may be somewhat long (90% of cycles: 21–45 days). By the third year after menarche, 60%–80% of menstrual cycles are 21–34 days long, as is typical of adults.

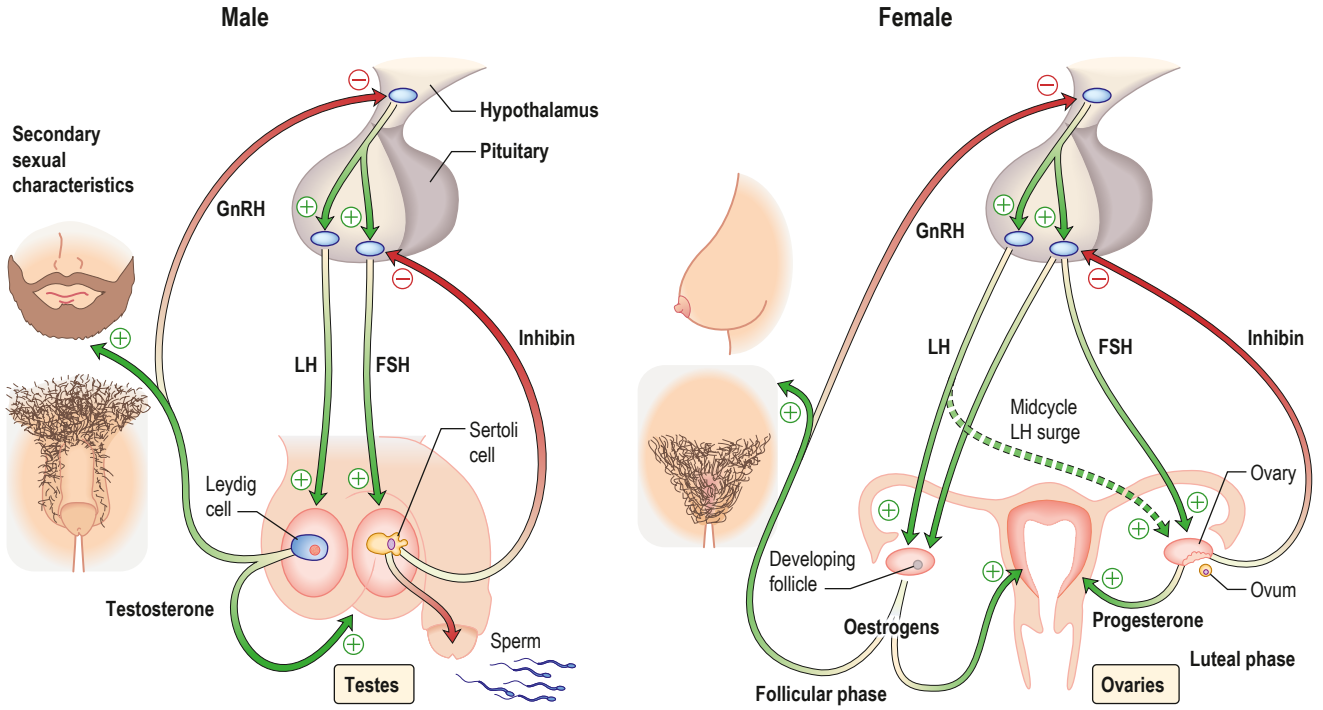


Fig. 34.27 Male and female hypothalamic–pituitary–gonadal axes. Luteinizing hormone (LH) and follicle-stimulating hormone (FSH) are secreted in pulses in response to hypothalamic gonadotrophin-releasing hormone (GnRH) and have different effects on the gonads (see text). Testosterone and oestrogens have multiple peripheral effects and exert negative feedback on LH secretion. Inhibin exerts negative feedback on FSH secretion.

4. LH stimulates ovarian androgen production by the ovarian theca cells.
5. FSH stimulates follicular development and activity of aromatase (an enzyme required to convert ovarian androgens to oestrogens) in the ovarian granulosa cells. FSH also stimulates release of inhibin from ovarian stromal cells, which inhibits FSH release. Activin counteracts inhibin (see Fig. 34.27).
6. Although many follicles are 'recruited' for development in early folliculogenesis, by day 8–10, a 'leading' (or 'dominant') follicle is selected for development into a mature Graafian follicle.
7. Oestrogens have a double feedback action on the pituitary (see Fig. 34.27). Initially, they inhibit gonadotrophin secretion (negative feedback) but, later, high-level exposure results in increased GnRH secretion and increased LH sensitivity to GnRH (positive feedback), which leads to the mid-cycle LH surge, inducing ovulation from the leading follicle (see Fig. 34.28).
8. The follicle then differentiates into a corpus luteum, which secretes both progesterone and oestradiol during the second half of the cycle (luteal phase).
9. Oestrogen (initially) and then progesterone cause uterine endometrial proliferation in preparation for possible implantation; if implantation does not occur, the corpus luteum regresses and progesterone secretion and inhibin levels fall so that the endometrium is shed (menstruation), allowing increased GnRH and FSH secretion.
10. If implantation and pregnancy follow, HCG production from the trophoblast maintains corpus luteum function until 10–12 weeks of gestation, by which time the placenta will be making sufficient oestrogen and progesterone to support itself.

Clinical features of disorders of sex and reproduction

A detailed history and examination of all systems are required (Box 34.36). A man having regular satisfactory intercourse or a woman with regular ovulatory periods is most unlikely to have significant endocrine disease, assuming the history is accurate.

Investigation of gonadal function

In adults much can be deduced from basal measurements of the gonadotrophins, oestrogens/testosterone, and prolactin:

- High gonadotrophins with low testosterone or oestradiol indicate primary gonadal disease.
- Low levels of testosterone/oestradiol with low or normal LH/FSH imply HP disease.
- Demonstration of ovulation (by measurement of luteal phase serum progesterone and/or by serial ovarian ultrasound in the follicular phase) or a healthy sperm count (20–200 million/mL, >60% grade I motility and <20% abnormal forms) provides absolute confirmation of normal female or male reproductive endocrinology, but these tests are not always essential.
- Pregnancy provides complete demonstration of normal male and female function.
- Hyperprolactinaemia can be confirmed or excluded by direct measurement. Levels may increase with stress; if this is suspected, a cannula should be inserted and samples taken through it 30 minutes later. More detailed tests are indicated in Box 34.37.

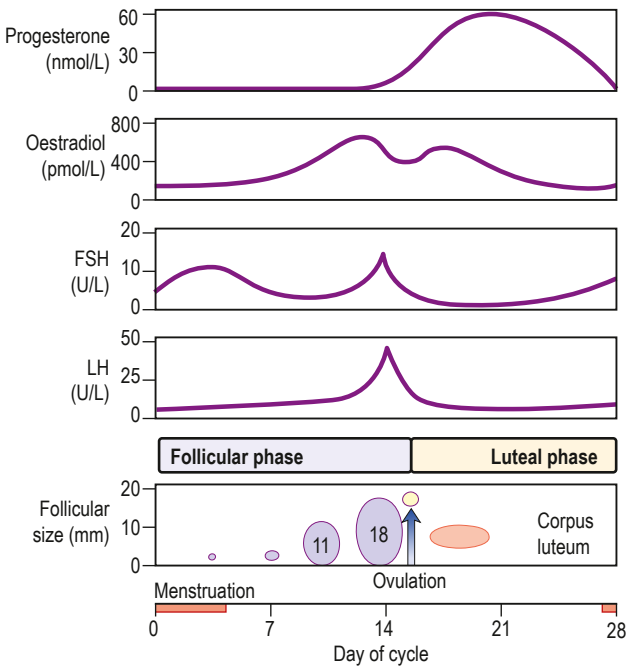


Fig. 34.28 Hormonal and follicular changes during the normal menstrual cycle. *FSH,* Follicle-stimulating hormone; *LH,* luteinizing hormone.

Box 34.35 Hormones and receptors of the hypothalamic–pituitary–gonadal axis

Hormone	Source	Site of action	Hormone structure	Receptor	Post-receptor activity
Gonadotrophin-releasing hormone (GnRH; LHRH)	Hypothalamus	Pituitary	Peptide—10 AA	Membrane—7TM GnRH-R1	G-proteins PLC/IP_3
Luteinizing hormone (LH)	Pituitary	Gonad	Glycoprotein: α and β subunits	Membrane—7TM LHhCGR	G-proteins cAMP
Follicle-stimulating hormone (FSH)	Pituitary	Gonad	Glycoprotein: α and β subunits	Membrane—7TM FSHR	G-proteins cAMP
Oestradiol	Ovary	Uterus, breast bone, vascular	Steroid ring	Nuclear ER α and β (ESR1/ESR2)	Homo-/heterodimer ERE Transcription
Testosterone	Testis	Many tissues	Steroid ring	Nuclear AR (NR3C4)	Dimer ARE Transcription
Inhibin and activin	Gonad	Pituitary–hypothalamus	Peptide dimers α and β subunits	Transmembrane dimerized	Phosphorylation by receptor

7TM, 7 transmembrane (G-protein coupled receptor); *AA,* amino acids; *ARE,* androgen response element; *cAMP,* adenylate cyclase → cyclic adenosine monophosphate; *ERE,* oestrogen response element; *LHhCGR,* luteinizing hormone human chorionic gonadotrophin receptor; *LHRH,* luteinizing hormone-releasing hormone; *PLC/IP_3,* phospholipase C/inositol triphosphate.

Box 34.36 Sexual and menstrual disorders

History
- Menstruation—timing of bleeding and cycle
- Relationship of symptoms to cycle
- Breasts (tenderness/galactorrhoea)
- Hirsutism and acne
- Libido and potency
- Problems with intercourse
- Past fertility and future plans

Physical signs
- Evidence of systemic disease
- Secondary sexual characteristics
- Extent/distribution of hair
- Genital size (testes, ovaries, uterus)
- Clitoromegaly
- Breast development, gynaecomastia
- Galactorrhoea

Box 34.37 Tests of gonadal function

Test	Uses/comments
Male	
Basal testosterone	Normal levels exclude hypogonadism
Sperm count	Normal count excludes deficiency Motility and abnormal sperm forms should be noted
Female	
Basal oestradiol	Normal levels exclude hypogonadism
Luteal phase progesterone (days 18–24 of cycle)	If >30 nmol/L, suggests ovulation
Ultrasound of ovaries	To confirm ovulation
Both sexes	
Basal LH/FSH	Demonstrates state of feedback system for hormone production (LH) and germ cell production (FSH)
HCG test (testosterone or oestradiol measured)	Response shows potential of ovary or testis; failure demonstrates primary gonadal problem
Clomifene test (LH and FSH measured)	Tests hypothalamic negative feedback system; clomifene is an oestrogen antagonist and causes LH/FSH to rise
LHRH test (rarely used)	Shows adequacy (or otherwise) of LH and FSH stores in pituitary

FSH, Follicle-stimulating hormone; *HCG,* human chorionic gonadotrophin; *LH,* luteinizing hormone; *LHRH,* luteinizing hormone-releasing hormone.

Box 34.38 Effects of androgens and consequences of androgen deficiency in the male

Physiological effect	Consequences of deficiency
General	
Maintenance of libido	Loss of libido
Deepening of voice	High-pitched voice (if prepubertal)
Frontotemporal balding	No temporal recession
Facial, axillary, and limb hair	Decreased hair
Maintenance of erectile and ejaculatory function	Loss of erections/ejaculation
Pubic hair	
Maintenance of male pattern	Thinning and loss of pubic hair
Testes and scrotum	
Maintenance of testicular size/consistency (needs gonadotrophins as well)	Small, soft testes
Rugosity of scrotum	Poorly developed penis/scrotum
Stimulation of spermatogenesis	Subfertility
Musculoskeletal	
Epiphyseal fusion	Eunuchoidism (if prepubertal)
Maintenance of muscle bulk and power	Decreased muscle bulk
Maintenance of bone mass	Osteoporosis

Box 34.39 Causes of male hypogonadism

- Reduced gonadotrophins (hypothalamic–pituitary disease):
  - Hypopituitarism
  - Selective gonadotrophic deficiency, Kallmann's syndrome, normosmic idiopathic hypogonadotropic hypogonadism
  - Severe systemic illness
  - Severe underweight
- Hyperprolactinaemia
- Primary gonadal disease (congenital):
  - Anorchia/Leydig cell agenesis
  - Cryptorchidism (testicular maldescent)
  - Chromosome abnormality (e.g., Klinefelter's syndrome)
  - Enzyme defects: 5α-reductase deficiency
- Primary gonadal disease (acquired):
  - Testicular torsion
  - Orchidectomy
  - Local testicular disease
  - Chemotherapy/radiation toxicity
  - Orchitis (e.g., mumps)
  - Chronic kidney disease
  - Cirrhosis/alcohol
  - Sickle cell disease
- Androgen receptor deficiency/abnormality

Disorders in the male

Hypogonadism

History and examination

How a male presents depends on age. Hypogonadism in an adolescent would present with delayed puberty and/or growth, whereas an adult may present with poor libido, erectile dysfunction, and loss of secondary sexual hair (Box 34.38). Hypogonadism may also be an incidental finding, such as one that emerges during investigation for subfertility. Semen volume is usually maintained as sperm makes up for only a very small proportion of seminal fluid volume; most is prostatic fluid. The testes may be small and soft, and there may be gynaecomastia. Except with subfertility, the symptoms are usually those of androgen deficiency.

Causes of male hypogonadism are shown in Box 34.39.

Investigations

Hypogonadism may be immediately apparent but basal levels of testosterone, LH, and FSH should be measured. These will allow the distinction to be made between primary gonadal (testicular) failure and HP disease.

If primary gonadal failure is identified, chromosomal analysis (e.g., to exclude Klinefelter's syndrome) is required. In secondary gonadal failure, gonadotrophin deficiency, pituitary MRI scan, prolactin levels, and other pituitary function tests are needed.

A frequent cause of referral is males with poor libido or erectile dysfunction with equivocal lowering of serum testosterone (7–10 nmol/L) without elevation of gonadotrophins. If the testosterone was taken in the morning, such test results can be interpreted as mild gonadotrophin deficiency or lower end of the normal but may also be seen in people with obesity and diabetes mellitus, who commonly have low circulating SHBG levels associated with insulin

Box 34.40 Androgen replacement therapy

Route	Preparation	Dose	Remarks
Dermal	Testosterone gel Testosterone patch	50–100 mg 300 μg/24 h	Rubs on shoulders Self-adhesive
Intramuscular	Testosterone enanthate Testosterone undecanoate Testosterone propionate	250 mg every 3 weeks 1 g every 3 months 50–100 mg every 2–3 weeks	Frequent injections Painful Large-volume injection Very frequent injections Short half-life
Oral	Testosterone undecanoate	120–160 mg daily, in divided doses	Variable dose, irregular absorption
Implant	Testosterone implant	600 mg every 4–5 months	Requires implant procedure Scarring, infection

resistance and therefore low total testosterone levels but normal free testosterone levels. Regular use of opiate analgesia may lead to gonadotrophin deficiency and should be considered. 'Anabolic' steroid (i.e., androgen) abuse causes similar biochemical findings and is likely if the patient appears well virilized. A therapeutic trial of testosterone replacement is often justified and forms part of the investigation in some patients; full pituitary evaluation may be required in such cases to exclude other pituitary disease.

Management

The cause of hypogonadism can rarely be reversed and testosterone replacement therapy should be commenced to control current symptoms and prevent osteoporosis in the long term. Replacement is usually given by transdermal gel or by intramuscular injection (**Box 34.40**). In gonadotrophin deficiency, LH and FSH (purified or synthetic) or pulsatile GnRH may be used when fertility is required.

Special instances of hypogonadism

Cryptorchidism

Cryptorchidism (or undescended testes) is usually treated by surgical exploration and orchidopexy in early childhood. After that age, the germinal epithelium is increasingly at risk, and lack of descent by puberty is associated with subfertility. A short trial of HCG occasionally induces descent; an HCG test with a testosterone response 72 hours later excludes anorchia. Intra-abdominal testes have an increased risk of developing malignancy; if presentation is after puberty, orchidectomy is advised. Patients may present in adulthood with a history of cryptorchidism with primary testicular failure (due to the testicular damage before or during surgery) or gonadotrophin deficiency (presumably the initial cause of maldescent).

Klinefelter's syndrome

Klinefelter's syndrome is a common chromosomal abnormality, affecting 1 in 1000 males. Patients are phenotypically male, but with an extra chromosome (47XXY and variants, e.g., 46XY/47XXY mosaicism). There is both a loss of Leydig cells and seminiferous tubular dysgenesis. If not identified on amniocentesis, only 10% are diagnosed prepuberty; early indications of the condition include cryptorchidism, behavioural problems, and learning disabilities as well as tall stature. In adolescence, there may be gynaecomastia and micro-orchidism (mostly from the age of 14 years), as well as a poor muscular bulk with excessive long legs (due to lack of epiphyseal closure); small shoulders and broad hips should raise suspicion. Most patients have a normal puberty. In adulthood, infertility is one of the most common presentations. There is also a predisposition to diabetes mellitus, breast cancer, emphysema, and bronchiectasis; these are all unrelated to the testosterone deficiency.

Clinical examination shows a wide spectrum of features with small, pea-sized but firm testes, usually gynaecomastia and other signs of ***androgen deficiency***. Confirmation is by chromosomal analysis.

Treatment is androgen replacement therapy if deficient in testosterone. The majority of men with Klinefelter's syndrome have relative testosterone deficiency (low-normal levels). The evidence that testosterone replacement in these men has positive benefits is lacking but should be discussed and trialled. No treatment is possible for the abnormal seminiferous tubules and infertility, although advances in fertility techniques and microdissection testicular sperm extraction ***(microTESE)*** offer some hope to men wanting to father their own child.

Kallmann's syndrome

This is isolated GnRH deficiency. It is associated with decreased or absent sense of smell (anosmia) and sometimes with other bony (cleft palate), renal, and cerebral abnormalities (e.g., colour blindness). It is often familial and is usually X-linked, resulting from a mutation in the *KAL1* gene, which encodes anosmin-1 (producing loss of smell); one sex-linked form is due to an abnormality of a cell adhesion molecule. Management is that of secondary hypogonadism. Fertility is possible with gonadotrophin therapy.

Normosmic idiopathic hypogonadotropic hypogonadism

This refers to isolated GnRH deficiency in the absence of anosmia. Known mutations account for less than 15% of normosmic idiopathic hypogonadotropic hypogonadism (nIHH). Mutations include the *KISS1* gene, which codes for kisspeptin, the protein that acts on the GPR54 receptor, and the *FGFR1* gene.

Oligospermia and azoospermia

These may be secondary to gonadotrophin deficiency and can be corrected by gonadotrophin therapy. More often, they result from primary testicular diseases, in which case they are rarely treatable, although advances in fertility techniques and microdissection testicular sperm extraction (microTESE) offer some hope to men with primary hypogonadism wanting to father their own child.

Azoospermia with normal testicular size and low FSH levels suggests a vas deferens block, which is sometimes reversible by surgical intervention.

Gynaecomastia

Gynaecomastia is development of breast tissue in the male. Causes are shown in **Box 34.41**. It is due to an imbalance between free oestrogen and free androgen effects on breast tissue.

Box 34.41 Causes of gynaecomastia

- Physiological:
  - Neonatal
  - Pubertal
  - Old age
- Hyperthyroidism
- Hyperprolactinaemia
- Renal disease
- Liver disease
- Hypogonadism (see Box 34.39)
- Oestrogen-producing tumours (testis, adrenal)
- HCG-producing tumours (testis, lung)
- Starvation/re-feeding
- Carcinoma of breast
- Drugs:
  - Oestrogenic: oestrogens, cannabis, digoxin, diamorphine
  - Antiandrogens: spironolactone, cimetidine, cyproterone
  - Others: gonadotrophins, cytotoxics

Pubertal gynaecomastia

This occurs in perhaps 50% of normal boys, often asymmetrically. It usually resolves spontaneously within 6–18 months, but after this duration, it may require surgical removal, as fibrous tissue will have been laid down. The cause is thought to be relative oestrogen excess, and the oestrogen antagonist tamoxifen is occasionally helpful.

Gynaecomastia in the older male

This requires a full assessment to exclude potentially serious underlying disease, such as bronchial carcinoma and testicular tumours (e.g., Leydig cell tumour). However, aromatase activity increases with age and may be the cause of gynaecomastia in this group. Aromatase is an enzyme of the cytochrome P450 family and converts androgens to produce oestrogens. Drug effects are common (especially digoxin and spironolactone), and once these and significant liver diseases are excluded, most cases have no definable cause. Painful gynaecomastia may be treated with a 3–6-month trial of tamoxifen, and surgery is occasionally necessary if the symptoms or cosmetic appearance are unacceptable.

The ageing male

In the male, there is no sudden menopause or 'change of life'. However, there is a progressive loss of sexual function with reduction in morning erections and frequency of intercourse.

The age of onset varies widely. Typically, overall testicular volume diminishes and SHBG and gonadotrophin levels gradually rise, but other men present with low or borderline testosterone without elevation of LH/FSH. Low testosterone certainly increases the risk of osteoporosis and, in some studies, is associated with increased cardiovascular risk. It remains unclear to what extent general symptoms of lack of energy, drive, muscle strength, and general well-being may relate to these hormonal changes, but a recent trial of testosterone replacement showed no benefit in general symptoms, although there was an improvement in sexual function. Loss of libido and erectile dysfunction are, however, common symptoms, even when hormones are normal, and long-term outcome studies of testosterone replacement are still awaited. Therefore, the decision to offer testosterone replacement to an ageing male is currently based on full clinical and biochemical assessment and full discussion of potential risks (including prostate disease) as well as benefits. If testosterone is unequivocally low (<7 pmol/L) and there are symptoms specific to androgen deficiency (low libido, erectile dysfunction and loss of early morning erections), most authorities would recommend replacement. However, few would treat if testosterone is higher than 12 pmol/L with normal LH/FSH. Clinically, a large proportion of cases are in the borderline range (7–12 pmol/L), which can lead to difficulties in reaching a firm diagnosis. There is an increasing move towards measuring fasting testosterone levels, as food intake may decrease testosterone, leading to an incorrect diagnosis of androgen deficiency.

Box 34.42 Effects of oestrogens and consequences of oestrogen deficiency

Physiological effect	Consequences of deficiency
Breast	
Development of connective and duct tissue	Small, atrophic breast
Nipple enlargement and areolar pigmentation	
Pubic hair	
Maintenance of female pattern	Thinning and loss of pubic hair
Vulva and vagina	
Vulval growth	Atrophic vulva
Vaginal glandular and epithelial proliferation	Atrophic vagina
Vaginal lubrication	Dry vagina and dyspareunia
Uterus and tubes	
Myometrial and tubal hypertrophy	Small, atrophic uterus and tubes
Endometrial proliferation	Amenorrhoea
Skeletal	
Epiphyseal fusion	Eunuchoidism (if prepubertal)
Maintenance of bone mass	Osteoporosis

Disorders in the female

Hypogonadism

Impaired ovarian function, whether primary or secondary, will lead to both oestrogen deficiency and abnormalities of the menstrual cycle. The latter is very sensitive to disruption, with cycles becoming anovulatory and irregular before disappearing altogether. Symptoms will depend on the age at which failure develops. Thus, before puberty, primary amenorrhoea will occur, possibly with delayed puberty; after puberty, there may be secondary amenorrhoea and symptoms of oestrogen deficiency.

Oestrogen deficiency

The physiological effects of oestrogens and symptoms/signs of deficiency are shown in Box 34.42.

Amenorrhoea

Absence of periods or markedly irregular infrequent periods (oligomenorrhoea) is the most common presentation of female gonadal disease. The clinical assessment of such patients is shown in Box 34.43, and common causes are listed in Box 34.44. Causes can be divided into whether amenorrhoea is primary or secondary, whether associated with oestrogen deficiency or not, whether oestrogen deficiency is due to gonadotrophin deficiency or ovarian insufficiency, whether it is associated with another condition, and whether amenorrhoea is due to structural problems.

Polycystic ovary syndrome

Polycystic ovary syndrome (PCOS) is an example of amenorrhoea that could either be primary or secondary and is not associated with

Box 34.43 Clinical assessment of amenorrhoea

History
- Ensure not pregnant
- Date of onset
- Age of menarche, if any
- Sudden or gradual onset
- General health
- Weight—absolute and changes in recent past
- Stress (job, lifestyle, examinations, relationships)
- Excessive exercise
- Drugs
- Hirsutism, acne, virilization
- Headaches/visual symptoms
- Sense of smell
- Past history of pregnancies
- Past history of gynaecological surgery

Examination
- General health
- Body shape and skeletal abnormalities
- Weight and height
- Hirsutism and acne
- Evidence of virilization
- Maturity of secondary sexual characteristics
- Galactorrhoea
- Normality of vagina, cervix and uterus

oestrogen deficiency. It is the most common cause of oligomenorrhoea and amenorrhoea (see below).

Hypothalamic and weight-related amenorrhoea

Amenorrhoea with low oestrogen and gonadotrophins in the absence of organic pituitary disease is described as hypothalamic amenorrhoea. This may be related to 'stress'; low body weight; excessive exercise or previous weight loss; stopping the contraceptive pill; or severe illness. However, some patients appear to have defective cycling mechanisms without apparent explanation.

A minimum body weight is necessary for regular menstruation. While anorexia nervosa represents an extreme form of weight loss (see Chapter 30, 'Anorexia nervosa' section), amenorrhoea is common and may also be seen at weights within the 'normal' range. It is possible that alterations in leptin levels are responsible for the hypothalamic dysfunction seen in this situation. Restoration of body weight to above the 50th centile for height is usually effective in restoring menstruation, but in the many cases where this cannot be achieved oestrogen replacement is necessary.

Premature ovarian insufficiency

Premature ovarian insufficiency is the preferred term to describe premature menopause or ovarian failure of onset before the age of 40. The cause may be autoimmune or may rarely be identifiable genetic causes such as the fragile X pre-mutation or Turner's syndrome; most commonly, however, it is of unknown aetiology, although often familial.

Repeat measurements of LH/FSH levels are necessary before a diagnosis of premature ovarian insufficiency is given because of the psychological impact of this diagnosis and the possibility that a single elevation of LH/FSH might simply be the mid-cycle ovulatory surge. Iatrogenic causes, such as bilateral oophorectomy or cancer therapy with radiotherapy and some chemotherapy agents, can also cause premature ovarian insufficiency. In some scenarios, ovarian function can fluctuate; therefore, in younger women, consideration needs to be given as to whether contraception is required and whether the combined oral contraceptive pill is the best option for hormone replacement therapy (HRT).

Turner's syndrome (caused by presence of a single X chromosome) is a cause of premature ovarian insufficiency, frequently with delayed puberty and primary amenorrhoea, although some girls do undergo menarche and establish a menstrual cycle but remain at risk of premature ovarian insufficiency. The phenotype is female with female external genitalia. There is gonadal dysgenesis with streak ovaries. Features include short stature, webbing of the neck (up to 40%), a wide carrying angle of the elbows, high-arched palate, and low-set ears. These patients also have an increased incidence of autoimmune disease (2%), bicuspid aortic valves, aortic coarctation and dissection, coronary artery disease, hypertension, type 2 diabetes, horseshoe kidneys, lymphoedema, reduced bone density, hearing problems, and inflammatory bowel disease.

Investigation of oligo-amenorrhoea

Basal levels of FSH, LH, oestrogen, and prolactin allow initial distinction between primary gonadal and HP causes (see **Box 34.44**). Elevation of LH and FSH to menopausal levels usually confirms the diagnosis. Subsequent investigations are also shown in **Box 34.44**.

Anti-Müllerian hormone (AMH) released from granulosa cells of preantral and small antral follicles in the ovary, a marker of *'ovarian reserve'*, is being increasingly used in the context of secondary amenorrhoea, especially in infertility centres: it is elevated in PCOS, low in premature ovarian insufficiency and after menopause, and either normal or low in hypogonadotropic hypogonadism. It is shown to be helpful in deciding the type of infertility treatment strategy to be adopted, specifically in premature ovarian insufficiency or hypogonadotropic hypogonadism cases where low AMH level indicates poor follicular reserve, when oocyte donation is recommended. Conversely, adequate AMH level means there is hope for oocyte retrieval following gonadotrophin stimulation protocols to facilitate *in vitro* fertilization (IVF).

Management

Treatment is that of the cause wherever possible (e.g., hypothyroidism, low weight, stress, excessive exercise). Hyperprolactinaemia should be corrected (see below). PCOS is discussed in detail below.

When oestrogen deficiency is not reversed, HRT should almost always be given, as the risk of osteoporosis and other conditions related to oestrogen deficiency almost always outweighs the risks of HRT at this younger age. HRT may still also be actively recommended when normal menopause occurs relatively early (e.g., before the age of 50).

In younger and older women with premature ovarian insufficiency ovarian function may recover or fluctuate, and therefore consideration should be given to using the combined oral contraceptive pill if fertility is not desired.

When fertility is desired, in gonadotrophin deficiency, gonadotrophins can be given to achieve ovulation. Ovarian insufficiency is rarely treatable, and often the option for pregnancy is donor eggs. There is a rare condition called 'resistant' ovary, where high-dose gonadotrophin therapy can occasionally lead to folliculogenesis.

Hirsutism

Normal hair versus hirsutism

The extent of normal hair growth varies between individuals, families, and races, being more extensive in the Mediterranean and some Asian subcontinent populations. These normal variations in body hair, and the more extensive hair growth seen in patients complaining of hirsutism, represent a continuum from no visible hair to extensive cover with thick, dark hair. It is therefore impossible to draw an absolute dividing line between 'normal' and 'abnormal' degrees of facial and body hair in the female. Soft vellus hair is normally present all over the body; this type of hair on the face and elsewhere is 'normal' and is not sex hormone dependent. Hair in the beard, moustache, breast, chest, axilla, abdominal midline, and pubic and thigh areas is sex hormone dependent. Any excess in the latter regions is

Box 34.44 Differential diagnosis and investigation of amenorrhoea

Hormone results					Possible diagnoses	Secondary tests
LH	FSH	E2	PRL	T		
↑	↑	↓	N	N	Ovarian failure	
					Ovarian dysgenesis[a]	Karyotype
					Premature ovarian failure[a]	Ultrasound of ovary/uterus
					Steroid biosynthetic defect[a]	Laparoscopy/biopsy of ovary
					(Oophorectomy)	HCG stimulation
					(Chemotherapy)	
					Resistant ovary syndrome	
↓	↓	↓	N	N	Gonadotrophin failure	
					Hypothalamic–pituitary disease[a]	Pituitary MRI if diagnosis unclear
					Kallmann's syndrome/nIHH[a]	Possibly LHRH test
					Possible hypothalamic causes:	Serum free T_4
					Hypothalamic amenorrhoea[a]	Possibly full assessment of pituitary function
					Weight-related amenorrhoea[a]	
					Exercise-induced amenorrhoea and anorexia[a]	
					Post-pill amenorrhoea	
					General illness[a]	
↓	↓	↓	↑/↑↑	N	Hyperprolactinaemia	
					Prolactinoma[a]	See 'Hyperprolactinaemia' section
					Idiopathic hyperprolactinaemia[a]	Serum free T_4/TSH
					Hypothyroidism[a]	Pituitary MRI
					Polycystic ovarian disease[a]	Other tests for PCOS
					Physiological in lactation	
					Dopamine-antagonist drugs	
↑/N	N	N	N/↑	N/↑	Polycystic ovary syndrome[a]	
					Rarely, Cushing's syndrome	Androstenedione, DHEAS
						SHBG
						Ultrasound of ovary
						Progesterone challenge
						See 'Cushing's syndrome' section (Cushing's syndrome)
N/↓	N/↓	N/↓	N	↑↑	Androgen excess	
					Gonadal or adrenal tumour	Imaging ovary/adrenal
					Congenital adrenal hyperplasia[a]	17a-OH-progesterone
N/↓	N/↓	↑	↑	N	Pregnancy	Pregnancy test
N	N	N	N	N	Uterine/vaginal abnormality	
					Imperforate hymen[a]	Examination findings (?EUA)
					Absent uterus[a]	Ultrasound of pelvis
					Lack of endometrium	Progesterone challenge
						Hysteroscopy

[a]These conditions may present as primary amenorrhoea.

DHEAS, Dehydroepiandrosterone sulphate; *E2,* oestradiol; *EUA,* examination under anaesthetic; *FSH,* follicle-stimulating hormone; *HCG,* human chorionic gonadotrophin; *LH,* luteinizing hormone; *LHRH,* luteinizing hormone-releasing hormone; *nIHH,* normosmic idiopathic hypogonadotropic hypogonadism; *PCOS,* polycystic ovary syndrome; *PRL,* prolactin; *SHBG,* sex hormone-binding globulin; *T,* testosterone.

thus a marker of increased ovarian or adrenal androgen production: most commonly, PCOS, but occasionally other rarer causes.

Aetiology

- ***Polycystic ovarian syndrome (PCOS).*** This is the most common cause of hirsutism; see below.
- ***Idiopathic hirsutism.*** People with hirsutism, no elevation of serum androgen levels, and no other clinical features are sometimes labelled as having 'idiopathic hirsutism'. However, studies suggest that most people with 'idiopathic hirsutism' have some radiological or biochemical evidence of PCOS on more detailed investigation, and several studies have demonstrated

evidence of mild PCOS in up to 20% of the normal female population. Familial or idiopathic hirsutism does occur but usually involves a distribution of hair growth that is not typically androgenic.

- ***Ovarian hyperthecosis.*** This is a non-malignant ovarian disorder characterized by luteinized thecal cells in the ovarian stroma, which secrete testosterone. The clinical features are similar to those of PCOS but tend to present in perimenopausal women, and serum testosterone levels are higher than typically seen in PCOS.
- ***Iatrogenic hirsutism.*** This occurs after treatment with androgens, or more weakly androgenic drugs such as progestogens or danazol.
- ***Non–androgen-dependent hair growth (hypertrichosis).*** This occurs with drugs such as phenytoin, diazoxide, minoxidil, and ciclosporin.
- ***Other causes.*** Rarer and more serious endocrine causes of hirsutism and virilization include CAH (see 'Congenital adrenal hyperplasia' section), Cushing's syndrome (see 'The adrenal cortex' section), and androgen-secreting (virilizing) tumours of the ovary and adrenal.

Polycystic ovary syndrome

Polycystic ovary syndrome (PCOS) is the most common cause of hirsutism in clinical practice, affecting about 1 in 5 women worldwide. It is characterized by multiple small cysts within the ovary (Fig. 34.29) (which represent arrested follicular development) and by excess androgen production from the ovaries (and to a lesser extent from the adrenals).

Measured levels of androgens in blood vary widely from patient to patient and may remain within the normal range, but SHBG levels are often low (due to insulin resistance and high insulin levels), leading to high free androgen levels. In PCOS, there is thought to be an increased frequency of the GnRH pulse generator, causing an increase in LH pulses and androgen secretion. The response of the hair follicle to circulating androgens also seems to vary between individuals with otherwise identical clinical and biochemical features, and the reason for this variation in end-organ response remains poorly understood.

Fig. 34.29 Polycystic ovary syndrome. Ultrasound of the ovary, revealing multiple cysts with central ovarian stroma showing increased echo texture.

PCOS is frequently associated with:

- ***hyperinsulinaemia and insulin resistance***, the prevalence of type 2 diabetes being 10 times higher than in normal women
- ***hypertension, hyperlipidaemia, and increased cardiovascular risk*** ('metabolic syndrome'), which is 2–3 times higher in PCOS. Studies suggest that PCOS *per se* confers an absolute increase in cardiovascular end-points.

Obesity (see Chapter 37) with PCOS is an additional risk factor for insulin resistance. The precise mechanisms that link the aetiology of polycystic ovaries, hyperandrogenism, anovulation, and insulin resistance are still to be elucidated, and whether the basic defect is in the ovary, adrenal, or pituitary, or is a more generalized metabolic defect, remains unknown. Frequently, there is a family history of either PCOS or type 2 diabetes, suggesting a genetic component.

In routine clinical practice, the majority of people with objective signs of androgen-dependent hirsutism will have PCOS, and investigation is mainly required to exclude rarer and more serious causes of virilization.

History and examination

Most patients with PCOS present with amenorrhoea/oligomenorrhoea, hirsutism, and acne, shortly after menarche. Frequently, patients are also obese with evidence of insulin resistance. Clinical, biochemical, and radiological features of PCOS merge imperceptibly into those of the normal populations.

- ***Hirsutism.*** This should be recorded objectively, ideally using a scoring system, both to document the problem and to monitor treatment. The method and frequency of physical removal (e.g., shaving, plucking) should also be recorded. The development of hirsutism commonly provokes severe distress in young women and may lead to avoidance of normal social activities. Most patients who complain of hirsutism will have an objective excess of hair on examination, but occasionally very little will be found (and appropriate counselling is then indicated).
- ***Age and speed of onset.*** Hirsutism related to PCOS usually begins around the time of the menarche and increases slowly and steadily in the teens and twenties. Rapid progression and prepubertal or late onset suggest a more serious cause.
- ***Accompanying virilization.*** Hirsutism due to PCOS may be severe and affect all androgen-dependent areas on the face and body. However, more severe virilization (clitoromegaly, recent-onset frontal balding, male phenotype) implies substantial androgen excess and usually indicates a rarer cause rather than PCOS. Thinning of head hair in a male pattern—androgenic alopecia—occurs in a proportion of women with uncomplicated PCOS, typically with a familial tendency for premature androgen-related hair loss in both sexes.
- ***Menstruation.*** Most people with PCOS will have some disturbance of menstruation, typically oligo-/amenorrhoea, although more frequent erratic bleeding can also occur. However, PCOS can present as hirsutism with regular periods or as irregular periods, with no evidence of hirsutism or acne.
- ***Weight.*** Many people with PCOS are also overweight or obese. The obesity worsens the underlying androgen excess and insulin resistance and inhibits the response to treatment; it is an indication for appropriate advice on diet and exercise. In severe cases, the insulin resistance may have a visible manifestation as acanthosis nigricans on the neck and in the axillae. In addition to worsening the symptoms of PCOS, central obesity in PCOS significantly increases the likelihood of developing diabetes and cardiovascular disease.

Investigations and differential diagnosis

- ***Serum total testosterone.*** This is often elevated in PCOS and is invariably substantially raised in virilizing tumours (usually >5 nmol/L). People with hirsutism and normal testosterone levels frequently have low levels of SHBG, leading to high free androgen levels. The ***free androgen index*** ([testosterone/SHBG] *100) is often used and is high; free testosterone is difficult to measure directly.
- ***Other androgens.*** Androstenedione and DHEA sulphate are frequently elevated in PCOS, and even higher in CAH and virilizing tumours.
- ***17α-Hydroxyprogesterone.*** This is elevated in classical CAH but may be apparent in late-onset CAH only after stimulation tests.
- ***Gonadotrophin levels.*** LH hypersecretion is a frequent feature of PCOS, but the pulsatile nature of secretion of this hormone means that a 'classic' increased LH/FSH ratio is not always observed on a random sample.
- ***Oestrogen levels.*** Oestradiol is usually normal in PCOS, but oestrone levels (which are rarely measured) are elevated because of peripheral conversion. Levels are variable in other causes.
- ***Ovarian ultrasound.*** This is a useful investigation (see Fig. 34.28). Typical features are those of a thickened capsule, multiple 3–5 mm cysts, and a hyperechogenic stroma. Prolonged hyperandrogenization from any cause may lead to polycystic changes in the ovary. Ultrasound may also reveal virilizing ovarian tumours, although these are often small.
- ***Serum prolactin.*** Mild hyperprolactinaemia is common in PCOS but rarely exceeds 1500 mU/L.

If an androgen-secreting tumour is suspected clinically or after investigation, then more complex investigations include dexamethasone suppression tests, CT or MRI of adrenals, and selective venous sampling.

Diagnosis

Most patients presenting with a combination of hirsutism and menstrual disturbance will be shown to have PCOS, but the rarer alternative diagnoses should be excluded; the latter include late-onset CAH (early onset, raised serum 17α-hydroxyprogesterone), Cushing's syndrome (look for other clinical features), and virilizing tumours of the ovary or adrenals (severe virilization, markedly elevated serum testosterone).

The consensus ***(Rotterdam) criteria 2003*** retain international support for diagnosis of PCOS, requiring at least two of the following:

- clinical and/or biochemical evidence of hyperandrogenism
- oligo-ovulation and/or anovulation
- polycystic ovaries on ultrasound.

Management

Local therapy for hirsutism

Regular plucking, bleaching, depilatory cream, waxing, or shaving is used. Such removal neither worsens nor improves the underlying severity of hirsutism. More 'permanent' solutions include electrolysis and a variety of 'laser' hair removal systems; all appear effective but have not been evaluated in long-term studies, but they are expensive, and still often require repeated long-term treatment. ***Eflornithine cream*** (an antiprotozoal) inhibits hair growth by inhibiting ornithine decarboxylase but is effective in only a minority of cases and should be discontinued if there is no improvement after 4 months.

Systemic therapy for hirsutism

This always requires a year or more of treatment for maximal benefit, and long-term therapy is frequently required as the problem tends to recur when treatment is stopped. The patient must therefore always be an active participant in the decision to use systemic therapy and must understand the rare risks as well as the benefits. Weight loss (at least 5%) should also be encouraged, as many patients have improvement in symptoms.

- ***Oestrogens*** (e.g., oral contraceptives) suppress ovarian androgen production and reduce free androgens by increasing SHBG levels. Combined hormone pills will produce a slow improvement in hirsutism in a majority of cases and should normally be used first unless there is a contraindication, such as a history of thrombosis. After menopause, HRT preparations that contain medroxyprogesterone (rather than more androgenic progestogens) may be helpful.
- ***Cyproterone acetate*** (50–100 mg daily) is an antiandrogen but is also a progestogen, teratogen, and a weak glucocorticoid. Given continuously, it produces amenorrhoea and so is normally given for days 1–14 of each cycle. In women of childbearing age, contraception is essential.
- ***Spironolactone*** (200 mg daily) also has antiandrogen activity and can cause useful improvements in hirsutism. In women of childbearing age, contraception is essential.
- ***Finasteride*** (5 mg daily), a 5α-reductase inhibitor that prevents the formation of dihydrotestosterone in the skin, has also been shown to be effective but long-term experience is limited. In women of childbearing age, contraception is essential.
- ***Flutamide***, another antiandrogen, is less commonly used owing to the high incidence of hepatic side-effects. In women of childbearing age, contraception is essential.

Treatment of menstrual disturbance

- ***Cyclical oestrogen/progestogen*** administration will regulate the menstrual cycle and remove the symptom of oligo- or amenorrhoea. This is most frequently an additional benefit of the treatment of hirsutism but may also be used when menstrual disturbance is the only symptom.
- ***Metformin*** (500 mg three times daily) is commonly used in this condition because of the recognized association between PCOS and insulin resistance. It may improve menstrual cyclicity and ovulation; some patients also report improvement in hirsutism and ease of weight loss, but gastrointestinal upset may limit use.

Treatment for fertility in PCOS

- ***Clomifene or letrozole*** is given daily on days 2–6 of the cycle and is effective in 75% of women in achieving ovulation. It can occasionally cause ***ovarian hyperstimulation syndrome***, an iatrogenic complication of ovulation induction therapy, consisting of ovarian enlargement, oedema, hypovolaemia, acute kidney injury, and possibly shock; specialist supervision is essential. It is recommended that clomifene and letrozole should not normally be used for more than six cycles (owing to a possible increased risk of ovarian cancer in patients treated for longer than recommended).
- ***Low-dose FSH*** is used for non-responders to clomifene.
- ***Metformin*** alone may improve ovulation and achieve conception.

Wedge resection of the ovary was a traditional therapy but is now rarely required, although laparoscopic ovarian electrodiathermy may be helpful.

It is recognized that patients with PCOS often require psychological input to their condition.

Disorders of sex development

Disorders of sex development (DSDs) are a rare group of conditions that can be divided into 46XY and 46XX DSD and encompass:

- conditions of abnormalities in androgen synthesis, action, and excess
- disorders of gonadal development
- persistent Müllerian duct syndrome
- sex chromosomal DSDs (**Box 34.45**).

Such cases always require extensive, multidisciplinary clinical management. There is a fuller discussion on the interaction between sex and gender in Chapter 47.

Further reading

Argente J, Dunkel L, Kaiser UB et al. Molecular basis of normal and pathological puberty: from basic mechanisms to clinical implications. *Lancet Diabetes Endocrinol* 2023; 11(3):203–216.
Basaria S. Male hypogonadism. *Lancet* 2014; 383:1250–1263.
Jayasena CN, Anderson RA, Llahana S et al. Society for Endocrinology guidelines for testosterone replacement therapy in male hypogonadism. *Clin Endocrinol* (Oxf) 2022; 96(2):200–219.
Mani H, Levy MJ, Davies MJ et al. Diabetes and cardiovascular events in women with polycystic ovary syndrome: a 20-year retrospective cohort study. *Clin Endocrinol* 2013; 78:926–934.
Rotterdam ESHRE/ASRM-Sponsored PCOS Consensus Workshop Group. Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome. *Fertil Steril* 2004; 81:19–25.
Snyder PJ, Bhasin S, Cunningham AM et al. Effects of testosterone treatment in older men. *N Engl J Med* 2016; 374:611–624.
Teede HJ, Tay CT, Laven JJE et al. Recommendations from the 2023 International evidence-based guideline for the assessment and management of polycystic ovary syndrome. *J Clin Endocrinol Metab* 2023; 108(10):2447–2469.
Webber L, Vermeulen N. ESHRE guideline: management of women with premature ovarian insufficiency. *Hum Reprod* 2016; 31:926–937.

Acromegaly and gigantism

Disorders of childhood growth and pubertal development are managed by paediatric endocrinologists. Since GH stimulates skeletal and soft tissue growth, excess produces gigantism in children (if acquired before epiphyseal fusion) and acromegaly in adults. Both are due to a GH-secreting pituitary tumour (somatotroph adenoma) in almost all cases. Hyperplasia due to ectopic GHRH excess is very rare. The overall incidence is approximately 3–4/million per year and the prevalence is 50–80/million worldwide. Acromegaly usually occurs sporadically, although gene mutations can rarely give rise to familial acromegaly: typically, the *AIP* gene in familial isolated pituitary adenoma. In familial acromegaly, there is an increased onset before puberty compared with sporadic cases, leading to a higher prevalence of gigantism.

Physiology

Growth hormone (GH) is the pituitary factor responsible for stimulation of body growth in humans. Its secretion is stimulated by GHRH, released into the portal system from the hypothalamus; it is also under inhibitory control exerted by somatostatin (**Fig. 34.30**). A separate GH-stimulating system involves a distinct receptor (GH secretagogue receptor), which interacts with ghrelin. It is not known how these two systems interact but, because ghrelin is synthesized in the stomach, a nutritional role is suggested for GH.

- ***GH*** acts by binding to a specific (single transmembrane) receptor located mainly in the liver (**Box 34.46**). This induces an intracellular phosphorylation cascade involving the Janus kinase/signal transducing activators of transcription (JAK/STAT) pathway. STAT proteins are translocated from the cytoplasm into the cell nucleus and cause GH-specific effects by binding to nuclear DNA.
- ***Insulin-like growth factor-1 (IGF-1)***, a somatomedin, stimulates growth. Its hepatic secretion is stimulated by a tissue-specific effect of GH on the liver. There are multiple IGF-binding proteins (IGF-BP) in plasma; IGF-BP3 can be measured clinically to improve assessment of GH status, particularly in children.

The metabolic actions of the system are:

- increasing collagen and protein synthesis
- promoting retention of calcium, phosphorus, and nitrogen, necessary substrates for anabolism
- opposing the action of insulin (a 'counter-regulatory' hormone effect).

GH release is intermittent and mainly nocturnal, especially during REM sleep. The frequency and size of GH pulses increase

Box 34.45 Disorders of sexual differentiation

Condition	Chromosomes	Gonads	Phenotype	Remarks
Turner's syndrome	45X (50%) 46,X,i (Xq) (5%–10%) 45,X, mosaicism (remainder)	Streak	Female	Often morphological features (e.g., short stature, web neck, coarctation of aorta)
Gonadal dysgenesis	46XY	Streak or minimal testes[a]	Immature female	
Congenital adrenal hyperplasia	46XX	Ovary	Female with variable virilization	Obvious androgen excess
Virilizing tumour	46XX	Ovary	Female with variable virilization	Obvious androgen excess
True hermaphroditism	46XX/XY or mosaic	Testis and ovary	Male or ambiguous	
Klinefelter's syndrome	47XXY	Small testes	Male, often with gynaecomastia	Many are hypogonadal
Testicular feminization	46XY	Testes[a]	Ambiguous or infantile female	Androgen receptor defective
Testicular synthetic defects	46XY	Testes[a]	Cryptorchid, ambiguous	
5α-Reductase deficiency	46XY	Testes	Cryptorchid, ambiguous	Impaired conversion of testosterone to dihydrotestosterone
Anorchia	46XY	Absent	Immature female	

[a]Gonadectomy advised because of high risk of malignancy.
i, Isochromosome.

during the growth spurt of adolescence and decline thereafter. Acute stress and exercise both stimulate GH release, while, in the normal subject, hyperglycaemia suppresses it.

IGF-1 may, in addition, play a major role in maintaining neoplastic growth. A relationship has been shown between circulating IGF-1 concentrations and breast cancer in premenopausal women and prostate cancer in men.

History and examination

Symptoms and signs of acromegaly are shown in Fig. 34.31. One-third of patients present with changes in appearance, and one-quarter with visual field defects or headaches; in the remainder, the diagnosis is made by an alert observer in another clinic: for example, in general practice, or in diabetes, hypertension, dental, or dermatology departments. Sleep apnoea is common and requires investigation and treatment if there are suggestive symptoms (see Chapter 20, 'Obstructive sleep apnoea' section). Sweating, headaches, and soft tissue swelling are particularly useful symptoms of persistent GH secretion. Headache is very common in acromegaly and may be severe, even with small tumours; it is often improved after surgical cure or with somatostatin analogues.

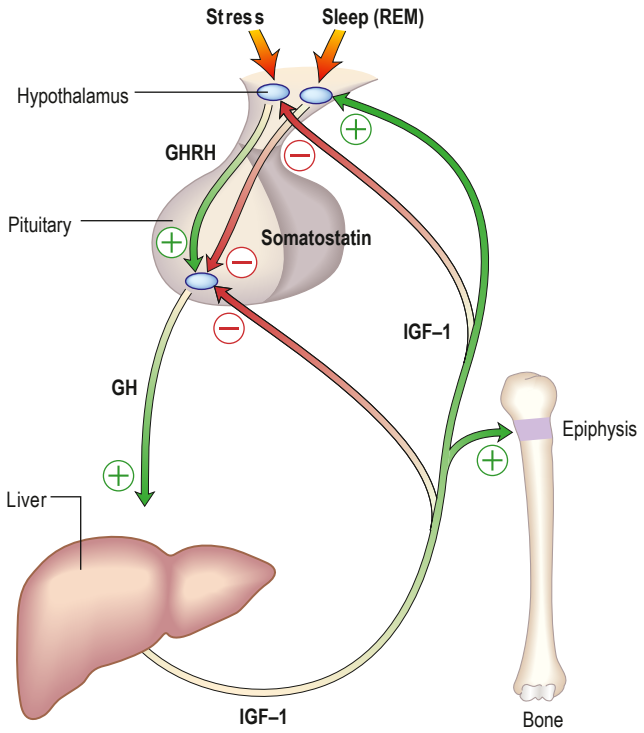


Fig. 34.30 The control of growth hormone (GH) and insulin-like growth factor-1 (IGF-1). Pituitary GH is secreted under dual control of growth hormone-releasing hormone (GHRH) and somatostatin and stimulates release of IGF-1 in the liver and elsewhere. IGF-1 has peripheral actions, including bone growth, and exerts negative feedback to the hypothalamus and pituitary. *REM,* Rapid eye movement.

Investigations

- ***GH levels*** may exclude acromegaly if they are undetectable, but a detectable value is non-diagnostic taken alone. Normal adult levels are below 0.5 μg/L for most of the day, except during stress or a 'GH pulse'.
- ***A glucose tolerance test*** is diagnostic if there is no suppression of GH. Acromegalics fail to suppress GH below 0.3 μg/L and some show a paradoxical rise; about 25% of acromegalics have a positive diabetic glucose tolerance test.
- ***IGF-1 levels*** are almost always raised in acromegaly; a single plasma level of IGF-1 reflects mean 24-hour GH levels and is useful in diagnosis. A normal IGF-1, together with random GH of less than 1 μg/L, may be taken to exclude acromegaly if the diagnosis is clinically unlikely.
- ***Visual field examination*** commonly reveals defects, such as bitemporal hemianopia.
- ***MRI scan*** of the pituitary is carried out if the above tests are abnormal. This will almost always reveal the pituitary adenoma. In cases where no clear lesion is present, or in postoperative cases where residual disease is being sought, there is increasing interest in the use of functional imaging such as methionine PET.
- ***Pituitary function*** commonly demonstrates partial or complete anterior hypopituitarism.
- ***Prolactin*** levels show mild to moderate hyperprolactinaemia in 30% of patients (Fig. 34.32). In some, the adenoma secretes both GH and prolactin.

Management

Untreated acromegaly results in markedly reduced survival. Most deaths occur from heart failure, coronary artery disease, and hypertension-related causes. In addition, there is an increase in deaths due to neoplasia, particularly large bowel tumours; guidance advocates regular colonoscopy to detect and remove colonic polyps in order to reduce the risk of colonic cancer, although the evidence is inconclusive. Treatment is therefore indicated in all except the elderly or those with minimal abnormalities. The aim of therapy is to achieve a mean GH level below 2.5 μg/L; this is not always

Box 34.46 Hormones and receptors of the hypothalamic–pituitary–GH axis

Hormone	Source	Site of action	Hormone structure	Receptor	Post-receptor activity
Growth hormone-releasing hormone (GHRH)	Hypothalamus	Pituitary	Peptide—44 AA	Membrane—7TM	G-proteins cAMP
Somatostatin (inhibitory GHRIH)	Hypothalamus	Pituitary	Cyclic peptide—14 or 28 AA	Membrane—7TM SST2 + SST5	G-proteins Inhibit cAMP
Growth hormone	Pituitary	Liver and other tissues	Peptide—191 AA	Transmembrane Dimerized GHR	JAK2 STAT
Insulin-like growth factor 1 (IGF-1)	Liver and locally elsewhere	Many tissues	Peptide—70 AA 3 S=S bridges	Transmembrane—IGFR 2α + 2β subunits	Receptor tyrosine kinase

7TM, 7 transmembrane (G-protein coupled receptor); *AA,* amino acids; *cAMP,* adenylate cyclase → cyclic adenosine monophosphate; *GHRIH,* growth hormone-releasing inhibitory hormone; *JAK,* Janus kinase; *STAT,* signal transducers and activators of transcriptions; *SST,* somatostatin receptor subtypes.

Symptoms

Head and neck
Change in facial appearance
Headaches
Visual deterioration
Deep voice
Goitre

General
Tiredness
Weight gain
Breathlessness
Excessive sweating
Muscular weakness
Joint pain

Hormonal
Amenorrhoea or oligomenorrhoea in women
Galactorrhoea
Impotence or poor libido

Increased glove or hat size

Signs

Facial
Prominent supraorbital ridge
Prognathism
Interdental separation
Large tongue
Visual field defects

General
Hirsutism
Thick greasy skin

Hands
Spade-like hands and feet
Tight rings on fingers
Carpal tunnel syndrome

Other
Galactorrhoea
Hypertension
Heart failure
Oedema
Arthopathy
Proximal myopathy
Glycosuria (plus possible signs of hypopituitarism)

Fig. 34.31 Acromegaly: symptoms and signs. *Bold italic* type indicates signs of greater discriminant value.

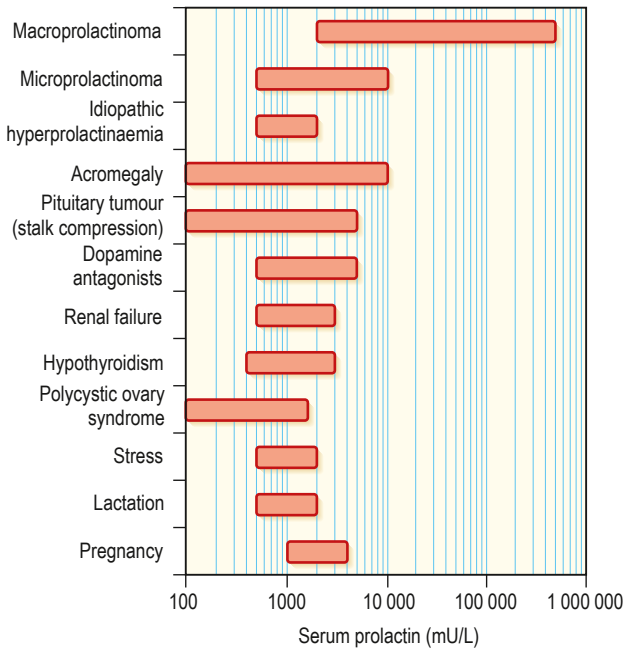


Fig. 34.32 Range of serum prolactin seen in common causes of hyperprolactinaemia.

'normal' but has been shown to reduce mortality to normal levels and is therefore considered a 'safe' GH level. A normal IGF-1 is also a goal of therapy. Occasionally, there can be discordance between GH and IGF-1 levels, which can create management dilemmas.

When present, hypopituitarism should be corrected and concurrent diabetes and/or hypertension should be treated conventionally; both usually improve with treatment of the acromegaly.

The general advantages and disadvantages of surgery, radiotherapy, and medical treatment are discussed in the 'Management' section of the 'Pituitary space-occupying lesions and tumours' section. Progress can be assessed by monitoring GH and IGF-1 levels.

Surgery

Trans-sphenoidal surgery is the appropriate first-line therapy. It will result in clinical remission in a majority of cases (60%–90%) with pituitary microadenoma but in only 50% of those with macroadenoma. Very high preoperative GH and IGF-1 levels are also poor prognostic markers of surgical cure. Surgical success rates are variable and highly dependent upon experience, and a specialist pituitary surgeon is essential. Transfrontal surgery is rarely required except for massive macroadenomas. There is a recurrence rate of approximately 10%.

Pituitary radiotherapy

External radiotherapy is normally used after pituitary surgery fails to normalize GH levels rather than as primary therapy. It is often combined with medium-term treatment using a somatostatin analogue, dopamine agonist, or GH antagonist because of the slow biochemical response to radiotherapy, which may take 10 years or more; it is also often associated with hypopituitarism, which makes it unattractive in patients of reproductive age. Stereotactic (gamma-knife) radiotherapy is used in some centres, as it delivers a more concentrated field of radiation.

Medical therapy

There are three receptor targets for the treatment of acromegaly: pituitary somatostatin receptors and dopamine (D_2) receptors, and GH receptors in the periphery.

Somatostatin receptor agonists

Octreotide and ***lanreotide*** are synthetic analogues of somatostatin, which act selectively on somatostatin receptor subtypes

(SSTR2 and SSTR5) and are highly expressed in GH-secreting tumours. These drugs were used as a short-term treatment while other modalities became effective, but now are sometimes used as primary therapy. They reduce GH and IGF levels in most patients but often do not achieve treatment targets. Both drugs are typically administered as monthly depot injections and are generally well tolerated; they are, however, associated with an increased incidence of gallstones and are expensive. Pasireotide (acting on SSTR1, 2, 3, and 5) is a newer agent but its role in acromegaly is still being evaluated.

Dopamine agonists

Dopamine agonists act on D_2 receptors and can be given to shrink tumours prior to definitive therapy or to control symptoms and persisting GH secretion; they are probably most effective in mixed GH-producing (somatotroph) and prolactin-producing (mammotroph) tumours. Typical doses are bromocriptine 10–60 mg daily or cabergoline 0.5 mg daily (higher than for prolactinomas). Given alone, they reduce GH to 'safe' levels in only a minority of cases but are useful for mild residual disease or in combination with somatostatin analogues. Drugs with combined somatostatin and dopamine receptor activity are under development.

Growth hormone antagonists

Pegvisomant (a genetically modified analogue of GH) is a GH receptor antagonist that exerts its effect by binding to and preventing dimerization of the GH receptor. It does not lower GH levels or reduce tumour size but has been shown to normalize IGF-1 levels in 90% of patients. It is given by daily injection and its main role at the present time is treatment of patients in whom GH and IGF levels cannot be reduced to safe levels with somatostatin analogues alone, or by surgery or radiotherapy.

Hyperprolactinaemia

The HP control of prolactin secretion is illustrated in Fig. 34.33. Prolactin is a large peptide secreted in the pituitary and acts via a transmembrane receptor stimulating JAK2 and other pathways (Box 34.47).

Prolactin is under tonic dopamine inhibition; factors known to increase prolactin secretion (e.g., TRH) are probably of less relevance unless the patient has primary hypothyroidism. Prolactin stimulates milk secretion (but not breast tissue development) but also inhibits gonadal activity. It decreases GnRH pulsatility at the hypothalamic level and, to a lesser extent, blocks the action of LH on the ovary or testis, producing hypogonadism even when the pituitary–gonadal axis itself is intact.

The role of prolactin outside pregnancy and lactation is not well defined, although there is some epidemiological evidence of a link between high prolactin levels and breast cancer, which has led to an interest in the development of prolactin receptor antagonists.

The range of serum prolactin seen in common causes of hyperprolactinaemia is illustrated in Fig. 34.32. Mildly increased prolactin levels (400–600 mU/L) may be physiological and asymptomatic but higher levels require a diagnosis. Levels above 5000 mU/L always imply a prolactin-secreting pituitary tumour.

Aetiology

Hyperprolactinaemia has many causes:

- ***Physiological.*** These include pregnancy, lactation, stress, sleep, exercise, and coitus.
- ***Pathological.*** These include prolactinoma, co-secretion of prolactin in tumours causing acromegaly, stalk compression due to pituitary adenomas and other pituitary masses, polycystic ovarian syndrome, primary hypothyroidism, and idiopathic hyperprolactinaemia. Rarer causes include renal failure, liver failure, postictal state, and chest wall injury (e.g., following herpes zoster).
- ***Drug-induced.*** Drug causes include ***oestrogens*** (e.g., contraceptive pill), ***dopamine antagonists*** (e.g., antipsychotics such as phenothiazines), ***antidepressants*** (e.g., tricyclics, selective serotonin reuptake inhibitors [SSRIs]), ***antiemetics*** (e.g., metoclopramide, domperidone), and ***others*** (e.g., verapamil, cimetidine, methyldopa).

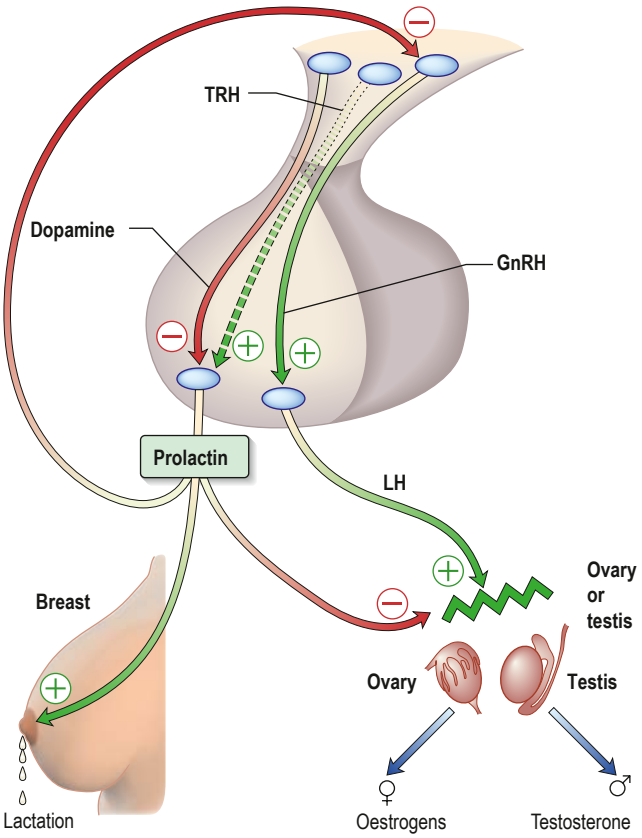


Fig. 34.33 The control and actions of prolactin. Prolactin is mainly controlled by tonic inhibition by hypothalamic dopamine. Prolactin stimulates lactation but also inhibits both hypothalamic gonadotrophin-releasing hormone (GnRH) secretion and the gonadal actions of luteinizing hormone (LH). *TRH,* Thyrotrophin-releasing hormone.

History and examination

Patients may present with features of hyperprolactinaemia or structural symptoms from a pituitary tumour with headaches and visual loss. This latter presentation is more common in males.

Common symptoms include:

- Galactorrhoea, spontaneous or expressible (60% of cases). Note that many people with galactorrhoea do not have hyperprolactinaemia. This is termed 'normoprolactinaemic galactorrhoea' and the causes are poorly understood.
- Oligomenorrhoea or amenorrhoea.
- Decreased libido in both sexes.
- Male impotence.
- Subfertility.
- Symptoms or signs of oestrogen or androgen deficiency—in the long term osteoporosis may result, especially in women.

Box 34.47 Hormones and receptors of the prolactin axis

Hormone	Source	Site of action	Hormone structure	Receptor	Post-receptor activity
Dopamine	Hypothalamus	Pituitary	Amine	Membrane—7TM D2 receptor	G-proteins Inhibit cAMP
Prolactin	Pituitary	Breast Other tissues	Peptide—199 AA	Transmembrane PRLR Class 1 cytokine	JAK2 and other pathways

7TM, 7 transmembrane (G-protein coupled receptor); *AA,* amino acids; *cAMP,* adenylate cyclase → cyclic adenosine monophosphate; *JAK,* Janus kinase.

- Delayed or arrested puberty in the peripubertal patient.
- Mild gynaecomastia—often seen in men due to the associated hypogonadism rather than being a direct effect of prolactin.

Additionally, headaches and/or visual field defects occur if there is a pituitary macroadenoma (more common in men).

Investigations

Hyperprolactinaemia should be confirmed by repeat measurement. It should be excluded in all patients presenting with a pituitary mass.

If there are no clinical features of hyperprolactinaemia, the possibility of ***macroprolactinaemia*** should be considered. This is a higher-molecular-weight complex of prolactin bound to IgG, which is physiologically inactive but occurs in a small proportion of normal people and can therefore lead to unnecessary treatment. Macroprolactinaemia can be diagnosed in the laboratory by precipitation of immunoglobulin G (IgG) with polyethylene glycol, after which prolactin levels will be normal on testing; most laboratories will do this routinely.

Further tests are appropriate after physiological and drug causes have been excluded:

- ***Visual fields*** should be checked.
- ***Primary hypothyroidism*** must be excluded since this is a cause of hyperprolactinaemia.
- ***Anterior pituitary function*** should be assessed if there is any clinical evidence of hypopituitarism or radiological evidence of a pituitary tumour.
- ***MRI of the pituitary*** is necessary if there are any clinical features suggestive of a pituitary tumour, and it is desirable in all cases when prolactin is significantly elevated (>1000 mU/L).

In the presence of a pituitary mass on MRI, the level of prolactin helps determine whether the mass is a prolactinoma or a non-functioning pituitary tumour causing stalk-disconnection hyperprolactinaemia: levels of above 5000 mU/L in the presence of a macroadenoma, or above 2000 mU/L in the presence of a microadenoma (or with no radiological abnormality) strongly suggest a prolactinoma. Macroprolactinoma refers to tumours above 10 mm in diameter, microprolactinoma to smaller ones. Tumours that are approximately 10 mm are termed mesoadenomas.

Occasionally, very large prolactinomas can be associated with such high serum prolactin levels that some assays give an artefactual falsely low result (known as the 'hook effect'). If suspected, this can be excluded by serial dilutions of the serum sample.

Management

Treatment is recommended in all patients with a macroprolactinoma to reduce the size of the tumour. It is also recommended in those with troublesome galactorrhoea. In those with sex steroid deficiency, lowering prolactin levels should reverse hypogonadism and avoid the long-term effects of deficiency, although for those intolerant of prolactin-lowering medication, sex steroid replacement is an alternative approach. Therefore some patients (for example, postmenopausal women with microprolactinomas) may not require any treatment.

Medical treatment

Hyperprolactinaemia is controlled with a dopamine agonist.

- ***Cabergoline*** (500 μg once or twice a week, judged on clinical response and prolactin levels) is the best-tolerated and longest-acting drug and is the drug of choice.
- ***Bromocriptine*** is the longest-established therapy and is therefore preferred if pregnancy is planned; initial doses should be small (e.g., 1 mg), taken with food and gradually increased to 2.5 mg two or three times daily. Side-effects, which prevent effective therapy in a minority of cases, include nausea and vomiting, dizziness and syncope, constipation, and cold peripheries.
- ***Quinagolide*** (75–150 μg once daily) is an alternative.

Complications (similar to those seen when cabergoline is used in higher doses in Parkinson's disease) include pulmonary, retroperitoneal, and pericardial fibrotic reactions and cardiac valve lesions. Patients need monitoring, although studies suggest that adverse effects appear to be very rare in those on lower, 'endocrine' doses. In addition, patients can also develop neuropsychiatric side-effects, including compulsive behaviours such as gambling (termed 'impulse control disorders').

In most cases, a dopamine agonist will be the first and only therapy and can be used in the long term, although a trial of discontinuation can be considered in microprolactinomas after 2–3 years, as a significant percentage may not recur. Prolactinomas usually shrink in size on a dopamine agonist, and in macroadenomas any pituitary mass effects commonly resolve (Fig. 34.34). Macroprolactinomas may not recur after several years of dopamine agonist therapy in a minority of cases, but in the majority, hyperprolactinaemia will return if treatment is stopped.

In patients planning pregnancy, it is useful to know the size of the pituitary lesion before starting dopamine agonist therapy. Rarely, tumours enlarge during pregnancy to produce headaches and visual field defects. Dopamine agonists, which are traditionally stopped after conception, particularly when used in the treatment of microprolactinomas, can be restarted if there are any signs of tumour growth during pregnancy.

Trans-sphenoidal surgery

Surgery may rarely be needed in patients who are intolerant of or unresponsive to dopamine agonists. Surgery may restore normoprolactinaemia in people with microadenoma but is rarely completely successful in those with macroadenomas and risks damage to normal pituitary function. Therefore, most patients and physicians elect to continue medical therapy rather than proceed to surgery. Prolactin should therefore always be measured before surgery on any mass in the pituitary region. Some surgeons believe that long-term bromocriptine increases the hardness of the adenoma and makes resection more difficult, but others dissent from this view.

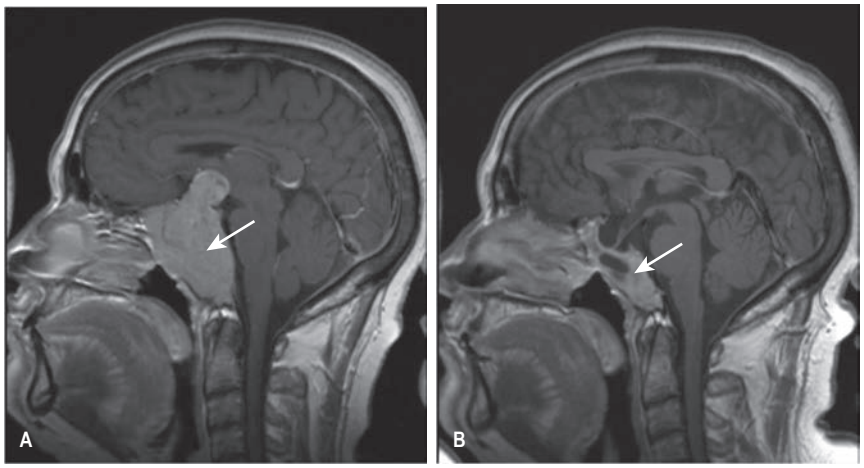


Fig. 34.34 Macroprolactinoma *(arrowed)*. **(A)** Before treatment. **(B)** After 2 years' treatment with the dopamine agonist cabergoline, showing marked adenoma shrinkage.

Box 34.48 Hormones and receptors of the posterior pituitary axes

Hormone	Source	Site of action	Hormone structure	Receptor	Post-receptor activity
Vasopressin (arginine vasopressin, AVP)	Hypothalamus → pituitary	Kidney	Peptide—9 AA	Membrane—7TM AVPR2	G-proteins cAMP Aquaporin-2
	Hypothalamus → pituitary	Vascular	Peptide—9 AA	Membrane—7TM AVPR1A	G-proteins PLC/IP_3
	Hypothalamus → portal veins	Pituitary (ACTH secretion)	Peptide—9 AA	Membrane—7TM AVPR1B	G-proteins PLC/IP_3
Oxytocin	Hypothalamus → pituitary	Uterus and breast	Peptide—9 AA	Membrane—7TM OXTR	G-proteins

7TM, 7 transmembrane (G-protein coupled receptor); *AA,* amino acids; *ARE,* androgen response element; *AVPR,* arginine vasopressin receptor; *cAMP,* adenylate cyclase → cyclic adenosine monophosphate; *OXTR,* oxytocin receptor.

Radiotherapy

Also rarely needed, radiotherapy usually controls adenoma growth and is slowly effective in lowering prolactin but causes progressive hypopituitarism. It may be advocated after medical tumour shrinkage or after surgery in larger tumours, especially when patients have completed their family or if the drug treatment is poorly tolerated, but most simply advocate continuation of dopamine agonist therapy in responsive cases.

THE THIRST AXIS

Thirst and water regulation are largely controlled by vasopressin (also known as anti-diuretic hormone, ADH; or arginine vasopressin, AVP). This hormone is synthesized in the hypothalamus and then migrates in neurosecretory granules along axonal pathways to the posterior pituitary. Pituitary disease alone without hypothalamic involvement therefore does not lead to AVP-deficiency (also referred to as 'cranial diabetes insipidus'), as the hormone can still 'leak' from the damaged end of the intact axon.

At normal concentrations, the kidney is the predominant site of action of vasopressin. Vasopressin stimulation of the V_2 receptors (**Box 34.48**) allows the collecting ducts to become permeable to water via the migration of aquaporin-2 water channels, thus permitting reabsorption of hypotonic luminal fluid. Vasopressin therefore reduces diuresis and results in overall retention of water. At high concentrations, vasopressin also causes vasoconstriction via the V_1 receptors in vascular tissue.

Changes in plasma osmolality are sensed by osmoreceptors in the anterior hypothalamus. Vasopressin secretion is suppressed at levels below 280 mOsm/kg, thus allowing maximal water diuresis. Above this level, plasma vasopressin increases in direct proportion to plasma osmolality. At the upper limit of normal (295 mOsm/kg), maximum antidiuresis is achieved and thirst is experienced at about 298 mOsm/kg (**Fig. 34.35**).

Other factors affecting vasopressin release are shown in **Box 34.49**.

Disorders of vasopressin secretion or activity include:

- AVP-deficiency (previously 'cranial diabetes inspidus') as a result of hypothalamic disease
- AVP-resistance (previously 'nephrogenic DI')—where the renal tubules are insensitive to vasopressin, and an example of a receptor abnormality
- inappropriate excess of the hormone (syndrome of inappropriate diuresis, SIAD), also called syndrome of inappropriate antidiuretic hormone (SIADH), a common cause of hyponatraemia in hospital patients.

While AVP-deficiency and AVP-resistance (previously diabetes insipidus) is uncommon, patients need to be distinguished from the occasional patient with 'primary polydipsia' and from those whose renal tubular function has been impaired by electrolyte abnormalities, such as hypokalaemia or hypercalcaemia. Similarly other causes of hyponatraemia need to be excluded prior to diagnosing and investigating SIAD.

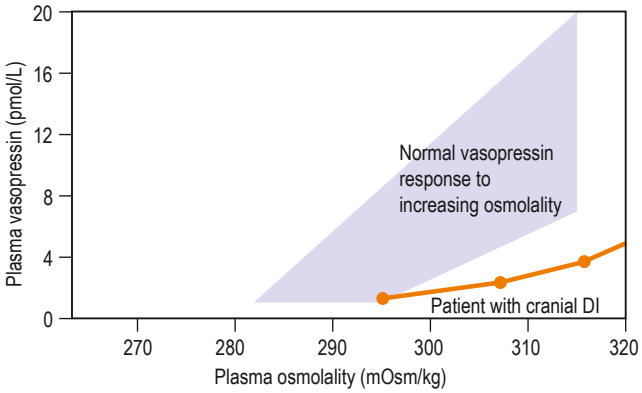


Fig. 34.35 **Plasma vasopressin response to increasing osmolality in normal subjects and in a patient with diabetes insipidus (DI).**

Box 34.49 Factors affecting vasopressin (AVP) release

Increased by
- Increased osmolality
- Hypovolaemia
- Hypotension
- Nausea
- Hypothyroidism
- Angiotensin II
- Adrenaline (epinephrine)
- Cortisol
- Nicotine
- Antidepressants

Decreased by
- Decreased osmolality
- Hypervolaemia
- Hypertension
- Ethanol
- α-Adrenergic stimulation

Diabetes insipidus

History and examination

Deficiency of vasopressin (AVP) or insensitivity to its action leads to excess excretion of dilute urine with a compensatory increase in thirst (polydipsia). Daily urine output may reach as much as 10–15 L, leading to dehydration that may be very severe if the thirst mechanism or consciousness is impaired or the patient is denied fluid (see also Chapter 27).

Diabetes insipidus (DI) may be masked by simultaneous cortisol deficiency; cortisol replacement allows a water diuresis and DI then becomes apparent.

Aetiology

Disorders of the hypothalamo-pituitary axis cause the condition of ***vasopressin deficiency (AVP-deficiency***), while renal disorders and electrolyte disturbance lead to vasopressin resistance (AVP-resistance). Causes of ***vasopressin deficiency (AVP-deficiency)*** are listed in **Box 34.50**. The most common is HP surgery, following which transient DI is common, frequently remitting after a few days or weeks. Typically, inflammatory lesions of the pituitary stalk are associated with DI (**Fig. 34.36**).

Familial isolated vasopressin deficiency causes DI from early childhood and is dominantly inherited, caused by a mutation in the *AVP-NPII* gene. ***DIDMOAD (Wolfram's) syndrome*** is a rare autosomal recessive disorder comprising DI, diabetes mellitus, optic atrophy, and deafness and is caused by mutations in the *WFS1* gene on chromosome 4. MRI may show an absent or poorly developed posterior pituitary.

Biochemistry

- High or high-normal plasma osmolality with low urine osmolality (in primary polydipsia, plasma osmolality tends to be low).

Box 34.50 Causes of diabetes insipidus

AVP-deficiency
- Familial (e.g., DIDMOAD)
- Idiopathic (often autoimmune)
- Tumours:
  - Craniopharyngioma
  - Hypothalamic tumour, e.g., glioma, germinoma
  - Metastases, especially breast
  - Lymphoma/leukaemia
  - Pituitary with suprasellar extension (rare)
- Infections:
  - Tuberculosis
  - Meningitis
  - Cerebral abscess
- Infiltrations:
  - Sarcoidosis
  - Langerhans' cell histiocytosis
- Inflammatory:
  - Hypophysitis
- Post-surgical:
  - Transfrontal
  - Trans-sphenoidal
- Vascular:
  - Haemorrhage/thrombosis
  - Sheehan's syndrome
  - Aneurysm
- Trauma (e.g., head injury)

AVP-resistance
- Familial (e.g., vasopressin receptor gene, aquaporin-2 gene defect)
- Idiopathic
- Renal disease (e.g., renal tubular acidosis)
- Hypokalaemia
- Hypercalcaemia
- Drugs (e.g., lithium, demeclocycline, glibenclamide)
- Sickle cell disease
- Prolonged polyuria due to any cause, including AVP-deficiency and primary polydipsia—can cause mild temporary AVP-resistance

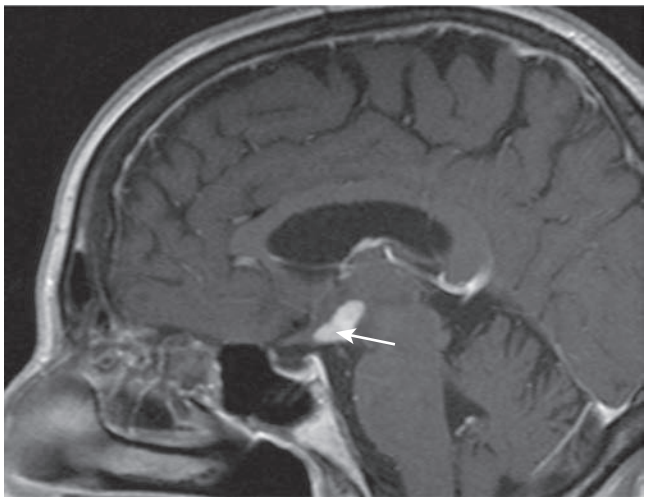


Fig. 34.36 **MRI showing a thickened pituitary stalk in hypophysitis.**

- Resultant high or high-normal plasma sodium (hypernatraemia).
- High 24-hour urine volumes (<2 L excludes the need for further investigation).
- Failure of urinary concentration with fluid deprivation.
- Restoration of urinary concentration with vasopressin or an analogue.

The latter two points are studied with a formal water deprivation test (**Box 34.51**). In normal subjects, plasma osmolality remains normal while urine osmolality rises above 600 mOsm/kg. In DI, plasma osmolality rises while the urine remains dilute, only concentrating after exogenous vasopressin is given (in AVP-deficiency), or not concentrating after vasopressin if AVP-resistance is present. An alternative is measurement of plasma vasopressin during hypertonic saline infusion, but these measurements are not widely available.

Management

The synthetic vasopressin (AVP) analogue, desmopressin, is the treatment of choice in AVP-deficiency. It has a longer duration of

Box 34.51 Water deprivation test

Indications
- Diagnosis or exclusion of diabetes insipidus (DI).

Procedure
- Fasting and no fluids from 07:30 h (or overnight if only mild DI is expected and polyuria is only modest).
- Monitor serum and urine osmolality, urine volume, and weight hourly for up to 8 h.
- Abandon fluid deprivation if weight loss >3% occurs.
- If serum osmolality is >300 mOsm/kg and/or urine osmolality <600 mOsm/kg, give desmopressin 2 μg i.m. at end of test. Allow free fluid but measure urine osmolality for 2–4 h.

Interpretation
- ***Normal response***—serum osmolality remains within normal range (275–295 mOsm/kg). Urine osmolality rises to >600 mOsm/kg.
- ***Diabetes insipidus (DI)***—serum osmolality rises above normal without adequate concentration of urine osmolality (i.e., serum osmolality >300 mOsm/kg; urine osmolality <600 mOsm/kg).
- ***AVP-resistance***—if desmopressin does not concentrate urine.
- ***AVP-deficiency***—if urine osmolality rises by >50% after desmopressin.

action than vasopressin and has no vasoconstrictive effects. It can be given intranasally as a spray 10–40 μg once or twice daily, orally as 100–200 μg three times daily, or intramuscularly 2–4 μg daily. Response is variable and must be monitored carefully with enquiry about fluid input/output and plasma osmolality measurements. The main problem is avoiding water overload and consequent hyponatraemia. Where there is an underlying cause (e.g., a hypothalamic tumour), this should be investigated and treated.

Alternative agents in mild DI, probably working by sensitizing the renal tubules to endogenous vasopressin, include thiazide diuretics, carbamazepine (200–400 mg daily), or chlorpropamide (200–350 mg daily) but these are rarely used.

AVP-resistance

In this condition, renal tubules are resistant to normal or high levels of plasma vasopressin (AVP). AVP-resistance may be inherited as a rare sex-linked recessive disorder, with an abnormality in the vasopressin-2 receptor, or as an autosomal post-receptor defect in an AVP-sensitive water channel, aquaporin-2. More commonly, it can be acquired as a result of renal disease, sickle cell disease, drug ingestion (e.g., lithium), hypercalcaemia, or hypokalaemia. Wherever possible, the cause should be reversed (see **Box 34.50**). Polyuria is helped by thiazide diuretics.

Other causes of polyuria and polydipsia

Diabetes mellitus, hypokalaemia, and hypercalcaemia should be excluded. In the case of diabetes mellitus, the cause is an osmotic diuresis secondary to glycosuria, which leads to dehydration and an increased perception of thirst owing to hypertonicity of the extracellular fluid.

Primary polydipsia

This is a relatively common cause of thirst and polyuria. It is a psychiatric disturbance characterized by the excessive intake of water. Plasma sodium and osmolality fall as a result and the urine produced is appropriately dilute. Vasopressin levels become virtually undetectable. Prolonged primary polydipsia may lead to the phenomenon of 'renal medullary washout', with a fall in the concentrating ability of the kidney.

Characteristically, the diagnosis is made by a water deprivation test. A low plasma osmolality is usual at the start of the test, and since vasopressin secretion and action can be stimulated, the patient's urine becomes concentrated (although 'maximum' concentrating ability may be impaired); the initially low urine osmolality gradually increases with the duration of the water deprivation.

Syndrome of inappropriate antidiuresis (SIAD)

History and examination

Inappropriate secretion of AVP (also called vasopressin) leads to retention of water and hyponatraemia. The presentation of the syndrome of inappropriate antidiuresis (SIAD, also known as the syndrome of inappropriate antidiuretic hormone, SIADH) is usually vague, with confusion, nausea, irritability, and, later, fits and coma. There is no oedema. Mild symptoms usually occur with plasma sodium levels below 125 mmol/L and serious manifestations are likely below 115 mmol/L. The elderly may show symptoms with milder abnormalities.

The syndrome must be distinguished from dilutional hyponatraemia due to excess infusion of glucose/water solutions or diuretic administration. The clinical approach to the patient with hyponatraemia is discussed further in Chapter 27.

Investigations

The usual features are:
- dilutional hyponatraemia due to excessive water retention
- euvolaemia (in contrast to hypovolaemia of sodium and water depletion states)
- low plasma osmolality with 'inappropriate' urine osmolality >100 mOsm/kg (and typically higher than plasma osmolality)
- continued urinary sodium excretion >30 mmol/L (lower levels suggest sodium depletion or 'hypovolaemic hyponatraemia' and should respond to 0.9% saline infusion)
- absence of hypokalaemia (or hypotension)
- normal renal and adrenal and thyroid function.

The causes are listed in **Box 34.52**.

Hyponatraemia is very common during illness in frail elderly patients and it may sometimes be clinically difficult to distinguish SIAD from salt and water depletion, particularly when mixed clinical features are present. Under these circumstances, a trial infusion of 1–2 L 0.9% saline is given. SIAD will not respond (but will excrete the sodium and water load effectively); sodium depletion will respond. ACTH deficiency can give a very similar biochemical picture to SIAD; therefore it is necessary to ensure that the HPA axis is intact, particularly in neurosurgical patients, in whom ACTH deficiency may be relatively common.

Management

The underlying cause should be corrected where possible. Symptomatic relief can be obtained by the following measures:
- ***Fluid intake*** should be restricted to 500–1000 mL daily. If tolerated and complied with, this will correct the biochemical abnormalities in almost every case.
- ***Frequent measurement*** of plasma osmolality, serum sodium, and body weight is needed.
- ***Demeclocycline*** (600–1200 mg daily) is given if water restriction is poorly tolerated or ineffective; this inhibits the action of vasopressin on the kidney, causing a reversible form of AVP-resistance. However, it often causes photosensitive rashes.
- ***Hypertonic saline*** may be indicated when the syndrome is very severe (i.e., acute and symptomatic), but this is potentially dangerous and should only be used with extreme caution.

Box 34.52 Common causes of the syndrome of inappropriate antidiuresis (SIAD)

Tumours
- Small-cell carcinoma of lung
- Prostate
- Thymus
- Pancreas
- Lymphomas

Pulmonary lesions
- Pneumonia
- Tuberculosis
- Lung abscess

Central nervous system causes
- Meningitis
- Tumours
- Head injury
- Subdural haematoma
- Cerebral abscess
- Systemic lupus erythematosus
- Vasculitis

Metabolic causes
- Alcohol withdrawal
- Porphyria

Drugs
- Chlorpropamide
- Carbamazepine
- Cyclophosphamide
- Vincristine
- Phenothiazines

- ***Vasopressin V_2 antagonists***, e.g., tolvaptan 15 mg daily, are being used with good results.

Further reading

Atila C, Loughrey PB, Garrahy A et al. Central diabetes insipidus from a patient's perspective: management, psychological co-morbidities, and renaming of the condition: results from an international web-based survey. *Lancet Diabetes Endocrinol* 2022; 10(10):700–709.
Fleseriu M, Biller BMK, Freda PUA et al. Pituitary society update to acromegaly management guidelines. *Pituitary* 2021; 24(1):1–13.
Spasovski G, Vanholder R, Allolio B et al. Clinical practice guideline on diagnosis and treatment of hyponatraemia. *Eur J Endocrinol* 2014; 170:G1–G47.

DISORDERS OF CALCIUM METABOLISM

Serum calcium levels are mainly controlled by parathyroid hormone (PTH) and vitamin D. Hypercalcaemia is much more common than hypocalcaemia and is frequently detected incidentally with multichannel biochemical analysers. Mild asymptomatic hypercalcaemia occurs in about 1 in 1000 of the population, with an incidence of 25–30 per 100 000 population. It occurs mainly in elderly females and is usually due to primary hyperparathyroidism.

Parathyroid hormone

There are normally four parathyroid glands, which are situated posterior to the thyroid; occasionally, additional glands exist or they may be found elsewhere in the neck or mediastinum. PTH, an 84-amino-acid hormone derived from a 115-residue pre-prohormone, is secreted from the chief cells of the parathyroid glands. PTH levels rise as serum ionized calcium falls. The latter is detected by specific G-protein-coupled, calcium-sensing receptors on the plasma membrane of the parathyroid cells. PTH has several major actions, all serving to increase plasma calcium by:
- increasing osteoclastic resorption of bone (occurring rapidly)
- increasing intestinal absorption of calcium (a slow response)
- increasing synthesis of 1,25-dihydroxyvitamin D_3
- increasing renal tubular reabsorption of calcium
- increasing excretion of phosphate.

PTH effects are mediated at specific membrane receptors on the target cells, resulting in an increase of adenyl cyclase messenger activity.

Vitamin D metabolism is discussed in Chapter 33 ('Vitamin D metabolism' section).

PTH measurements use two-site immunometric assays that measure only the intact PTH molecule; interpretation requires a simultaneous calcium measurement in order to differentiate most causes of hyper- and hypocalcaemia.

Box 34.53 Causes of hypercalcaemia

Excessive parathyroid hormone (PTH) secretion
- Primary hyperparathyroidism (most common by far), adenoma (common), hyperplasia, or carcinoma (rare)
- Tertiary hyperparathyroidism
- Ectopic PTH secretion (very rare indeed)

Malignant disease—low PTH levels (second most common cause)
- Myeloma
- Secondary deposits in bone
- Production of osteoclastic factors by tumours
- PTH-related protein secretion

Excess action of vitamin D
- Iatrogenic or self-administered excess
- Granulomatous diseases, e.g., sarcoidosis, tuberculosis
- Lymphoma

Excessive calcium intake
- 'Milk-alkali' syndrome

Other endocrine disease (mild hypercalcaemia only)
- Thyrotoxicosis
- Addison's disease

Drugs
- Thiazide diuretics
- Vitamin D analogues
- Lithium administration (chronic)
- Vitamin A

Miscellaneous
- Long-term immobility
- Familial hypocalciuric hypercalcaemia

Hypercalcaemia

Pathophysiology and aetiology

The major causes of hypercalcaemia are listed in **Box 34.53**; primary hyperparathyroidism and malignancies are by far the most common (>90% of cases). Hyperparathyroidism itself may be primary, secondary or tertiary. Primary hyperparathyroidism is caused by single (>80%) parathyroid adenomas or by diffuse hyperplasia of all the glands (15%–20%); multiple parathyroid adenomas are rare. Involvement of multiple parathyroid glands may be part of a familial syndrome (e.g., MEN type 1 or 2a). Parathyroid carcinoma is rare (<1%), though it usually produces severe and intractable hypercalcaemia. Hyperparathyroidism–jaw tumour syndrome is a rare familial cause of hyperparathyroidism that may be associated with parathyroid carcinoma and maxillary or mandibular tumours.

Primary hyperparathyroidism

Primary hyperparathyroidism is of unknown cause, though it appears that adenomas are monoclonal. Hyperplasia may also be monoclonal. Chromosomal rearrangements in the 5′ regulatory region of the parathyroid hormone gene have been identified, and inactivation of some tumour suppressor genes at a variety of sites may also be involved.

Secondary hyperparathyroidism

Secondary hyperparathyroidism (see Chapter 26, 'CKD mineral and bone disorder [CKD-MBD]' section) is physiological compensatory hypertrophy of all parathyroids because of hypocalcaemia, such as occurs in chronic kidney disease or vitamin D deficiency. PTH levels are raised but calcium levels are low or normal, and PTH falls to

normal after correction of the cause of hypocalcaemia where this is possible.

Tertiary hyperparathyroidism

Tertiary hyperparathyroidism is the development of apparently autonomous parathyroid hyperplasia after longstanding secondary hyperparathyroidism, most often in renal failure. Plasma calcium and phosphate are both raised, the latter often grossly so. Use of calcimimetic medications, or surgical parathyroidectomy, is generally necessary at this stage.

Clinical features

Mild hypercalcaemia

Mild hypercalcaemia (e.g., adjusted calcium <3 mmol/L) is frequently asymptomatic, but more severe hypercalcaemia can produce a number of symptoms:

- ***General.*** There may be tiredness, malaise, dehydration, and depression.
- ***Renal.*** Renal colic occurs from stones, and primary hyperparathyroidism is present in about 5% of patients who present with renal calculi. Polyuria or nocturia, haematuria, and hypertension are seen. The polyuria results from the effect of hypercalcaemia on renal tubules, reducing their concentrating ability—a form of mild AVP-resistance.
- ***Bones.*** There may be bone pain. Hyperparathyroidism mainly affects cortical bone, and bone cysts and locally destructive 'brown tumours' occur but only in advanced disease. Only 5%–10% of all cases have definite bony lesions even when sought. Bone disease may be more apparent when there is coexisting vitamin D deficiency.
- ***Abdomen.*** There may be abdominal pain.
- ***Chondrocalcinosis*** and ectopic calcification. These are occasional features.
- ***Corneal calcification.*** This is a marker of longstanding hypercalcaemia but causes no symptoms.

There may also be symptoms from the underlying cause. Malignant disease is usually advanced by the time hypercalcaemia occurs, typically with bony metastases. The common primary tumours are bronchus, breast, myeloma, oesophagus, thyroid, prostate, lymphoma, and renal cell carcinoma. True 'ectopic PTH secretion' by tumours is very rare, and most cases are associated with raised levels of PTH-related protein. This is a 144-amino-acid polypeptide that is necessary in fetal development but does not have a clearly defined role in the adult, the initial sequence of which shows an approximate homology with the biologically active part of PTH. Bone-resorbing cytokines and prostaglandins may be involved locally where there are metastatic skeletal lesions, leading to local mobilization of calcium by osteolysis with subsequent hypercalcaemia.

Severe hypercalcaemia

Severe hypercalcaemia (>3 mmol/L) is usually associated with malignant disease, hyperparathyroidism, chronic kidney disease, or vitamin D therapy.

Investigations and differential diagnosis

Biochemistry

A combination of serum calcium, phosphate, PTH, and vitamin D concentrations is required to define the clinical phenotype.

- ***Serum PTH.*** The hallmark of primary hyperparathyroidism is hypercalcaemia and hypophosphataemia with detectable or elevated intact PTH levels during hypercalcaemia. When this combination is present in an asymptomatic patient, then further investigation is usually unnecessary. However, undetectable PTH level in the context of hypercalcaemia always requires further investigation to exclude malignancy or other pathology (see **Box 34.53**).
- ***Renal function.*** This is usually normal but should be measured as a baseline. Advancing CKD leads to the development first of secondary and then tertiary hyperparathyroidism, as the PTH glands initially function physiologically in the face of hypocalcaemia associated with CKD, before then demonstrating hypertrophy and autonomous secretion in the face of a rising serum calcium.
- ***24-h urinary calcium*** or single calcium creatinine ratio. This should be measured in a young patient with modest elevation in calcium and PTH to exclude familial hypocalciuric hypercalcaemia (see Chapter 26, 'Hypercalciuria' section).
- ***Elevated serum alkaline phosphatase.*** This is found in severe parathyroid bone disease, but otherwise, it suggests an alternative cause for hypercalcaemia.

Where PTH is undetectable or equivocal, a number of other tests may lead to the diagnosis:

- Protein electrophoresis and serum light chain immunofixation: to exclude myeloma.
- Serum TSH: to exclude hyperthyroidism.
- 9 a.m. cortisol and/or ACTH test: to exclude Addison's disease.
- Serum ACE: may be helpful in the diagnosis of sarcoidosis.

Imaging

The success of parathyroid imaging is highly operator-dependent and choice therefore depends on local skills and experience. Imaging is frequently far less accurate than parathyroid exploration by an expert surgeon, where the success rate is at least 90%. Methods include:

- Radioisotope scanning using ^{99m}Tc-sestamibi, which is approximately 90% sensitive in detecting adenomas; it is enhanced with single-photon emission computed tomography (SPECT). Choline-PET scanning is a newer technique which may prove superior to more traditional approaches.
- Ultrasound, which is simple and safe, although insensitive for small tumours.
- High-resolution CT scan or MRI (more sensitive). A CT scan of the thorax, abdomen, and pelvis will often show malignancy or granulomatous disease in patients with hypercalcaemia and a suppressed PTH level.

Dual-energy X-ray absorptiometry (DXA) bone density scanning is useful to detect bone effects in asymptomatic people with hyperparathyroidism in whom conservative management is planned.

Abdominal X-rays may show renal calculi or nephrocalcinosis. High-definition hand X-rays can show subperiosteal erosions in the middle or terminal phalanges. Neither is required as part of diagnosis unless symptoms of renal calculi are present or there is evidence of renal failure.

Management of hypercalcaemia

Details of emergency treatment for severe hypercalcaemia are given in **Box 34.54**. This should be followed by oral therapy unless the underlying disease can be treated.

Management of primary hyperparathyroidism

Medical management

A high fluid intake should be maintained. Calcimimetic agents (that target the calcium-sensing receptors in the parathyroid gland, e.g.,

Box 34.54 Management of acute severe hypercalcaemia

Acute hypercalcaemia often presents with dehydration, nausea and vomiting, nocturia and polyuria, drowsiness, and altered consciousness. The serum Ca^{2+} is >3 mmol/L and sometimes as high as 5 mmol/L. While investigation of the cause is underway, immediate treatment is mandatory if the patient is seriously ill or if the Ca^{2+} is >3.5 mmol/L.

- ***Rehydrate*** using at least 4–6 L of 0.9% saline on day 1, and 3–4 L for several days thereafter. Central venous pressure (CVP) may need to be monitored to control the hydration rate.
- ***Intravenous bisphosphonates*** are the treatment of choice for the hypercalcaemia of malignancy or of undiagnosed cause. Pamidronate is preferred (60–90 mg as an intravenous infusion in 0.9% saline or glucose over 2–4 h or, if less urgent, over 2–4 days). Levels fall after 24–72 h, lasting for approximately 2 weeks. Zoledronic acid is an alternative.
- ***Prednisolone*** (30–60 mg daily) is effective in some instances (e.g., in myeloma, sarcoidosis and vitamin D excess) but in most cases is ineffective.
- ***Calcitonin*** (200 units i.v. 6-hourly) has a short-lived action and is little used.
- ***Oral phosphate*** (sodium cellulose phosphate 5 g three times daily) produces diarrhoea.

cinacalcet) are of proven value in parathyroid carcinoma and in dialysis patients (see Chapter 26, 'CKD mineral and bone disorder [CKD-MBD]' section) and are increasingly used in primary hyperparathyroidism as an alternative to surgical intervention.

Surgery

Surgery is indicated in primary hyperparathyroidism for:

- symptoms of hypercalcaemia such as thirst, frequent or excessive urination, or constipation or
- end-organ disease (renal stones, fragility fractures, or osteoporosis) or
- an albumin-adjusted serum calcium level of 2.85 mmol/L or above.

The situation where plasma calcium is mildly raised (2.65–2.85 mmol/L) is more controversial. Most authorities feel that young patients should be operated on, as should those who have reduced cortical bone density or significant hypercalciuria, as this is associated with stone formation.

In older patients without these problems, or in those unfit for or unwilling to have surgery, conservative management is indicated. Regular measurement of serum calcium and of renal function is necessary. Bone density of cortical bone should be monitored if conservative management is used. Hyperparathyroidism can cause non-specific symptoms of weakness, fatigue, and depression, and it can be difficult to determine whether these symptoms are related to hypercalcaemia or coincidental.

Surgical technique and complications

Parathyroid surgery should be performed only by experienced surgeons, as the minute glands may be very difficult to define, and it is difficult to distinguish between an adenoma and normal parathyroid. In expert centres, over 90% of operations are successful, involving removal of the adenoma or removal of all four hyperplastic parathyroids. Minimal access surgery is used, and some centres measure PTH levels intraoperatively to ensure that the adenoma has been removed.

Other than postoperative hypocalcaemia (see below), the other rare complications are those of thyroid surgery: bleeding and recurrent laryngeal nerve palsies (<1%). Vocal cord function should be checked preoperatively.

If initial exploration is unsuccessful, a full work-up, including venous catheterization and detailed scanning, is essential, remembering that parathyroid tissue can be ectopic.

Postoperative care

The major danger after operation is hypocalcaemia, which is more common in patients who have significant bone disease and/or vitamin D deficiency—the 'hungry bone' syndrome. Some authorities pre-treat such patients with alfacalcidol 2 μg daily from 2 days preoperatively for 10–14 days, and routine vitamin D replacement (preferably without calcium) is always indicated if deficiency is diagnosed. Chvostek's and Trousseau's signs (see below) are monitored, as well as biochemistry. Plasma calcium measurements are performed at least daily until stable, with or without replacement; a mild transient hypoparathyroidism often continues for 1–2 weeks. Depending on its severity, oral or intravenous calcium should be given temporarily, as only a few patients (<1%) will develop long-standing surgical hypoparathyroidism.

Familial hypocalciuric hypercalcaemia

This uncommon autosomal dominant, and usually asymptomatic, condition demonstrates increased renal reabsorption of calcium despite hypercalcaemia. PTH levels are normal or slightly raised and urinary calcium is low. Familial hypocalciuric hypercalcaemia is caused by loss-of-function mutations in the gene on the long arm of chromosome 3 encoding for the calcium ion–sensing G-protein-coupled receptor in the kidney and parathyroid gland. Family members are often affected and this is detected by genetic analysis. The course is generally benign. This diagnosis can be differentiated from hyperparathyroidism in an isolated case by the calcium:creatinine ratio in blood and urine.

Hypocalcaemia

Pathophysiology

Hypocalcaemia may be due to deficiencies of calcium homeostatic mechanisms, or secondary to high phosphate levels or other causes of hypocalcaemia (Box 34.55). All forms of hypoparathyroidism, except transient surgical effects, are uncommon.

Aetiology

- ***Chronic kidney disease*** is the most common cause of hypocalcaemia.
- ***Severe vitamin D deficiency*** may cause mild, and occasionally severe, hypocalcaemia.
- ***Hypocalcaemia after thyroid or parathyroid surgery*** is common but usually transient; fewer than 1% of thyroidectomies leave permanent damage (see above).
- ***Idiopathic hypoparathyroidism*** is one of the rarer autoimmune disorders and is often accompanied by vitiligo, cutaneous candidiasis, and other autoimmune disease.
- ***DiGeorge's syndrome*** is a familial condition in which hypoparathyroidism is associated with intellectual impairment, cataracts, and calcified basal ganglia, and occasionally with specific autoimmune disease.
- ***Pseudohypoparathyroidism*** is a syndrome of end-organ resistance to PTH owing to a mutation in the $G_S\alpha$-protein (GNAS1), which is coupled to the PTH receptor. It is associated with short stature, short metacarpals, subcutaneous calcification, and sometimes intellectual impairment. Variable degrees of resistance involving other G-protein-linked hormone receptors may also be seen (TSH, LH, FSH).
- ***Pseudo-pseudohypoparathyroidism*** describes the phenotypic defects but without any abnormalities of calcium metabolism.

Box 34.55 Causes of hypocalcaemia

Increased phosphate levels
- Chronic kidney disease (common)
- Phosphate therapy

Hypoparathyroidism
- Surgical—after neck exploration (thyroidectomy, parathyroidectomy—common)
- Congenital deficiency (DiGeorge's syndrome)
- Idiopathic hypoparathyroidism (rare)
- Severe hypomagnesaemia

Vitamin D deficiency
- Osteomalacia/rickets
- Vitamin D resistance

Resistance to PTH
- Pseudohypoparathyroidism

Drugs
- Calcitonin
- Bisphosphonates
- Other
  - Acute pancreatitis (quite common)
  - Citrated blood in massive transfusion (not uncommon)
  - Low plasma albumin, e.g., malnutrition, chronic liver disease
  - Malabsorption, e.g., coeliac disease

Individuals with this condition may share the same gene defect as those with pseudohypoparathyroidism and be members of the same families.

Clinical features

Hypoparathyroidism presents as neuromuscular irritability and neuropsychiatric manifestations. Paraesthesiae, circumoral numbness, cramps, anxiety, and tetany (Box 34.56) are followed by convulsions, laryngeal stridor, dystonia, and psychosis. Two signs of hypocalcaemia are Chvostek's sign (gentle tapping over the facial nerve causes twitching of the ipsilateral facial muscles) and Trousseau's sign (inflation of the sphygmomanometer cuff above systolic pressure for 3 min induces tetanic spasm of the fingers and wrist). Severe hypocalcaemia may cause papilloedema and, frequently, a prolonged QT interval on the ECG.

Investigations

The clinical history and picture is usually diagnostic and is confirmed by a low serum calcium (after correction for any albumin abnormality). Additional tests include:
- ***Serum and urine creatinine.*** These test for renal disease.
- ***PTH levels*** in the serum. These are absent or inappropriately low in hypoparathyroidism, high in other causes of hypocalcaemia.
- ***Parathyroid antibodies.*** These are present in idiopathic hypoparathyroidism.
- ***25-hydroxyvitamin D serum level.*** This is low in vitamin D deficiency.
- ***Magnesium level.*** Severe hypomagnesaemia results in functional hypoparathyroidism, which is reversed by magnesium replacement.
- ***X-rays of metacarpals.*** Short fourth metacarpals occur in pseudohypoparathyroidism.

Management

In vitamin D deficiency, colecalciferol is the most appropriate treatment. In other cases, alpha-hydroxylated derivatives of vitamin D are preferred for their shorter half-life, and especially in renal disease, as the others require renal hydroxylation. Usual daily maintenance doses are 0.25–2 μg for alfacalcidol (1α-OH-D_3). During treatment, plasma calcium must be monitored frequently to detect hypercalcaemia. Oral calcium supplements may be used in the early stages of treatment, and severe hypocalcaemia presenting as an emergency may occasionally require replacement with intravenous calcium gluconate. In hypoparathyroidism, PTH therapy is being used in clinical trials.

Box 34.56 Causes of tetany

In the presence of alkalosis
- Hyperventilation
- Excess antacid therapy
- Persistent vomiting
- Hypochloraemic alkalosis, e.g., primary hyperaldosteronism

In the presence of hypocalcaemia
See Box 34.54

Further reading

Bilezikian JP, Bandeira L, Khan A et al. Hyperparathyroidism. *Lancet* 2018; 391(10116):168–178.
Bilezikian JP, Khan AA, Silverberg SJ et al. Evaluation and Management of Primary Hyperparathyroidism: summary statement and guidelines from the fifth International workshop. International Workshop on Primary Hyperparathyroidism. *J Bone Miner Res* 2022; 37(11):2293–2314.
Khan AA, Bilezikian JP, Brandi ML et al. Evaluation and Management of Hypoparathyroidism Summary Statement and guidelines from the second International workshop. *J Bone Miner Res* 2022; 37(12):2568–2585.
Society for Endocrinology Endocrine Emergency Guidance. *Emergency Management of Acute Hypercalcaemia in Adult Patients* 2016.

HYPOGLYCAEMIA IN THE NON-DIABETIC PATIENT

Hypoglycaemia develops when hepatic glucose output falls below the rate of glucose uptake by peripheral tissues. Hepatic glucose output may be reduced by:
- inhibition of hepatic glycogenolysis and gluconeogenesis by insulin
- depletion of hepatic glycogen reserves by malnutrition, fasting, exercise, or advanced liver disease
- impaired gluconeogenesis (e.g., following alcohol ingestion).

In the first of these categories, insulin levels are raised, the liver contains adequate glycogen stores, and the hypoglycaemia can be reversed by injection of glucagon. In the other two situations, insulin levels are low and glucagon is ineffective. Peripheral glucose uptake is accelerated by high insulin levels and by exercise, but these conditions are normally balanced by increased hepatic glucose output.

The most common symptoms and signs of hypoglycaemia are neurological. The brain consumes about 50% of the total glucose produced by the liver. This high energy requirement is needed to generate ATP, used to maintain the potential difference across axonal membranes.

Insulinomas

Insulinomas are pancreatic islet cell tumours that secrete insulin. Most are sporadic but some patients have multiple tumours arising from neural crest tissue (MEN). Some 95% of these tumours are benign. The classic presentation is with fasting hypoglycaemia, but early symptoms may also develop in the late morning or afternoon. Recurrent hypoglycaemia is often present for months or years before the diagnosis is made, and the symptoms may be atypical or even bizarre; the presenting features in one series are given in Box 34.57. Common misdiagnoses include psychiatric disorders,

Box 34.57 Presenting features of insulinoma

- Diplopia
- Sweating, palpitations, weakness
- Confusion or abnormal behaviour
- Loss of consciousness
- Grand mal seizures

particularly pseudodementia in elderly people, epilepsy, and cerebrovascular disease.

Whipple's triad remains the basis of clinical diagnosis. This is satisfied when:

- symptoms are associated with fasting or exercise
- hypoglycaemia is confirmed during these episodes
- glucose relieves the symptoms.

A fourth criterion—demonstration of inappropriately high insulin levels during hypoglycaemia—may usefully be added.

The diagnosis is confirmed by the demonstration of hypoglycaemia in association with inappropriate and excessive insulin secretion. Hypoglycaemia is demonstrated by:

- Measurement of overnight fasting (16 h) glucose and insulin levels on three occasions. About 90% of patients with insulinomas will have low glucose and non-suppressed (normal or elevated) insulin levels.
- A prolonged 72-hour supervised fast if overnight testing is inconclusive and symptoms persist.

Autonomous insulin secretion is demonstrated by lack of normal feedback suppression during hypoglycaemia. This may be shown by measuring insulin, C-peptide, or proinsulin during a spontaneous episode of hypoglycaemia.

Management

The most effective therapy is surgical excision of the tumour but insulinomas are often very small and difficult to localize. Many techniques can be used to attempt to localize insulinomas. Sensitivity and specificity vary between centres and between operators. These include highly selective angiography, contrast-enhanced high-resolution CT scanning, scanning with radiolabelled somatostatin (some insulinomas express somatostatin receptors), and endoscopic and intraoperative ultrasound scanning. Venous sampling for the detection of 'hot spots' of high insulin concentration in the various intra-abdominal veins is still used occasionally.

Medical treatment with diazoxide is useful when the insulinoma is malignant, when a tumour cannot be located, and when elderly patients have mild symptoms. Symptoms may also remit on treatment with a somatostatin analogue (octreotide or lanreotide).

Hypoglycaemia with other tumours

Hypoglycaemia may develop in the course of advanced neoplasia and cachexia and has been described in association with mesenchymal tumours such as sarcomas including tumours involving gastrointestinal tract, classically termed ***non-islet cell tumour hypoglycaemia*****.** Hypoglycaemia is due to excessive secretion of IGF-2 and pro-IGF-2 leading to activation of insulin receptor. True ectopic insulin secretion is extremely rare.

Postprandial hypoglycaemia

If frequent venous blood glucose samples are taken following a prolonged glucose tolerance test, about 1 in 4 subjects will have at least one value below 3 mmol/L. The arteriovenous glucose difference is quite marked during this phase, so very few are truly hypoglycaemic in terms of arterial (or capillary) blood glucose content. Failure to appreciate this simple fact led some authorities to believe that postprandial (or reactive) hypoglycaemia was a potential 'organic' explanation for a variety of complaints that might otherwise have been considered psychosomatic. Later work showed a poor correlation between symptoms and biochemical hypoglycaemia. Even so, a number of otherwise normal people occasionally become pale, weak, and sweaty at times when meals are due and report benefit from advice to take regular snacks between meals.

True postprandial hypoglycaemia may develop in the presence of alcohol, which 'primes' the cells to produce an exaggerated insulin response to carbohydrate. The person who substitutes alcoholic beverages for lunch is particularly at risk. Postprandial hypoglycaemia sometimes occurs after gastric surgery, owing to rapid gastric emptying and mismatching of nutrient absorption and insulin secretion. This is referred to as 'dumping' but it is now rarely encountered.

Hepatic and renal causes of hypoglycaemia

The liver can maintain a normal glucose output despite extensive damage, and hepatic hypoglycaemia is uncommon. It is, however, a particular problem with acute hepatic failure.

The kidney has a subsidiary role in glucose production (via gluconeogenesis in the renal cortex), and hypoglycaemia is sometimes a problem in terminal renal failure.

Endocrine causes of hypoglycaemia

Deficiencies of hormones antagonistic to insulin are rare but well-recognized causes of hypoglycaemia. These include hypopituitarism, isolated ACTH deficiency, and Addison's disease.

Drug-induced hypoglycaemia

Many drugs have been reported to produce isolated cases of hypoglycaemia, including sulfonylureas, quinine (during treatment for falciparum malaria), and propranolol.

Alcohol-induced hypoglycaemia

Alcohol inhibits gluconeogenesis. Alcohol-induced hypoglycaemia occurs in poorly nourished chronic alcohol users, binge drinkers, and children who have taken relatively small amounts of alcohol, since they have a diminished hepatic glycogen reserve. They present with coma and hypothermia (hypothermia is a feature of hypoglycaemia, due to the suppression of central thermoregulation, particularly the shivering response; children manifest hypothermia more frequently due to their high ratio of surface area to body mass).

Factitious hypoglycaemia

This is a relatively common variant of self-induced disease and is more common than an insulinoma. Hypoglycaemia may be produced by surreptitious self-administration of insulin or sulphonylureas. Many patients in this category have been extensively investigated for an insulinoma. Measurement of C-peptide levels during hypoglycaemia should identify patients who are injecting insulin; sulphonylurea abuse can be detected by chromatography of plasma or urine.

OTHER ENDOCRINE DISORDERS

Diseases of many glands

Polyglandular autoimmune syndromes

Primary endocrine gland failure is commonly caused by autoantibodies (see **Box 34.4**) and there may be multiple gland deficiencies. Most common are the associations of primary hypothyroidism and type 1 diabetes, and either of these with Addison's disease or pernicious anaemia. IgG4 disease may affect multiple endocrine glands (see Chapter 6).

Autoimmune polyendocrinopathy type 1

Autoimmune polyendocrinopathy type 1 (APS-1) is an autosomal recessive disorder and is caused by *AIRE* gene mutations. This condition is also associated with non-endocrine manifestations. The *AIRE* gene is present in the epithelium of the thymus and is involved in the presentation of self-antigens to thymocytes. Mutations will allow persistence of thymic lymphocytes, which react against self-antigens and cause development of autoimmune disorders. Mucocutaneous candidiasis often develops before the onset of endocrine deficiencies, such as hypothyroidism, Addison's disease, type 1 diabetes, hypoparathyroidism, primary hypogonadism, nail dystrophy, vitiligo, and dental enamel hypoplasia.

Autoimmune polyendocrinopathy type 2

Autoimmune polyendocrinopathy type 2 (APS-2) is not associated with candidiasis and is also known as Schmidt's syndrome, typically when hypothyroidism, Addison's disease, type 1 diabetes, myasthenia gravis, and primary hypogonadism are present in combination; coeliac disease is also an association.

Multiple endocrine neoplasias

Multiple endocrine neoplasia (MEN) is the name given to the simultaneous or metachronous occurrence of tumours involving a number of endocrine glands (**Box 34.58**). The condition is inherited in an autosomal dominant manner and arises from the expression of recessive oncogenic mutations, most of which have been isolated. Affected persons may pass on the mutation to their offspring in the germ cell, but for the disease to become evident, a somatic mutation must also occur, such as deletion or loss of a normal homologous chromosome.

MEN 1

The defect in MEN 1 is in a novel gene *(menin)* on the long arm of chromosome 11, which encodes for a 610-amino-acid protein. *Menin* represses a transcription factor *(JunD)*, and lack of *JunD* suppression leads to decreased apoptosis and oncogenesis. People with the *MEN1* gene carry one mutant gene and a wild-type gene (i.e., are heterozygous). When the wild-type gene undergoes a random somatic mutation during life, this leads to loss of heterozygosity and explains the late onset of tumours at any stage (the 'two hit' hypothesis). MEN 1 is classically associated with pancreatic, parathyroid, and pituitary tumours, although other glands may be affected (see **Box 34.58**; **Fig. 34.37**).

MEN 2a and 2b

MEN 2a and 2b are caused by mutations of the *RET* proto-oncogene on chromosome 10 (see medullary thyroid cancer, 'Medullary carcinoma' section). This gene encodes for a transmembrane glycoprotein receptor. For MEN 2a, the mutation is in the extracellular domain; for 2b, it is in the intracellular domain. MEN 2 is classically associated with parathyroid tumours, phaeochromocytoma, and MTC (see **Box 34.58**). Unlike MEN 2a, MEN 2b is associated with a marfanoid phenotype and intestinal and visceral ganglioneuromas, as well as neuromas around the lips and tongue.

Box 34.58 Multiple endocrine neoplasia (MEN) syndromes

Organ	Frequency	Tumours/manifestations
Type 1		
Parathyroid	95%	Adenomas/hyperplasia
Pituitary	70%	Adenomas—prolactinoma, ACTH- or GH-secreting (acromegaly)
Pancreas	50%	Islet cell tumours (secreting insulin, glucagon, somatostatin, VIP, pancreatic polypeptide, GH-releasing factor) Zollinger–Ellison syndrome (gastrinoma) Non-functional tumour
Adrenal	40%	Non-functional adenoma
Thyroid	20%	Adenomas—multiple or single
Type 2a		
Adrenal	Most	Phaeochromocytoma (70% bilateral) Cushing's syndrome
Thyroid	Most	Medullary carcinoma (calcitonin-producing)
Parathyroid	60%	Hyperplasia
Type 2b		
Type 2a with marfanoid phenotype and intestinal and visceral ganglioneuromas but not hyperparathyroidism. Neuromas also present around lips and tongue		

ACTH, Adrenocorticotrophic hormone; *GH*, growth hormone; *VIP*, vasoactive intestinal peptide.

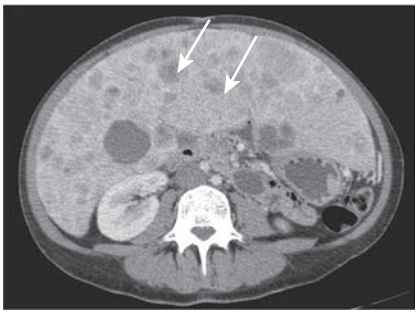


Fig. 34.37 Multiple liver metastases *(arrowed)* from a pancreatic endocrine tumour in multiple endocrine neoplasia type 1.

Management

Management of established tumours in MEN is largely the same as treatment for similar tumours occurring sporadically. In MEN 1, four-gland parathyroidectomy is usually recommended when surgery is needed since all glands are typically involved. However, the essence of management in MEN is annual screening to detect tumours at an early, treatable stage.

Screening

A careful family history is essential. If the precise gene mutation has been identified in a particular family, then family members at risk

can be offered genetic screening for the presence of the mutation, ideally in childhood. In affected individuals, biochemical screening and periodic imaging are then required.

Screening for MEN 1

Hyperparathyroidism is usually the first manifestation, and serum calcium is the simplest screening test in families with no identified mutation. In an established case (or gene-positive family member), other screening bloods include prolactin, GH/IGF-1, and 'gut hormones'. Periodic imaging of pancreas, adrenals, and pituitary is usually performed. People with MEN 1 can develop metastases to the liver from non-functional pancreatic tumours that are clinically silent; this emphasizes the need for regular screening imaging.

Screening for MEN 2

Serum calcium levels will easily detect hyperparathyroidism.

- ***Medullary carcinoma of thyroid (MCT).*** With the known presence of the gene defect, total thyroidectomy is recommended in early childhood or as soon as the gene defect is identified. Calcitonin is a useful tumour marker.
- ***Phaeochromocytoma.*** Metanephrine or catecholamine estimations are required.

McCune–Albright syndrome

This condition is associated with autonomous hypersecretion of a number of endocrine glands at a young age. Gonadotrophin-independent puberty occurs, with Leydig cell hyperplasia in males and ovarian oestrogen production in girls. Pituitary hypersecretion may lead to hyperprolactinaemia, acromegaly, or gigantism. Cushing's syndrome due to nodular hyperplasia of the adrenal cortex is observed, as well as autonomous functioning thyroid nodules. Non-endocrine manifestations include café-au-lait patches and increased bone deformity and fractures due to polyostotic fibrous dysplasia. The pathological basis is a point mutation of the *GNAS1* gene that inhibits GTPase activity, leading to persistent activation of cAMP-mediated endocrine secretion.

Ectopic hormone secretion

This term refers to hormone synthesis, and normally secretion, from a neoplastic non-endocrine cell, most usually seen in tumours that have some degree of embryological resemblance to specialist endocrine cells. The clinical effects are those of the hormone produced, with or without manifestations of systemic malignancy. The most common situations seen are the following:

- ***Hypercalcaemia of malignant disease***, often from squamous cell tumours of lung and breast, often with bone metastases. Where metastases are not present, most cases are mediated by secretion of PTH-related protein (PTHrP), which has considerable sequence homology to PTH; a variety of other factors may sometimes be involved but very rarely PTH itself. Treatment is discussed in Chapter 11 ('Acute hypercalcaemia' section).
- ***SIAD*** (see 'Syndrome of inappropriate antidiuresis [SIAD]' section). Again, this is most commonly caused by a primary lung tumour.
- ***Ectopic ACTH syndrome*** (see 'The adrenal cortex' section). Small-cell carcinoma of the lung, carcinoid tumours, and MTCs are the most common causes, though many other tumours rarely cause it.
- ***Production of insulin-like activity.*** This may result in hypoglycaemia.
- ***Carcinoid syndrome.*** This is the collection of symptoms (diarrhoea, abdominal pain and loss of appetite; flushing of the skin, particularly the face; tachycardia; breathlessness and wheezing) when a neuroendocrine tumour, usually one that has spread to the liver, releases hormones such as serotonin into the bloodstream. Measurement of 24-hour urinary 5HIAAs is required for diagnosis.

Bibliography

Jameson JL, DeGroot L. *Endocrinology*, 7th edn. Philadelphia: WB Saunders; 2015.

Pieterman CRC, Valk GD. Update on the clinical management of multiple endocrine neoplasia type 1. *Clin Endocrinol* (Oxf) 2022; 97(4):409–423.

Significant websites

http://www.addisons.org.uk/ *Addison's Self Help Group (UK): information and guidelines on Addison's and steroid replacement.*
http://www.endocrine.niddk.nih.gov/ *US National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases.*
http://www.endocrineweb.com *Endocrine web resources.*
http://www.endocrinology.org *UK Society for Endocrinology.*
http://www.endo-society.org *Endocrine Society.*
http://www.endotext.org/ *Online endocrinology textbook.*
http://www.medicalert.org.uk *Emergency identification system for people with hidden medical conditions.*
http://www.pituitary.org.uk *Pituitary Foundation (UK charity): comprehensive information for patients and GPs.*
http://www.thyroidmanager.org/ *Online thyroid disease textbook.*
http://www.tss.org.uk *UK Turner Syndrome Support Society.*

Questions for this chapter can be found on Elsevier eBooks+.

35 Diabetes mellitus

Richard Ian Gregory Holt

Clinical skills 1200
Background 1201
Physiology of insulin secretion and action 1203
Glucose metabolism 1204
Hormonal regulation 1204
Classification of diabetes 1204
Type 1 diabetes— 1206
Type 2 diabetes— 1209
Monogenic diabetes 1210
Clinical approach to the person with diabetes 1211
Presentation 1211
Diagnosis and investigations 1211
Prevention of type 2 diabetes 1212
Management of diabetes 1212
Aims of diabetes care 1212
Lifestyle measures 1213
Medication for diabetes 1214
Insulin 1214
Non-insulin antidiabetes treatments 1219
First-line treatments 1219
Second-line treatments 1220
Which drug and when? 1221
Pancreatic islet and whole-pancreas transplantation 1222
Monitoring glucose levels in people with diabetes 1222
Diabetes-related metabolic emergencies 1224
Complications of diabetes 1224
Microvascular complications 1224
Diabetic retinopathy 1224
Diabetic nephropathy 1227
Diabetic neuropathy 1229
Diabetes-related foot problems 1231
Macrovascular complications 1232
Other complications of diabetes 1233
Psychosocial implications of diabetes 1234
Diabetes in special situations 1235
Organization of diabetes care 1238

CORE SKILLS AND KNOWLEDGE

Diabetes is the most prevalent and clinically important endocrine disorder, affecting around 10% of the adult population. Many people with diabetes go on to develop life-limiting complications, including cardiovascular, renal, eye and neurological disease. Most consultant diabetologists are trained in both diabetes and endocrinology as well as general internal medicine, although many diabetes specialists focus solely on diabetes. Some subspecialize, focusing on areas of practice including inpatient care, diabetes in pregnancy, technology (insulin pumps and continuous glucose monitoring [CGM]), and pancreas or islet cell transplantation.

Most people with diabetes are seen in outpatient clinics either in primary or secondary care, although approximately 15% of all inpatients have diabetes. Many people are admitted for other reasons, but diabetes may complicate their hospital stays.

Key skills within diabetes include:

- Listening to people with diabetes (as with any person with a long-term condition), involving them actively in the management of their illness, and understanding the role of the multidisciplinary team in preventing and managing long-term diabetes-related complications
- Learning about the acute management of diabetic emergencies (hypoglycaemia, diabetic ketoacidosis (DKA), and hyperosmolar hyperglycaemic state [HHS]); see Chapter 11
- Developing confidence in prescribing insulin therapy and the wide range of oral and injectable antidiabetes agents.

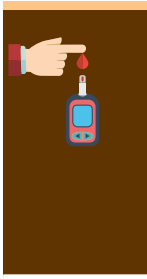
Diabetes is best understood in outpatient settings by listening to people with diabetes. Because diabetes is a chronic condition, there is an opportunity to build up long-term patient relationships. Diabetic emergencies are common reasons for admission to hospital, and people presenting with these may be found in the hospital emergency department or acute admissions unit.

CLINICAL SKILLS

Diabetes mellitus is a chronic condition, and people with diabetes may present to healthcare services for routine monitoring of their condition or for a host of reasons related or unrelated to their diabetes. Treatment for diabetic emergencies, which may be the presentation when diabetes is first diagnosed, is covered in the chapter on acute medicine (see Chapter 11). A typical approach to people with diabetes in non-emergency settings is shown in **Box 35.1**.

Routine checks in people with diabetes

All people with diabetes should be screened regularly to assess their glycaemic levels and for the emergence of any complications. **Box 35.2** shows the checks (covering history, examination, and



selected investigations) that should form part of this screening process, which should be undertaken at least every year.

Examination

The physical examination carried out in people with diabetes should be determined by the clinical context and any presenting complications. Fig. 35.1 presents a range of physical signs seen in people with diabetes: some are a direct result of complications of hyperglycaemia, and some are conditions associated with diabetes (e.g., vitiligo, an autoimmune disease sometimes found in people with type 1 diabetes).

Box 35.3 lists features to look for when examining the feet of a person with diabetes.

Language matters

There has been a recognition that the way language is used by healthcare professionals has a profound effect on how people with diabetes (and their carers) experience the condition and feel about living with diabetes. This convention has been adopted in the chapter and is an important change in focus, reflecting the primary role of putting people with diabetes at the centre of the management of their condition.

BACKGROUND

Diabetes mellitus is a complex metabolic disorder characterized by chronic hyperglycaemia caused by relative insulin deficiency, resistance, or both. In 2021, the International Diabetes Federation estimated that 537 million adults aged 20 to 79 years (10.5% of the global population) had diabetes, and this number is projected to reach 643 million by 2030 and 783 million by 2045 (Fig. 35.2).

Diabetes is associated with a number of short- and long-term complications that reduce quality of life and life expectancy and are associated with major health costs, estimated globally to be US$ 966 billion in 2021. Complications include acute metabolic disturbance; macrovascular disease (leading to an increased prevalence of coronary artery disease, peripheral vascular disease, and stroke); and microvascular damage causing retinopathy, nephropathy, and neuropathy.

Box 35.1 Taking a history in diabetes mellitus (see also Box 1.7)

Presenting complaint
- Why has the person with diabetes attended?
- Could their presenting complaints be related to a complication of diabetes?

Diagnosis of diabetes
- When were they diagnosed?
- Did they present as an emergency or with symptomatic hyperglycaemia, or was diabetes picked up without symptoms through routine screening?
- What type of diabetes do they have?
  - Might it be secondary to another disease process?
  - Is there any reason to suspect monogenic diabetes?

Management of the disease
- What do they understand about the role of self-management in diabetes?
- Do they monitor their capillary or interstitial glucose?
  - Do they keep a diary of the results?
  - What are the usual readings?
  - Is there evidence of hypoglycaemia or hyperglycaemia?
- Have they ever required hospital (or intensive care) admission for diabetic emergencies?
- Have they ever experienced hypoglycaemic episodes?
  - Do they have impaired awareness if they are becoming hypoglycaemic?
  - Have they needed external help to manage an episode of hypoglycaemia?
- Do they drive or engage in any other activities that may raise safety concerns in the context of hypoglycaemia?

Medications
- Do they see a primary care or hospital team to manage their diabetes?
  - Are they in contact with a specialist nurse or dietitian?
- What medication do they use to manage their blood glucose?
- If they inject insulin, do they rotate injection sites?
  - What size needles do they use?
- Are they able to adjust doses of insulin or other medications according to blood glucose levels and carbohydrate intake?
- What other medications are they taking?

Complications
- Have they had any cardiovascular events?
  - What is their cardiovascular risk, and is this being appropriately managed?
- Are they undergoing regular retinal screening?
  - Have they required any retinal intervention?
- Do they supply regular urine samples to screen for diabetic nephropathy?
  - Are they taking angiotensin-converting enzyme inhibitors or angiotensin receptor blockers or other renoprotective medication?
- Have they had any problems with neuropathy or vascular disease affecting their feet?
  - Have they experienced a foot ulcer?
  - Have they needed an amputation?
  - Have they experienced any adverse psychological sequelae of diabetes, for example diabetes distress, depression, anxiety, or disordered eating?

Social history
- Do they smoke? If so, have they tried to quit?
- Are they aware of key health behavioural interventions around healthy eating, weight reduction, and exercise?

Box 35.2 Screening checks in people with diabetes (performed at least annually)

General health
- Level of knowledge about diabetes
- Weight and body mass index
- Diet and eating habits
- Exercise

Glycaemic management
- Glycated haemoglobin
- Glucose monitoring
- Medication
- Review of insulin injection sites

Development of complications
- Retinal screening, ideally by retinal photography
- Visual acuity
- Urine albumin:creatinine ratio
- Serum urea and electrolytes; estimated glomerular filtration rate
- Examination of feet for vascular disease and neuropathy
- Full blood count
- Thyroid function
- Liver function

Cardiovascular risk
- Smoking status
- Blood pressure
- Lipid profile
- Overall cardiovascular risk using validated risk model (e.g., QRISK)

General observation
Does patient look well/ unwell?
Weight loss (T1DM)
Weight gain (T2DM)
Dehydrated?
Breathing (air hunger, Kussmaul breathing)

Face
Cranial nerve palsy, particularly CN III
Eye movements
Ptosis

Skin
Vitiligo (autoimmune)
Pigmentation (e.g., axillary acanthosis nigricans in insulin resistance)
Granuloma annulare
Bullosis

Vitiligo

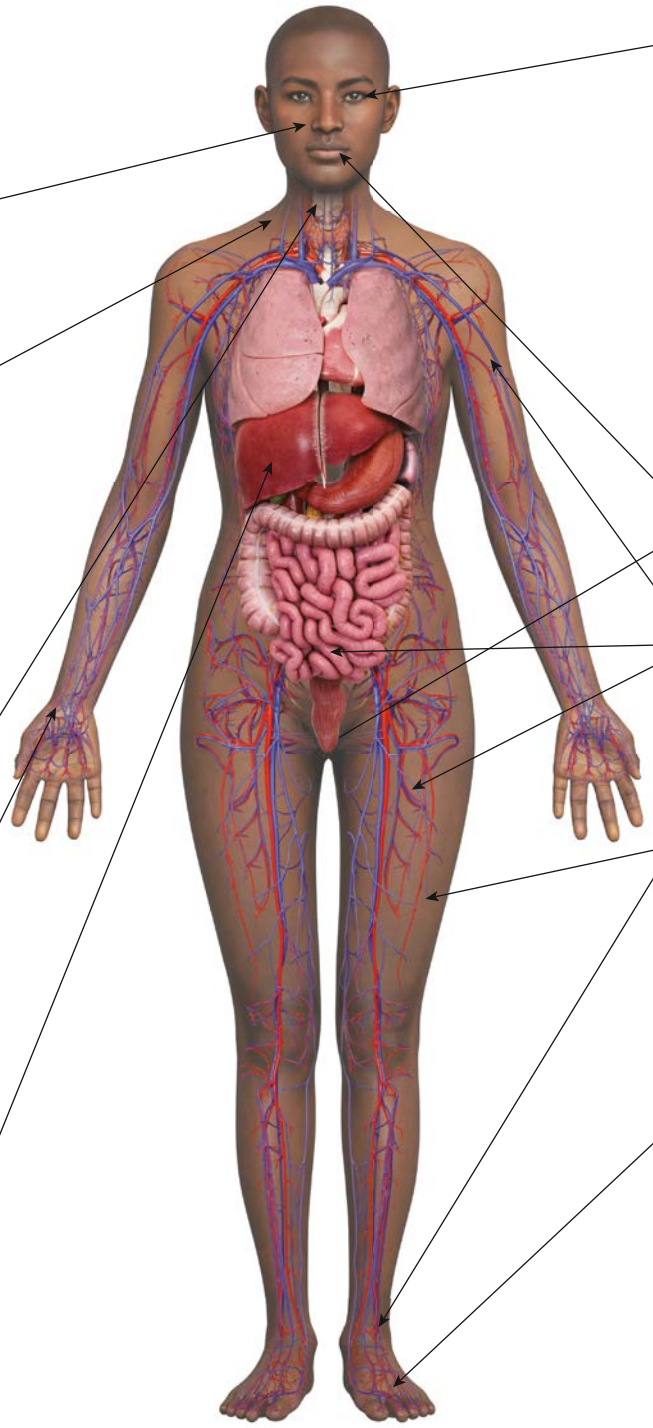
Neck
Carotid pulses/bruits
Check thyroid gland for goitre (autoimmune)

Hands
Carpal tunnel
Dupuytren's contracture
Muscle wasting
Limited joint movement

Dupuytren's contracture

(From Palastanga N. Anatomy and Human Movement, 6th edn, with permission.)

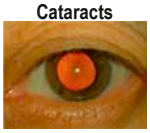
Abdomen
Hepatomegaly (fatty liver)



Eyes
Fundoscopy
- Cataracts, against red reflex
- Retinopathy (see Chapter 31)
- Visual acuity

Eyelids – xanthelasma

Cataracts



Retinopathy

Mouth and genitalia
Candidiasis

Insulin injection sites
Bruising
Lipohypertrophy
Lipoatrophy (rare)

Legs
Muscle wasting
Hair loss
Sensory neuropathy (glove and stocking)
Reflexes (lost in sensorimotor neuropathy)
Necrobiosis lipoidica

Feet
- Feel for peripheral pulses
- Skin – colour, ulcers, gangrene
- Look between toes for infection
- Sensory loss
  - Neuropathic foot ulcer
  - Charcot neuroarthropathy

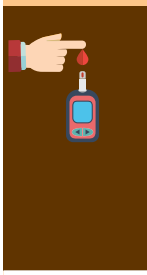
Neuropathic ulcer

Charcot joints

(From Miller MD. Presentation, Imaging and Treatment of Common Musculoskeletal Conditions, Saunders 2011, with permission.)

Aims of physical examination in people with diabetes:
* Assessing for diabetic emergencies — diabetic ketoacidosis, hypoglycaemia.
* Establishing the presence of complications — neuropathy, eyes and retinopathy, kidney disease, skin changes including ulceration, cardiovascular disease including peripheral vascular disease, foot complications, soft tissue infection
* Assessment of cardiovascular risk factors — blood pressure, stigmata of hyperlipidaemia, weight, and Body Mass Index
* Revealing signs of autoimmune disease — vitiligo, thyroid disease
* Assessment of injection sites

Fig. 35.1 Physical examination of people with diabetes. This is determined by the clinical context and any presenting complications. *CN,* Cranial nerve; *T1DM,* type 1 diabetes mellitus; *T2DM,* type 2 diabetes mellitus.



Box 35.3 Examination of the feet

Features of vascular disease

- Pale discoloration
- Loss of hair
- Cool temperature
- Absent pulses (begin in the foot and proceed up the legs)
- Reduced capillary refill time
- Evidence of gangrene or infection

Features of neuropathy

- Clawing of toes, loss of plantar arch
- Neuropathic ulcers
- Joint deformity (Charcot's joint)
- Glove and stocking sensory loss (check with monofilament)
- Loss of vibration sense, proprioception, and pain
- Loss of ankle jerk

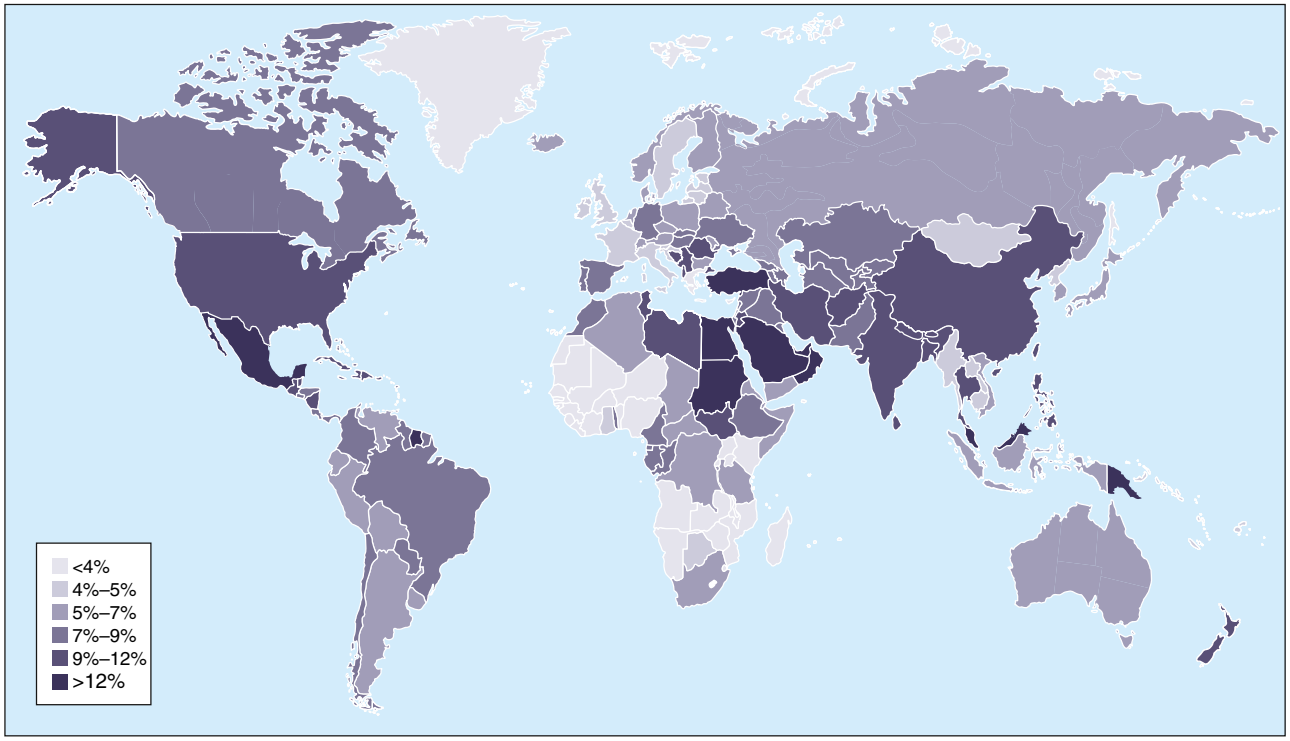


Fig. 35.2 Estimated age-adjusted prevalence of diabetes in adults aged 20 to 79 years in 2021. *(From IDF Diabetes Atlas, 10th edition.)* Map lines delineate study areas and do not necessarily depict accepted national boundaries.

Diabetes was responsible for approximately 6.7 million deaths or 12.2% of all deaths in 2021, outnumbering the combined number of global deaths from HIV, tuberculosis, and malaria. Globally, diabetes is among the top 10 causes of death. Approximately one-third of all diabetes-related deaths occur in people of working age (younger than 60 years of age). Although the number of years lost to diabetes is decreasing with time, a diagnosis of diabetes in a man or woman at the age of 18 years reduces life expectancy by 6.5 and 6.0 years, respectively. By contrast, diabetes diagnosed at the age of 60 years reduces life expectancy by 2.4 years in a man and 6.5 years in a woman; a diagnosis after 80 years has a limited effect on life expectancy. For many years, heart disease was the commonest cause of death accouting for approximately two-thirds of all deaths in people with diabetes. The large decline in vascular disease in the first two decades of the 21st century now means that cancers are the leading cause of death in people with diabetes.

PHYSIOLOGY OF INSULIN SECRETION AND ACTION

Insulin is the key hormone involved in the regulation of cellular energy supply and macronutrient balance derived from food. It is a 51–amino acid peptide hormone consisting of two polypeptide chains, the A and B chains, and is synthesized in the β cells of the pancreatic islets of Langerhans (Fig. 35.3). The synthesis, intracellular processing, and secretion of insulin by the β cells is typical of the way that the body produces and manipulates many peptide hormones. Fig. 35.4 illustrates the cellular events triggering the release of insulin-containing granules. After secretion, insulin enters the portal circulation and is carried to the liver, its prime target organ. About 50%–70% of secreted insulin is extracted and degraded in the liver; the remainder is broken down by the kidneys. C-peptide is only partially extracted by the liver (and hence provides a useful index of the rate of insulin secretion) but is mainly degraded by the kidneys.

Insulin is secreted at a slow background rate throughout the day, resulting in a low plasma insulin concentration between meals and overnight. In response to eating, there is a rapid rise in circulating insulin concentration, which falls back to baseline within 2 hours (Fig. 35.5).

The insulin receptor

Insulin exerts its actions through binding to a receptor, which straddles the cell membrane of many cells (Fig. 35.6). The receptor comprises a dimer with two α-subunits, which include the binding sites

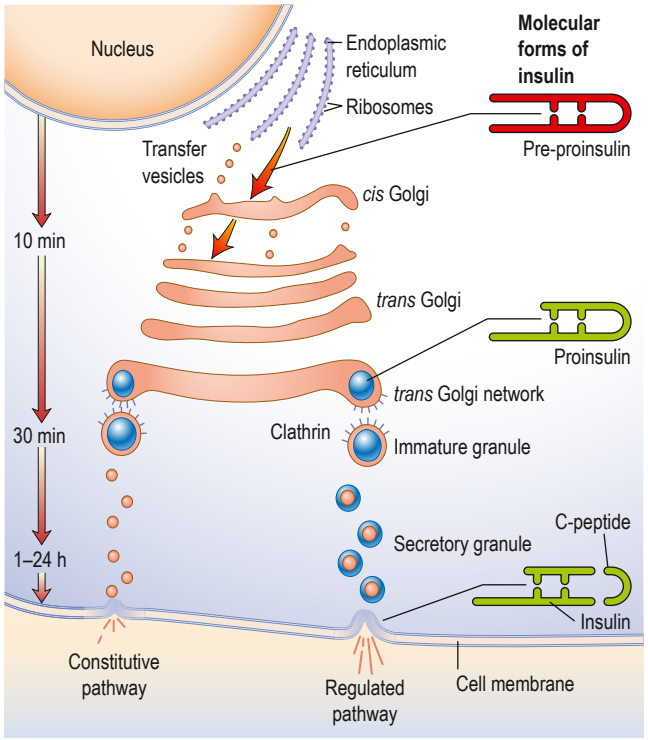


Fig. 35.3 Part of a β cell. The ribosomes manufacture pre-proinsulin from insulin messenger RNA (mRNA). The hydrophobic 'pre' portion of pre-proinsulin allows it to transfer to the Golgi apparatus and is subsequently enzymatically cleaved off. Proinsulin is parceled into secretory granules in the Golgi apparatus. These mature and pass towards the cell membrane, where they are stored before release. The proinsulin molecule folds back on itself and is stabilized by disulphide bonds. The biochemically inert peptide fragment known as connecting (C-)peptide splits off from proinsulin in the secretory process, leaving insulin as a complex of two linked peptide chains. Equimolar quantities of insulin and C-peptide are released into the circulation via the 'regulated pathway'. A small amount of insulin is secreted by the β cell directly via the 'constitutive pathway', which bypasses the secretory granules.

for insulin, and two β-subunits, which traverse the cell membrane. When insulin binds to the α-subunits, it induces a conformational change in the β-subunits, resulting in activation of tyrosine kinase and initiation of a cascade response involving a host of other intracellular substrates. The insulin–receptor complex is then internalized by the cell, insulin is degraded, and the receptor is recycled to the cell surface.

Glucose metabolism

Blood glucose levels are tightly regulated in health and rarely stray outside the range of 3.5 to 8.0 mmol/L (63–144 mg/dL) despite the widely varying effects of food, fasting, and exercise. Tight regulation of glucose is necessary because some tissues, particularly the brain, are highly dependent on glucose as an energy source, but a high glucose concentration irreversibly damages cellular proteins.

The principal organ of glucose homeostasis is the liver, which absorbs and stores glucose (as glycogen) in the post-absorptive state and releases it into the circulation between meals to match the rate of glucose utilization by peripheral tissues. The liver also combines three carbon molecules derived from breakdown of fat (glycerol), muscle glycogen (lactate), and protein (e.g., alanine) into the six-carbon glucose molecule by the process of gluconeogenesis.

Glucose production

More than 90% of the approximately 200 g of glucose used daily is derived from liver glycogen and hepatic gluconeogenesis, with the remainder coming from renal gluconeogenesis.

Glucose utilization

The major consumer of glucose is the brain, whose function depends on an uninterrupted supply of this substrate. Its requirement is 1 mg/kg body weight per minute, or approximately 100 g daily in a 70-kg person. Glucose uptake by the brain is obligatory and is not dependent on insulin, and the glucose used is oxidized to carbon dioxide and water. Tissues such as muscle and fat have insulin-responsive glucose transporters and absorb glucose in response to postprandial peaks in glucose and insulin. At other times, energy requirements are largely met by fatty acid oxidation. Glucose taken up by muscle is stored as glycogen or metabolized to lactate or carbon dioxide and water. Fat uses glucose as a substrate for triglyceride synthesis; lipolysis releases fatty acids from triglyceride together with glycerol, a substrate for hepatic gluconeogenesis.

Glucose transport

Cell membranes are not inherently permeable to glucose. A family of specialized glucose transporter (GLUT) proteins carry glucose through the membrane into cells. The function of GLUT-1 to -3 is insulin independent, but insulin stimulates glucose uptake into muscle and adipose tissue through GLUT-4. GLUT-4 is normally present in the cytoplasm, but after insulin binds to its receptor, GLUT-4 moves to the cell surface, where it creates a pore for glucose entry (Fig. 35.6).

Hormonal regulation

Insulin is a major regulator of intermediary metabolism, but its actions in the fasting and postprandial states differ (Fig. 35.7). In the fasting state, insulin's main action is to regulate glucose release by the liver; in the postprandial state, it additionally promotes glucose uptake by fat and muscle.

A number of 'counter-regulatory hormones' that antagonize the action of insulin are also important in maintaining normoglycaemia. As glucose concentration falls below the normal range, glucagon is secreted from the pancreatic α-cells. At the same time, several other hormones, including noradrenaline (norepinephrine), cortisol, and growth hormone, are released. These counter-regulatory hormones increase hepatic glucose production and reduce its utilization in fat and muscle for any given insulin concentration.

CLASSIFICATION OF DIABETES

Diabetes may be primary (idiopathic) or secondary (Box 35.4). Gestational diabetes refers to glucose intolerance appearing for the first time in pregnancy and is described in more detail later in the chapter (see 'Diabetes in pregnancy' section). Primary diabetes is classified into:

- ***Type 1 diabetes***, which usually has an immune pathogenesis and is characterized by severe insulin deficiency
- ***Type 2 diabetes***, which results from a combination of insulin resistance and less severe insulin deficiency

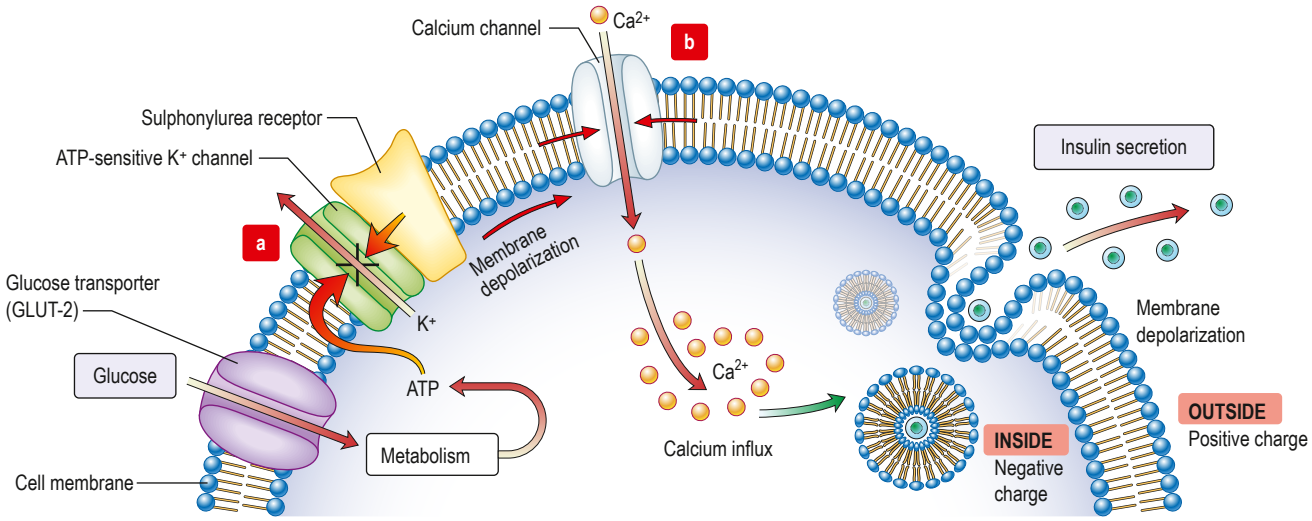


Fig. 35.4 The local regulation of insulin secretion from β cells. Glucose *(left)* enters the β cell via the glucose transporter 2 (GLUT-2) transporter protein, which is closely associated with the glycolytic enzyme glucokinase. Metabolism of glucose within the β cell generates adenosine triphosphate (ATP). ATP closes potassium channels in the cell membrane **(a)**. If a sulphonylurea binds to its receptor, this also closes potassium channels. Closure of potassium channels predisposes to cell membrane depolarization, allowing calcium ions to enter the cell via calcium channels in the cell membrane **(b)**. The rise in intracellular calcium triggers activation of calcium-dependent phospholipid protein kinase, which, via intermediary phosphorylation steps, leads to fusion of the insulin-containing granules with the cell membrane and exocytosis of the insulin-rich granule contents. Similar mechanisms produce hormone-granule secretion in many other endocrine cells.

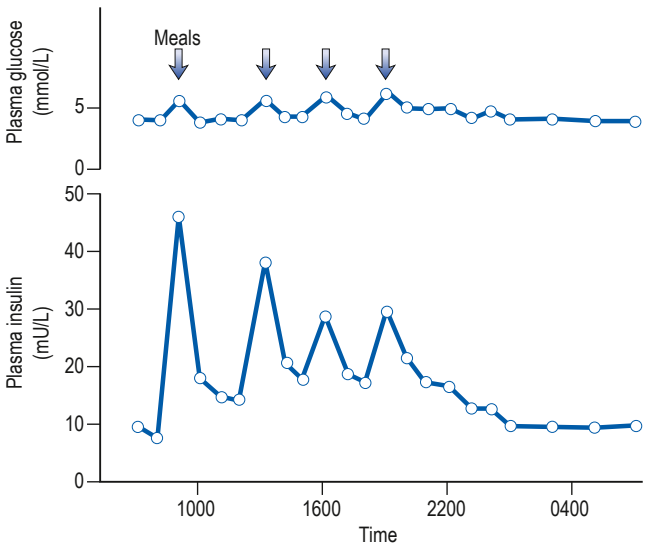


Fig. 35.5 Glucose and insulin profiles in people without diabetes.

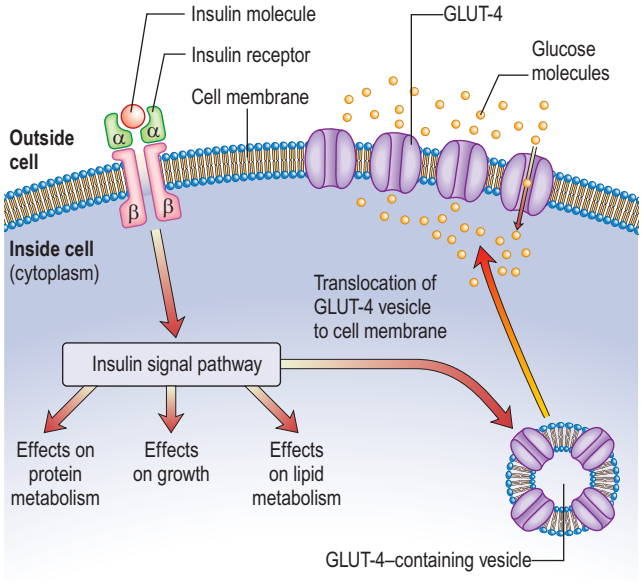


Fig. 35.6 Insulin signalling in peripheral cells (e.g., muscle and adipose tissue). The insulin receptor consists of α- and β-subunits linked by disulphide bridges *(top left)*. The β-subunits straddle the cell membrane. The transporter protein glucose transporter 4 (GLUT-4) *(bottom right)* is stored in intracellular vesicles. The binding of insulin to its receptor initiates many intracellular actions, including translocation of these vesicles to the cell membrane, carrying GLUT-4 with them; this allows glucose transport into the cell.

- ***Hybrid forms of diabetes***, which have features of both type 1 diabetes and type 2 diabetes
- ***Secondary diabetes*** can be subdivided into:
  - Diabetes secondary to single gene (monogenic) defects
  - Diabetes secondary to exocrine pancreatic disease
  - Diabetes secondary to endocrine disease
  - Diabetes secondary to drugs and chemicals
  - Diabetes secondary to infection
- **Uncommon forms** of immune-mediated diabetes
- **Other genetic syndromes** sometimes associated with diabetes
- **Unclassified diabetes.**

Box 35.5 lists the key features of the different types of diabetes. Type 1 diabetes and type 2 diabetes represent two distinct diseases from the epidemiological point of view, but from a clinical perspective, the two conditions appear as a spectrum, distinct at the two ends but overlapping in the middle.

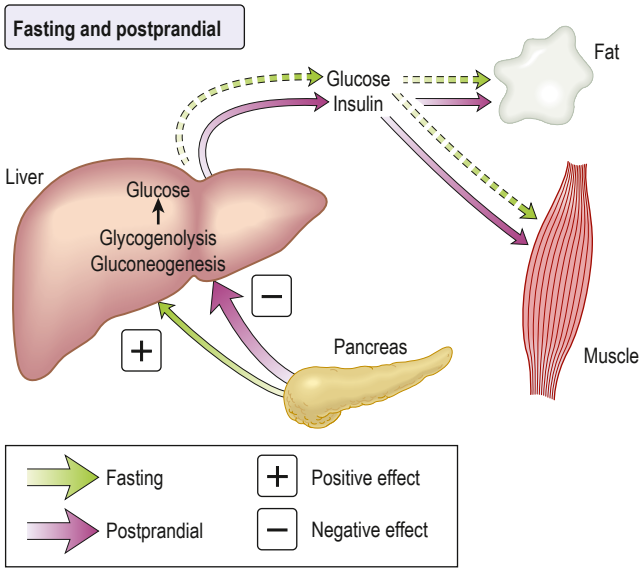


Fig. 35.7 Fasting and postprandial effects of insulin. In the fasting state, insulin concentrations are low, and it acts mainly as a hepatic hormone, modulating glucose production (via glycogenolysis and gluconeogenesis) from the liver. Hepatic glucose production rises as insulin levels fall. In the postprandial state, insulin concentrations are high, and it then suppresses glucose production from the liver and promotes the entry of glucose into peripheral tissues, particularly skeletal muscle and fat (increased glucose utilization).

Approximately 1% to 3% of cases of diabetes diagnosed before the age of 30 years are caused by mutations in a single gene (monogenic mutation). Although uncommon, it is important to recognize because the treatments and implications are different from those for type 1 diabetes or type 2 diabetes.

All forms of diabetes result from inadequate insulin secretion relative to the body's needs, and progressive insulin secretory failure is characteristic of both common forms of diabetes. Thus, a few people with immune-mediated type 1 diabetes may not require insulin at the point of diagnosis, whereas many with type 2 diabetes will eventually require it.

Type 1 diabetes

Epidemiology

Type 1 diabetes is a disease of insulin deficiency and accounts for 5% to 10% of all cases of diabetes. It typically presents in childhood and young adulthood, reaching a peak incidence around the time of puberty, but it can present at any age. In 2021, globally, more than 1.2 million children and adolescents had type 1 diabetes with approximately 184 100 new cases every year. The incidence of type 1 diabetes varies dramatically throughout the world, with the highest rates being in populations of Northern European origin and in a number of countries in the Middle East and North Africa.

Type 1 diabetes is subdivided into type 1A (immune-mediated) and type 1B (non–immune-mediated). The vast majority of those affected, especially in Western countries, have type 1A disease. A 'slowly evolving' immune-mediated variant with slower progression to insulin deficiency occurring in adults, often with features of the metabolic syndromes in later life, is termed *latent autoimmune diabetes in adults*.

Aetiology

Type 1 diabetes belongs to a family of immune-mediated organ-specific diseases, which include autoimmune thyroid disease, coeliac disease, Addison's disease, and pernicious anaemia. The precise molecular mechanisms that lead to type 1 diabetes are incompletely understood but involve the triggering of a selective autoimmune destruction of the insulin producing cells of a genetically predisposed individual. Initially, autoantibodies directed against pancreatic islet constituents appear in the circulation and often predate clinical onset by many years (stage 1). The islet antigens include insulin itself, the enzyme glutamic acid decarboxylase, protein tyrosine phosphatase (IA-2, also known as ICA512), the cation transporter ZnT8, and tetraspanin 7 (Fig. 35.8).

This is followed by a phase of asymptomatic loss of β-cell secretory capacity (stage 2); histologically, this is characterized by a chronic inflammatory mononuclear cell infiltrate of T lymphocytes and macrophages in the islets, known as insulitis (Fig. 35.9). Eventually, when the remaining β cells are no longer able to produce enough insulin to meet the body's needs, diabetes symptoms start to develop (stage 3).

Some recovery of endogenous insulin secretion may occur over the first few months after diagnosis and treatment initiation (the 'honeymoon period'). During this time, the insulin dose may need to be reduced. Some people with type 1 diabetes continue to produce small amounts of insulin for many decades after diagnosis. It is unclear why this occurs, but it seems that strict glucose management from diagnosis can prolong β-cell function.

The ability to detect autoantibodies in children before the development of type 1 diabetes is paving the way for trials to prevent the disease by immune modulation, but these interventions are still in early development.

Genetic susceptibility and inheritance

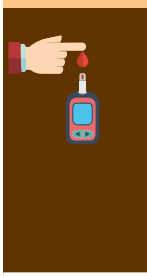
Approximately 10% to 20% of children with newly diagnosed type 1 diabetes have an affected first-degree relative. An increased susceptibility to type 1 diabetes is inherited, but the disease is not genetically predetermined. The identical twin of a person with type 1 diabetes has a 30% to 50% chance of developing the disease, which implies that non-genetic factors must also be involved. The risk of developing diabetes by age 20 years is greater with a father with diabetes (5%–8%) than with a mother with diabetes (2%–5%). If one child in a family has type 1 diabetes, each sibling has a 4% to 6% risk of developing diabetes. This risk rises to about 20% in siblings with the same human leucocyte antigen (HLA) genotype as the proband.

Human leucocyte antigen system

Genetic susceptibility is polygenic but the greatest contribution comes from polymorphisms in the HLA region. HLA genes are highly polymorphic and modulate the body's immune defence system. More than 90% of people with type 1 diabetes carry HLA-DR3-DQ2, HLA-DR4-DQ8, or both compared with about 35% to 40% of the background population. By contrast, certain HLA alleles confer protective effects (e.g., DQB1*0602).

Other genes or gene regions

Genome-wide association studies have greatly broadened our understanding of the genetic background of type 1 diabetes and have identified more than 250 non-HLA genes or gene regions that influence risk to a minor extent. These include the gene encoding insulin and other genes involved in immune responses. This has allowed the creation of genetic risk scores comprising between 9 and 67 single nucleotide polymorphisms that can usefully discriminate between type 1 diabetes and type 2 diabetes or those without diabetes in European populations.



Box 35.4 Aetiological classification of diabetes mellitus, based on the classification by the 2019 World Health Organization

Type of diabetes	Brief description
Type 1 diabetes	β-cell destruction (mostly immune mediated) and absolute insulin deficiency; onset peaks in childhood and early adulthood
Type 2 diabetes	Most common type; various degrees of β-cell dysfunction and insulin resistance; commonly associated with overweight and obesity
Hybrid forms of diabetes	
Slowly evolving, immune-mediated diabetes of adults	Similar to slowly evolving type 1 diabetes in adults but more often has features of the metabolic syndrome, a single glutamic acid decarboxylase autoantibody, and retains greater β-cell function
Ketosis-prone type 2 diabetes	Presents with ketosis and insulin deficiency but later does not require insulin; common episodes of ketosis; not immune mediated
Secondary forms of diabetes	
Monogenic defects of β-cell function Maturity-onset diabetes of the young Glucokinase mutations Hepatic nuclear factor mutations Neonatal diabetes Mitochondrial diabetes	Caused by specific gene mutations; has several clinical manifestations requiring different treatment; some occurring in the neonatal period and others by early adulthood
Monogenic defects in insulin action Leprechaunism Type A insulin resistance Rabson–Mendenhall syndrome Lipoatrophic diabetes	Caused by specific gene mutations; has features of severe insulin resistance without obesity; diabetes develops when β cells do not compensate for insulin resistance
Diseases of the exocrine pancreas Chronic pancreatitis Haemochromatosis Pancreatic surgery or trauma Cystic fibrosis Neoplasia Fibrocalculous pancreatopathy	Various conditions that affect the pancreas can result in hyperglycaemia (e.g., trauma, tumour, inflammation)
Endocrine disorders Acromegaly Cushing's syndrome Phaeochromocytoma Glucagonoma Hyperthyroidism Somatostatinoma Aldosteronoma	Occur in diseases with excess secretion of hormones that are insulin antagonists
Drug or chemical induced Glucocorticoids Thiazide diuretics Antipsychotics β-Adrenergic receptor blockers	Some medicines and chemicals impair insulin secretion or action; some can destroy β cells
Infection-related diabetes Congenital rubella Cytomegalovirus Mumps COVID-19	Some viruses have been associated with direct β-cell destruction
Uncommon specific forms of immune-mediated diabetes Stiff man syndrome Anti-insulin receptor antibodies	Associated with rare immune-mediated diseases
Other genetic syndromes sometimes associated with diabetes Down's syndrome Klinefelter's syndrome Turner's syndrome Prader–Willi syndrome Wolfram's syndrome Friedreich's ataxia Huntington's chorea Laurence–Moon–Biedl syndrome Myotonic dystrophy Porphyria	Many genetic disorders and chromosomal abnormalities increase the risk of diabetes
Unclassified diabetes	Used to describe diabetes that does not clearly fit into other categories; this category should be used temporarily when there is not a clear diagnostic category, especially close to the time of diagnosis
Hyperglycaemia first detected during pregnancy	
Diabetes in pregnancy	Type 1 diabetes or type 2 diabetes first diagnosed during pregnancy
Gestational diabetes	Hyperglycaemia below diagnostic thresholds for diabetes in pregnancy

Box 35.5 Spectrum of diabetes: a comparison of the different types of diabetes mellitus

Features	Type 1	Type 2	Monogenic	Secondary
Age	Younger (usually younger than 30 years)	Older (usually older than 30 years)	Neonates to early adulthood	Usually middle or older age
Weight loss	Usually present	Usually no	No	Depends on underlying cause
Obesity	Population frequency	Common	Population frequency	Depends on underlying cause
Time course of symptoms duration	Weeks	Months to years	Months	Weeks or months
Severity of symptoms	Can be marked	Variable but not usually severe	Not usually severe	Depends on underlying cause
Seasonal onset	Yes	No	No	No
Heredity	10%–20% have an affected first-degree relative	Heritability estimated to be 25% to 80% No HLA links	Present in almost all with onset in childhood or early adulthood	Unusual unless diabetes secondary to genetic conditions such as haemochromatosis
Pathogenesis	Autoimmune disease	No immune disturbance	No immune disturbance	No immune disturbance
Ketone formation	Yes	Unusual but may occur in ketosis-prone type 2 diabetes	No or minimal	May be present
Biochemical	C-peptide disappears	C-peptide persists	C-peptide persists	C-peptide persists

HLA, Human leucocyte antigen.

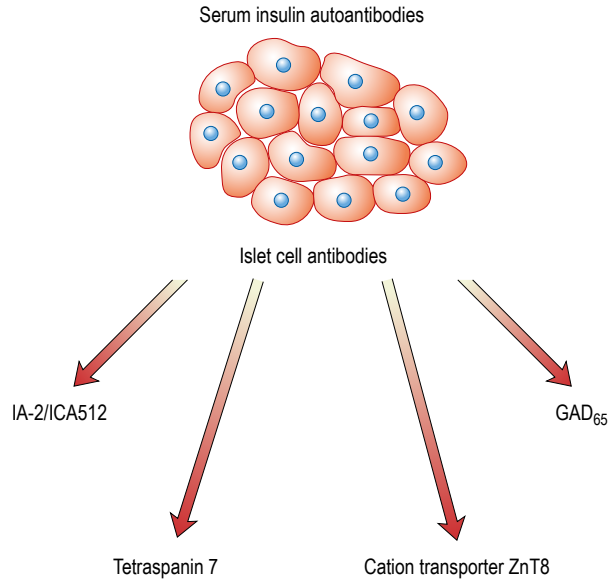


Fig. 35.8 Islet autoantibodies. Islet cell antibodies (ICAs) are detected by a fluorescent antibody technique that detects binding of autoantibodies to islet cells. Much of this staining reaction is caused by antibodies specific for glutamic acid decarboxylase (GAD) and protein tyrosine phosphatase (IA-2, also known as ICA512). Other islet autoantibodies involved are the cation transporter ZnT8 and the recently identified tetraspanin 7. Insulin autoantibodies also appear in the circulation but do not contribute to the ICA reaction.

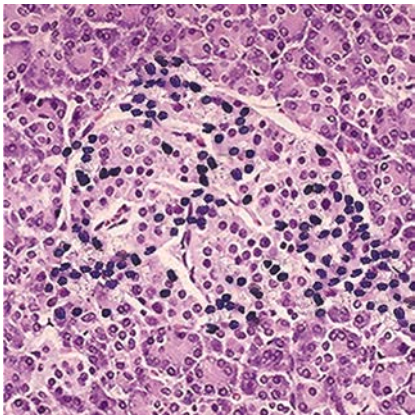


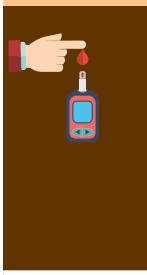
Fig. 35.9 Pancreatic islet showing infiltration by chronic inflammatory cells (insulitis).

Environmental factors

Despite the genetic susceptibility to type 1 diabetes, 80% to 90% of people with newly diagnosed type 1 diabetes do not have a close family history of diabetes, and only 10% of individuals with HLA-susceptible genes develop type 1 diabetes. This implies that genetic factors do not account entirely for the development of type 1 diabetes, but environmental triggers are also important in initiating the autoimmune process, modulating the progression of islet autoimmunity to clinical type 1 diabetes and precipitating the disease in those with advanced pre-clinical disease.

Many putative environmental factors have been identified, but none of these is either necessary or sufficient to cause type 1 diabetes. They include:

- ***Maternal factors***, such as gestational infection and older age
- ***Perinatal factors***, such as birth order, mode of delivery, and birth weight
- ***Viral infections***, including enteroviruses such as coxsackie B4
- ***Exposures to dietary constituents***, such as early introduction of cow's milk and relative deficiency of vitamin D
- ***Environmental toxins***, such as alloxan or Vacor
- ***Childhood obesity***
- ***Psychological stress.***



How these environment factors interact with the immune system is uncertain, and there are several theories to explain this:

- The environmental factor damages the β cells, leading to presentation of self-antigens to the T-helper cells, which trigger the autoimmune response against the remaining healthy β cells.
- The self-antigens may be modified and become antigenic.
- An immune response against the environment factor may cross-react with self-antigens, so-called 'molecular mimicry'.
- The 'hygiene hypothesis' proposes that a cleaner environment with less early stimulation of the immune system in childhood may lead to a relatively immature immune system prone to autoimmunity.
- Early life antibiotic use is associated with the subsequent development of type 1 diabetes, perhaps because depletion of the gut microbiome compromises its role in inducing immune tolerance.

Type 2 diabetes

Epidemiology

Type 2 diabetes is the most common form of diabetes, accounting for around 95% of all cases; it is also one of the most common non-communicable diseases. The prevalence of type 2 diabetes is rising rapidly because of a combination of population growth, an ageing population and longer survival with type 2 diabetes, earlier age at onset, and better diagnosis. Type 2 diabetes is a disease of nutrient excess and the incidence has increased with the obesity epidemic (see Chapter 37), poor-quality diet, and reduced physical activity.

The prevalence of type 2 diabetes varies markedly across the world with the highest rates being in the Middle East and Pacific Islands and the lowest in Africa and Europe. Although often viewed as a disease of affluence, around three-quarters of those with type 2 diabetes live in low- and middle-income countries, and the most rapid rise in incidence is occurring in countries with rapidly growing economies, such as India and China. Two-thirds of people with diabetes live in urban areas.

Aetiology

Genetic susceptibility and inheritance

Type 2 diabetes clusters in families, and the risk of developing type 2 diabetes is dependent on genetic factors, with heritability estimates of between 25% and 80% in different countries. Identical twins of people with type 2 diabetes have about a 70% chance of developing diabetes; whereas the risk to non-identical twins or siblings is approximately 20% to 30%, the lifetime risk of developing type 2 diabetes is 40% for those with one parent with type 2 diabetes and about 70% if both parents are affected.

Type 2 diabetes is a polygenic disorder, and genome-wide studies of association between common DNA variants and disease have allowed identification of more than 1000 susceptibility loci. The vast majority of type 2 diabetes–associated risk variants (e.g., *TCF7L2*, *CDKAL*, *SLC30A*) appear to influence β-cell function. Most of the identified genetic markers exert very modest risk and together explain less than 20% of the heritability of type 2 diabetes.

Ageing

Pancreatic β-cell function declines with age, so the incidence of type 2 diabetes increases with age; most people are diagnosed after the age of 40 years, and one-third of those living with type 2 diabetes are older than 65 years. However, type 2 diabetes is becoming increasingly common in children and young adults.

Fetal origins of diabetes

There is a J-shaped relationship between low weight at birth and at 12 months of age and glucose intolerance later in life, particularly in those who gain excessive weight in adulthood. The concept is that poor nutrition early in life impairs β-cell development and function, predisposing to diabetes later on. Low birth weight also predisposes to a number of other chronic diseases in adulthood, including heart disease, hypertension, and osteoporosis.

Obesity

Obesity increases the risk of type 2 diabetes up to 80- to 100-fold and accounts for 80% to 85% of the overall risk of developing type 2 diabetes. In Western Europe and North America, 80% to 90% of adults with type 2 diabetes are overweight, and nearly a half have obesity. A central distribution of fat increases the risk of type 2 diabetes, so for any given level of obesity, the more visceral fat an individual has, the higher the risk of type 2 diabetes.

Diet

Although diet can affect the risk of diabetes through changes in body weight, certain dietary patterns are associated with higher or lower risks of type 2 diabetes. Components that increase the risk include dietary fat, particularly saturated fat; red meat; ultra-processed foods; fried foods, including French fries; white rice; and sugar-sweetened beverages. Wholegrains, increased fruit and vegetable intakes, nuts and seeds, pulses, fermented dairy products, oily fish, and a Mediterranean dietary pattern are associated with lower rates of type 2 diabetes.

Physical inactivity

Physical inactivity and sedentary behaviour are also associated with an increased risk of diabetes.

Other factors

Other risk factors include urbanization, poverty, abnormal sleep patterns, environmental toxins, and mental illness.

Pathogenesis

The relative role of defects in insulin secretion and action in the pathogenesis of type 2 diabetes has been much debated and likely differs between different people. However, both defects are necessary to develop diabetes.

Abnormalities of insulin action

Insulin action is diminished in type 2 diabetes through the development of insulin resistance, which is defined as the inability of insulin to produce its usual biological effects at physiological concentrations. It is characterized by an impaired ability of insulin to:

- Inhibit hepatic glucose output
- Stimulate glucose uptake into skeletal muscle
- Suppress lipolysis in adipose tissue.

The underlying mechanisms of insulin resistance are not fully understood but result from nutrient excess. Intracellular triglyceride accumulates in the liver and skeletal muscle and impedes the phosphorylation of the post-receptor insulin receptor substrate. Although this defect reduces insulin action with regards to glucose and lipid metabolism, insulin is still able to activate the mitogen-activated protein kinase pathway, which regulates a number of intracellular pathways involved in inflammation, cellular proliferation, and atherosclerosis.

Abnormalities of insulin secretion

Insulin resistance alone does not explain the development of diabetes because only one-fifth of those with the degree of insulin resistance characteristic of diabetes develop the condition. As insulin resistance develops, the body's response is to increase insulin secretion, so early type 2 diabetes is often associated with insulin hypersecretion. This observation led some to question the role of β-cell function in the aetiology of diabetes. Nevertheless, insulin secretory abnormalities manifest early in the course of type 2 diabetes and progress with time; an early sign is loss of the first phase of the normal biphasic insulin secretion. Even though circulating insulin concentrations are higher than in people without diabetes, they are still inadequate to restore glucose homeostasis. By the time of diagnosis, at least 50% of β-cell mass and function has been lost. Relative insulin lack is associated with increased glucose production from the liver (owing to inadequate suppression of gluconeogenesis) and reduced insulin-mediated glucose uptake by peripheral tissues. Hyperglycaemia and lipid excess are toxic to β cells, at least *in vitro*, a phenomenon known as glucotoxicity, which is thought to cause further β-cell loss and further deterioration of glucose homeostasis. With time, insulin secretion declines, an observation referred to as the 'Starling curve' of the pancreas. This time course varies widely between individuals.

Type 2 diabetes is thus a condition in which insulin deficiency relative to increased demand leads to insulin hypersecretion by a depleted β-cell mass and progression towards absolute insulin deficiency, requiring insulin therapy.

Other hormonal abnormalities

Glucagon secretion is increased in type 2 diabetes, likely because of diminished intra-islet insulin, and leads to increased hepatic glucose output.

The insulin response to oral glucose is greater than the response to intravenous (i.v.) glucose, a phenomenon known as the incretin effect. The effect is mostly mediated by two hormones, glucagon-like peptide-1 (GLP-1) and glucose-dependent insulinotropic polypeptide (GIP), which are released by the gastrointestinal tract after eating. Their major action is to increase glucose-induced β-cell insulin secretion while suppressing glucagon secretion, but they also slow gastric emptying and induce satiety. Both hormones have short half-lives in the circulation, being degraded within minutes predominantly by the enzyme dipeptidyl peptidase-4 (DPP-4). The incretin effect is impaired in type 2 diabetes.

Glucose reabsorption in the kidney

The sodium–glucose transporter 2 (SGLT2) is a sodium-dependent glucose transport protein located in the proximal renal tubules, whose function is to reabsorb glucose from the renal filtrate and restore it to the circulation. Its activity thus determines the renal threshold for glucose, which normally averages approximately 10 to 11 mmol/L (180–200 mg/dL). This prevents urinary glucose loss but is maladaptive in type 2 diabetes because this process is upregulated and serves to maintain hyperglycaemia.

Monogenic diabetes

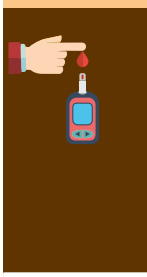
Approximately 1% to 3% of people with diabetes diagnosed before the age of 30 years have monogenic diabetes, previously called 'maturity-onset diabetes of the young' (**Box 35.6**). Monogenic diabetes is caused by a single gene mutation, which is dominantly inherited and predominantly affects β-cell function. Extra-pancreatic features may also be present. Many people with monogenic diabetes are misdiagnosed with either type 1 diabetes or type 2 diabetes, so monogenic diabetes should be considered in people presenting with early-onset diabetes in association with an affected parent and early-onset diabetes in approximately 50% of relatives. It is likely that as awareness improves, the diagnosis rate will increase.

? Box 35.6 Monogenic causes of diabetes[a]

Type	Features
Glucokinase	• 44% of all cases of MODY • Glucokinase gene is involved in the glucose-sensing mechanism within the pancreatic β cell; mutations lead to a change in the set-point in glucose with mild fasting hyperglycaemia • Non-progressive • Present from birth • No treatment needed except during pregnancy
HNF1A	• 24% of all cases of MODY • Transcription factor variants reduce insulin secretion • Progressive deterioration of glycaemia with age presenting during adolescence or early adulthood • Sensitive to sulfonylurea treatment • Associated with development of microvascular complications • Extrapancreatic features include low renal glucose threshold (glycosuria), raised HDL cholesterol, raised cardiovascular risk (in excess of type 2 diabetes), and liver adenomas
HNF4A	• 10% of all cases of MODY • Transcription factor variants reduce insulin secretion • Progressive deterioration of glycaemia with age presenting during adolescence or early adulthood • Sensitive to sulfonylurea treatment • Associated with development of microvascular complications • Extrapancreatic features include increased birth weight or macrosomia, neonatal hypoglycaemia, low HDL cholesterol, low lipoprotein A1 and A2, and raised LDL cholesterol
HNF1B	• 6% of all cases of MODY • Transcription factor regulates gene expression in several tissues, including the pancreas, kidney, liver, genital tract, and gut • Progressive deterioration of glycaemia with age presenting during adolescence or early adulthood • Initially sensitive to sulfonylurea treatment, but most require insulin • Associated with development of microvascular complications • Extrapancreatic features include renal cysts, renal development disorders, gout, liver enzyme derangement, short stature, and neurodevelopmental disorders
Maternally inherited diabetes and deafness	• 8% of monogenic diabetes • Maternally inherited • Variant in mitochondrial DNA affects the mitochondrial respiratory chain and results in cellular energy deficiency, leading to β-cell dysfunction and insulin deficiency • Mean age of diabetes onset is 37 years (range, 11–68 years) • Insulin treatment is usually required • Extrapancreatic features include deafness, short stature, cardiomyopathy, focal segmental glomerulosclerosis, and myopathy

[a]Other genes accounting for the remaining 8% of case include *ABCC8, WFS1, INSR, PPARG, INS, LMNA, PDX1, KCNJ11, NEUROD1, GATA6, SLC19A2*, and *TRMT10A*.

HDL, High-density lipoprotein; *LDL*, low-density lipoprotein; *MODY*, maturity-onset diabetes of the young.



Type 1 diabetes very rarely presents in children before 6 months of age, so infants who develop diabetes at this age are likely to have a monogenic defect. Transient neonatal diabetes mellitus occurs soon after birth and resolves at a median of 12 weeks; around 50% of people ultimately relapse later in life. Most have an abnormality of imprinting of the *ZAC* and *HYMAI* genes on chromosome 6q. The most common cause of permanent neonatal diabetes mellitus is mutations in the *KCNJ11* gene encoding the Kir6.2 subunit of the β-cell potassium–adenosine triphosphate (ATP) channel.

Further reading

International Diabetes Federation. *IDF Diabetes Atlas*, 10th edn. https://www.diabetesatlas.org/.
World Health Organization (2019). *Classification of Diabetes Mellitus*. Geneva: World Health Organization. Licence: CC BY-NC-SA 3.0 IGO.

CLINICAL APPROACH TO THE PERSON WITH DIABETES

Presentation

The presentation of diabetes may be acute, subacute, or asymptomatic, or an individual may present with a complication of diabetes.

Acute presentation

Young adults often present with a 2- to 6-week history of the classic triad of symptoms plus tiredness.

- ***Polyuria*** caused by the osmotic diuresis that results when blood glucose levels exceed the renal threshold
- ***Thirst and polydipsia*** caused by the resulting loss of fluid and electrolytes
- ***Weight loss*** caused by fluid depletion and accelerated breakdown of fat and muscle secondary to insulin deficiency.

These can be remembered by the four Ts (toilet, thirst, thin, tired). Ketonuria is often present in young people and may progress to ketoacidosis if these early symptoms are not recognized and treated.

Subacute presentation

The clinical onset may be prolonged over several months or years, particularly in older people. Thirst, polyuria, and weight loss are usually present, but the individual may complain of other symptoms such as lack of energy, visual blurring (owing to glucose-induced changes in refraction), or pruritus vulvae or balanitis caused by *Candida* infection.

Complications as the presenting feature

These include:

- Staphylococcal skin infections
- Retinopathy noted during a visit to the optician
- A polyneuropathy causing tingling and numbness in the feet
- Erectile dysfunction
- Arterial disease, resulting in myocardial infarction or peripheral gangrene.

Asymptomatic diabetes

It is estimated that approximately half of adults with diabetes are unaware of their condition. This proportion varies across the world, from about one-third in high-income countries to three-quarters in some low-income countries. Consequently, up to one-third of diagnoses are made as an incidental finding, and several countries have introduced screening programmes to identify those with asymptomatic undiagnosed diabetes.

Physical examination at diagnosis

Evidence of weight loss and dehydration may be present, and the breath may smell of ketones. Older people may present with established complications. Occasionally, there will be physical signs of an illness causing secondary diabetes (see **Boxes 23.4 and 23.5**). People with severe insulin resistance may have acanthosis nigricans, which is characterized by blackish pigmentation at the nape of the neck and in the axillae (see **Fig. 22.37**). Hypertension is present in 50% of people with type 2 diabetes and in a higher proportion of people of African and Caribbean ethnicity.

Diagnosis and investigations

Once considered, diabetes is easy to diagnose. This may be by a laboratory measurement of fasting plasma glucose, random glucose, or 2-hour plasma glucose after a 75-g oral glucose tolerance test (OGTT). The use of glycated haemoglobin (HbA_{1c}; also called A_{1c} in the USA) was introduced as an alternative method in 2011. HbA_{1c} is an integrated measure of an individual's prevailing blood glucose concentration over the previous 6 to 8 weeks and is also used to guide treatment decisions.

Diabetes can be diagnosed if the fasting plasma glucose is higher than or equal to 7.0 mmol/L (126 mg/dL) or the random or 2-hour OGTT plasma glucose is more than or equal to 11.1 mmol/L (200 mg/dL) or the HbA_{1c} is higher than or equal to 48 mmol/mol (6.5%).

When overt symptoms are present, only one biochemical test is needed, but in the absence of clear symptoms, two abnormal glucose or HbA_{1c} tests are required (**Boxes 35.7 and 35.8**). Fasting, 2-hour glucose, and HbA_{1c} identify slightly different populations of people with diabetes; that is, someone may be diagnosed by one criterion but not another. However, with time, people will eventually cross the diagnostic threshold for the other tests.

The diagnostic criteria recognize two further categories of abnormal glucose concentrations: impaired fasting glycaemia (IFG) and impaired glucose tolerance (IGT). Collectively, these have been described as 'pre-diabetes', but this is not strictly accurate because the majority will never develop diabetes. Neither IFG nor IGT are clinical entities in their own right, but they identify people who are at high risk of diabetes and cardiovascular disease. Individuals with IGT have a similar risk of cardiovascular disease as those with frank diabetes but do not develop microvascular complications. IFG only overlaps with IGT to a limited extent, and the associated risks of cardiovascular disease and future diabetes are not directly comparable.

Box 35.7 World Health Organization diagnostic criteria for diabetes[a]

- Fasting plasma glucose ≥7.0 mmol/L (126 mg/dL)
- 2-hour plasma glucose ≥11.1 mmol/L (200 mg/dl) during OGTT
- HbA_{1c} ≥6.5% (48 mmol/mol)
- Random plasma glucose ≥11.1 mmol/L (200 mg/dL) in a person with classic symptoms of hyperglycaemia

[a]Any one of these tests can be used to diagnose diabetes. Fasting is defined as no caloric intake for at least 8 hours. Glycated haemoglobin (HbA_{1c}) should be performed by a standardized laboratory method. One abnormal laboratory value is diagnostic in symptomatic individuals; two values are needed in asymptomatic people. The oral glucose tolerance test (OGTT; **Box 35.8**) is only required when there is diagnostic uncertainty and for diagnosis of cystic fibrosis diabetes and gestational diabetes.

Box 35.8 The 75-g oral glucose tolerance test: World Health Organization criteria[a]

Timing of test	Normal	Impaired glucose tolerance	Diabetes mellitus
Fasting	<6.0 mmol/L <106 mg/dL	<7.0 mmol/L <126 mg/dL	≥7.0 mmol/L ≥126 mmol/L
2 hours after glucose	<7.8 mmol/L <140 mg/dL	7.8–11.0 mmol/L 140–200 mg/dL	≥11.1 mmol/L ≥200 mg/dL

[a]An oral glucose tolerance test should be performed using a standardized protocol. The test should not be performed in people who are under physical stress (e.g., postsurgery, trauma, infection, within 6 weeks of a myocardial infarction or major surgery). Drugs that may interfere with the test (glucocorticoids, beta-blockers, thiazide, and loop diuretics) should be withdrawn for 3 days before the test if possible. Smoking and physical exercise should not be undertaken in the morning before and during the test. The person undergoing the test eat an unrestricted carbohydrate diet (>150 g per day) for at least 3 days before the test. The test should be performed in the morning after an overnight fast of at least 8 hours (water is allowed). The person should rest for at least 30 minutes before the test. After a fasting blood sample is collected, 75 g of glucose in 300 mL of water should be consumed over 5 minutes. The patient should rest and not smoke, eat, or drink (except sips of water). The second sample is collected 2 hours later. Results are for venous plasma; whole-blood values are lower.

IGT can only be diagnosed with an OGTT (**Box 35.8**) and is complicated by poor reproducibility of the key 2-hour value in this test. IFG is defined as a fasting plasma glucose between 6.1 and 6.9 mmol/L (110–126 mg/dL) and has the practical advantage that it avoids the need for an OGTT. The American Diabetes Association (ADA) definition of IFG differs slightly from the World Health Organization criteria in that the ADA threshold is 100 mg/dL (5.6 mmol/L).

Glycosuria cannot be used to diagnose diabetes but requires further investigation. It is commoner in older people, who have an altered renal threshold for glucose. It may also occur in familial renal glycosuria, which is a monogenic disorder affecting function of the SGLT2 and found in about 1 in 400 of the population.

Determining the type of diabetes

Whereas type 1 diabetes is frequently mistaken for type 2 diabetes in adulthood, other types of diabetes (e.g., monogenic diabetes) can be misdiagnosed as type 1 or type 2 diabetes. A misdiagnosis can cause inappropriate treatment choices and distress for the patient. No single clinical feature differentiates type 1 diabetes from type 2 diabetes, but the most discriminative feature is younger age at diagnosis (younger than 35 years), with lower body mass index (BMI) (<25 kg/m^2), unintentional weight loss, ketoacidosis, and glucose greater than 20 mmol/L (>360 mg/dL) at presentation all pointing to type 1 diabetes. In those with clinical features of type 1 diabetes, the measurement of islet autoantibodies at diagnosis is recommended. The presence of one or more of these antibodies is highly predictive of rapid progression and severe insulin deficiency, and these individuals should be considered to have type 1 diabetes.

However, the absence of autoantibodies does not exclude type 1 diabetes because approximately 5% to 10% of White European people with new-onset type 1 diabetes have negative islet antibodies, and further consideration of the diagnosis is necessary. Furthermore, antibodies may disappear with time. Here clinical judgement is needed to decide whether to initiate insulin therapy.

Insulin secretion continues to diminish with duration of diabetes to the point that C-peptide is less than 200 pmol/L after 3 years in those with type 1 diabetes, and this test may help differentiate type 1 diabetes from type 2 diabetes if uncertainty remains.

Genetic testing may be appropriate if there are features of monogenic diabetes.

Other investigations

Other routine investigations include urine testing for microalbuminuria (urine albumin:creatinine ratio [uACR]), a full blood count, urea and electrolytes, liver biochemistry, and random lipids. The latter test is useful to exclude associated hyperlipidaemia (see Chapter 19), and if elevated, it should be repeated as a fasting measurement after diabetes has been brought under control. Investigations of secondary causes of diabetes or associated autoimmune disease may be appropriate (**Box 35.4**).

Further reading

American Diabetes Association. Diagnosis and classification of diabetes mellitus. *Diabetes Care* 2021; 44(Suppl 1):S15–S33.
Holt RIG et al. The management of type 1 diabetes in adults. A consensus report by the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). *Diabetologia* 2021; 64(12):2609–2652.

Prevention of type 2 diabetes

A dramatic reduction in the incidence of diabetes in adults was documented during the Second World War when food was scarce. More recently, clinical trials in individuals with IGT have shown that it is possible to prevent or at least delay the development of type 2 diabetes by lifestyle or pharmacological interventions (or both). A range of diet, exercise, and medical and surgical interventions are described in detail in Chapter 37, which have been demonstrated to reduce the incidence of frank diabetes in high-risk individuals.

Given limited healthcare resources, interventions are targeted at those at highest risk and who will benefit most. Several risk assessment algorithms are available to identify those at highest risk of diabetes. Measurement of blood glucose identifies people with IFG or IGT.

Further reading

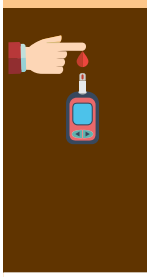
Diab M, Barhoosh HA, Daoudi B et al. Prevention and screening recommendations in type 2 diabetes: review and critical appraisal of clinical practice guidelines. *Prim Care Diabetes* 2019; 13:197–203.
National Institute of Diabetes and Digestive and Kidney Diseases. *Preventing Type 2 Diabetes*; https://www.niddk.nih.gov/health-information/diabetes/overview/preventing-type-2-diabetes.
National Institute for Health and Care Excellence. *NICE Public Health Guideline 38: Type 2 Diabetes: Prevention in People at High Risk.* NICE 2012; https://www.nice.org.uk/guidance/ph38.
NHS England. *NHS Diabetes Prevention Programme*. https://www.england.nhs.uk/diabetes/diabetes-prevention/.

MANAGEMENT OF DIABETES

Aims of diabetes care

For most people, diabetes is a life-long condition. Except for a few hours a year spent in contact with healthcare professionals, the majority of people with diabetes look after their condition themselves for most of the time. It is therefore vital that people develop the skills to manage their condition effectively. The aims of diabetes care and management are fivefold:

- Prevention of life-threatening diabetes emergencies such as DKA and hypoglycaemia, with effective management when they do occur
- Treatment of hyperglycaemic symptoms
- Minimization of long-term complications through screening and effective management of hyperglycaemia and other cardiovascular risk factors
- Avoidance of iatrogenic side-effects, such as hypoglycaemia
- Minimizing the psychosocial burden of living with diabetes, while promoting psychological well-being.



Fortunately, individuals with diabetes usually do not experience life-threatening diabetes emergencies frequently (the incidence of DKA is ≈28 episodes per 1000 person-years, and the rate of severe hypoglycaemia is ≈1–4 events per person-year), and hyperglycaemic symptoms are easily treated. Therefore, most of the discussions regarding diabetes management revolve around maintaining optimal glycaemia to prevent long-term complications while balancing the short-term adverse events associated with treatment. Hyperglycaemia drives microvascular complications, and it is unsurprising that the impact of improved glucose levels on small-vessel complications, such as retinopathy and nephropathy, has been compellingly demonstrated.

The Diabetes Control and Complications Trial (DCCT) compared standard and intensive insulin therapy in a large cohort of young people with type 1 diabetes, some who had no evidence of retinopathy or microalbuminuria (primary prevention group) at the beginning of the trial and others who already had evidence of these complications (secondary prevention group). Despite intensive therapy, mean blood glucose levels were still 40% above the normal physiological range, but even at this level of glycaemia, the risk of primary progression to retinopathy was reduced by 60%, nephropathy by 30%, and neuropathy by 20% over 7 years with intensive insulin therapy. These benefits persisted beyond the end of the trial despite equivalent glucose levels in the post-trial period.

The UK Prospective Diabetes Study (UKPDS) compared standard and intensive treatment in a large trial in people with recently diagnosed type 2 diabetes. Intensive treatment was associated with a 25% overall reduction in microvascular disease end-points, a 33% reduction in albuminuria, and a 30% reduction in the need for laser treatment for retinopathy in the more intensively treated people. Similar to the DCCT, these benefits persisted for many years after conclusion of the trial (the 'legacy effect'). This study also showed blood pressure management to be equally necessary in the prevention of retinopathy, but there is no legacy effect, and optimal blood pressure needs to be maintained.

Although there is an association between glucose and large-vessel disease, which extends well into the glucose range of the general population, with no clear cut-off for risk, the benefits of reducing hyperglycaemia are less certain for macrovascular disease. This may be because diabetes affects not only carbohydrate metabolism but also metabolism of lipid and protein. People with diabetes often have other cardiovascular risk factors, such as dyslipidaemia, hypertension, and obesity, and these risk factors may play a more important role than hyperglycaemia. Both the DCCT and UKPDS studies showed a reduction in cardiovascular events in the post-trial follow-up period in the intensive treatment groups, suggesting that early management of hyperglycaemia may improve the long-term risk of cardiovascular disease.

Risks of tight glycaemic management

In both the DCCT and UKPDS, the achieved glucose levels were above those of people without diabetes, so further trials were undertaken in people with type 2 diabetes to assess whether further lowering of glucose to normoglycaemia would yield greater benefits. These trials confirmed the benefits for microvascular disease but did not improve cardiovascular outcomes, and in the Action to Control Cardiovascular Risk in Diabetes (ACCORD) study, more people died in the intensively treated group than the standard arm. Exactly why this occurred is uncertain, but it seems likely that any benefit of improved glucose levels was outweighed by iatrogenic side-effects, such as hypoglycaemia and weight gain.

Box 35.9 Factors to consider with individualizing glycaemic targets for people with diabetes

Stringent glycaemic targets
- No hypoglycaemia
- Less complexity or polypharmacy
- Lifestyle or metformin only
- Short disease duration
- Long life expectancy
- No cardiovascular disease

Less stringent targets
- History of severe hypoglycaemia
- High burden of therapy
- Longer disease duration
- Limited life expectancy
- Extensive co-morbidity
- Cardiovascular disease

Most guidelines advocate HbA_{1c} targets between 48 and 58 mmol/mol (6.5%–7.5%), but because the benefits and risks of improved glucose levels differ between people with diabetes, the actual glycaemic targets should be determined for each individual, taking into account the factors shown in **Box 35.9**. Realistic goals should be agreed with each person, considering what is likely to be achievable. This applies particularly to older people and those with a short life expectancy.

Role of self-management education

Because the onus of diabetes management lies with the individual with diabetes, it is essential that the individual is equipped with the skills for this task. People with diabetes need to understand the risks of diabetes and the potential benefits of improving their glucose levels and other measures such as maintaining a lean weight, stopping smoking, and taking care of their feet. If healthcare professionals do not provide accurate information, misinformation from others will take its place. To address this need, diabetes self-management education has become an important cornerstone of diabetes care, delivered by all members of the diabetes multidisciplinary team (**Fig. 35.10**). The education needs to be tailored to the individual by embracing a patient-centred approach.

Structured educational programmes with a clear philosophy, such as Dose Adjustment for Normal Eating (DAFNE), using trained facilitators using a written curriculum are seen as the gold standard. These programmes improve diabetes knowledge and quality of life, glycaemic levels and reduce mortality. Education should not be viewed as a once-off inoculation, and regular ongoing support and updates are needed.

Lifestyle measures

Diet

Diet plays a key role in the management and clinical care of all people with diabetes. However, there has been a lot of controversy about what constitutes an ideal or optimal diet for a person with or at risk of diabetes. Previous guidance focused on the ideal macronutrient composition of the diet, but the latest guidance highlights the role of certain foods and dietary patterns. In common with the general population, people should be encouraged to eat more vegetables, fruits, wholegrains, fish, nuts, and pulses while reducing the consumption of red and processed meat, refined carbohydrates, and sugar-sweetened beverages. Because no single diet is effective in managing diabetes, it is important to adopt an individualized approach, taking into consideration the person's personal and

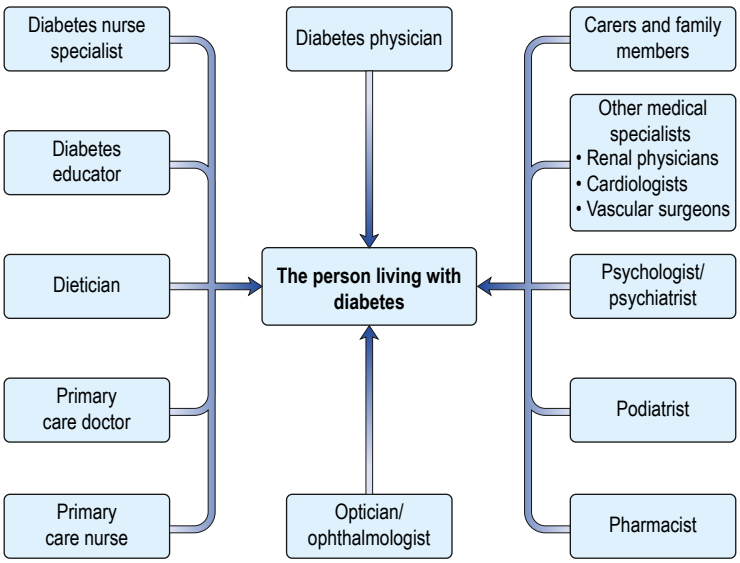


Fig. 35.10 **Members of the diabetes team.**

Box 35.10 Abbreviated summary of the evidence-based nutrition guidelines for the management of diabetes, hyperlipidaemia, and hypertension

- Nutrition advice should be offered to all people with diabetes, hyperlipidaemia, and hypertension and those at risk of developing type 2 diabetes as part of an integrated package of education and clinical care.
- Dietary advice should be individualized and culturally sensitive.
- People with type 1 diabetes should be supported to identify and quantify their dietary carbohydrate intake to support the adjustment of insulin dose to carbohydrate intake.
- People with type 2 diabetes, hyperlipidaemia, or hypertension or at risk of developing type 2 diabetes who are overweight should be supported to lose at least 5% of their weight by reducing calorie intake and increasing energy expenditure.
- Foods with a low glycaemic index are encouraged.
- A Mediterranean-style diet or equivalent healthy eating pattern should be encouraged. Key features of these diets include:
  - Decreasing salt intake to <6 g/day
  - Eating two portions of oily fish each week
  - Eating more wholegrains, fruits and vegetables, fish, nuts, and legumes (pulses)
  - Consuming less red and processed meat, refined carbohydrates, and sugar-sweetened beverages
  - Replacing saturated fats with unsaturated fats and limiting intake of trans-fatty acids
  - Limiting alcohol intake to ≤14 units/week

Based on the Diabetes UK evidence-based nutrition guidelines for the management of diabetes.

cultural preferences. **Box 35.10** summarizes the latest Diabetes UK dietary recommendations for people with diabetes.

One exception to the food-based recommendations concerns carbohydrate intake, which is important for people using intensive insulin regimens, particularly those with type 1 diabetes. In people without diabetes, the amount of insulin produced after eating depends on the carbohydrate consumed. To optimize glycaemia, people need to match their insulin dose to the carbohydrate in the meal, so it is important to teach people how to 'carbohydrate count'. Several books and apps are available to help with this process.

The peak blood glucose after a meal varies with the type of carbohydrate. For example, the glucose profile after eating pasta is much flatter than that seen after eating the same amount of carbohydrate as a sugary drink because pasta has a lower 'glycaemic index'. Foods with a low glycaemic index are encouraged to prevent rapid swings in plasma glucose.

Weight loss

For many years, it was thought that type 2 diabetes was an incurable progressive condition; however, there is now significant evidence to challenge this view. People who can be supported to lose 10 to 15 kg of body weight through lifestyle or pharmacological or surgical treatment can enter remission. Chapter 37 describes various weight-reduction interventions (including lifestyle modification, pharmacological treatments such as GLP-1 receptor agonists, and bariatric surgery), some of which have been shown to induce full remission of diabetes.

Physical activity

No treatment for diabetes is complete without exercise because physical activity has profound benefits, including improved fitness, reduced insulin requirement and improved glycaemic levels, lower cardiovascular risk, and greater life expectancy. The maxim that *some exercise is better than no exercise and more exercise is better than some exercise* applies. People with diabetes should be advised to take at least 150 minutes of aerobic exercise and resistance training per week, spread over a minimum of 3 days.

People with type 1 diabetes should be advised that physical activity can reduce the insulin requirement and lead to unstable glucose levels during and immediately after exercise and a later risk of severe hypoglycaemia.

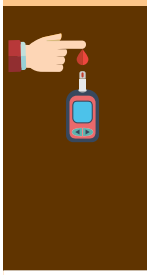
Tobacco smoking

Smoking can adversely affect glucose management and increase the risk of a number of medical conditions, so healthcare professionals should support people who smoke to quit.

MEDICATION FOR DIABETES

Insulin

Before discovery of insulin in 1921, the life expectancy of people with type 1 diabetes was only several months. Many people died from DKA, but it was possible to survive for several years on diets



of around 500 to 700 calories/day; these individuals had a wretched existence until either the person broke their diet or died from starvation. The discovery of insulin is rightly considered one of the greatest medical advances of the 20th century because the lives of millions of people with diabetes have been transformed by its discovery. Insulin is always indicated in people with type 1 diabetes and is often needed in those with type 2 diabetes as the condition progresses.

The philosophy of insulin therapy is to mimic the normal physiological secretion of insulin as closely as possible (Fig. 35.5). This involves the use of both long-acting insulin to replicate the basal secretion of insulin and short-acting insulin to cover mealtimes.

Insulin was historically derived from beef or pig pancreas, but this has been replaced by highly purified biosynthetic insulin, which is produced by adding a DNA sequence coding for insulin or proinsulin into cultured yeast or bacterial cells.

Short-acting insulins

Soluble human insulin is absorbed slowly, reaching a peak 60 to 90 minutes after subcutaneous injection. Its action tends to persist after meals, predisposing to hypoglycaemia. Absorption is delayed because soluble insulin forms stable hexamers (six insulin molecules around a zinc core), which need to dissociate to monomers or dimers before it can enter the circulation.

This delay is inconvenient for people with diabetes because the insulin should be injected 20 to 30 minutes before a meal, which often is not feasible. ***Short-acting insulin analogues*** (***insulin lispro***, ***insulin aspart***, and ***insulin glulisine***) have been engineered that dissociate more rapidly after injection without altering the biological effect (Fig. 35.11). These ***insulin analogues*** enter and disappear from the circulation more rapidly than soluble insulin. Newer formulations of insulin aspart and insulin lispro have been developed that dissociate and enter the circulation even more rapidly.

The use of short-acting insulin analogues in people with type 1 diabetes reduces total and nocturnal hypoglycaemic episodes and improves glycaemic levels as judged by postprandial glucose concentrations and HbA_{1c}. The benefits in people with type 2 diabetes are more modest, but short-acting insulin analogues seem to improve HbA_{1c} and postprandial glucose compared with soluble human insulin.

Intermediate- and longer-acting insulins

The action of synthetic human insulin can be prolonged by the addition of zinc or protamine. The most widely used form is NPH (neutral protamine Hagedorn), or 'isophane insulin', but its use is hampered by variability from one injection to another, the need to re-suspend the insulin before injection, and a peak action that generally occurs in the middle of the night.

Long-acting analogues have been developed to delay absorption and prolong the duration of action (Fig. 35.11). Insulin glargine is soluble in the vial as a slightly acidic (pH, 4) solution but precipitates at subcutaneous pH, thus prolonging its duration of action. A concentrated formulation of insulin glargine is now available with a longer and flatter duration of action than the standard preparation. ***Insulin detemir*** and ***insulin degludec*** have a fatty acid 'tail' that binds to serum albumin; the slow dissociation from the bound state prolongs the duration of action. In the case of insulin degludec, it forms long multihexamer chains at the site of injection, which only dissociate very slowly. Long-acting insulin analogues reduce hypoglycaemia for people with both type 1 diabetes and type 2 diabetes, particularly at night. A once-weekly basal insulin, insulin icodec, is currently in clinical development.

Because insulin analogues are more expensive than NPH insulin, in the case of type 2 diabetes, the National Institute for Health and Care Excellence (NICE) recommends that insulin analogues are reserved for people who do not respond well to NPH insulin. The European Association of the Study of Diabetes and ADA consensus report on the management of type 2 diabetes notes the benefits of insulin analogues but highlights that the way in which insulin is used has a greater impact on the effectiveness and adverse effects of insulin than differences between insulin formulations.

Strengths of insulin

Most insulin is formulated as 100 units/mL, but more concentrated forms are available, for example, insulin glargine is available as U300 (300 units/mL) as described earlier. Insulin degludec is available in two concentrations: 100 and 200 units/mL. The pen devices for these insulins indicate how many units of insulin are given, but there is a significant danger of overdose if the insulin is withdrawn from the pen and administered using an insulin syringe. A number of warnings have been issued to prevent this practice.

Some people with diabetes experience severe insulin resistance and require massive insulin doses. U500-strength insulin is available for these individuals, allowing them to inject the same dose of insulin in one-fifth of the usual volume. Rare syndromes of insulin resistance exist, but most cases are unexplained (Box 35.4). Insulin resistance associated with antibodies directed against the insulin receptor has been reported in people with acanthosis nigricans.

Insulin regimens

Many different combinations of insulin are used to treat people with diabetes, but several regimens are more commonly used. The possible treatment options should be discussed with the individual to find the one that is most suitable to their needs and lifestyle.

Basal-bolus regimen

The basal-bolus regimen involves the administration of both short- and long-acting insulin. Long-acting insulin is injected one or two times per day to provide the background (basal) insulin to keep the glucose concentration consistent during periods of fasting (Fig. 35.12). Short-acting insulin is given shortly before eating as a bolus to control glucose concentration after a meal. This regimen most closely mimics normal insulin physiology and is more flexible than other regimens because each injection can be adjusted individually.

This regimen is the treatment of choice for people with type 1 diabetes but is increasingly used in younger people with type 2 diabetes. The flexibility is of great value to people who have busy jobs and work shifts and travel regularly. The main disadvantage of a basal-bolus regimen is the number of injections.

Twice-daily mixed insulin regimen

A mixture of short- and long-acting insulin is injected before breakfast and the evening meal (Fig. 35.12). There are a number of insulin preparations with premixed insulin. This regimen is used more commonly for people with type 2 diabetes but may be used by people with type 1 diabetes when it is not possible to inject four times a day. The main disadvantage of this regimen is the lack of a lunchtime bolus and higher basal levels between meals, which increase the risk of hypoglycaemia.

Basal-only and basal-plus insulin regimens

Because people with type 2 diabetes are able to produce insulin, it may be possible to use a less intensive insulin regimen. Basal long-acting insulin at night, together with other non-insulin treatments during the

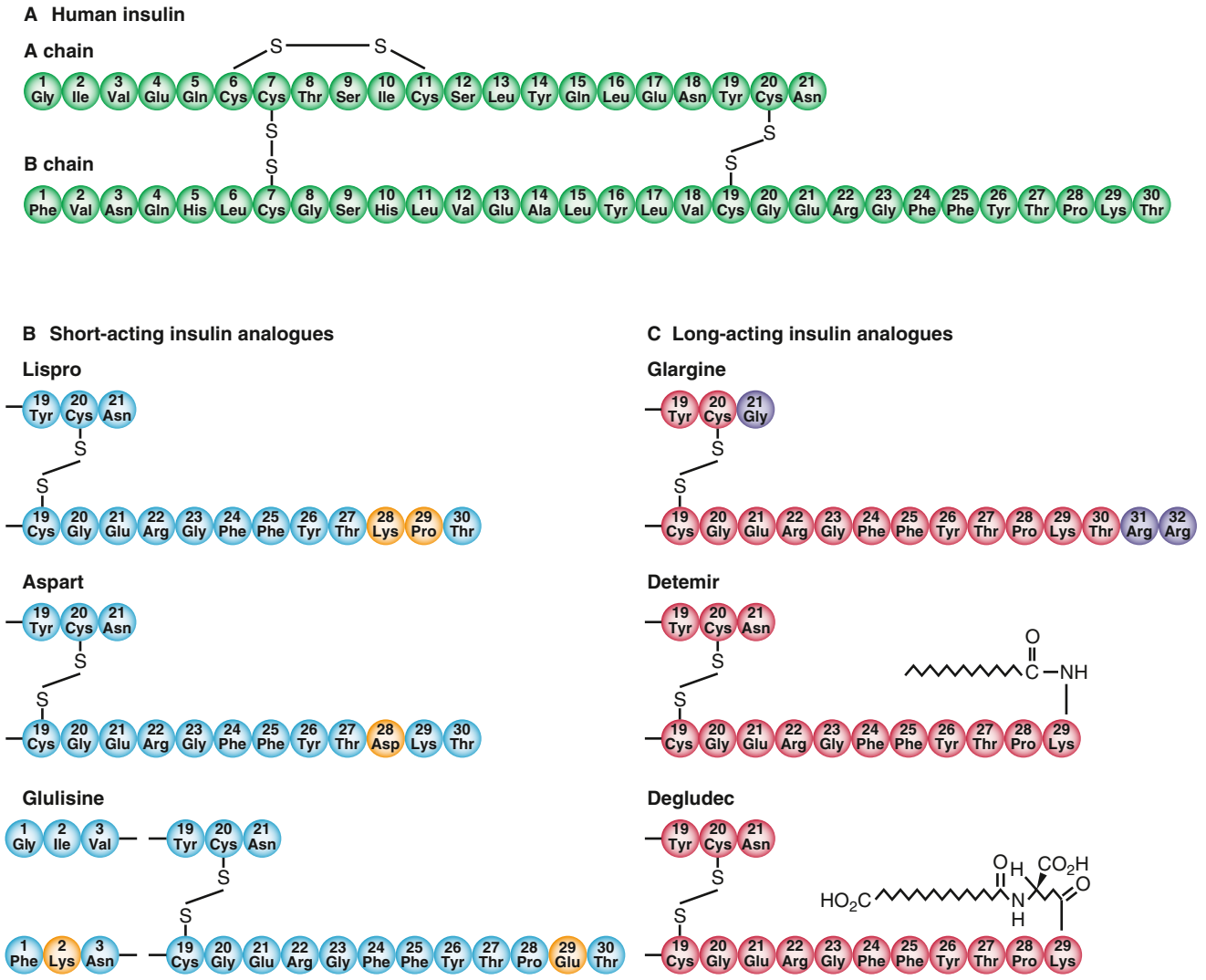


Fig. 35.11 Amino acid structure of insulin. **(A)** Human insulin. **(B)** Modification of human insulin produces rapid-acting insulin analogues. Lispro is created by reversing the order of the amino acids proline and lysine in positions 28 and 29 of the B chain. Aspart is a similar analogue created by replacing proline at position 28 of the B chain with an aspartic acid residue. Glulisine is a similar analogue created by replacing asparagine at position B3 with lysine and the lysine in position B29 is replaced by glutamic acid. **(C)** Modification of human insulin produces long-acting insulin analogues. Insulin glargine is created by replacing asparagine in position 21 of the A chain with a glycine residue and adding two arginines to the end of the B chain. Detemir discards threonine in position 30 of the B chain and adds a fatty acyl chain to lysine in position B29. Degludec is created by removing the threonine at position 30 of the B chain and the attachment via a glutamic acid linker of a 16-carbon fatty diacid to the lysine at position 29 of the B chain.

day, is initially as effective as multidose insulin regimens and is less likely to promote weight gain. Over time, addition of mealtime insulin may become necessary to prevent postprandial hyperglycaemia.

Insulin administration

Insulin is administered subcutaneously by intermittent injection or by insulin pumps (continuous subcutaneous insulin infusion [CSII]).

Insulin injection

The usual injection sites are the front and upper outer side of the thigh, abdominal wall, buttocks, and upper outer arms. Injection technology has improved dramatically since insulin was first discovered. Although insulin may be drawn up from a vial into special plastic insulin syringes, many people now use insulin pens, which may be disposable with prefilled insulin or reusable by changing the insulin cartridge when empty. The pens are portable, use a fine 4- to 5-mm needle, and have a simplified procedure for dialing the insulin dose. Compared with needles and syringes, pens are more convenient, more accurate, and easier to use, and they produce a less painful injection. Even though most injections are virtually painless, people are understandably apprehensive and must learn how to give insulin safely and effectively. Modern pens record when the last dose was given and how much insulin was administered and can connect with glucose monitoring systems.

Insulin pumps

Introduced nearly 40 years ago, insulin pumps are small battery-operated programmable devices that continuously infuse insulin

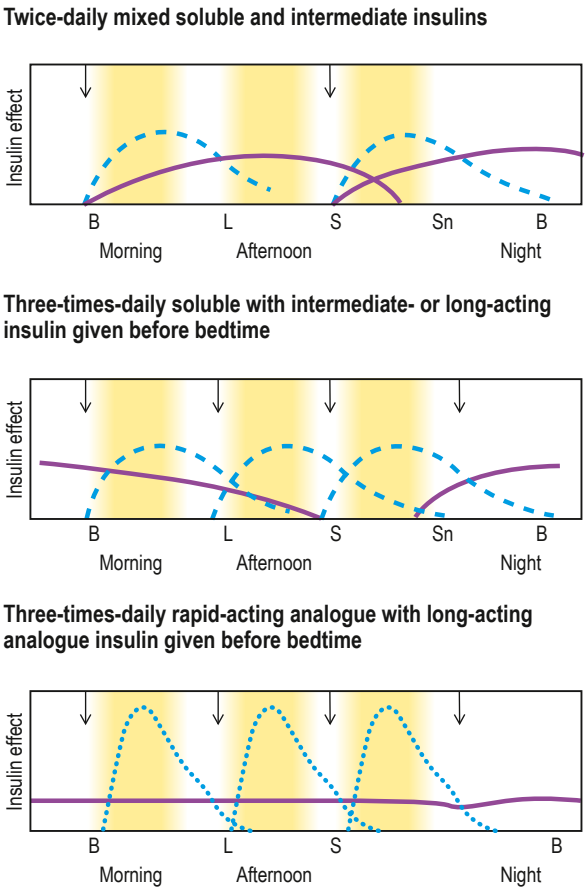


Fig. 35.12 Insulin regimens. Profiles of soluble insulins are shown as *dashed lines*, intermediate- and long-acting insulin as *solid lines (purple)*, and rapid-acting insulin as *dotted lines (blue)*. The *arrows* indicate when the injections are given. *B,* Breakfast; *L,* lunch; *S,* supper; *Sn,* snack (bedtime).

into the subcutaneous tissue. Conventional pumps deliver insulin through a catheter inserted into the subcutaneous tissue, but tubeless pumps (also called patch pumps) that do not use plastic tubing and adhere directly to the skin are increasingly available.

The technology has progressively improved and allows the user to give varying amounts of basal insulin throughout the day with mealtime boluses, guided by the pump software. Newer models are smaller and more portable and have greater technological capabilities, including bolus calculators, advanced bolus delivery options, and the ability to integrate with continuous glucose monitors.

- *Sensor-augmented pumps* alter insulin delivery in response to changes or predicted changes in glucose without input from the user.
- *Hybrid closed-loop systems* are currently the most advanced form of insulin delivery; they use a control algorithm to automatically adjust basal insulin delivery every few minutes based on sensor glucose levels but still require the user to calculate and deliver meal-related boluses.

Fully automated closed-loop insulin delivery systems have been developed by members of the diabetes community, but these devices are not approved, commercialized, or regulated.

Pump therapy in people with type 1 diabetes improves glycaemic levels, reduces hypoglycaemia, and leads to better quality of life compared with multiple daily injections. These benefits have led to an increasing number of pump users, although the uptake of CSII varies across the world; up to 40% of people with type 1 diabetes use pumps in the USA compared with 11.7% in the UK.

The disadvantages of pump therapy include the nuisance of being attached to a gadget, skin reactions, the risk of ketoacidosis if the flow of insulin is interrupted (because there is no protective reservoir of injected basal insulin), and the cost.

Injection site complications

Fatty lumps, known as lipohypertrophy, may occur after the overuse of a single injection site with any type of insulin. As well as being unsightly, lipohypertrophy interferes with insulin absorption, thereby compromising insulin action, and is a cause of glycaemic variability, leading to both hyper- and hypoglycaemia.

Local allergic responses sometimes occur early in therapy but usually resolve spontaneously. Formerly, insulin could produce local atrophy of fat tissue (lipoatrophy), but this is rare because modern insulin is highly purified. Insulin allergy is rare, but reactions may occur to the excipients.

Generalized allergic responses and injection site abscesses are exceptionally rare.

Alternative routes of insulin administration

The first inhaled insulin was withdrawn from the market in 2007 on the grounds of limited clinical demand and safety concerns, but the US Food and Drug Administration approved a new formulation (Afrezza) in July 2014. Oral insulin preparations are in development, but they must overcome the challenges of peptide digestion by the gastrointestinal tract and barriers to absorption.

Challenges of insulin therapy

Although insulin and the means of administering insulin continue to improve, achieving optimal glycaemic levels is extremely challenging for people living with diabetes for a number of reasons:

- The amount of insulin needed to reduce glucose levels in the postprandial state causes hypoglycaemia in the fasted state. If too little insulin is administered, glucose concentrations rise outside the normal range, but if too much insulin is administered, glucose concentrations fall into the hypoglycaemic range.
- Insulin requirements are highly variable from one person to another and vary within the same individual within a day and between days.
- In people without diabetes, insulin is secreted directly into the portal circulation and reaches the liver in high concentration, where 50% is cleared on first passage. By contrast, insulin injected subcutaneously passes into the systemic circulation. People treated with insulin therefore have lower portal insulin levels and higher systemic levels relative to normal physiology.
- Subcutaneous insulin takes longer to enter and achieve peak plasma levels and is slower to leave the circulation.
- The absorption of subcutaneous insulin into the circulation from one injection to the next is variable.
- Basal insulin levels adapt to the metabolic needs of people without diabetes, but injected insulin peaks and declines, with resulting swings in glucose levels in those with diabetes.
- Bolus insulin doses should ideally be adjusted to the carbohydrate content of the meal.
- Carbohydrate counting is difficult because many foods do not list the carbohydrate content.
- It is not always possible to determine how much will be eaten because the appetite may change.

- The rate of absorption of glucose is affected by the glycaemic index and the amount of protein and fat in the meal, as well as the timing of the meal.
- Alcohol and caffeine consumption affect insulin requirements.
- Calculating the appropriate dose of insulin requires advanced mathematical skills.
- Several environmental factors, including temperature and altitude, affect insulin requirements
- The type, duration and timing of exercise, as well as sleep and stress, affect glucose levels.
- There are no holidays from diabetes; an individual using a basal bolus regimen needs to inject approximately 1500 times a year.
- The price of giving too much insulin is hypoglycaemia, which is unpleasant and often embarrassing.
- The pens or devices can fail.

It is critically important that people are trained to use their insulin effectively and safely, and this forms a central component of diabetes self-management education for people with diabetes. As their insulin requirement changes, people should be encouraged to adjust their insulin doses with a view to maintaining blood glucose values between 4 and 7 mmol/L (70–126 mg/dL) before meals and between 4 and 10 mmol/L (70–180 mg/dL) after meals, assuming that this can be achieved without troublesome hypoglycaemia (**Box 35.11**).

Professional advice should be available to support people with diabetes who experience problems; alongside this is a responsibility for healthcare professionals to prescribe insulin safely (**Box 35.12**). Insulin prescribing errors, with potentially fatal consequences, are some of the most common causes of iatrogenic harm. There are many different types of insulin, often with similar sounding names, that are used in different ways and contribute to the confusion surrounding prescription.

Box 35.11 Guide to adjusting insulin dosage according to blood glucose test results[a]

Time	Blood glucose persistently too high	Blood glucose persistently too low
Before breakfast	Increase evening long-acting or premixed insulin	Reduce evening long-acting or premixed insulin
Before lunch	Increase morning short-acting or premixed insulin	Reduce morning short-acting or premixed insulin or eat a mid-morning snack
Before evening meal	Increase morning long-acting or premixed insulin or lunch short-acting insulin	Reduce morning long-acting or premixed insulin or lunch short-acting insulin or increase mid-afternoon snack
Before bed	Increase evening short-acting or premixed insulin	Reduce evening short-acting or premixed insulin

[a]Adjustment should be approximately 10% of the current dose.

Box 35.12 The six steps to insulin safety

- The right person
- The right insulin
- The right dose
- The right device
- The right way
- The right time

Hypoglycaemia

The clinical features and acute management of hypoglycaemia are covered in Chapter 11. After the acute event has been treated, it is important to ascertain why the hypoglycaemia occurred to try to prevent future episodes.

Hypoglycaemia (low blood glucose) is the most common side-effect of insulin therapy and is the major limitation to what can be achieved with insulin treatment. A recent consensus report defined a glucose concentration below 3.0 mmol/L (54 mg/dL) as 'clinically significant hypoglycaemia' but recognized that any value below 4.0 mmol/L (70 mg/dL) represented an 'alert value' that required treatment. 'Severe hypoglycaemia' is defined as hypoglycaemia leading to severe cognitive impairment and requiring external help for recovery.

Most people treated with insulin experience several episodes of symptomatic hypoglycaemia per week and one to two severe episodes per year. Despite the advances in insulin therapy, the incidence of hypoglycaemia has not changed over the past 25 years. Interestingly, with the advent of CGM, it is apparent that asymptomatic hypoglycaemia occurs much more frequently than previously thought. Hypoglycaemia is more common in people with type 1 diabetes, when there is an absolute insulin deficiency, and with longer duration of diabetes. It occurs more frequently in young children and those trying to achieve normal or near-normal glucose levels.

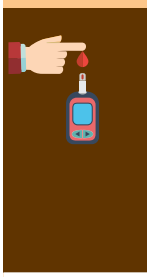
Hypoglycaemia greatly impairs quality of life and induces fear and anxiety in the person with diabetes and their family and carers. In an older study, the fear of severe hypoglycaemia was worse than the fear of developing blindness or end-stage renal disease. Nocturnal hypoglycaemia can be particularly disruptive. In addition to the psychological effects, hypoglycaemia is associated with a number of long-term medical consequences, including cardiovascular events and falls.

Hypoglycaemia is caused when more insulin is injected than is needed. Irregular eating habits, unusual exertion, and alcohol excess may all precipitate hypoglycaemic episodes, but insulin errors and variation in insulin absorption are also important. The times of greatest risk are before meals, during the night, and during and after exercise.

The mechanisms that protect against hypoglycaemia in people without diabetes are compromised because of abnormal islet cell function or defective counter-regulatory hormone secretion. The delivery of insulin after injection cannot be halted, and glucagon secretion is impaired because its main regulator is a fall of intra-islet insulin. Even the third line defence of counter-regulatory hormones and other sympathetic nervous system responses to hypoglycaemia may become diminished in people with longstanding diabetes or recurrent hypoglycaemia.

Weight gain

Because insulin is an anabolic hormone and the absence of insulin leads to profound weight loss, it is unsurprising that many people gain weight while treated with insulin. Restoration of normal weight is a therapeutic effect, but some individuals experience excessive weight gain, especially if the insulin dose is increased inappropriately. Excessive weight gain may be the consequence of frequent hypoglycaemia if an individual takes excess carbohydrate to treat or prevent hypoglycaemia (so-called 'defensive eating'). Weight gain can be offset, to some extent, by dietary measures and exercise.



Non-insulin antidiabetes treatments

Unlike type 1 diabetes, there are alternative pharmacological treatments for people with type 2 diabetes and other types of diabetes when dietary and other lifestyle changes do not adequately bring the glucose into the target range (**Box 35.13**). These treatments are adjuncts rather than alternatives to dietary and lifestyle management. Metformin, SGLT2 inhibitors (or 'flozins'), and GLP-1 receptor agonists have strong evidence supporting health benefits (e.g., mortality benefit and reduction of cardiovascular disease) that extend well beyond their glucose-lowering effects. Most clinical guidelines therefore recommend these agents as first-line antidiabetes medications for people with type 2 diabetes.

First-line treatments

Biguanides: metformin

Metformin is currently the only available biguanide. It was introduced into clinical practice in the 1950s alongside two other biguanides, phenformin and buformin, but these latter two agents were withdrawn because of a high risk of lactic acidosis.

Mode of action

The precise mechanism of action of metformin remains unclear but involves a range of insulin-dependent and insulin-independent actions that improve insulin sensitivity and reduce blood glucose. Metformin enhances post-receptor signaling of insulin and alters cellular energy metabolism by activating the enzyme adenosine monophosphate kinase. Metformin reduces the rate of gluconeogenesis and hence hepatic glucose output. It does not affect insulin secretion, does not induce hypoglycaemia, and does not predispose to weight gain.

Box 35.13 Examples of oral and injectable antidiabetes drugs other than insulin[a]

Drug class	Examples
First-line treatments	
Biguanides	Metformin
SGLT2 inhibitors	Dapagliflozin Empagliflozin Canagliflozin
GLP-1 receptor agonists	Exenatide Liraglutide Lixisenatide Dulaglutide Semaglutide
Second-line treatments	
DPP-4 inhibitors	Sitagliptin Linagliptin Vildagliptin Alogliptin Saxagliptin
Sulphonylureas	Gliclazide Glimepiride Glibenclamide Glipizide Tolbutamide
Thiazolidinediones	Pioglitazone
Meglitinides	Repaglinide Nateglinide
Alpha-glucosidase inhibitors	Acarbose Miglitol Voglibose

[a]This list is not exhaustive, and other drugs are in development in many of these classes.
DPP-4, Dipeptidyl peptidase-4; *GLP-1,* glucagon-like peptide-1; *SGLT2,* sodium–glucose transporter 2.

Clinical use

Metformin is the best-validated treatment for type 2 diabetes and appears as a first-line pharmacological agent in type 2 diabetes guidelines. Metformin may be given in combination with all other oral diabetes treatments as well as GLP-1 receptor agonists or insulin. It lowers fasting plasma glucose by 2 to 4 mmol/L (36–72 mg/dL), corresponding to a fall in HbA_{1c} of 11 to 22 mmol/mol (1%–2%). In UKPDS, metformin treatment led to an unexpected reduction in cardiovascular events. Metformin has been used alongside insulin in people with type 1 diabetes but only with limited efficacy.

Adverse effects

Unwanted effects include gastrointestinal side-effects (10%–20%), such as anorexia, nausea, abdominal discomfort, and diarrhoea. The effects can be mitigated by starting at low dose and gradually increasing it until the desired therapeutic effect is achieved. A slow-release preparation of metformin appears to be better tolerated.

Metformin is contraindicated in renal impairment, cardiac failure, and hepatic failure because of the risk of lactic acidosis. Metformin should not be started in someone whose estimated glomerular filtration rate (eGFR) is less than 45 mL/min per 1.73 m^2 and should be stopped if the eGFR falls below 30 mL/min per 1.73 m^2.

Metformin must be stopped before intravascular administration of iodinated contrast agent because of the risk of renal failure and subsequent lactic acidosis. Treatment can be restarted 48 hours after the test if there has been no deterioration in renal function. Similarly, metformin should be stopped temporarily around surgery or during other intercurrent illnesses that may affect lactate clearance. Metformin should be avoided in people with a history of alcohol misuse.

Metformin reduces gastrointestinal vitamin B_{12} absorption but only rarely causes anaemia.

Sodium-glucose transporter 2 inhibitors ('flozins')

SGLT2 inhibitors have been available since 2011 (**Box 35.13**). Because of strong evidence of their cardioprotective and renoprotective effects, they are recommended as second-line agents or first-line agents in people with diabetes with established or at high risk of atherosclerotic cardiovascular disease or heart failure.

Mode of action

SGLT2 inhibitors lower the renal threshold for glucose, consequently increasing urinary glucose excretion. This has the effect of removing both glucose and calories from the circulation, thus lowering blood glucose by 7 to 13 mmol/mol (0.6%–1.2%) and facilitating weight loss (≈2–4 kg over 6–12 months). They alter glomerular haemodynamics, which may underlie the observed benefits on renal function.

Exactly how SGLT2 inhibitors reduce the risk of myocardial infarction, stroke, cardiovascular death, and heart failure is uncertain; however, several hypotheses have been suggested, including:

- Increased natriuresis leading to fluid loss and small reductions in systolic blood pressure
- Increased production of ketones acting as a 'superfuel' for the heart
- Vasodilation through blockade of the renin–angiotensin system
- Direct effects on the heart, including improved myocardial energetics.

Clinical use

SGLT2 inhibitors can be used as monotherapy, but because their blood glucose–lowering effects glucose are relatively modest, they are used more typically in combination with all other antidiabetes drugs. They are licensed to slow renal progression in people with diabetic nephropathy and are recommended for other cardiac and renal indications in people without diabetes. The glucose-lowering effect is dependent on good renal function, but the renoprotective benefits are seen in people with lower eGFR.

Adverse effects

The most common adverse effects are genital candidiasis and dehydration. Rarer side-effects include DKA, Fournier's gangrene, and lower limb amputation. DKA may occur in people with diagnosed type 2 diabetes. In general, these people have insulin deficiency, and some may have unrecognized type 1 diabetes. A diagnostic trap is that glucose concentrations may be normal or minimally raised in this situation.

GLP-1 receptor agonists

GLP-1 receptor agonists are a heterogeneous class of drugs that act by enhancing the incretin effect. The class is divided into two main groups; the first are drugs that are derived from the protein exendin, which is found in the saliva of Arizonian desert lizards. Exendin has 53% sequence homology with human GLP-1 and is resistant to cleavage by DPP-4, the enzyme that rapidly degrades native GLP-1. The second group are analogues of human GLP-1.

The pharmacokinetics have been modified to prolong the duration of action. The first GLP-1 receptor agonists were given twice daily, but now most are given once-weekly. Most GLP-1 receptor agonists must be given by injection, but there is an oral formulation of semaglutide that has similar effects on blood glucose, body weight, cardio-renal protection, and adverse event profile as injectable semaglutide. To prevent breakdown in the stomach and facilitate absorption across the gastric epithelium, semaglutide has been combined with the absorption enhancer sodium hydroxybenzoylamino-caprylate. Oral semaglutide is taken daily on an empty stomach, and food should be avoided for at least 30 minutes to allow adequate drug absorption.

There is strong clinical trial evidence that GLP-1 receptor agonists reduce the risk of myocardial infarction, stroke, cardiovascular death, and heart failure. The GLP-1 receptor agonists may also have reno- and neuroprotective effects, and may have a major role in the treatment of people with obesity (see Chapter 37 for details of GLP-1 agonists in weight loss programmes).

Mode of action

GLP-1 receptor agonists enhance the incretin effect by activating the GLP-1 receptor. Unlike DPP-4 inhibitors that restore physiological GLP-1 levels (see later), GLP-1 receptor agonists achieve supraphysiological levels and are therefore more potent than DPP-4 inhibitors. In addition to their effects on the pancreas to increase insulin secretion and decrease glucagon, they also act on the hypothalamus to reduce appetite and food intake, leading to weight loss. The size of effect has led to the licensing of liraglutide and semaglutide as anti-obesity treatments (see Chapter 37).

GLP-1 receptors are widely expressed in the cardiovascular system, and GLP-1 receptor agonists appear to protect the heart against ischaemic damage, both through direct effects on the heart as well as indirect effects, such as lowered blood pressure and free fatty acids.

Clinical use

Despite the need for injection, clinical guidelines presently endorse use of GLP-1 receptor agonists in combination with other antidiabetes agents, as second-line therapies. They should not be combined with DPP-4 inhibitors because DPP-4 inhibitors do not confer any additional benefit. Clinical trials have shown reductions of HbA_{1c} in the ranges 9 to 19 mmol/mol (0.8%–1.8%) and reduction of weight from –0.6 to 4.3 kg. As increasing evidence builds of the effectiveness of GLP-1 receptor agonists in reducing weight and cardiovascular risk, they are recommended as first-line agents in those with established or high risk of cardiovascular disease.

Adverse effects

The most common side-effects of GLP-1 receptor agonists are gastrointestinal and include nausea and vomiting, bloating, and diarrhoea. These tend to diminish with time and occur less frequently with longer-lasting GLP-1 receptor agonists. There is a low risk of hypoglycaemia, but this may occur if GLP-1 receptor agonists are combined with insulin or sulphonylureas. GLP-1 receptor agonists should not be used in people with a history of pancreatitis because of a risk of acute pancreatitis. Recent studies have not confirmed earlier concerns about links with pancreatic and thyroid cancer.

Dual GIP-GLP-1-receptor agonists

The dual GIP–GLP-1 receptor agonist tirzepatide has been developed for once-weekly subcutaneous injection and exerts greater reductions in HbA_{1c} and body weight in individuals with type 2 diabetes than GLP-1 receptor agonists. Further dual or triple agonists are in clinical development.

Second-line treatments

The following classes of antidiabetes agents may be used in people unable to tolerate first-line agents or when additional glucose management is required.

DPP-4 or 'gliptins'

DPP-4 inhibitors have been used since 2006 and improve glycaemia by enhancing the incretin effect.

Mode of action

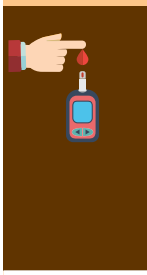
These drugs inhibit the enzyme DPP-4, which prevents the rapid inactivation of GLP-1, which in turn increases insulin secretion and reduces glucagon secretion.

Clinical use

DPP-4 inhibitors are most effective in the early stages of type 2 diabetes, when insulin secretion is relatively preserved, and are currently recommended for second-line use in combination with metformin. Although they only have relatively modest effects on glucose lowering, they are included in most type 2 diabetes treatment algorithms because they are well tolerated. DPP-4 inhibitors do not promote weight gain, have a low risk of hypoglycaemia, and can be used safely in people with renal impairment.

Adverse effects

Nausea may occur, and there have been occasional reports of acute pancreatitis. DPP-4 inhibitors do not alter the incidence of myocardial infarction, but saxagliptin may increase the risk of heart failure.



Sulphonylureas

Sulphonylureas were originally derived from sulfonamide antibiotics after it was recognized that the latter caused hypoglycaemia during a typhoid epidemic in Marseilles, France, in 1942 (**Box 35.13**).

Mode of action

Sulphonylureas act on the β cell to induce insulin secretion. They are ineffective in people without a functional β-cell mass and therefore have no effect in people with type 1 diabetes. Sulphonylureas bind to the sulphonylurea receptor 1 on the β-cell membrane, which closes ATP-sensitive potassium channels and blocks potassium efflux. The resulting depolarization promotes calcium influx and stimulates insulin release (**Fig. 35.4**).

Clinical use

As monotherapy, sulphonylureas typically reduce fasting plasma glucose by 2 to 4 mmol/L (36–72 mg/dL), corresponding to a fall in HbA_{1c} of 11 to 22 mmol/mol (1–2%). Although this is more effective than many other oral agents in the short term (1–3 years), the sulphonylurea effect wears off as the β-cell mass declines. Sulphonylureas can also be used in combination with other oral antidiabetes agents or basal insulin, although it is usually stopped when the individual requires short-acting insulin at mealtimes.

Adverse effects

Weight gain, typically 1 to 4 kg, and hypoglycaemia are the most common side-effects. Hypoglycaemia affects about one-fifth of those taking the drug, but severe hypoglycaemia is uncommon. The risk of hypoglycaemia is increased with longer-acting sulphonylureas, excessive alcohol intake, and older age and during intercurrent infection. Severe sulphonylurea-induced hypoglycaemia should be managed in hospital for up to 48 hours with glucose support until the drug has cleared from the circulation. Sulphonylureas should be used with care in people with liver or renal disease.

Thiazolidinediones or 'glitazones'

Three thiazolidinediones have been launched, but troglitazone was withdrawn because of liver toxicity, and rosiglitazone has been withdrawn or restricted because of concerns about cardiovascular safety. In many countries, pioglitazone is the only available thiazolidinedione.

Mode of action

Thiazolidinediones reduce insulin resistance by interaction with peroxisome proliferator-activated receptor-gamma, a nuclear receptor that regulates large numbers of genes, including those involved in lipid metabolism and insulin action. Thiazolidinediones increase fatty acid uptake by adipocytes and increase lipogenesis. This lowers circulating non-esterified free fatty acids, which in turn promotes glucose utilization in muscle and liver via the glucose–fatty acid (Randle) cycle. Thiazolidinediones reduce hepatic glucose production, an effect that is synergistic with that of metformin, and enhance peripheral glucose uptake. Like metformin, the glitazones potentiate the effect of endogenous or injected insulin.

Clinical use

Thiazolidinediones can be used as monotherapy if first-line agents are not tolerated or in combination with other antidiabetes drugs, including insulin. Thiazolidinediones do not cause hypoglycaemia as monotherapy, and pioglitazone may specifically benefit people with metabolic dysfunction-associated steatotic liver disease, a frequent co-morbidity of type 2 diabetes. Because the effect on plasma glucose is indirect, thiazolidinediones may take up to 3 months to reach their maximal effect.

Adverse effects

The most common adverse effect is weight gain of 5 to 6 kg, but pioglitazone may also cause fluid retention precipitating heart failure. Mild anaemia and osteoporosis resulting in peripheral bone fractures have been reported.

Other antidiabetes agents

Meglitinides or postprandial insulin releasers

Meglitinides are short-acting agents that promote insulin secretion in response to meals. Like sulphonylureas, meglitinides act by closing the K^+-ATP channel in the β cells (**Fig. 35.4**). They are rapidly metabolized and have a short duration of action of less than 3 hours. They may be used to treat people with postprandial hyperglycaemia and normal fasting glucose levels. They may be useful in older frail people in whom there is an imperative to avoid hypoglycaemia. However, meglitinides are less effective than other drugs, and their role in diabetes management is not well established.

Alpha-glucosidase inhibitors

Alpha-glucosidase inhibitors, such as acarbose, are drugs that reduce carbohydrate absorption after a meal by preventing the action of α-glucosidase, the last enzyme involved in carbohydrate digestion, from breaking down disaccharides to monosaccharides. This slows the absorption of glucose after a meal and lowers postprandial glucose.

Although α-glucosidase inhibitors can be used as monotherapy or in combination with all other antidiabetes drugs, they are not widely used because of their limited efficacy and gastrointestinal side-effects.

Quick-release bromocriptine

Abnormal circadian rhythm is associated with the development of insulin resistance, obesity, and diabetes. When administered at daybreak, quick-release bromocriptine appears to reset hypothalamic dopamine circadian rhythms. It lowers HbA_{1c} by 5 to 8 mmol/mol (0.5%–0.7%) either as monotherapy or in combination with other antidiabetes medications. The drug is well tolerated.

Colesevelam

Colesevelam is a bile acid–binding resin that lowers cholesterol and can reduce blood glucose concentrations by an unknown gastrointestinal mechanism.

Amylin analogues

Pramlintide is an analogue of amylin and is approved in the USA for use in people with either type 1 diabetes or insulin-treated type 2 diabetes. Amylin is co-secreted with insulin and delays gastric emptying, suppresses postprandial glucagon secretion, and increases satiety. It is not widely used.

Which drug and when?

With the introduction of increasing numbers of drugs to treat type 2 diabetes, the choice of which drug and when becomes more and more relevant. Because type 2 diabetes is characterized by progressive β-cell failure and glucose levels deteriorate over time, this requires a progressive and pre-emptive escalation of diabetes therapy. There is strong evidence of delays in treatment escalation,

Box 35.14 Considerations when choosing treatment for a person with type 2 diabetes

Characteristics of the person with diabetes
- Current lifestyle
- Co-morbidities
  - Atherosclerotic vascular disease
  - Heart failure
  - Chronic kidney disease
- Clinical characteristics
  - Age
  - Weight
  - Current HbA_{1c}
- Psychosocial issues such as motivation and depression
- Cultural and socioeconomic context

Specific factor
- Individualized HbA_{1c} target
- Impact on weight and hypoglycaemia
- Medication side-effects
- Treatment complexity
- Choose regimen to optimize medication taking
- Access, cost and medication availability

Shared decision-making
- Involves an educated and informed person with diabetes (and their family or caregiver)
- Seeks patient preferences
- Ensures effective consultation
- Empowers the person with diabetes
- Ensures access to diabetes self-management education

HbA_{1c}, Glycated haemoglobin.

which exposes the individual with diabetes to prolonged periods of hyperglycaemia and the inherent risks of complications.

Decisions about treatment should be individualized and shared with the person with diabetes. The importance of diabetes self-management education is highlighted to enable the person with diabetes to make an informed choice. Clinicians are advised to consider specific patient characteristics and factors that will inform the discussion (**Box 35.14**).

Metformin is recommended as the first-line treatment for most people with type 2 diabetes whose glucose levels remain above target despite optimal lifestyle management. However, for those with or at high risk of atherosclerotic cardiovascular disease, SGLT2 inhibitors or GLP-1 receptor agonists with proven cardiovascular disease benefit are preferred. SGLT2 inhibitors are also preferred for individuals with heart failure or chronic kidney disease (CKD).

If there is a pressing need to minimize weight gain or promote loss, again GLP-1 receptor agonists or SGLT2 inhibitors are recommended. Where there is a need to avoid hypoglycaemia, GLP-1 receptor agonists, SGLT2 inhibitors, a DPP-4 inhibitor, and thiazolidinediones may be used. Sulphonylureas or thiazolidinediones are preferred when cost is a major issue.

When dual or triple therapy is needed, other recommended drugs can be added. For example, if a GLP-1 receptor agonist is used first in someone with established atherosclerotic cardiovascular disease, an SGLT2 inhibitor would be an appropriate second-line choice or vice versa. Beyond triple therapy, insulin is required because further additional treatments have only a minimal effect, if any.

Several fixed-dose oral combinations and fixed-ratio injectable combinations of insulin and GLP-1 receptor agonists are available and reduce the pill or injection burden for the person with diabetes as treatment is escalated.

PANCREATIC ISLET AND WHOLE-PANCREAS TRANSPLANTATION

Islet transplantation

Since 2000, there has been remarkable progress in the technical aspects of islet cell processing and the outcomes of islet cell transplantation. The process involves harvesting pancreatic islets from cadavers (two or three pancreata are usually needed), which are then injected into the portal vein and seed themselves into the liver. It is one of the safest and least invasive transplant procedures. Islet cell transplantation is indicated for people with type 1 diabetes who have problematic hypoglycaemia, defined as two or more episodes of severe hypoglycaemic episodes per year, impaired awareness of hypoglycaemia, or persistent hyperglycaemia despite optimal diabetes management. The goal of treatment is to prevent severe hypoglycaemia, but about 30% of people receiving islet cell transplants maintain insulin independence after 5 years.

Whole-pancreas transplantation

Worldwide, more than 63 000 whole-pancreas transplantations have been performed. They are most often undertaken in people with type 1 diabetes and end-stage kidney disease at the same time as kidney transplantation ('simultaneous pancreas–kidney' transplants). Pancreas transplant may be performed less frequently after a kidney transplant or alone. Pancreatic transplantation can establish normoglycaemia and greatly improve quality of life. There is some evidence of protection against or reversal of some diabetes complications, but this comes at the cost of long-term immunosuppression.

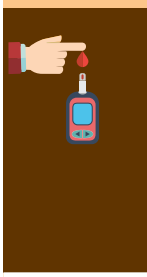
MONITORING GLUCOSE LEVELS IN PEOPLE WITH DIABETES

Measures of glucose levels are needed to help people with diabetes and their healthcare professionals make rational choices about therapy. CGM of interstitial glucose is now the measure of choice for people with type 1 diabetes and many people with insulin treated type 2 diabetes. Measurements of glucose levels can be divided into short-term measures that provide an almost instantaneous record of the current glucose concentration and long-term measures that provide an assessment of average glucose concentration over the preceding weeks or months.

Short-term measures of glucose levels

Self-monitoring of capillary blood glucose

Handheld meters allow people with diabetes to measure their capillary blood glucose concentration regularly throughout the day wherever they are. Blood is taken from the side of a fingertip (not from the top, which is densely innervated) using a special lancet. Single blood glucose concentrations are of limited value because of their wide variability, but repeated glucose measurements offer invaluable feedback that permits appropriate adjustment of treatment to be made. Some meters include bolus calculators that advise the individual on the appropriate dose of mealtime insulin based on the current glucose and estimated carbohydrate of the meal. The parameters of the automated bolus calculator (the ratio of insulin dose to carbohydrate intake and the amount by which one unit of insulin will reduce glucose levels) need to be programmed into the meter, usually during a consultation.



Self-monitoring of capillary blood glucose is recommended for everyone treated with insulin, but some people with diet- or tablet-treated type 2 diabetes may also find testing a useful guide to therapy. Like any investigation, measuring capillary blood glucose should inform future management, so if individuals are unable or unwilling to adjust their treatment, then 'testing for testing's sake' should be discouraged. Blood glucose monitoring does not, in itself, improve glucose levels but is essential to those wishing to achieve it.

Continuous glucose monitoring

CGM systems measure interstitial glucose every few minutes, allowing the user to see trends in glucose levels throughout the day, in particular at times when testing is not usually done, for example, at night. By tracking the glucose, the user can act earlier if the glucose starts to rise or fall. These devices have three parts, a sensor that sits just underneath the skin, a transmitter attached to the sensor, and a display device that shows the glucose level. The sensors need to be changed frequently, but some devices can be worn for up to 14 days.

The technology is developing extremely rapidly; some systems have alarms that alert the user to either high or low glucose levels. CGM devices can be integrated with pumps, allowing the insulin infusion to be suspended if the glucose levels are low or predicted to become low. *Hybrid closed-loop systems* (see 'Insulin pumps' section, earlier in the chapter) use a control algorithm to automatically adjust basal insulin delivery every few minutes based on sensor glucose levels. Results can be shared with healthcare professionals or others via the cloud to allow remote monitoring.

Clinical studies have shown that the use of CGM improves HbA_{1c} while reducing glucose variability and thereby lowering the risk of hypoglycaemia in type 1 diabetes.

The main disadvantage of CGM is the lag of 15 to 20 minutes between changes in blood glucose and interstitial glucose. Caution is needed if the glucose levels are changing rapidly, and if an individual experiences hypoglycaemic symptoms, they should test their capillary glucose even if the CGM reading is normal. Although many people with type 1 diabetes find CGM incredibly helpful, the wealth of data and alarms can overburden others.

CGM is recommended for people with type 1 diabetes. It is also recommended for people with type 2 diabetes on multiple daily injections if they have recurrent or severe hypoglycaemia, impaired aware of hypoglycaemia, or a condition or disability that means that they cannot self-monitor or if they would otherwise need to test eight times a day.

Long-term measures of glycaemia

Glycated haemoglobin

HbA_{1c} is a measure of an individual's average blood glucose concentration over the previous 6 to 8 weeks. The glycation occurs as a two-step reaction, resulting in the formation of a covalent bond between the glucose molecule and the terminal valine of the β chain of the haemoglobin molecule. The rate at which this reaction occurs is related to the prevailing glucose concentration. HbA_{1c} is expressed as the mmol concentration of HbA_{1c} per mole of normal haemoglobin (standardized range, 20–42 mmol/mol); formerly, it was expressed as the percentage of total haemoglobin that is glycated (normal range, 4%–6%).

Because HbA_{1c} provides an integrated measure of blood glucose concentration over the life of the haemoglobin molecule (≈6 weeks), the result may be misleading if the lifespan of the red blood cell is altered (**Box 35.15**).

Box 35.15 Limitations of HbA_{1c}

Analytical variability

- Different methods for measuring HbA_{1c} may give different results:
  - Largely addressed by standardization of the reference method
- Molecular variants of haemoglobin:
  - Fetal haemoglobin

Biological variability

- Interindividual variability:
  - An individual with a mean glucose of 10 mmol/L (180 mg/dL) may have an HbA_{1c} value that ranges from 6.0% to 9.0% (42–75 mmol/mol), probably reflecting differences in rates of protein glycation
  - Ageing
  - Ethnicity
  - Variation in erythrocyte lifespan:
- Shortened lifespan can give spuriously low results:
  - Haemolytic anaemia
  - Acute or chronic blood loss
  - Pregnancy
- Longer lifespan can give spuriously high results:
  - Iron deficiency
- Renal failure
- HIV infection

Clinical variability

- Predictive link between HbA_{1c} and clinical outcomes is not clear-cut

HbA_{1c}, Glycated haemoglobin.

Fructosamine

Glycation albumin measured as fructosamine relates to glycaemic levels over the preceding 2 to 3 weeks. It may be useful in people with anaemia or haemoglobinopathy or during pregnancy but is only rarely used.

Further reading

American Diabetes Association. Standards of care in diabetes—2023. *Diabetes Care* 2023; 46(Suppl 1):S1–S285.

Battelino T, Danne T, Bergenstal R et al. Clinical targets for continuous glucose monitoring data interpretation: recommendations from the international consensus on time in range. *Diabetes Care* 2019; 42:1593–1603.

Davies MJ, Aroda VR, Collins BS et al. Management of hyperglycaemia in type 2 diabetes, 2022. A consensus report by the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). *Diabetologia* 2022; 65:1925–1966 and *Diabetes Care* 2022; 45:2753–2786.

DCCT/Epidemiology of Diabetes Interventions and Complications Research Group. Retinopathy and nephropathy in patients with type 1 diabetes four years after a trial of intensive insulin therapy. *N Engl J Med* 2000; 342:381–389.

Dyson PA, Twenefour D, Breen C et al. Diabetes UK evidence-based nutrition guidelines for the prevention and management of diabetes. *Diabet Med* 2018; 35:541–547.

Holman RR, Paul SK, Bethel MA et al. 10 year follow-up of intensive glucose control in type 2 diabetes. *N Engl J Med* 2008; 359:1577–1589.

Holt RIG, DeVries JH, Hess-Fischl A et al. The management of type 1 diabetes in adults. A consensus report by the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). *Diabetologia* 2021; 64:2609–2652 and *Diabetes Care* 2021; 44:2589–2625.

Kilpatrick ES, Atkin SL. Using haemoglobin A(1c) to diagnose type 2 diabetes or to identify people at high risk of diabetes. *BMJ* 2014; 348:g2867.

Lean MEJ, Leslie WS, Barnes AC et al. Durability of a primary care-led weight-management intervention for remission of type 2 diabetes: 2-year results of the DiRECT open-label, cluster-randomised trial. *Lancet Diabetes Endocrinol* 2019; 7:344–355.

National Institute for Health and Care Excellence. *NICE Guideline 17: Type 1 Diabetes in Adults: Diagnosis and Management*. NICE 2016; https://www.nice.org.uk/guidance/ng17.

National Institute for Health and Care Excellence. *NICE Guideline 28: Type 2 Diabetes in Adults: Management*. NICE 2019; https://www.nice.org.uk/guidance/ng28.

Nguyen NT, Varela JE. Bariatric surgery for obesity and metabolic disorders: state of the art. *Nat Rev Gastroenterol Hepatol* 2017; 14:160–169.

Pickup JC. Is insulin pump therapy effective in type 1 diabetes? *Diabet Med* 2019; 36:269–278.

DIABETES-RELATED METABOLIC EMERGENCIES

The two main diabetes-related hyperglycaemic emergencies are DKA and HHS. Together with severe hypoglycaemia, these are covered in detail in Chapter 11 on acute medicine, alongside the emergency management of hypoglycaemia (see Chapter 11). DKA and HHS are both life-threatening emergencies that require prompt treatment to reduce morbidity and mortality. Although often considered as clinically distinct phenomena, DKA and HHS exist on a spectrum. Both are characterized by hyperglycaemia with insulin deficiency, with only the degree of hyperglycaemia, level of associated dehydration, and severity of the associated metabolic acidosis distinguishing them.

COMPLICATIONS OF DIABETES

When effective treatment allowed people to survive the acute metabolic consequences of diabetes, it became apparent that long-term diabetes was associated with the development of a number of complications. These can be broadly divided into two groups: the microvascular complications that affect the capillaries and arterioles throughout the body but particularly involve the eyes (retinopathy), kidneys (nephropathy), and nerves (neuropathy), and macrovascular complications, which increase the risk of myocardial infarction, stroke, and peripheral vascular disease. With improved treatment of people with diabetes, the incidence of diabetes complications has fallen for each individual, but overall the prevalence has gone up as the number of people with diabetes has increased.

Microvascular complications

Microvascular complications are specific to diabetes and affect more than 80% of individuals with diabetes. They are unusual in the first 10 years after the diagnosis of type 1 diabetes but are found in 20% to 50% of people with newly diagnosed type 2 diabetes as a result of the preceding undiagnosed hyperglycaemia. With the advent of screening for type 2 diabetes, the numbers with complications at presentation are falling.

Pathology of microvascular complications

Thickening of the capillary and arteriole basement membrane is the cardinal feature of microvascular complications. Although basement membrane thickening is seen in most people with diabetes, organ damage occurs much less frequently; however, with time, the small vessels become progressively narrower and eventually become blocked, leading to ischaemia and tissue dysfunction. The tissues damaged by hyperglycaemia are those that have high blood flow and cannot downregulate glucose uptake in the presence of hyperglycaemia.

The mechanisms leading to damage are ill-defined but likely to be multifactorial. There is no doubt that the duration and degree of hyperglycaemia play a major role in the development of complications. However, other factors also seem to play a role because some people are relatively protected against microvascular complications despite longstanding diabetes. Twin studies suggest that genetic polymorphisms may affect the risk of complications. There are also racial differences in the overall rates of nephropathy; in the USA, the prevalence is highest in Pima Native American people followed by Hispanics and Mexican Americans, then Black Americans with the lowest rates in people of White Northern European ancestry.

The following are consequences of hyperglycaemia and likely contribute to the development of microvascular complications. The common feature of each stems from a single hyperglycaemia-induced process of overproduction of superoxide by the mitochondrial electron chain.

Formation of advanced glycation end-products

When a wide variety of proteins are exposed to increased glucose concentrations, glucose binds irreversibly to the protein to form advanced glycation end-products (AGEs). These accumulate in proportion to hyperglycaemia and time; one example of this is HbA_{1c}, which is used to diagnose diabetes and monitor treatment. AGEs cause tissue injury and inflammation via stimulation of pro-inflammatory factors, such as complement and cytokines.

Increased flux of glucose through the sorbitol–polyol pathway

When hyperglycaemia occurs, excess glucose is metabolized to sorbitol via the polyol pathway. This leads to accumulation of sorbitol and fructose, which cause changes in vascular permeability, cell proliferation, and capillary structure via stimulation of protein kinase C and transforming growth factor-beta (TGF-β).

Abnormal microvascular blood flow

Hyperglycaemia swamps the normal autoregulatory mechanisms that limit tissue blood flow and the high flow rates damage the tissue, particularly when hypertension is present. Endothelial damage occurs and vasoconstrictors, such as endothelins and thrombogenesis, are released that can lead to microvascular occlusion. Nutrients and oxygen supply are impeded.

Blockage of the renin–angiotensin system slows the progression of microvascular complications to a greater extent than other blood pressure–lowering agents, suggesting the renin–angiotensin system may have a role in the development of microvascular complications.

Growth factors and cytokines

Through increased expression of protein kinase C β (PKCβ), a number of mitogenic cytokines and growth factors, including TGF-β, tumour necrosis factor, and vascular endothelial growth factor (VEGF), are upregulated.

Growth hormone–insulin-like growth factor axis

The reduction in portal insulin concentrations impairs the ability of growth hormone to generate insulin-like growth factor 1 in the liver, which in turn leads to growth hormone hypersecretion by the pituitary gland. Growth hormone excess has been implicated in the development of microvascular complications, and indeed, pituitary ablation was used to treat patients with retinopathy before laser therapy became available.

Diabetic retinopathy

Diabetic retinopathy (damage to the retina) is the most commonly diagnosed diabetes-related complication (**Fig. 35.13**, see Chapter 31). Its prevalence increases with the duration of diabetes and affects around one-third of all people with diabetes (**Fig. 35.14**). Approximately 20% of people with type 1 diabetes will have retinal changes after 10 years, rising to 90% after 20 years (**Box 35.16**); 20% to 30% of people with type 2 diabetes have retinopathy at diagnosis, 10% of people with diabetes develop sight-threatening retinopathy,

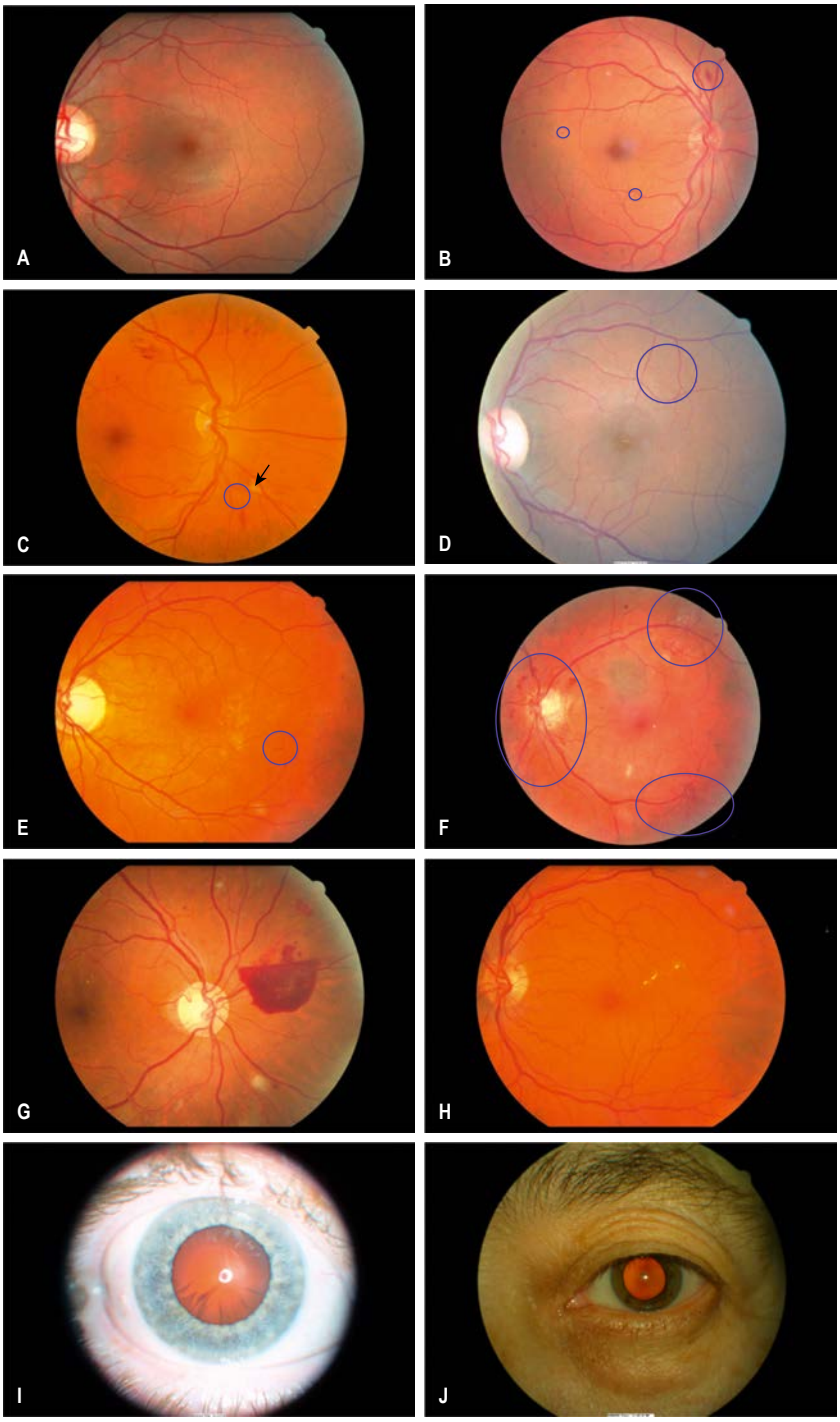


Fig. 35.13 Features of diabetic eye disease. **(A)** The normal macula *(centre)* and optic disc *(to left)*. **(B)** Microaneurysms *(small circles)* and blot haemorrhage *(larger circle)* indicating early non-proliferative retinopathy. **(C)** Hard exudates *(circled)* and single cotton wool spot *(arrow)* in addition to multiple blot haemorrhages in non-proliferative retinopathy. **(D)** Intraretinal microvascular abnormalities (IRMAs) indicating pre-proliferative retinopathy *(circled)*. **(E)** Venous loop *(circled)* also indicates pre-proliferative change. **(F)** Fronds of new vessels on the disc and elsewhere (proliferative; *circled*). **(G)** Pre-retinal haemorrhage in proliferative disease. **(H)** Hard exudates within a disc-width of the macula (maculopathy). **(I)** Cortical cataracts and **(J)** central cataracts can be seen against the red reflex with the ophthalmoscope.

and worldwide diabetic retinopathy is the most common cause of blindness in people of working age. However, the incidence and prevalence vary considerably between countries. In countries with good diabetes healthcare and systematic population screening, retinopathy develops less frequently and is more likely to be detected before sight-threatening changes have occurred. Approximately 1% to 7% of people with diabetes will require laser photocoagulation to prevent or limit progression to proliferative retinopathy.

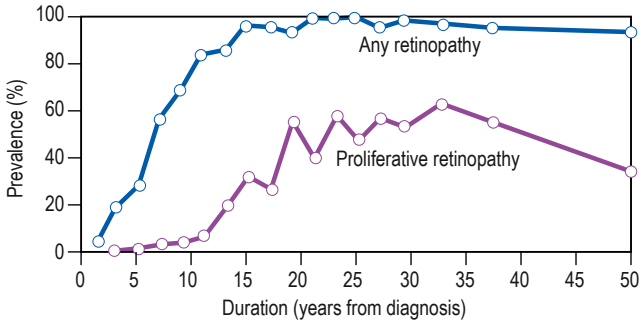


Fig. 35.14 Prevalence of retinopathy. Prevalence is shown in relation to duration of the disease in people with type 1 diabetes mellitus diagnosed before the age of 33 years. Almost all eventually develop background change, and 60% progress to proliferative retinopathy. The incidence of retinopathy and progression to proliferative retinopathy has declined substantially with improved glycaemic management. (*Data from* Archives of Ophthalmology. *1984;102:520.*)

Early changes without vision loss

The earliest changes in the retina are known as non-proliferative or background retinopathy (**Box 35.16**). Damage to the wall of small vessels causes microaneurysms (small red dots) within the retina. When vessel walls are breached, superficial (blot) haemorrhages occur in the ganglion cell and outer plexiform layers. Damaged blood vessels leak fluid into the retina. The fluid is cleared into the retinal veins, leaving behind protein and lipid deposits, which are seen as hard exudates. These are eventually cleared by macrophages. Micro-infarcts within the retina caused by occluded vessels cause cotton wool spots. The spot is swelling of the retinal nerve fibers because normal axoplasmic transport is disrupted, leading to the accumulation of axoplasmic debris. This debris is removed by macrophages. As this occurs, there may be white dots at the site of the previous cotton wool spot (cytoid bodies).

Sequential retinal photographs have demonstrated that all these lesions can heal, and it is not uncommon for changes to be seen on one occasion but not the next.

Transition to sight-threatening retinopathy

The next stage of retinopathy is known as pre-proliferative retinopathy (**Box 35.16**). Damage to the walls of veins causes their calibre to vary (venous beading) and elongation to occur, causing venous loops.

Sight-threatening retinopathy

With further progression, proliferative retinopathy develops (**Box 35.16**). Blockage of blood vessels leads to areas of capillary non-perfusion. Ischaemia in these areas causes the release of vascular growth factors such as VEGF. These factors cause new blood vessels to grow in the retina (neovascularization).

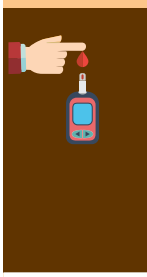
Some of these new blood vessels are inside the retina and are helpful. These new intraretinal vessels and other vessels whose walls are damaged and dilated give the appearance of intraretinal microvascular abnormalities (IRMAs). IRMAs are a feature of pre-proliferative retinopathy.

Other new vessels emerge through the retina and lie on its surface, usually at the margin of an area of capillary closure. The normal shearing stresses that occur within the eye can cause these poorly supported new vessels to bleed. Small haemorrhages give rise to pre-retinal haemorrhages (boat-shaped haemorrhages). With further bleeding, vitreous haemorrhage may occur with sudden loss of vision. Later, collagen tissue grows along the margins of the new vessels and gives rise to fibrotic bands. These bands may contract and pull on the retina, causing further haemorrhage and ***retinal detachment***. Sometimes vessels may be induced to grow on the pupil margin ***(rubeosis)*** and in the angle of the anterior chamber of the eye, giving rise to a rapid increase in intraocular pressure ***(rubeotic glaucoma)***.

Box 35.16 Grading and management of pathological changes in diabetic retinopathy

Retinopathy grade	Retinal abnormality (cause)	Action needed
Peripheral retina		
Nonproliferative (R1)	Dot haemorrhages (capillary microaneurysms) usually appear first Blot haemorrhages (leakage of blood into deeper retinal layers) Hard exudates (exudation of plasma rich in lipids and protein)[a] Cotton wool spots or cytoid bodies[b]	Annual screening only
Pre-proliferative (R2)	Venous beading or loops Intraretinal microvascular abnormalities Multiple deep, round haemorrhages	Non-urgent referral to an ophthalmologist
Proliferative (R3)	New blood vessel formation or neovascularization Pre-retinal or subhyaloid haemorrhage Vitreous haemorrhage	Urgent referral to an ophthalmologist
Advanced retinopathy	Retinal fibrosis Traction retinal detachment	Urgent referral to an ophthalmologist, but much vision has already been lost
Central retina		
Maculopathy (M1)	Hard exudates within 1 disc width of the macula Lines or circles of hard exudates within 2 disc widths of the macula Microaneurysms or retinal haemorrhages within 1 disc width of the macula if associated with an unexplained visual acuity of 6 of 12 or worse	Prompt referral to an ophthalmologist

[a]Hard exudates have a bright yellowish-white colour and are often irregular in outline with a sharply defined margin.
[b]Cotton-wool spots are greyish-white and have indistinct margins and a dull matt surface, unlike the glossy appearance of hard exudates.



Fluorescein angiography (a fluorescent dye is injected into an arm vein and photographed in transit through the retinal vessels) may be needed to define the extent of the potentially sight-threatening diabetic retinopathy.

Maculopathy

The progression of early diabetic retinopathy may result in the development of retinal oedema or exudates within the macular area, which is responsible for fine vision (**Box 35.16**). Fluid from leaking vessels is cleared poorly in this area because its anatomy differs from that of the rest of the retina. Above a certain rate of formation, clearance fails, and macular oedema occurs. This distorts and thickens the retina at the macula. If sustained, this distortion causes loss of central vision. Capillary occlusion in the macular area also causes loss of central vision.

Macular oedema is not visible with the ophthalmoscope or with retinal photography, and ocular coherence tomography is used to image the content of the layers of the macula and measure retinal thickness.

Management of retinopathy

Prevention is the best way of managing retinopathy; the DCCT and UKPDS showed conclusively that the risk of the development and progression of diabetic retinopathy can be reduced by optimal management of both glucose and blood pressure. However, a short-term deterioration of retinopathy may occur after rapid improvement in glycaemic levels, during pregnancy, and by nephropathy. The lipid-lowering drug fenofibrate has been shown to slow the progression of retinopathy, including macular oedema. Smoking cessation is recommended.

Intravitreal injection

Repeated injections of anti-VEGF drugs, such as bevacizumab, aflibercept, and ranibizumab, can control proliferative diabetic retinopathy and sight-threatening maculopathy. Studies have shown benefit over laser phototherapy for this type of maculopathy, but side-effects include infection, glaucoma, cataract, and retinal detachment. The corticosteroid triamcinolone reduces macular oedema in the short term and may be useful while other treatments exert a more permanent effect.

Laser photocoagulation

Laser photocoagulation therapy is used to treat the new vessels of proliferative retinopathy. New vessels on the disc carry the worst prognosis and warrant urgent laser therapy. The laser should be directed at the new vessels and at the associated areas of capillary non-perfusion (ischaemia). If the proliferative retinopathy has progressed to development of new vessels on the optic disc, then a technique known as panretinal photocoagulation is carried out. Rubeosis is also treated with panretinal photocoagulation. This involves multiple laser burns to the peripheral retina, especially in the areas of capillary non-perfusion. The principle is that by destroying the peripheral retina, the ischaemic stimulus for new vessel formation is diminished. Patients with maculopathy may also be treated with localized laser photocoagulation.

Although laser treatment prevents blindness, the main adverse effects result from the destruction of retinal tissue. The visual field becomes permanently smaller, and there is reduced dark adaptation. In essence, peripheral vision is sacrificed to maintain central vision.

Vitreoretinal surgery

Vitreoretinal surgery is used if bleeding is recurrent and preventing laser therapy. It is also used to try to salvage some vision if an intravitreal haemorrhage fails to clear and to treat fibrotic traction retinal detachment in advanced retinopathy.

Other ways in which diabetes can affect the eye

Cataract

Cataracts are caused by the denaturation of the protein and other components of the lens of the eye, which renders it opaque (see Chapter 31). Cataracts develop earlier in people with diabetes than in the general population. Sustained marked hyperglycaemia with a degree of ketosis can cause an acute cataract (snowflake cataract) to develop rapidly.

Extraction and intraocular lens implantation are indicated if the cataract is causing visual disability or an inability to view the retina adequately. Cataract extraction is straightforward if there is no retinopathy present. Pre-existing retinopathy may worsen after cataract extraction.

Refractory defects

The lens can also be affected by fluctuations in blood glucose concentration, which cause refractive variability, as a result of osmotic changes within the lens (the absorption of water into the lens causes temporary hypermetropia). This presents as fluctuating difficulty in reading, but people should be reassured that this resolves with better metabolic control of the diabetes.

External ocular palsies

These most commonly affect the sixth and the third nerves and are described in the section on neuropathies (see Chapter 28). These nerve palsies usually recover spontaneously within a period of 3 to 6 months.

Glaucoma

The prevalence of open-angle glaucoma is increased in people with diabetes.

Blindness

If sight loss occurs, individuals should be advised to register as blind and may require additional aids to help them to manage their diabetes. Less severe visual impairment is a bar to driving.

Eye screening

In the UK, the National Screening Committee has established universal digital photography–based screening across the country based on a national set of standards. As a result of the early detection of retinopathy, diabetes is no longer the most common cause of blindness in those younger than 65 years of age in the UK. All people with diabetes older than the age of 12 years are offered annual measurement of their acuity and retinal photographs. There are ongoing discussions about whether the screening interval might be extended in those with no retinopathy while more frequent screening is required in some groups (e.g., pregnant women). **Box 35.17** shows standardized criteria for screening schemes; these are regularly inspected.

Diabetic nephropathy

Classical diabetic nephropathy is characterized by gradually increasing urinary albumin excretion and blood pressure as the GFR falls insidiously towards end-stage renal disease. It is slowly progressive, and diabetic nephropathy usually manifests 15 to 25 years after the diagnosis of diabetes but affects 25% to 35% of people diagnosed before the age of 30 years. It is the leading cause of

Box 35.17 Criteria for a successful local screening scheme for sight-threatening diabetic retinopathy

- Clearly defined geographical area for the screening programme
- Adequate number of people with diabetes for viability (>12 000)
- Identified screening programme manager
- Identified clinical screening lead
- Identified hospital eye service for diagnosis and laser treatment
- Computer software capable of supporting call and recall of patients and image grading
- Centralized appointment administration
- Single collated list of all people with diabetes in the area older than the age of 12 years
- Equipment to obtain adequate disc- and macula-centred images of each eye
- Single image grading centre
- Process to manage people with poor-quality images
- Clear route of referral for treatment and for feedback from the treatment centre to the screening unit
- Accreditation of screening staff
- Annual reporting of service performance

premature death in young people with diabetes. A small number of people with type 1 diabetes and about 30% of those with type 2 diabetes who develop CKD have a non-classical natural history with progressive fall in GFR with no or minimal albuminuria.

The incidence of end-stage kidney disease has fallen in recent decades among people with type 1 diabetes, probably because of better management of blood glucose and blood pressure, but has increased in people with type 2 diabetes. Most people commencing renal replacement therapy have type 2 diabetes. The proportion of individuals commencing kidney replacement therapy because of diabetes varies across the world, ranging from 13% to 66%. Diabetic nephropathy increases the risk of cardiovascular risk, and most people with nephropathy die from cardiovascular disease before progressing to end-stage renal disease.

Pathophysiology

The earliest functional abnormality in the diabetic kidney is renal hypertrophy associated with a raised GFR ('glomerular hyperfiltration'). This appears soon after diagnosis and is related to hyperglycaemia. As the kidney becomes damaged by diabetes, the afferent arteriole (leading to the glomerulus) becomes vasodilated to a greater extent than the efferent glomerular arteriole. This increases the intraglomerular filtration pressure, further damaging the glomerular capillaries. Raised intraglomerular pressure also leads to increased local shearing forces, which likely contribute to mesangial cell hypertrophy and increased secretion of extracellular mesangial matrix material. This process eventually leads to glomerular sclerosis. The initial structural lesion in the glomerulus is thickening of the basement membrane. Associated changes result in disruption of the protein cross-linkages that normally make the membrane an effective filter. In consequence, there is a progressive leak of large molecules (particularly protein) into the urine.

Light-microscopic changes of both diffuse and nodular glomerulosclerosis become manifest; the latter is known as the Kimmelstiel–Wilson lesion. At the later stage of glomerulosclerosis, the glomerulus is replaced by hyaline material.

Albuminuria

Gradually increasing urinary albumin excretion is the hallmark of classical diabetic nephropathy. The earliest evidence of this is 'microalbuminuria', so called because the amounts of urinary albumin are so small as to be undetectable by standard dipsticks and only demonstrable by measurement of uACR. Microalbuminuria may, after some years, progress to persistent proteinuria (Fig. 35.15).

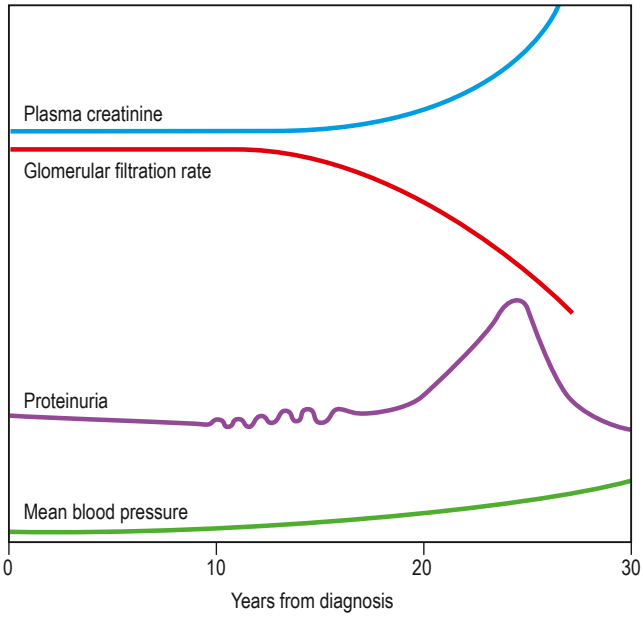


Fig. 35.15 Schematic representation of the natural history of nephropathy. The typical onset is 15 years after diagnosis. Intermittent proteinuria leads to persistent proteinuria. In time, the plasma creatinine rises as the glomerular filtration rate falls.

At the stage of persistent proteinuria, the plasma creatinine is normal, but when this stage is reached, end-stage kidney disease ensues within 5 to 10 years, although the rate of progression varies widely between individuals. Hypertension is a common development and may itself damage the kidney still further. Rising plasma creatinine is a late feature.

The proteinuria may become so heavy as to induce a transient nephrotic syndrome, with peripheral oedema and hypoalbuminaemia.

Hypertension may also contribute to the progression of CKD in people with diabetes, and they are also more prone to urinary tract infections. Autonomic neuropathy may result in bladder dysfunction and urinary retention.

Screening for diabetic nephropathy

The urine of all people with diabetes should be checked regularly (at least annually) for the presence of microalbuminuria. The uACR, tested on a midstream first morning urine sample, should be less than 2.5 mg/mmol in healthy men and less than 3.5 mg/mmol in healthy women. If microalbuminuria is detected, the test should be repeated because false-positive readings are common.

Investigation of microalbuminuria and proteinuria

When proteinuria is present, other possible causes should be considered, but after these are excluded, a presumptive diagnosis of diabetic nephropathy can be made. Clinical suspicion of a non-diabetic cause of kidney disease may be provoked by an atypical history, the absence of diabetic retinopathy (usually, but not invariably, present alongside diabetic nephropathy) and the presence of microscopic haematuria. Renal biopsy should be considered in

such cases but is rarely necessary or helpful. Renal function should be measured regularly.

Management of nephropathy

The management of diabetic nephropathy is similar to that of other causes of CKD, with a number of provisos. Although diabetic nephropathy is progressive, the time course can be markedly slowed by aggressive early therapy. Mainstays of treatment include:

- **Use of angiotensin-converting (ACE) inhibitors or angiotensin receptor blockers**, which reduce proteinuria through a reduction of intraglomerular pressure. They also contribute to blood pressure control but should be prescribed in people with diabetic nephropathy and persistent microalbuminuria even if the blood pressure is normal.
- **Use of SGLT2 inhibitors**, which slow the progression of diabetic nephropathy, even in those with low eGFR in whom the glucose-lowering effect is diminished. The explanation for these benefits is unclear but may involve effects on intraglomerular pressure, hypoxia, blood pressure, obesity, and other metabolic effects. There have been no studies of SGLT2 in diabetic nephropathy in people with type 1 diabetes, but the benefit is likely to be as in type 2 diabetes.
- **Use of non-steroidal mineralocorticoid receptor antagonists** such as finerenone, which are also effective in reducing albumin excretion and have been shown to reduce end-stage kidney disease, renal death, and cardiovascular events.
- **Tight blood pressure control**, with a target of less than 130/80 mmHg.
- **Meticulous glycaemic management.**

The development of nephropathy also has implications for glycaemic management. Metformin should be stopped when the eGFR drops below 30 mL/min. Because insulin clearance is reduced in advanced renal disease, the insulin dosage is usually reduced.

Management of end-stage renal disease

Managing end-stage renal disease is made more difficult by the fact that patients often have other complications of diabetes, such as blindness, autonomic neuropathy, or peripheral vascular disease. The failure rate of renal transplants is higher than in people without diabetes.

Diabetic neuropathy

Diabetes can damage peripheral nervous tissue in several ways. The vascular hypothesis postulates occlusion of the vasa nervorum as the prime cause. This seems likely in isolated mononeuropathies, but the diffuse symmetrical nature of the common forms of neuropathy implies a metabolic cause. Because hyperglycaemia leads to increased formation of sorbitol and fructose in Schwann cells, accumulation of these sugars may disrupt function and structure.

The earliest functional change in nerves of people with diabetes is delayed nerve conduction velocity; the earliest histologic change is segmental demyelination caused by damage to Schwann cells. In the early stages, axons are preserved, implying prospects of recovery; at a later stage, irreversible axonal degeneration develops. Unlike retinopathy or nephropathy in which there are good tools to assess early complications, the field of neuropathy is hampered by the lack of such investigative methods. However, recently, corneal confocal microscopy has been developed as a rapid, non-invasive, objective, and reproducible ophthalmic imaging technique that evaluates corneal nerve morphology and can be used as a surrogate marker for peripheral neuropathies.

Diabetic neuropathy can manifest in a number of different ways (Fig. 35.16).

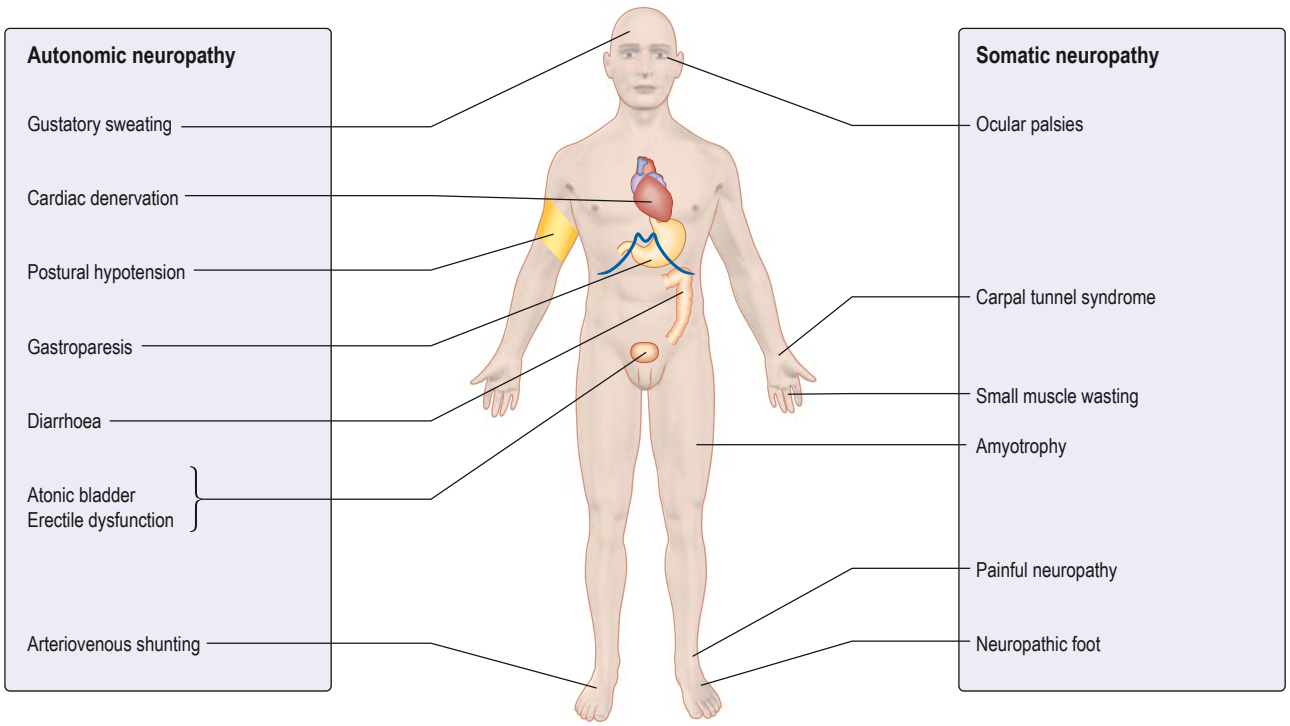


Fig. 35.16 **The effects of neuropathy in people with diabetes.**

Symmetrical distal polyneuropathy

Peripheral neuropathy is often unrecognized by the person with diabetes in its early stages. Early clinical signs are mainly sensory and include loss of vibration sense, pain sensation (deep before superficial), and temperature sensation in the feet. At later stages, people may complain of a feeling of 'walking on cotton wool' and can lose their balance when washing their faces or walking in the dark owing to impaired proprioception. Early involvement of the hands is less common and should prompt a search for non-diabetic causes. Complications include unrecognized trauma because of the loss of pain sensation.

Involvement of motor nerves to the small muscles of the feet gives rise to interosseous wasting. Unbalanced traction by the long flexor muscles leads to a characteristic shape of the foot, with a high arch and clawing of the toes, which in turn causes abnormal distribution of pressure on walking, resulting in callus formation under the first metatarsal head or on the tips of the toes. The hands show small-muscle wasting, as well as sensory changes, but these signs and symptoms must be differentiated from those of carpal tunnel syndrome, which occurs with increased frequency in diabetes and may be amenable to treatment (see Chapter 28).

Acute painful neuropathy

A diffuse, painful neuropathy is less common. The individual describes burning or crawling pains in the feet, shins, and anterior thighs. These symptoms are typically worse at night, and pressure from bedclothes may be intolerable. The condition may present at diagnosis or after sudden improvement in glycaemic levels (e.g., when insulin is started). It usually remits spontaneously after 3 to 12 months if near-normal glucose levels are maintained. A more chronic form, developing later in the course of the disease, is sometimes resistant to almost all forms of therapy. Neurological assessment is difficult because of the hyperaesthesia experienced by the individual, but muscle wasting is not a feature and objective signs can be minimal.

It is important to explore non-diabetic causes (see Chapter 28). However, if no other cause is found, explanation and reassurance about the likelihood of remission within months may be all that is needed. Duloxetine, venlafaxine, pregabalin, oxcarbazepine, tricyclic antidepressants, atypical opioids, and botulinum toxin all reduce the perception of neuropathic pain but usually not as much as patients hope for. The limited benefit of any one agent alone reflects the complex aetiology of neuropathic pain. Transepidermal nerve stimulation benefits some people. Topical capsaicin-containing creams or local anaesthetics help occasionally. A few report that acupuncture has been of value. There are several emerging treatments in clinical development.

Mononeuritis and mononeuritis multiplex (multiple mononeuropathy)

Any nerve in the body can be involved in diabetic mononeuritis; the onset is typically abrupt and sometimes painful. Radiculopathy (i.e., involvement of a spinal root) may also occur.

Isolated palsies of nerves to the external eye muscles, especially the third and sixth nerves, are more common in people with diabetes. A characteristic feature of diabetic third nerve lesions is that they are painless, and the pupillary reflexes are retained owing to sparing of pupillomotor fibres. Full spontaneous recovery is the rule for most episodes of mononeuritis over 3 to 6 months. Lesions are more likely to occur at common sites for external pressure palsies or nerve entrapment (e.g., the median nerve in the carpal tunnel).

Diabetic amyotrophy

This condition is usually seen in older men with diabetes. Presentation is with painful wasting, usually asymmetrical, of the quadriceps muscles. The wasting may be very marked, and knee reflexes are diminished or absent. The affected area is often extremely tender. Extensor plantar responses sometimes develop, and cerebrospinal fluid protein content is elevated. Diabetic amyotrophy is usually associated with periods of more severe hyperglycaemia and may be present at diagnosis. It often resolves over time with improved glycaemic levels.

Autonomic neuropathy

Asymptomatic autonomic disturbances can be demonstrated on testing in many people with diabetes (**Box 35.18**), but symptomatic autonomic neuropathy is rare. It affects both the sympathetic and parasympathetic nervous systems and can cause disabling postural hypotension.

Cardiovascular system

Vagal neuropathy results in tachycardia at rest and loss of sinus arrhythmia. At a later stage, the heart may become denervated (resembling a transplanted heart). Cardiovascular reflexes, such as the Valsalva manoeuvre, are impaired. Postural hypotension occurs owing to loss of sympathetic tone to peripheral arterioles. A warm foot with a bounding pulse is often seen in a polyneuropathy as a result of peripheral vasodilation.

Gastrointestinal tract

Vagal damage can lead to gastroparesis, often asymptomatic but sometimes leading to intractable vomiting. Implantable devices that stimulate gastric emptying and injections of botulinum toxin into the pylorus (to paralyse the sphincter partly) have each shown benefit in cases of this previously intractable problem.

Other symptoms resulting from altered gastrointestinal motility include dysphagia, dyspepsia, abdominal pain, constipation, diarrhoea, and faecal incontinence. Autonomic diarrhoea characteristically occurs at night, accompanied by urgency and incontinence.

Small intestinal dysmotility predisposes to bacterial overgrowth, which can cause bile salt deconjugation, fat malabsorption, and diarrhoea. Bacterial overgrowth may also lead to features such as anaemia and macrocytosis; treatment is with antibiotics such as tetracycline.

Bladder involvement

Loss of tone, incomplete emptying, and stasis (predisposing to infection) can occur and may ultimately result in an atonic, painless,

Box 35.18 Bedside testing of autonomic function

Test	Normal	Abnormal
Supine to erect BP		
Systolic BP fall (mmHg)	10	≥30
Heart rate responses to:		
Deep breathing (6 breaths over 1 min) maximum to minimum. HR	≥15	≤10
Valsalva manoeuvre (15 s): ratio of longest to shortest R–R interval	≥1.21	≤1.20
Lying-to-standing ratio of R–R interval of 30th to 15th beats	≥1.04	≤1.00

BP, Blood pressure; *HR*, heart rate; *R–R*, time between R wave and next R wave on electrocardiography.

Fig. 35.17 How diabetes can affect the foot. **(A)** High arch and clawing of the toes. **(B)** Typical neuropathic plantar ulceration. **(C)** Vascular pattern of ulceration.

distended bladder. Treatment is with intermittent self-catheterization, permanent catheterization if that fails, and prophylactic antibiotic therapy for those prone to recurrent infections.

Sexual dysfunction

Sexual dysfunction is common in both men and women with diabetes. The overall prevalence in men is 35% to 50% but increases with age. Approximately 60% of men with diabetes aged 60 years or older are affected. The first manifestation is incomplete erection, which may, in time, progress to total failure; retrograde ejaculation also occurs in men with autonomic neuropathy.

The mechanisms involved in erectile dysfunction in diabetes are multifactorial and include inadequate vascular supply owing to atheroma in pudendal arteries, autonomic neuropathy, primary or secondary gonadal failure, hypothyroidism, anxiety, depression, alcohol excess, and drugs (e.g., thiazides and beta-blockers).

When a man presents with erectile dysfunction, a detailed history and examination is needed to search for these causes. Blood is taken for luteinizing hormone, follicle-stimulating hormone, testosterone, prolactin, and thyroid function.

Treatment should ideally include sympathetic counselling of both partners. Phosphodiesterase type 5 inhibitors (sildenafil, tadalafil, vardenafil, avanafil), which enhance the effects of nitric oxide on smooth muscle and increase penile blood flow, are the first-line treatment in those who do not take nitrates for angina. Although less effective in people with diabetes, approximately 60% of men benefit from this therapy.

Alternative treatments include:

- Apomorphine 2 mg or 3 mg sublingually 20 minutes before sexual activity
- Alprostadil (prostaglandin E1 preparation), given as a small pellet inserted with a device into the urethra (125 μg initially to a maximum of 1 mg). If the partner is pregnant, barrier contraception must be used to prevent fetal exposure to the prostaglandin
- Intracavernosal injection of alprostadil (2.5 μg initially to a maximum of 60 μg). Side-effects include priapism, which needs urgent treatment if erection lasts longer than 3 hours
- Vacuum devices
- Penile prostheses.

Women with diabetes may complain of vaginal dryness and impaired sexual arousal.

Diabetes-related foot problems

Up to 50% of older people with type 2 diabetes have risk factors for foot problems, and 10% to 15% of people with diabetes develop foot ulcers at some stage in their lives. Diabetes-related foot problems (Fig. 35.17) are responsible for nearly 50% of all diabetes-related hospital admissions. Diabetes is the most common cause of non-traumatic lower limb amputation, but many amputations could be delayed or prevented by more effective self-management education and medical supervision.

Ischaemia, resulting from peripheral vascular disease, neuropathy, and infection, produces tissue necrosis and ulceration. Peripheral vascular disease tends to affect more distal vessels and occurs at a younger age in people with diabetes. Ischaemia compromises the ability to heal after minor trauma or infection.

Pain is a protective mechanism, and the diminished sensation that results from peripheral neuropathy means that the individual is less able to perceive trauma and may continue to walk on a wounded foot, thereby worsening the injury. Autonomic neuropathy reduces sweating and alters blood flow, resulting in dry skin that is prone to crack and fissure.

Although ischaemia and neuropathy may coexist, certain features allow an ischaemic to be distinguished from a predominantly neuropathic aetiology (Box 35.19).

Charcot neuroarthropathy

Charcot arthropathy is a complication of severe neuropathy. It occurs in a well-perfused foot and can be divided into three phases:

- Acute onset
- Bony destruction
- Radiological consolidation and stabilization.

Individuals present with an acutely swollen, hot foot, and about one-third have pain. At this stage, it may be difficult to differentiate between Charcot arthropathy and cellulitis; if in doubt, both conditions should be treated. Acute gout and deep vein thrombosis may also masquerade as Charcot arthropathy. If treatment is delayed, the foot can become deformed as bone is destroyed, often very rapidly over a few weeks. After 6 to 12 months, the destructive process stabilizes. Rehabilitation is always necessary after a long period in a cast, and reconstructive surgery may be needed.

The aim of treatment is to prevent or minimize bony destruction and deformity. Any resulting deformity can alter the pressure distribution across the foot and predisposes the foot to future ulceration. Immobilization in a non–weight-bearing cast is the treatment of choice and should be continued until the swelling and temperature in the foot has resolved.

Management of diabetes-related foot problems

The key to managing diabetes-related foot problems is prevention. Learning about the principles of foot care is essential (Box 35.20). Healthcare professionals should screen for foot problems at least once a year; this process should include questions about past or

Box 35.19 Distinguishing features between ischaemia and neuropathy

Feature	Ischaemia	Neuropathy
Symptoms	Claudication	Usually painless
	Rest pain	Sometimes painful neuropathy
Inspection	Dependent rubor	High arch
	Trophic changes	Clawing of toes
		No trophic changes
Palpation	Cold	Warm
	Pulseless	Bounding pulses
Ulceration	Painful	Painless
	Heels and toes	Plantar

Box 35.20 Principles of foot care for people with diabetes

- Inspect the feet daily.
- Moisturize dry skin.
- Seek early advice for any damage.
- Take care if walking barefoot.
- Check shoes inside and out for sharp bodies or areas before wearing them.
- Ensure shoes fit well with plenty of room for the toes.
- Keep feet away from sources of heat (hot sand, hot-water bottles, radiators, fires).
- Check the bath temperature before stepping in.
- Do not treat corns or callosities without professional help.
- Attend a podiatrist regularly:
  - Older people with diabetes may require help to cut their toenails.

present ulceration, an examination of the foot to detect structural abnormalities or callus formation, and an assessment of neuropathy and peripheral vascular disease. Although a clinical examination may reveal a distal loss of sensation, the use of monofilament, applied perpendicular to the foot and that buckles at a force of 10 g, has allowed reproducible assessments. The ability to feel this level of pressure provides protective sensation against foot ulceration. Both the VibraTip 128-Hz (a device which tests vibration perception) and the Ipswich Touch Test (a quick test involving lightly touching the first, third, and fifth toes of both feet) are reliable and sensitive tests for identifying those at risk of foot ulceration. Doppler ultrasound is a useful addition to a clinical examination of the lower limb pulses. If neuropathy or impaired blood supply is detected, the individual should be warned that their foot is at increased risk of ulceration.

Foot ulceration

The key to healing an ulcer is to ensure that the foot is non–weight bearing. Resting the affected leg may need to be supplemented with special shoes and insoles to move pressure away from critical sites or by removable or non-removable leg casts. After healing, special shoes and insoles are likely to be needed in the long term to protect the feet and prevent abnormal pressure repeating damage to a healed area. In neuropathic feet particularly, sharp surgical debridement by a podiatrist is necessary to prevent callus distorting the local wound architecture and causing damage through abnormal pressure on normal skin nearby.

Ischaemia

Clinical signs of ischaemia may be confirmed by femoral angiography, which is used to localize areas of occlusion that may be amenable to bypass surgery or angioplasty.

Infection

Foot infections can take hold rapidly and should be treated as medical emergencies. Some people have termed this a 'foot attack' akin to heart attacks to emphasize the seriousness of this condition. Clinical signs may include purulent discharge, erythema, local warmth, and swelling.

Early broad-spectrum antibiotics are essential, with therapy adjusted in the light of culture results. The organism grown from the skin surface may not be the organism causing deeper infection. Collections of pus are drained and excision of infected bone may be needed if osteomyelitis develops and does not respond to appropriate antibiotic therapy. Regular X-rays of the foot are needed to check on progress.

Wound environment

Dressings are used to absorb or remove exudate, maintain moisture, and protect the wound from contaminating agents; they should be easily removable. Various newer approaches to aid healing of diabetes-related foot ulcers include new dressings containing growth factors and other biologically active agents, hyperbaric oxygen, negative-pressure wound therapy, and bioengineered skin substitutes.

Multidisciplinary diabetes foot team

A multidisciplinary diabetes foot team, involving diabetologists working with podiatrists, orthopaedic and vascular surgeons, specialist nurses, orthotists, and other healthcare professionals, improves the outcome for people with diabetes-related foot ulcers. Rapid access to services is needed and people with diabetes should know how to contact the foot team if they develop problems.

Macrovascular complications

Atherosclerotic vascular disease

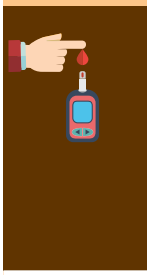
Diabetes is a risk factor for the development of atherosclerosis. Myocardial infarction, stroke, peripheral vascular disease, and cardiovascular death are increased two- or threefold compared with the background population. This risk is increased up to 10-fold in premenopausal women, who lose their normal protection against cardiovascular disease.

Pathogenesis

Atherosclerotic vascular disease in people with diabetes tends to occur at an earlier age, progress more rapidly, and is more distal and diffuse than in people without diabetes. Although hyperglycaemia contributes to the development of macrovascular disease, unlike microvascular disease, other more important haemodynamic and metabolic mechanisms are involved. The relationship between glycaemic levels and cardiovascular disease is less strong than microvascular disease but continues across the normal glucose range. Other more important factors include hypertension, dyslipidaemia (characterized by low high-density lipoprotein cholesterol and high triglyceride concentrations), obesity, hyperinsulinaemia, and proteinuria.

Management of atherosclerotic cardiovascular risk

Addressing cardiovascular risk involves the aggressive and systematic management of each of the predisposing factors (**Box 35.21**). As discussed earlier, early tight glycaemic management reduces the long-term risk of cardiovascular disease. In people with type 2 diabetes with established or at risk of cardiovascular disease, the use



Box 35.21 Target goals for cardiovascular risk factor management of people with diabetes

Parameter	Ideal	Reasonable but not ideal
HbA_{1c}	<7% (53 mmol/mol)	<8% (64 mmol/mol)
Blood pressure (mmHg)	<130/80	<140/85
Total cholesterol (mmol/L)	<4.0	<5.0
Low-density lipoprotein (mmol/L)	<2.0	<3.0
High-density lipoprotein (mmol/L)	>1.1 men >1.3 women	>0.8
Triglycerides (mmol/L)	<1.7	<2.0

Standards from the American Diabetes Association (2023).
HbA_{1c}, Glycated haemoglobin.

of SGLT2 inhibitors and GLP-1 receptor agonists is recommended because of their protective effect. Antihypertensive treatment produces a marked reduction in adverse cardiovascular outcomes as well as microvascular events. To achieve the target for blood pressure, most people with diabetes need combination drug therapy, usually beginning with an ACE inhibitor or angiotensin receptor blocker. Treating people with diabetes and at least one other major cardiovascular risk factor with an ACE inhibitor produces a 25% to 35% lowering of the risk of heart attack, stroke, overt nephropathy, or cardiovascular death. Angiotensin II receptor antagonists are sometimes preferred initially and are also used for those who are intolerant to ACE inhibitors.

Statins are recommended for those with diabetes over the age of 40 years or after a 10-year history of diabetes if microvascular complications are present as cardiovascular events are reduced by 20% for every 1-mmol/L reduction in low-density lipoprotein cholesterol. Other lipid-lowering agents, such as ezetimibe or PCSK9 inhibitors, are indicated if statins are not tolerated or do not bring the cholesterol to target.

As in the general population, smoking cessation is important, and help should be offered to those wishing to quit.

Low-dose aspirin reduces macrovascular risk but is associated with a morbidity and mortality from bleeding. The benefits of aspirin outweigh the bleeding risk when used for secondary prevention, but in Europe, it is not recommended for primary prevention.

The management of established cardiovascular disease by angioplasty or surgery is discussed in Chapter 17.

Heart failure

Heart failure is more common than atherosclerotic cardiovascular disease in people with diabetes, and both heart failure with preserved ejection fraction and heart failure with reduced ejection fraction occur more frequently. Around two-thirds of people with type 2 diabetes have evidence of left ventricular dysfunction 5 years after the diagnosis of diabetes. The prognosis after the development of heart failure is poor, with only 10% alive after 5 years.

Despite this, heart failure has received less attention than atherosclerotic cardiovascular disease and is frequently under-diagnosed. The prevalence of undiagnosed heart failure is higher in those with increasing age, obesity, hypertension, dyspnoea, or fatigue and in women.

The diagnosis and management of heart failure is similar to the general population, but it is important to consider the effects of diabetes medications on the heart failure. Pioglitazone and saxagliptin, a DPP-4 inhibitor, increased the risk of heart failure in clinical trials, but epidemiological studies show an association between the use of insulin and sulphonylureas with heart failure. GLP-1 receptor agonists have no effect on heart failure, but SGLT2 inhibitors substantially reduce the risk and have been used to treat heart failure in people without diabetes.

Other complications of diabetes

Metabolic dysfunction-associated steatotic liver disease and other gastrointestinal manifestations of diabetes

Metabolic dysfunction-associated steatotic liver disease (MASLD), formerly called non-alcoholic fatty liver disease, is characterized by intrahepatic fat accumulation with varying degrees of hepatic inflammation and fibrosis. A small proportion of people progress to cirrhosis. MASLD is present in up to one-fifth of the general population, but the prevalence is increased in people with type 2 diabetes. More than one-third of people with MASLD have diabetes, and MASLD is an independent predictor of the development of diabetes and cardiovascular disease. It is associated with central obesity and insulin resistance (see Chapter 37). Management is through weight loss and exercise. Pioglitazone may also reduce intrahepatic fat.

Malabsorption secondary to gastrointestinal dysmotility is described earlier, but people with diabetes may also develop malabsorption through chronic exocrine pancreatic insufficiency. This is unusual because only 10% of pancreatic function is sufficient for normal digestion. Treatment is with pancreatic enzymes. People with type 1 diabetes have a higher prevalence of coeliac disease.

Infections

Although people with well-managed diabetes are not more prone to infection, hyperglycaemia leads to more frequent infection through multiple disturbances in innate immunity; by contrast, humoral immunity appears relatively unaffected. Microvascular disease may further impede the immune system and likely contributes to the severity of certain infections, including malignant otitis externa, emphysematous pyelonephritis, and necrotizing fasciitis. Some microorganisms, such as *Klebsiella* serotypes and *Burkholderia pseudomallei*, are more virulent in hyperglycaemic environments.

Urinary tract infections and asymptomatic bacteriuria are more common in people with diabetes, with autonomic neuropathy a common and important underlying factor. Mycotic genital infections are more common, particularly in women. Skin and soft tissue infections are more common. The risk of tuberculosis is increased in people with diabetes. Certain viral infections, such as hepatitis C, and treatment for HIV/AIDS increase the risk of diabetes.

Cancer

In several high-income countries, cancer has become the leading cause of death among people with diabetes as deaths from cardiovascular diseases decline. Certain types of cancer, including hepatocellular carcinoma, endometrial cancer, and pancreatic cancer, are more common in those with type 2 diabetes. These associations appear to be mediated, at least in part, by obesity, which confers similar levels of risk in the absence of hyperglycaemia, although there is also an element of reverse causation with carcinoma of the pancreas, which can precipitate or cause diabetes.

Skin and joints

Diabetic cheiroarthropathy refers to joint contractures in the hands as a result of longstanding diabetes. This may be diagnosed by asking the individual to join the hands as if in prayer; the metacarpophalangeal and interphalangeal joints cannot be opposed. Thickened, waxy skin can be noted on the backs of the fingers. These features may be caused by glycation of collagen and are not progressive. Frozen shoulder, Dupuytren's contracture, trigger finger, and carpal tunnel syndrome all occur more commonly in those with diabetes.

Fracture risk is increased in those with type 1 diabetes and type 2 diabetes. Although bone mineral density is decreased in type 1 diabetes, it is increased in type 2 diabetes, so other mechanisms are at play. Thiazolidinediones decrease bone formation and bone mineral density and increase fracture risk in type 2 diabetes. After a fracture, healing may be delayed in people with diabetes. The risk of gout is also increased in people with type 2 diabetes.

Further reading

Antonetti DA, Klein R, Gardner TW. Diabetic retinopathy. *N Engl J Med* 2012; 366:1227–1239.
de Boer I, Khunti K, Sadusky T et al. Diabetes management in chronic kidney disease: a consensus report by the American Diabetes Association (ADA) and Kidney Disease: Improving Global Outcomes (KDIGO). *Diabetes Care* 2022; 45:3075–3090.
Diabetes UK. *How to Spot a Foot Attack*; https://www.diabetes.org.uk/get_involved/volunteer/involve-newsletter/new-booklet-for-people-at-increased-risk-of-a-foot-attack.
International Working Group on the Diabetic Foot. *International Consensus and Practical Guidelines on the Management and the Prevention of the Diabetic Foot*. https://iwgdfguidelines.org/prevention-guideline-2023/.
Martin DF, Maguire MG. Treatment choice for diabetic macular edema. *N Engl J Med* 2015; 372:1260–1261.
National Institute for Health and Care Excellence. *NICE Guideline 19: Diabetic Foot Problems: Prevention and Management*. NICE 2019; https://www.nice.org.uk/guidance/ng19.
White NH, Sun W, Cleary PA. Prolonged effect of intensive therapy on the risk of retinopathy complications in patients with type 1 diabetes mellitus: 10 years after the diabetes control and complications trial. *Arch Ophthalmol* 2008; 126:1707–1715.

Psychosocial implications of diabetes

The interaction between mind and body is central to the management and outcome of all chronic diseases but is particularly fascinating in the case of diabetes. People with diabetes have an increased risk of developing a number of psychological conditions, and a range of psychiatric disorders are associated with an increased risk of diabetes (**Box 35.22**). People with chronic stress, low stress resilience, sleeping problems, depression, or severe mental illness are at increased risk of developing type 2 diabetes. Diabetes affects the quality of many aspects of life, including emotional wellbeing, financial situation, leisure and work activities, and relationships with friends and families.

Psychological problems associated with diagnosis

Most people show remarkable psychological resilience to a diagnosis of diabetes despite the implications for self-management and long-term health. In some people, the diagnosis may lead to symptoms of anxiety and depression, but others develop an 'adjustment disorder'. Transient psychological symptoms begin within 3 months of the diagnosis of diabetes but usually resolve within 6 months. These symptoms are seen as part of the normal 'grieving process' that often accompanies the development of any chronic illness.

Diabetes-related distress

Living with diabetes is demanding; it is estimated that people with type 1 diabetes need to devote up to 2 hours a day to self-management, but around half of people with diabetes report that taking medication interferes with their ability to live a normal life. Diabetes self-management is often frustrating when blood glucose levels remain stubbornly out of target despite the individual's best efforts. More than 50% of people develop psychological symptoms, which are collectively known as diabetes distress, in response to living with the constant burden of diabetes. These include:

- Experiencing worries about the future and the possibility of serious complications
- Feeling guilty or anxious when their diabetes management goes awry
- Feeling scared or depressed when thinking about living with diabetes
- Feeling discouraged with their diabetes regimen.

Fear of hypoglycaemia

Episodes of hypoglycaemia are frequently unpredictable and are associated with unpleasant, embarrassing symptoms associated with activation of the nervous system. It is perhaps unsurprising that many people worry about future occurrences of hypoglycaemia, and some people develop significant fear of hypoglycaemia. This fear may also be experienced by other family members. Significant fear of hypoglycaemia is more common in people who have experienced loss of consciousness during a previous episode.

Depression

The prevalence of depression is increased twofold in people with diabetes. Significant depressive symptoms affect approximately one in four adults with diabetes, and a formal diagnosis of depressive disorders is made in 10% to 15% of people with diabetes. Depression is more common in women, those who are divorced, and those who have experienced childhood adversity and social deprivation. In addition, a number of diabetes-specific risk factors are associated with depression, including the development of complications and after initiation of insulin in type 2 diabetes.

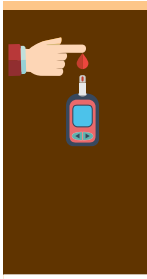
Box 35.22 Bidirectional relationship between diabetes and psychiatric and psychological conditions

Psychiatric and psychological conditions that occur more frequently in people with diabetes

- Adjustment disorders
- Diabetes-related distress
- Fear of hypoglycaemia
- Stigma and discrimination
- Anxiety
- Depression
- Phobia
- Eating disorders

Psychiatric and psychological conditions that increase the risk of diabetes

- Chronic stress
- Low resilience
- Sleeping problems
- Anxiety
- Depression
- Bipolar disorder
- Schizophrenia
- Other psychotic conditions treated with antipsychotics



Disordered eating

Eating disorders and disordered eating in diabetes encompass a broad spectrum of conditions ranging from well-defined eating disorders such as anorexia or bulimia nervosa to harmful diabetes self-care behaviours driven by fear of weight gain, with direct and indirect insulin restriction being the most common one.

Consequences of psychological and psychiatric disorders of diabetes

As well as affecting quality of life, the psychological sequelae of diabetes often impair the individual's ability to manage their diabetes. They are less likely to follow a healthy diet and exercise and less likely to take medication as prescribed, and it is therefore no wonder that psychological problems are associated with higher HbA_{1c}, more frequent admissions with hypoglycaemia and DKA, a higher incidence of complications, and shortened life expectancy.

Management of psychological issues

The key step to managing the psychological sequelae of diabetes is identifying the problem. It is important to ask the person with diabetes how they feel about their diabetes and whether they are experiencing difficulties with it. The diabetes specialist team will undertake much of the psychological management, but this should include liaison with the mental health teams to ensure appropriate referral of those with more complex problems. Some diabetes teams now include a psychologist.

A range of psychological interventions, including cognitive behavioural therapy and problem solving, has been used successfully to improve the psychological wellbeing of people with diabetes.

For people with depression, antidepressants are effective as long as they are used in adequate doses.

Social aspects of diabetes

Diabetes may affect the safety of driving for people who have hypoglycaemia, visual impairment (from either retinopathy or cataract), or disability from leg amputation or neuropathy. Because hypoglycaemia is the most common cause of accidents in people with insulin-treated diabetes, people with diabetes must take precautions to avoid hypoglycaemia while driving. In most countries, a driver with diabetes will need to inform the relevant authorities and insurance companies of their diagnosis.

Diabetes should not prevent access to most occupations, but certain occupations, including the armed forces, emergency services, commercial pilots, prison and security services, and jobs in potentially dangerous areas (e.g., at heights, underwater, and offshore) are barred or require special precautions. Shift work can make insulin management more complex, but modern long-acting insulin analogues with durations of action longer than 24 hours have made life easier.

Further reading

Fisher L, Polonsky WH, Hessler D. Addressing diabetes distress in clinical care: a practical guide. *Diabet Med* 2019; 36.

Hanlan ME, Griffith J, Patel N et al. Eating disorders and disordered eating in type 1 diabetes: prevalence, Screening, and treatment options. *Curr Diab Rep* 2013; 10.1007/s11892-013-0418-4.

Holt RI, de Groot M, Golden SH. Diabetes and depression. *Curr Diab Rep* 2014; 14:491.

Holt RI, Nicolucci A, Kovacs Burns K et al. Diabetes Attitudes, Wishes and Needs second study (DAWN2™): cross-national comparisons on barriers and resources for optimal care—healthcare professional perspective. *Diabet Med* 2013; 30:789–798.

Kovacs Burns K, Nicolucci A, Holt RI et al. Diabetes Attitudes, Wishes and Needs second study (DAWN2™): cross-national benchmarking indicators for family members living with people with diabetes. *Diabet Med* 2013; 30:778–788.

DIABETES IN SPECIAL SITUATIONS

Diabetes in early adulthood

More than 1 million children have diabetes worldwide with approximately 132 600 new cases diagnosed annually. The incidence has been steadily rising over the past 50 years. Most children with diabetes have type 1 diabetes, although type 2 diabetes is becoming more common in older children. As the child grows up, the responsibility for managing the diabetes will shift from parent to child. The period of transition between paediatric and adult services is challenging, and many young adults with diabetes lose contact with healthcare professionals if this process is not managed effectively (see Chapter 46). The physical transfer between paediatric and adult services is only a small component of transition, and the process can take a number of years as the paediatric team, working with the adult team, provides the young adult with the information and skills needs to negotiate an adult health service independently. Although many young adults move to adult services around the age of 18 to 19 years, the decision to transfer should be based on skills and competencies rather than chronological age.

Emerging adulthood is a new concept that describes the early years of adulthood when an individual may be entering the world of work and moving away from home for the first time, including going to university. As well as providing many new opportunities, it is often a time for risk-taking. Many centres have now developed young adult clinics to care for the specific needs of this age group.

Diabetes in older people and at the end of life

The prevalence of diabetes in older people is increasing dramatically because of an ageing population and age-related challenges in body composition and insulin resistance. The lifetime risk of developing diabetes is 22% for women and 19% for men from the age of 60 years. More than half of those with diabetes are aged older than 65 years. Approximately 25% of older people living in care homes or nursing homes have diabetes.

Diabetes places an additional burden on older people, who frequently have other chronic conditions, with up to 40% having three or more other co-morbidities. Diabetes increases the risk of dementia, and cognitive decline compromises diabetes management. Frailty, which is characterized by weight loss, weakness, decreased physical activity, exhaustion, and slow gait speed, is more common in older people.

The management of diabetes in an older individual should consider the health status and needs of the person with diabetes. A thorough assessment should identify co-morbidities. Older people with diabetes are a highly heterogeneous population ranging from those living independently in the community to frail individuals with multiple co-morbidities in care homes. Glycaemic targets should therefore be individualized by taking into consideration the person's overall health, life expectancy, cognitive decline, and co-morbidities (**Box 35.23**). Maximizing quality of life should always be the focus of care.

Box 35.23 Glycaemic targets for older people

Independent or robust	Dependent	Dependent dementia or frail	End of life
7.0%–7.5%	7.0%–8.0%	<8.5%	Avoid symptomatic hypoglycaemia
53–59 mmol/mol	53–64 mmol/mol	<70 mmol/mol	

Dunning T, Sinclair A, Colagiuri S. New IDF Guideline for managing type 2 diabetes in older people. Diabetes Res Clin Pract. *2014;103(3):538–540.*

Box 35.24 Use of insulin in older people[a]

Insulin regimen	Independent or robust	Dependent frail	Dependent dementia	End of life
Basal insulin + oral antidiabetes agents	First choice	First choice	First choice	First choice but consider withdrawal
Premix insulin	Alternative first choice			
Basal-bolus insulin	Second-line choice	Avoid	Avoid	Avoid

[a]Choose the appropriate device, usually a prefilled pen.

Dunning T, Sinclair A, Colagiuri S. New IDF Guideline for managing type 2 diabetes in older people. Diabetes Res Clin Pract. *2014;103(3):538–540.*

When possible, drugs that avoid hypoglycaemia are preferred because the associated co-morbidities, geriatric syndromes, polypharmacy, and long duration of diabetes and the increased prevalence of hepatic and renal dysfunction increase the risk of hypoglycaemia. Insulin regimens should be simplified (**Box 35.24**). Furthermore, hypoglycaemia can be less easily recognized and is often misdiagnosed because of reduced autonomic symptoms.

As people approach the end of life, the goals of treatment shift from the prevention of long-term complications to the avoidance of symptomatic hyper- and hypoglycaemia, acute metabolic decompensation, and drug-related adverse events. Glycaemic targets should be relaxed with the aim of keeping the glucose between 6 and 15 mmol/L (110–270 mg/dL) to minimize the risk of hypoglycaemia and other drug-related adverse events. Insulin should usually be continued. It is important to take care to avoid foot ulceration in bedbound individuals.

Inpatient diabetes

Having diabetes more than doubles the risk of being hospitalized for any given condition. In 2019, the UK National Diabetes Inpatient Audit showed that 18.1% of hospital inpatients had diabetes. In the majority, diabetes was not the cause for their admission, but poor management of diabetes lengthens hospital stay and increases the risk of developing complications, such as hospital-acquired infection, or dying. This applies equally to medical and surgical inpatients.

Although the stress of acute illness can worsen hyperglycaemia, hypoglycaemia is also associated with increased morbidity and mortality. Because there is a lack of robust evidence demonstrating that intensive glucose management improves outcomes, the general consensus is that the main priority is to avoid hypoglycaemia, and diabetes treatments need to be adjusted accordingly.

Because people with diabetes may be less able to self-manage their diabetes while in hospital, it is important the diagnosis of diabetes is clearly identified in the medical records and glucose and HbA_{1c} measured on admission. Errors in insulin prescribing are common, and one-third of all fatal inpatient medical errors that cause death within 48 hours of the error involve insulin administration. An emphasis on insulin safety, particularly when an i.v. insulin is being used, is essential. There is an increasing move for people with insulin-treated diabetes to self-manage when they are able on the grounds that they know how to do so much better than many hospital staff.

Many hospitals have specialized diabetes inpatient teams that take responsibility for diabetes management, but all staff caring for people with diabetes should be appropriately trained and competent in the management of inpatient diabetes (**Box 35.12**).

Intravenous insulin is required for a number of specific conditions because this mode of administration provides the greatest flexibility to meet rapidly changing insulin requirements (**Box 35.25**). Intravenous insulin is usually administered as a variable rate infusion, but the insulin rate is fixed when treating DKA. Variable-rate or fixed-rate i.v. insulin infusion is preferred to the old but more ambiguous terminology 'sliding scale'. To maintain patient safety, the glucose should be monitored frequently.

Box 35.25 Indications for intravenous insulin in hospital

Variable rates intravenous insulin infusion

- Critical care illness
- Acute coronary syndromes
- Stroke
- 'Nil by mouth' status when more than one meal will be missed (type 1 diabetes)
- General preoperative, intraoperative, and postoperative care
- Stroke
- Hyperglycaemia during high-dose corticosteroid therapy
- Labour and delivery
- Other acute illness for which prompt glycaemic control is judged important for recovery

Fixed rates intravenous insulin infusion

- Diabetes emergencies:
  - Diabetic ketoacidosis
  - Hyperosmolar hyperglycaemic syndrome

Management of diabetes during an admission for surgery

Both hyperglycaemia and hypoglycaemia in the postoperative period worsen surgical outcomes by increasing the risk of infection, myocardial infarction, and death and are associated with an increased duration of stay on the intensive care unit and in hospital.

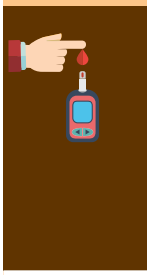
The patient journey begins in primary care and general practitioners should help optimize glucose management in the preoperative period. This reduces the likelihood of the surgery's being cancelled or postponed owing to hyperglycaemia. The preoperative assessment provides a further opportunity to improve diabetes management and allows the patient to be placed at the beginning of a morning list when possible to minimize the period of fasting and subsequent metabolic disturbance.

For people treated with insulin, unless the period of fasting is less than one missed meal, a variable-rate i.v. insulin infusion is needed to manage the glucose perioperatively. The long-acting insulin should be continued because it will prevent rebound hyperglycaemia when the i.v. insulin is stopped. Oral drugs that can cause hypoglycaemia should also be stopped for the dose immediately before surgery but can be resumed after the patient is eating and drinking normally. Metformin should be stopped if i.v. contrast is used.

Before discharge, a clear follow-up plan is needed, which should be communicated to the person with diabetes and their primary care team.

Diabetes in pregnancy

Diabetes is one of the most common medical conditions in pregnancy, with 15.8% of live births in 2019 being affected by some



form of diabetes in pregnancy. Approximately 85% of women with diabetes in pregnancy have gestational diabetes with the remainder having pre-existing diabetes. The epidemiology of pre-existing diabetes in pregnancy is changing as women are delaying pregnancy to an older age, and the prevalence of obesity and type 2 diabetes is increasing in women of childbearing age. Whereas 25 years ago, type 2 diabetes was rare in pregnant women, the latest UK national audit has indicated that there are now more pregnant women with type 2 diabetes than type 1 diabetes.

Adverse effects of diabetes in pregnancy

Before the discovery of insulin, pregnancy with diabetes was invariably fatal for both the mother and fetus. Even in the era up to the 1960s, pregnancy remained a risky enterprise, with perinatal mortality rates of up to 40%. However, with an understanding that maternal hyperglycaemia was the major cause of fetal loss, management changed to ensure as near normoglycaemia as possible, and outcomes began to improve. Nevertheless, pregnancy outcomes in women with diabetes still remain worse than for women without diabetes.

Diabetes in pregnancy is associated with risks to both the woman and to the developing fetus. Women with pre-existing diabetes may find it harder to conceive and in the first trimester have higher rates of miscarriage and congenital anomalies. The frequency of many developmental abnormalities, including cardiac and skeletal malformations, is increased. Although rare, the risk of caudal regression syndrome appears to be specifically increased in women with diabetes. Towards the end of pregnancy, women with either pre-existing diabetes or gestational diabetes have a higher risk of preterm labour, hypertension, and pre-eclampsia.

The risks to the fetus include higher rates of macrosomia and associated shoulder dystocia and birth injury, polyhydramnios, stillbirth and perinatal mortality, and impaired postnatal metabolic adaptation leading to higher rates of neonatal jaundice and respiratory distress. After birth, neonatal hypoglycaemia may occur as maternal glucose crosses the placenta but insulin does not; this leads the fetal islets to hypersecrete insulin to combat the rising hyperglycaemia, and rebound hypoglycaemia may occur when the umbilical cord is cut. In the longer term, offspring of mothers with diabetes have a higher risk of obesity and diabetes in adulthood.

Management of diabetes during pregnancy

As the adverse effects of diabetes are largely mediated by hyperglycaemia, the major aim of management in diabetic pregnancies is to maintain normoglycaemia by avoiding pre- and postprandial hyperglycaemia without causing severe hypoglycaemia. This is a demanding process, so pregnant women with diabetes should be in contact with a specialist multidisciplinary team jointly led by a physician and obstetrician at least every 2 weeks.

Preconception care

Pregnancy management should begin well before conception; because pregnancy can only be diagnosed a minimum of 2 weeks after conception, it is important that optimal glycaemic levels are achieved before conception. This requires planning on the part of both the woman with diabetes and the diabetes team. From puberty, girls with diabetes need to know about the importance of pregnancy planning and contraception until they are ready to start a family.

When a woman expresses a desire for pregnancy, referral to a specialist diabetes preconception clinic improves the pregnancy outcomes. This provides an opportunity to review diabetes management, discuss the need for high dose folic acid (5 mg daily) to reduce the risk of neural tube defects, ensure that microvascular screening is up to date, stop medications that are contraindicated in pregnancy, and check that the woman has been immunized against rubella.

The optimal target preconception HbA_{1c} is below 48 mmol/mol (6.5%) when this can be achieved safely, and women should be advised not to become pregnant if their HbA_{1c} is above 86 mmol/mol (10%). Target fasting plasma glucose levels are 5 to 7 mmol/L (90–126 mg/dL) and 4 to 7 mmol/L (70–126 mg/dL) before meals at other times of the day. This is challenging for women, and they should be offered the opportunity to attend structured education if they have not done so already. When necessary, insulin analogues should be used because they appear to convey advantages over human insulin. Metformin may be used while preparing for pregnancy, but other oral antidiabetes drugs and GLP-1 receptor agonists should be stopped. ACE inhibitors, angiotensin II receptor antagonists, and statins should be stopped before conception.

Antenatal management

The need for optimal glycaemic management continues when a female becomes pregnant; the target glucose levels are less than 5.3 mmol/L fasting (95 mg/dL), less than 7.8 mmol/L (140 mg/dL) 1 hour after meals, and less than 6.4 mmol/L (115 mg/dL) 2 hours after meals. Women need to monitor their glucose levels by capillary blood testing; when available, CGM is now preferred because it increases the time that women spend in the target glucose range and is associated with improved fetal outcomes.

These stringent glycaemic targets are made harder to achieve because pregnancy changes the insulin requirement; this tends to fall in the first trimester, partly as a result of the hormonal changes in pregnancy but also because pregnancy-associated nausea and vomiting can reduce appetite. This increases the risk and severity of hypoglycaemia; for reasons that are not entirely clear, early pregnancy is associated with impaired awareness of hypoglycaemia, further increasing the likelihood of hypoglycaemia. Women and their partners need to know about the risk and how to manage it. Interestingly, fetuses tolerate maternal hypoglycaemia relatively well.

During the second and third trimesters, pregnancy induces a state of insulin resistance, which typically increases insulin requirements by 50% to 100% in the latter half of pregnancy. Towards the end of pregnancy, insulin absorption is delayed, so women may need to increase the length of time between the short-acting insulin injection and eating in the third trimester.

Ketoacidosis in pregnancy carries a 50% fetal mortality rate, so if it occurs in pregnancy, it should be diagnosed and managed rapidly. To expedite the diagnosis, women should be taught to monitor blood ketones at the outset of pregnancy.

As pregnancy can accelerate the progression of retinopathy, pregnant women should be screened in the first and third trimesters and in the second trimester if any retinopathy is present at the outset.

Low-dose aspirin (150 mg daily) from the 12th week of pregnancy lowers the risk of hypertension and pre-eclampsia.

Enhanced fetal surveillance is needed and a detailed ultrasound scan for fetal anomalies including a four-chamber view of the fetal heart and outflow tracts should be offered at 18 to 20 weeks of gestation. Growth scans and other tests of fetal wellbeing should be performed during the third trimester.

Birth

Most babies of women with diabetes are born before term after an induction of labour or elective caesarean section to reduce the risk

of stillbirth. Glycaemic management during labour is usually maintained by a variable rate i.v. insulin infusion, but pumps are being increasingly used as an alternative in some centres.

Postnatal care

After birth, maternal insulin requirements drop rapidly to preconception levels within hours of the birth and may even be lower if a woman chooses to breast-feed. Because the neonate is at increased risk of hypoglycaemia, early feeding should be encouraged with breast-feeding being the preferred option, and the baby's blood glucose should be checked regularly until it is clear that the pre-feeding glucose levels are being maintained in the normal range.

Gestational diabetes

Gestational diabetes is defined as diabetes occurring for the first time in pregnancy that is clearly not overt diabetes before pregnancy. It is usually diagnosed in the second and third trimesters of pregnancy after a 75-g OGTT. Gestational diabetes occurs in women whose pancreatic β cells are unable to secrete sufficient insulin to meet the pregnancy-induced insulin resistance. The risk factors for gestational diabetes are the same as for type 2 diabetes, so gestational diabetes can be regarded as an early revelation of the metabolic abnormality brought on by the demands of pregnancy. Not all diabetes presenting in pregnancy is gestational. Screening may identify women with previously undiagnosed type 2 diabetes or monogenic diabetes, and in a few cases, gestational diabetes may unmask pre-clinical type 1 diabetes.

Diagnosis of gestational diabetes

There is a lack of international agreement regarding the diagnosis with varying diagnostic cut-off values and number of abnormal values needed between recommendations (**Box 35.26**). Universal screening for diabetes is recommended in many countries but in the UK, NICE recommends screening only women with risk factors for gestational diabetes; the advantage of the latter approach is that many low-risk women do not need the screening test. This costs less but it risks missing gestational diabetes in a significant proportion of women (**Box 35.27**).

Management of gestational diabetes

Glucose targets are the same as for pre-existing diabetes, but management begins with health behaviour modification, similar to the management of type 2 diabetes (see Chapter 48). If health behaviour modification does not achieve adequate maternal glycaemia within 1 to 2 weeks, pharmacological therapy is indicated. Although insulin is still widely used, studies have suggested that metformin or glibenclamide (a sulphonylurea) is an acceptable alternative.

Women with gestational diabetes have a 20%–50% chance of developing diabetes within 10 years of the pregnancy. The diabetes antenatal team should discuss this risk and the measures needed to reduce the likelihood of developing diabetes before discharge after the birth. An annual HbA_{1c} or glucose test is recommended to diagnose asymptomatic diabetes. At present, the uptake of postnatal interventions and screening is low, and more research is needed to find out how to manage this risk more effectively.

Further reading

International Association of Diabetes in Pregnancy Study Groups Consensus Panel. International Association of Diabetes in Pregnancy Study Groups recommendations on the diagnosis and classification of hyperglycemia in pregnancy. *Diabetes Care* 2010; 33:676–682.
International Diabetes Federation. *Managing Older People With Type 2 Diabetes*. Global guideline; https://www.idf.org/e-library/guidelines/78-global-guideline-for-managing-older-people-with-type-2-diabetes.html.
Joint British Diabetes Societies. *Management of Adults With Diabetes Undergoing Surgery and Elective Procedures: Improving Standards*. https://www.diabetes.org.uk/resources-s3/2017-09/Use%20of%20variable%20rate%20intravenous%20insulin%20infusion%20in%20medical%20inpatients_0.pdf.
Joint British Diabetes Societies. The Use of Variable Rate Intravenous Insulin Infusion (VRIII) in Medical Inpatients. https://www.diabetes.org.uk/resources-s3/2017-09/Use%20of%20variable%20rate%20intravenous%20insulin%20infusion%20in%20medical%20inpatients_0.pdf.

ORGANIZATION OF DIABETES CARE

The diabetes care team involves a multidisciplinary group of healthcare professionals who are available to support the person with diabetes (**Fig. 35.10**). Because the amount of time that each person with diabetes spends with a healthcare professional is small—on average, less than 6 hours a year—it is important that every contact counts and provides appropriate medical and psychosocial support. To ensure that the person with diabetes derives the maximum benefit from the time spent with their diabetes team, whether in a hospital or community setting, the consultation should be collaborative, patient-centred, and goal-focused and with clear aims. As well as traditional clinic visits, people with diabetes may access diabetes care virtually through phone, email, or Internet contact or through educational sessions.

Box 35.27 Risk factors for gestational diabetes: NICE recommendations for a 75-g oral glucose tolerance test

NICE recommends undertaking a 75-g OGTT at 24–28 weeks of gestation in females with the following risk factors:

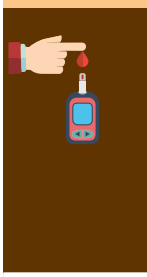
- Body mass index >30 kg/m^2
- Previous macrosomic baby weighing ≥4.5 kg
- Previous gestational diabetes[a]
- First-degree relative with diabetes
- Family origin with a high prevalence of diabetes: South Asian, Black Caribbean, or Middle Eastern

[a]Women with a previous history of gestational diabetes should be offered a 75 g 2-hour oral glucose tolerance test (OGTT) as soon as possible after booking (whether in the first or second trimester), and a further 75 g 2-hour OGTT at 24 to 28 weeks of gestation if the results of the first OGTT results are normal.

Box 35.26 Comparison of the NICE and IADPSG diagnostic criteria for gestational diabetes

Association	Glucose load (g)	No. of high readings needed	Fasting glucose	1-hour glucose	2-hour glucose
IADPSG	75	≥1	5.1 mmol/L 90 mg/dL	10.0 mmol/L 180 mg/dL	8.5 mmol/L 153 mg/dL
NICE	75	≥1	5.6 mmol/L 100 mg/dL		7.8 mmol/L 140 mg/dL

IADPSG, International Association of Diabetes and Pregnancy Study Groups; *NICE,* National Institute for Health and Care Excellence.



Box 35.28 Diabetes UK 15 essentials of diabetes care

1. Blood glucose test (HbA_{1c} test)
2. Blood pressure check
3. Cholesterol and lipid check
4. Eye screening
5. Foot and leg check
6. Kidney tests
7. Advice on diet
8. Emotional and psychological support
9. Diabetes education course
10. Care from diabetes specialists
11. Free flu jab
12. Good care if in hospital
13. Support with any sexual problems
14. Help to stop smoking
15. Specialist care while planning a pregnancy

HbA_{1c}, Glycated haemoglobin.

Much of care is focused towards minimizing the long-term complications while avoiding adverse effects from treatment. Diabetes UK has described the 15 essentials of diabetes care (**Box 35.28**) that should be assessed and discussed at least once a year.

The majority of diabetes care takes place in a primary care setting, but people with diabetes with complex medical or psychological needs require management and support in a specialist setting for some or all of their care. A close collaboration between primary and secondary healthcare professionals and among specialists is essential to ensure that care is integrated and coordinated across the wide range of disciplines involved.

People with diabetes need to be involved in decisions about their care because this improves outcomes and the likelihood that a specific management plan will be followed. It is good practice to provide the person with diabetes with copies of any letters written about them so that all decisions are clear, and it is easier to review the management plans at subsequent contacts.

Significant websites

http://www.diabetes.ca. *Canadian Diabetes Association site—well-designed, practical site with many links to other diabetes-related sites; a good starting point.*
http://www.diabetes.org. *American Diabetes Association—heavyweight and authoritative, with an American flavour.*
http://www.diabetes.org.uk. *Diabetes UK charity—information for patients, researchers and health professionals.*
http://www.dtu.ox.ac.uk. *Diabetes Trials Unit (University of Oxford)—research information, particularly the UK Prospective Diabetes Study results.*
http://www.idf.org. *International Diabetes Federation.*
http://www.sign.ac.uk. *Scottish Intercollegiate Guidelines Network—guidelines on a range of subjects including diabetes.*

Questions for this chapter can be found on Elsevier eBooks+.

36 Metabolic disease

Radha Ramachandran and Raphael Buttigieg

Background 1240
Approach to the patient 1240
History 1240
Examination 1241
Investigations 1241
Acute presentations of inherited metabolic diseases 1241
Known inherited metabolic disease 1241
Situations when an inherited metabolic disease should be suspected 1242
General principles of management of inherited metabolic diseases 1244
Specific inherited metabolic diseases 1244
Mitochondrial diseases 1244
Disorders of protein metabolism 1245
Lysosomal storage disorders 1246
Disorders of carbohydrate metabolism 1246
Disorders of fatty acid oxidation defects 1246
Porphyrias 1246
Peroxisomal diseases 1247

CORE SKILLS AND KNOWLEDGE

Inherited metabolic diseases (IMDs) are individually rare, although as a group, they are not altogether uncommon.

Many patients with IMDs are diagnosed in childhood, and transition in adulthood to the care of metabolic medicine specialists in tertiary referral centres, supported by dieticians and other healthcare professionals. However, they may also present for the first time as adults to general services or require general or emergency management in non-specialist settings (e.g., when a metabolic emergency is triggered by starvation, dehydration, or intercurrent illness). All doctors therefore must have at least a basic awareness of these conditions.

Key skills in this area include:

- Knowing when to suspect an IMD and understanding the principles of tailored therapy for rare diseases
- Recognizing the acute presentations of such conditions and knowing core steps in management of these emergencies.

Opportunities for learning in this specialty include observing clinics in specialist metabolic medicine centres and attending specialist conferences, courses, and workshops.

BACKGROUND

Inherited metabolic diseases (IMDs), previously known as inborn errors of metabolism, are a heterogeneous group of complex monogenic disorders that occur because of an impairment in the activity of an enzyme involved in one or more metabolic pathways. Hundreds of disorders exist, with the number of known disorders increasing yearly. The International Classification of Inborn Metabolic Disorders currently includes 1450 disorders. Thus, although individually rare, collectively, they are not that uncommon, with reported prevalence ranging from 1 in 800 to 1 in 2500 live births. IMDs are commoner in populations where consanguinity occurs. Specific founder mutations are found in societies derived from small original ancestral populations, where intermarriage and genetic drift further increase the prevalence of specific IMDs, especially in geographically or culturally isolated populations.

Clinical signs and symptoms occur because of accumulation of substrate precursors or deficiencies of enzyme products. The majority of IMDs present in childhood or are picked up through neonatal screening programmes. With early diagnosis and specific treatments, life expectancy is improved considerably, and patients may transition to adult metabolic services in late adolescence. Their risk of other common medical problems may be higher than that of patients in the general adult population.

Up to 40% of IMDs may be diagnosed in adults, with presentations that may differ from those seen in children. Patients may present to the emergency department with acute metabolic decompensation or to any specialist outpatient clinic with problems related either to a previously diagnosed IMD or an as-yet undiagnosed IMD.

APPROACH TO THE PATIENT

History

An IMD should be considered:

- In patients presenting with symptoms associated with particular IMDs, including encephalopathy, rhabdomyolysis or unexplained high creatine kinase (CK), acidosis or persistently high lactate of unknown aetiology, hypoglycaemia, or stroke
- When there is multisystem involvement
- When there is a history of metabolic triggers
- When there is a family history of similar symptoms, early death, or consanguinity
- When an alternative more common diagnosis is not forthcoming

It is important not to miss a diagnosis because the prognosis is often good after early institution of simple therapy or diet modulation.

A confirmed diagnosis also has implications for the extended family, who may benefit from early intervention and reproductive counselling.

Examination

Patients with IMD may have signs relating to their clinical presentation such as focal neurological symptoms, encephalopathy, retinopathy, or hearing impairment. In some of the storage diseases, hepatosplenomegaly (sometimes massive) or dysmorphia can be demonstrated.

Investigations

Patients with IMDs may require measurement of metabolites, including ammonia, lactate, glucose, organic acids, and amino acids in serum, urine, and cerebrospinal fluid. Basic and additional tests, including the role of genetic testing, are described in **Box 36.1**.

Special considerations

Patients with IMDs may have specific risks related to pregnancy, surgery, or anaesthesia. Hence, any pregnancy or surgery should be discussed with a specialist metabolic physician well in advance to enable a personalized birth or surgical treatment plan to be drawn up.

Genetic testing, including multigene panels and whole-exome and whole-genome sequencing, is becoming an almost first-line diagnostic tool and is likely to widen the spectrum of IMDs both in terms of phenotypic presentations and new diseases. Genetic testing is discussed in greater detail in Chapter 5.

ACUTE PRESENTATIONS OF INHERITED METABOLIC DISEASES

Patients with known or undiagnosed IMDs can present acutely to the emergency department as a result of metabolic decompensation precipitated by metabolic triggers, including:

- Catabolic states such as intercurrent illness or an inability to tolerate food and drink
- Labour and the postpartum period
- Non-compliance with IMD treatment
- Drugs (therapeutic or recreational), including alcohol
- Change in diet

Known inherited metabolic disease

Patients at risk of decompensation often carry with them an emergency management plan provided by their specialist metabolic centre. For many patients with inherited disorders of glucose, lipids, and protein metabolism, this may include a basic oral ***emergency regimen*** (ER) to use when unwell. This usually comprises 200 mL of 25% glucose polymer drink every 2 hours (50 g of carbohydrates or 200 kcal) with instructions to attend hospital for intravenous (i.v.) treatment and further management if they are unable to tolerate oral ER or fail to improve despite it. Intravenous ER for these conditions is usually an infusion of 10% glucose i.v. at 2 mL/kg per hour. The oral or i.v. glucose provides adequate calories and increases

Box 36.1 Basic and additional tests in known or suspected inherited metabolic diseases

Investigation	Timing of investigation and clinical significance
Basic metabolic screen	
FBC, renal, liver and bone profiles, CRP, ammonia, lactate, glucose, ketones (beta-hydroxybutyrate and free fatty acids), venous blood gases, CK, urate	At presentation Laboratories servicing emergency departments should return results for these tests within the hour, at any time of day or night, to enable prompt treatment.
Tests of vitamin B_{12} metabolism, preferably including active vitamin B_{12} (transcobalamin and methylmalonic acid), homocysteine, and folate	At presentation Abnormalities may be suggestive of IMDs in cobalamin and folate metabolism that are eminently treatable.
Two bottles of 5-mL lithium heparin preservative samples for plasma amino acids and acyl-carnitines	At presentation
Urine collected in a 15-mL universal container for urine organic acids	At presentation
Additional tests to consider	
If the patient is undergoing a lumbar puncture and IMD is suspected, send: • 1 aliquot of CSF in a fluoride–oxalate tube for lactate and glucose • 1 aliquot of CSF in a plain tube to the laboratory to be saved for further investigations if needed Paired plasma fluoride–oxalate and lithium heparin samples should be sent along with the CSF samples. CSF results can only be interpreted by comparison with serum and plasma levels.	Preferably discuss with a metabolic centre before taking samples. If unable to contact centre, take and send to the laboratory to be saved. Ensure prior consent taken as per local guidelines.
Consider sending two 5-mL EDTA preservative whole-blood samples for DNA analysis and WBC extraction.	Preferably discuss with a metabolic centre before taking samples. If unable to contact centre or if patient perimortem, take and send to laboratory to be saved. Ensure prior consent taken as per local guidelines.
Skin biopsy: DNA can be extracted from fibroblasts grown from a skin biopsy (even taken up to 24 hours postmortem if sterile); a postmortem diagnosis of individuals dying of a suspected IMD may allow screening of other family members, as well as confirm the cause of death.	Preferably discuss with a metabolic centre before taking samples. If unable to contact centre or if the patient is perimortem or immediate postmortem, take and send to the laboratory to be saved. Ensure prior consent taken as per local guidelines.
Tissue biopsy (if indicated): muscle, liver, renal, and cardiac	Preferably discuss with a metabolic centre before taking samples. If unable to contact centre or if the patient is perimortem or immediate postmortem, take and send to the laboratory to be saved. Ensure prior consent taken as per local guidelines.

CK, Creatine kinase; *CRP,* C-reactive protein test; *CSF,* cerebrospinal fluid; *EDTA,* ethylene-diamine-tetra-acetic acid; *FBC,* full blood count; *IMD,* inherited metabolic disease; *WBC,* white blood cell.

endogenous insulin production, and both help reverse the catabolic state. ER prevents or mitigates acute metabolic decompensation caused by a catabolic state, resulting in endogenous protein breakdown, lipolysis, and possible hypoglycaemia.

Emergency departments should ensure these patients are triaged as emergencies and that treatment and investigations are instituted without delay as per their emergency management plan. Although prompt treatment is often associated with good prognosis, a delay can lead to permanent significant neurological sequelae or death.

Patients should also be investigated and treated for metabolic triggers of acute decompensation. ERs for common metabolic conditions are available online, for example, on the British Inherited Metabolic Disease Group's (BIMDG) website (see Further reading). Every effort should be made to contact the patient's metabolic centre for advice as soon as possible, ideally within the first hour. (Many metabolic centres, including in the UK, offer a 24/7 on-call advice.)

Situations when an inherited metabolic disease should be suspected

An IMD should be suspected in patients presenting with one of the clinical syndromes described next, particularly if there is no obvious alternative diagnosis, if there is a suggestive family history, one or more metabolic triggers, a history of aversion to certain food groups (e.g., proteins), or a recent history of exclusion diets (low-carbohydrate, high-fat, high-protein diets). A basic metabolic screen should be sent off. Abnormalities in biochemistry may only be apparent during an acute episode, so samples for a basic acute metabolic screen (**Box 36.1**) should be taken at presentation. Specialist advice should be sought and emergency treatment initiated as soon as possible if an IMD is suspected.

Clinical presentations in which an IMD should be suspected include the disorders described next.

Encephalopathy and urea cycle disorders

Patients with a known and an as-yet undiagnosed IMD can present for the first time in adulthood with encephalopathy. Specific findings on brain imaging and electroencephalography may point towards a metabolic encephalopathy, which most commonly occurs secondary to hyperammonaemia. Plasma ammonia levels must therefore be measured urgently in all patients presenting with undiagnosed encephalopathy.

Ammonia is a by-product of protein catabolism (**Fig. 36.1**). It is detoxified by conversion to urea via the urea cycle in the liver (**Fig. 36.2**), and urea is subsequently excreted by the kidneys. A diagnosis of a urea

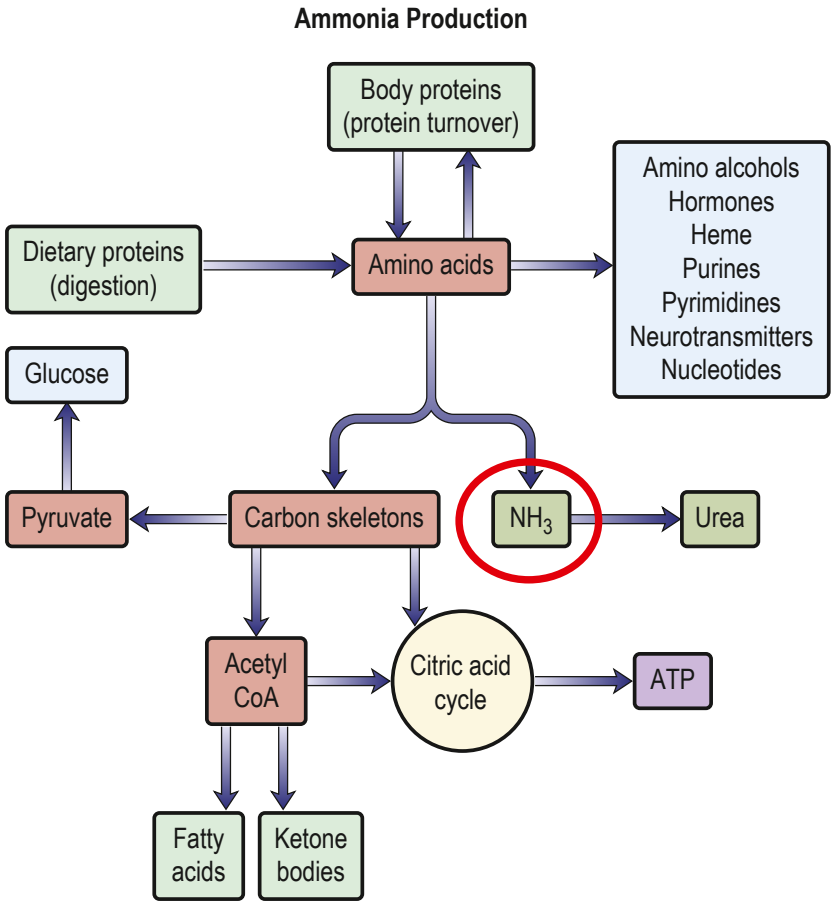


Fig. 36.1 Much of the body, including structural components, enzymes, and receptors, is made of proteins. Dietary proteins are broken down to constituent amino acids and then used for synthesis of body proteins. Amino acids derived from excess dietary protein, and during times of increased energy demands, from breakdown of body proteins, are used as metabolic fuel through decomposition to hydrocarbons skeletons and ammonia. Hydrocarbons are converted to pyruvate and acetyl CoA, which are in turn metabolized to glucose and fatty acids or ketones, respectively. Ammonia is converted to urea via the urea cycle in the liver and subsequently excreted in the urine. *ATP,* Adenosine triphosphate.

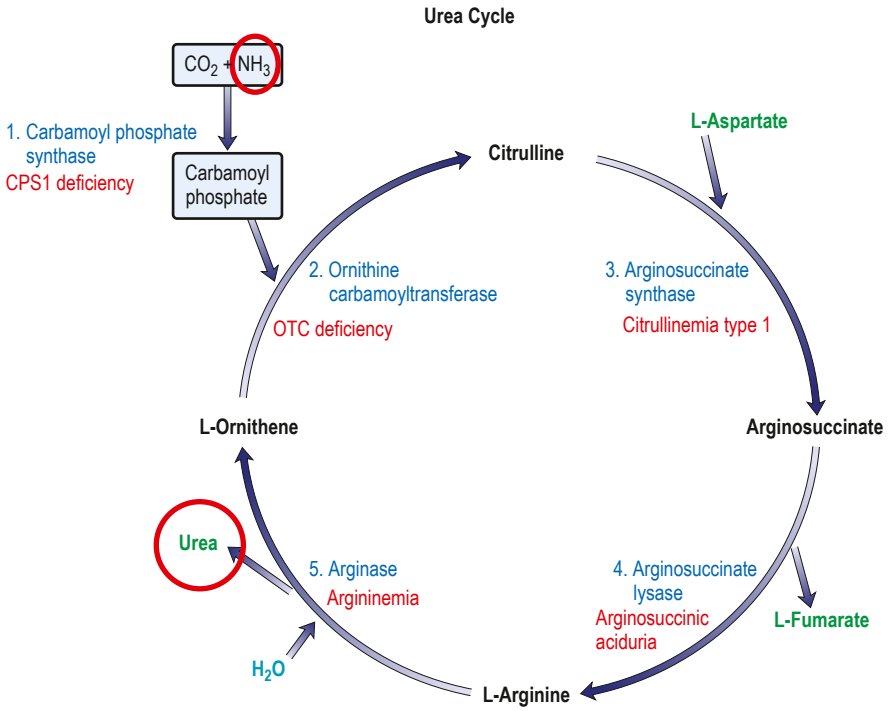


Fig. 36.2 Ammonia is detoxified by conversion to urea through the urea cycle in the liver hepatocytes. The urea cycle is a set of cytosolic and mitochondrial enzymatic reactions that convert ammonia (NH_3) and carbon dioxide (CO_2) to urea (enzymes are highlighted in blue). Briefly, carbamoyl phosphate synthetase 1 (CPS1) *(1)* catalyzes the condensation of NH_3 and CO_2 to carbamoyl phosphate, which in turn combines with ornithine in the presence of ornithine carbamoyltransferase *(2)* to citrulline. Citrulline is bound to aspartate by arginosuccinate synthetase *(3)* to form arginosuccinate. Arginosuccinate is split by arginosuccinate lyase *(4)* to form arginine and fumarate. Arginine is hydrolyzed by arginase *(5)* to the end products urea and ornithine, which are recycled back into the urea cycle. Inherited deficiencies in any of the enzymes in the urea cycle are collectively known as urea cycle disorders, labelled in red. *1,* CPS1 deficiency; *2,* ornithine transcarbamylase (OTC) deficiency; *3,* citrullinemia type 1; *4,* arginosuccinic aciduria; *5,* argininemia.

cycle defect, one form of IMD, should be sought in patients in whom other more common causes of hyperammonaemia such as decompensated liver disease and drugs have been excluded. Prolonged hyperammonaemia is associated with permanent neurological deficit and death. Ammonia should be included in the first-line screening investigations for all patients presenting with encephalopathic symptoms.

Ammonia samples are unstable and should be collected on ice and dispatched promptly to the laboratory. If hyperammonaemia is confirmed, treatment can and should be started promptly before the precise details of the IMD causing hyperammonaemia is known.

The main principles of treatment are to:

- Prevent catabolism by giving i.v. ER. (This can be done before ammonia levels are known if index of suspicion high.)
- Lower plasma ammonia levels as quickly as possible. This is most effectively done using haemofiltration or haemodialysis in symptomatic patients.
- Decrease protein load by temporarily stopping all external natural protein intake. (This can be done before ammonia levels are known if index of suspicion high.) Total protein restriction should not continue beyond 48 hours.
- Use ammonia scavengers such as sodium benzoate and phenylbutyrate as soon as available.

Published guidelines ('Urea cycle defects—acute emergency management') are available on the BIMDG's website (see Further reading). Advice from a specialist IMD centre should be sought for ongoing management and further investigations immediately on finding a high ammonia level.

Rhabdomyolysis

Although trauma and drugs are the commonest causes of rhabdomyolysis, IMDs such as glycogen storage disorder (GSD) type 5 (see **Box E36.3**; the eBoxes can be found in our ebook, please use the PIN code on the inside cover of the book for access) or disorders of fatty acid metabolism (see **Box E36.4**) should be considered in patients with recurrent episodes or a suggestive history. Send samples for an acute metabolic screen, including CK. In addition to standard treatment for rhabdomyolysis, patients with IMD-related rhabdomyolysis benefit from i.v. ER.

Stroke

Thrombotic strokes in younger adults or a cerebral sinus thrombosis should prompt measurement of plasma homocysteine levels to exclude homocysteine and folate metabolism disorders, particularly as they are treatable (see primary homocystinuria; see **Box E36.1**).

IMDs such as mitochondrial disorders and Fabry's disease should also be considered in patients with strokes who are young or in whom symptoms or imaging is inconsistent with a defined

vascular territory. The risk of stroke can be up to 12 times higher in young males with Fabry's disease than in the general population

Hypoglycaemia

Although hypoglycaemia is a common neonatal and childhood presentation of GSDs (see **Box E36.3**) and fatty acid oxidation disorders (FAODs) (see **Box E36.4**), presentations among adults are rare. Nevertheless, there are isolated reports of patients with IMDs such as glucokinase-activating mutations presenting for the first time in adulthood with hypoglycaemia. Because this is eminently treatable, the diagnosis should be considered and excluded. Insulin and C-peptide measurements and an acute metabolic screen at the time of hypoglycaemia provide important diagnostic clues.

Acidosis

A raised unexplained anion gap may be suggestive of an IMD. Hyperlactatemia, hyperammonaemia, or ketosis without hyperglycaemia may all be diagnostic clues. Clinicians should send samples for an acute metabolic screen, including blood gases, bicarbonate, glucose, and lactate, at presentation if an IMD is suspected.

GENERAL PRINCIPLES OF MANAGEMENT OF INHERITED METABOLIC DISEASES

There are several common mechanisms by which IMDs may cause symptoms, for instance, via accumulation of metabolic precursors or toxic waste products. These suggest general management principles applicable to the specific conditions described later in the chapter:

Dietary manipulation to bypass 'metabolic block'

'Metabolic block' describes accumulation of upstream precursors and deficiency of downstream products of a defective metabolic enzyme.

The principle of management of these conditions is to institute dietary changes to decrease the availability of the precursors and replace any crucial important products.

For example, amino acid metabolism disorders (described later) are generally managed by varying degrees of protein restriction. In phenylketonuria (PKU) and maple syrup urine disease (MSUD), this restriction may be severe (less than 10 g of protein a day). To prevent severe protein deficiency and micronutrient deficiencies as a result of food restriction, patients need supplementation with amino acid mixtures that contain all other essential amino acids other than the specific precursor (phenylalanine in those with PKU and leucine in those with MSUD). Patients should also be advised to have regular adequate calorie intake and avoid fasting to prevent endogenous protein breakdown and build-up of precursor amino acids.

Adult Refsum's disease and long-chain fatty acid disorders (covered later) are examples of diseases in which symptoms are caused by increase in circulating levels of long or very long-chain fats. Treatment principles here are similar and include restriction of very long- or long-chain fats in the diet and adequate calorie intake to prevent lipolysis.

Maintenance of regular and adequate calorie intake

Patients with disorders of glycolysis, glycogenolysis, or gluconeogenesis (e.g., GSD 1 and 3) are treated with regular and adequate calorie intake in the form of slow-release carbohydrates, or glucose, or both. Although patients with GSD 1 are unable to use fats and proteins as energy sources because of impaired gluconeogenesis, patients with conditions such as GSD 3 benefit from alternative energy sources such as protein and fats in the diet. Patients with glucose transporter 1 deficiency are unable to adequately transport glucose for brain energy metabolism; these patients benefit from alternative fuel sources such as ketone bodies produced endogenously when on a ketogenic diet.

Supplementation with vitamins, minerals, and co-factors

Vitamins and minerals are important co-factors for a number of intermediary energy metabolism pathways. A number of IMDs are related to impaired absorption, transport, or metabolism of these vitamins and minerals. Some of these disorders are managed with high-dose vitamins, often resulting in normalization of biochemistry and mitigation of symptoms.

Examples include pyridoxine responsive homocystinuria, vitamin B_{12}–responsive methylmalonic aciduria, and sapropterin responsive PKU.

Enzyme replacement therapy

Enzyme replacement therapy (ERT) is becoming available for patients with a number of conditions, most notably for lysosomal disorders. A range of 'orphan drugs' are available via government subsidy schemes that cover the pharmaceutical costs of development of drugs that would otherwise not be profitable to manufacturers because of low patient numbers. Examples include recombinant alpha-galactosidase (e.g., agalsidase alpha and beta) for use in Fabry's disease, or recombinant glucocerebrosidase (e.g., imiglucerase) for use in Gaucher's disease. Some of the main challenges for ERT are the need for regular lifelong parenteral administration, variable bioavailability, and development of antibodies to the drugs.

Newer therapies

Gene, mRNA, and oligonucleotide-based therapies for a number of IMDs are currently undergoing clinical trials. Advances in this field of emerging therapies are likely to have a significant positive impact on the prognosis and outcomes in patients with IMDs in the future and may even offer the prospect of a complete and lifelong cure.

Supportive management and surveillance

Despite recent advances in specific therapies, management of many IMDs remains largely supportive. Patients must be followed up regularly to ensure early detection and management of co-morbidities such as cardiomyopathy, diabetes, neurological or skeletal abnormalities, learning difficulties, and mental health problems. The spectrum of phenotype is wide, even within a single IMD, and therefore management plans and surveillance must be personalized to ensure best possible quality of life outcomes.

SPECIFIC INHERITED METABOLIC DISEASES

Mitochondrial diseases

Mitochondrial function is regulated by the interplay between nuclear, chromosomal genes and the 37 mitochondrial genes that together code for nearly 1500 mitochondrial proteins (see Chapter 5). Thus,

mitochondrial diseases (MDs) can be caused by maternally inherited defects in the mitochondrial genome (see Fig. 5.6 in Chapter 5) or by Mendelian inherited nuclear genome defects. *De novo* mutations are also known to occur.

Every nucleated cell in the body contains between 100 and 10 000 mitochondria that are responsible for adenosine triphosphate (ATP) generation through oxidative phosphorylation. Hence, mitochondrial dysfunction leads to clinical manifestations secondary to cellular dysfunction and premature cell death as a result of an energy deficit. MDs are not especially uncommon, with a reported incidence of 1 in 4300 in the UK. MDs presenting in adulthood were traditionally thought to be caused by mutations in the mitochondrial genome; however, it is now increasingly apparent that roughly 25% of MDs presenting for the first time in adulthood may be secondary to mutations in the nuclear genome.

Patients can present with a variable phenotype with one or more clinical features associated with MD, varying in severity and chronology (**Box 36.2**). Patients may sometimes present with well-defined clinical syndromes that are associated with a particular nuclear or mitochondrial mutation (**Box 36.3**). Investigation profiles should include glucose, lactate, acyl-carnitines, CK, muscle biopsy, and genetics.

Diagnosis

Although biochemistry may be indicative, confirmatory diagnosis requires specialist investigations, including tissue (most often muscle) biopsy and genetic testing.

Management

Although clinical trials are ongoing, specific therapies for patients with MD are currently not available in the clinical setting. Management is largely supportive, including surveillance for and treatment of MD-associated co-morbidities such as diabetes and cardiac disease. Patients should be offered genetic counselling, and a detailed pedigree history should be taken to ensure appropriate cascade testing of at-risk relatives. With the option of mitochondrial donation now becoming available, reproductive options should also be discussed.

Box 36.2 Clinical features that should prompt consideration of a mitochondrial disease[a]

System	Clinical feature
Neurological	Ataxia or stroke-like episodes (predominantly occipital), seizures, myoclonus, neuropathy, myopathy, migraines, encephalopathic symptoms, magnetic resonance imaging changes suggestive of mitochondrial disease
Cardiac	Cardiomyopathy, conduction defects, Wolff–Parkinson–White syndrome
Endocrine	Diabetes, particularly in patients with a strong maternal history of diabetes, sensorineural hearing loss or a phenotype, biochemistry not typical of insulin resistance
Eyes	Non-fatigable ptosis, ophthalmoplegia, pigmented retinopathy, cataracts, optic atrophy
Sensorineural hearing loss	Early onset, bilateral, with strong maternal family history
Gastrointestinal	Dysmotility presenting as vomiting or constipation
Other biochemistry	May have raised plasma lactate or creatine kinase

[a]Unless otherwise explained.

Disorders of protein metabolism

Phenylketonuria

PKU (incidence, 1 in 20 000 globally) is an autosomal recessive amino acid metabolism disorder caused by a bi-allelic mutation in the phenylalanine hydroxylase gene. It is the most common IMD encountered in specialist metabolic centres in Europe, where the incidence is highest (1 in 5000 to 15 000 in white babies of European descent). It is characterized by elevated levels of phenylalanine in the blood, and if treatment is delayed, it can result in severe intellectual disability. Early treatment within the first few days of life results in normal development. Many countries, including those in the UK, offer newborn screening for PKU, with the aim of starting therapy within the first 2 weeks of life. The aim of treatment is to maintain phenylalanine levels below therapeutic targets, which may vary slightly among countries. The mainstay of treatment is a diet restricted in natural proteins, combined with supplements (phenylalanine-free protein, vitamins, and minerals) and specialist low-protein foods available on prescription. More recently, more specific therapies, including enzyme and co-factor replacement therapy, have become available in some countries. High phenylalanine levels during pregnancy in patients with PKU can result in microcephaly, significant intellectual disability, cardiac defects, and growth retardation in the baby. Females with PKU are therefore

Box 36.3 Some well-described mitochondrial disease syndromes

Syndrome	Inheritance	Key feature	Additional features
Spastic paraplegia type 7 (SPG7) (6 in 100 000)	AR	Spastic paraparesis	Dysarthria, dysphagia, nystagmus
Progressive external ophthalmoplegia	AD or AR Mitochondrial Sporadic	Ophthalmoplegia, ptosis	Dysphagia, myopathy, exercise intolerance
Optic atrophy type 1	AD	Optic atrophy, ophthalmoplegia	Sensory neural hearing loss, ataxia, neuropathy
Mitochondrial encephalopathy, lactic acidosis, and stroke-like episodes	Mitochondrial	Diabetes, sensory neural hearing impairment, lactic acidosis, encephalopathy	
Leber's hereditary optic neuropathy	Mitochondrial	Optic atrophy in early adulthood	Movement disorders, tremors, cardiac conduction defects

AD, Autosomal dominant; *AR*, autosomal recessive.

advised against unplanned pregnancies and require close monitoring by specialist centres because target phenylalanine levels during pregnancy are much lower than baseline adult targets. Unplanned pregnancy in a female with PKU is a medical emergency and should be referred to a metabolic centre to enable treatment to be started or options discussed within 24 hours. Pregnancy-related hyperemesis in females with PKU needs aggressive and prompt management because if it is left untreated, it can lead to a rise in phenylalanine levels because of reduced calorie intake.

Some other common protein metabolism IMDs are listed in **Box E36.1**. They all have autosomal recessive inheritance, except for ornithine transcarbamylase deficiency, which is X-linked. All the conditions other listed in **Box E36.1** other than alkaptonuria can present acutely.

Lysosomal storage disorders

Lysosomes play an important role in cell turnover. Their main functions are to break down complex macromolecules to simple monomers and to transport these monomers out of the lysosome into the cytosol for recycling. Any disruption to this process results in lysosomal storage disease (LSD). Genetically inherited LSDs are some of the commonest IMDs encountered in metabolic centres treating adult patients, with Gaucher's and Fabry's disease being the most common among these. A number of these conditions can present for the first time in adulthood. The prognosis is variable, depending on residual enzyme activity. Specific therapies are now available for a number of LSDs, and hence it is important that clinicians consider LSDs in patients presenting with suggestive symptoms. A few common LSDs with specific therapies are listed in **Box E36.2**. All LSDs listed in the table have an autosomal recessive inheritance except for Fabry's disease, which has an X-linked inheritance.

Disorders of carbohydrate metabolism

Glycogen storage disorders

GSDs are a group of IMDs in which either glucose utilization (glycolysis) or glycogen mobilization (glycogenolysis or glycogen transport) is impaired. Glycogen is a glucose polymer stored most abundantly in the muscle and liver. Patients present with symptoms related to decreased availability of glucose for ATP generation (e.g., resulting in hypoglycaemia), abnormal glycogen accumulation (causing hepatosplenomegaly), and impairment in other linked metabolic pathways. Effective treatments for a number of GSDs exist, and early treatment significantly improves the prognosis. GSDs can be classified broadly based on the most affected organs. Some of the principal GSDs are listed in **Box E36.3**. A few other treatable IMDs of carbohydrate metabolism are listed in **Box E36.4**.

Disorders of fatty acid oxidation defects

The fatty acid oxidation pathway provides an important alternative fuel source by using fats in situations such as fasting, exercise, or illness in which the glucose supply may be inadequate. Briefly, it involves the following:

- Conversion of fatty acids to fatty acyl CoA on entry through the plasma membrane
- Entry of fatty acyl CoAs into the mitochondria via the carnitine shuttle
- Beta-oxidation of fatty acyl CoAs to acetyl-CoA by dehydrogenases in the mitochondria
- Transfer of electrons by dehydrogenases to the respiratory chain via electron carriers. Acetyl CoA is used either to generate ketone bodies (an alternate fuel for the brain) or for direct ATP production via Krebs cycle (e.g., in skeletal muscle).

Defects in any step in the pathway result in a FAOD. Characteristic features of FAODs include hypoketotic hypoglycaemia, cardiomyopathy, and myopathy with or without rhabdomyolysis. Although most present in childhood, some FAODs can present for the first time in adulthood (**Box E36.5**) with myopathy or rhabdomyolysis. Unlike in GSD 5 (McArdle's disease), symptoms usually occur after prolonged exercise, during a prolonged fast, or during an intercurrent illness that results in a catabolic state after glycogen reserves are depleted. When the diagnosis is known, management to avoid crisis is relatively simple (see **Box E36.5**). As with all other IMDs, symptoms may be unmasked (with catastrophic consequences) by a recent change in diet (e.g., a ketogenic diet) in addition to all the other metabolic triggers already described.

Porphyrias

Porphyrias are a group of disorders caused by defects in the haem synthesis pathway, and symptoms occur because of an accumulation of toxic metabolites. Whereas a block in the early stages of synthesis leads to accumulation of porphyrin precursors, including porphobilinogen (PBG) and α-aminolaevulinic acid (ALA), a block in the late stages results mainly in accumulation of porphyrins (**Fig. 36.3**). Porphyrin precursors are neurotoxic, but porphyrins themselves induce a photosensitive blistering rash. Thus, presentations may include acute neurovisceral crisis (in conditions such as acute intermittent porphyria in which there is a block in the early stages of haem synthesis) or non-acute cutaneous disease (in conditions such as porphyria cutanea tarda in which there is block in the late stages of haem synthesis). A block in the intermediate stages cause cutaneous and neurovisceral symptoms (e.g., in variegate porphyria and hereditary coproporphyria).

Acute porphyrias have an autosomal dominant inheritance; however, penetrance is incomplete. They almost always present in late childhood or adulthood. The prevalence is 5 per 1 000 000 in Europe, but it is much higher prevalence in South Africa (3 in 1000). Clinical features include abdominal pain, vomiting, neuropsychiatric symptoms, hypertension, seizures, hyponatraemia, and fever. In addition to the usual triggers for metabolic decompensation in IMDs, anaesthesia and drugs (including cytochrome P450 enzyme inducers) can precipitate an acute attack. Porphyria occurs more often in females, particularly in the luteal phase of the menstrual cycle and pregnancy. Symptoms may mimic an acute abdomen, and anaesthesia during an acute attack can be catastrophic. It is treatable, so a diagnosis of porphyria should be considered and sought early in the emergency department if symptoms are suggestive. The diagnosis is made by measuring ALA and PBG in a fresh urine sample (in a plain bottle with no preservatives and always protected from light). Normal levels exclude acute porphyria, and an alternative diagnosis for symptoms should be sought.

If porphyria is confirmed or likely, management should be initiated immediately.

- Remove all precipitating causes, including any triggering drugs (www.drugs-porphyria.org).
- Give supportive therapy for symptoms and an i.v. infusion for hydration and correction of electrolyte imbalance.
- Ensure a high carbohydrate intake of at least 600 g/day.
- Administer heme arginate (i.v.).
- Contact a porphyria reference centre for further advice. Published emergency management guidelines are available on the BIMDG's website. Recurrent severe attacks may merit consideration for liver transplantation.

Box E36.1 Common inherited metabolic diseases of protein metabolism

Urea cycle defects (A)[a]	Mutations in one of six urea cycle enzymes or two transporters (CPS1, OTC,[a] ASS1, ASL, ARG1, NAGS, SLC25A15, SLC25A13; most common, OTC deficiency) Incidence: 1 in 8000–45 000 Presentation in adulthood may be triggered by a change in diet (e.g., ketogenic diet, intercurrent illness, extreme metabolic stress). Progressive encephalopathy, cyclical vomiting, non-specific neuropsychiatric symptoms Diagnosis and biochemistry: characteristic abnormalities in AA or OA; when unwell: hyperammonaemia, genetics Management: as for hyperammonaemia in acute setting, lifelong protein restriction +/– ammonia scavengers or arginine, citrulline; liver transplant; gene therapy clinical trials ongoing Prognosis: good if treated promptly, fatal or catastrophic sequelae if diagnosis and treatment are delayed
Tyrosinaemia type 1A[a]	Fumaryl-acetoacetate hydroxylase deficiency Incidence: 1 in 100 000 More common in Quebec, Canada, because of founder mutation Presentation: liver failure, Fanconi's syndrome, tubulopathy, renal rickets, acute porphyria, gastrointestinal bleeding Guillain-Barré syndrome–like acute neurogenic crisis Hepatocellular carcinoma Diagnosis and biochemistry: raised tyrosine in serum AA and raised succinylacetone in urine OA, genetics Management: nitisinone (NTBC), tyrosine restriction, tyrosine-free amino acid supplementation Liver transplantation: less needed now because of success with nitosinone Prognosis: good if treated early If the patient is not compliant with medication or untreated, can present in adulthood with any of clinical features described
Alkaptonuria (A)	Homogentisate 1,2-dioxygenase deficiency Incidence: 1 in 250 000 Presentation: more often in adulthood (third to fifth decades of life); large joint arthritis, ochronotic pigmentation (sclera, ears), valvular heart disease Urine turns black on standing, black nappies in children Diagnosis and biochemistry: raised homogentisic acid on OA, genetics Management: nitisinone (NTBC) improves biochemical profile Prognosis: chronic progressive disease with variable prognosis
Classical homocystinuria (A)	Classical homocystinuria Cystathionine β synthase deficiency Incidence: 1 in 40 000–250 000; Qatar: 1 in 1800 Presentation: thromboembolic events (stroke or arterial or venous occlusive disease) Acute lens dislocation, visual loss, neuropsychiatric symptoms Developmental delay, seizures; acute, subacute, or chronic neurological deterioration; marfanoid habitus; osteoporosis Diagnosis and biochemistry: raised plasma homocysteine, (particularly if >50 μmol/L); raised plasma methionine; reduced plasma cysteine Management: pyridoxine, betaine, vitamin B_{12}, folic acid Methionine-restricted diet in pyridoxine-non-responsive homocystinuria with added vitamin supplements Genetic analysis may be useful in predicting response to pyridoxine in addition to confirming diagnosis Prognosis: variable depending on severity; early treatment (in neonatal period or early childhood) results in normal IQ and decreased risk of long-term morbidity Note: Folate or vitamin B_{12} deficiencies or other inherited metabolism defects in the folate–vitamin B_{12} pathway can also cause a rise in homocysteine levels; these must be excluded
Defects in metabolism of branched-chain amino acids Maple syrup urine disease (MSUD) Propionic aciduria (PA) Methylmalonic aciduria (MMA) (A)	MSUD: branched-chain α-ketoacid dehydrogenase deficiency PA: propionyl CoA carboxylase deficiency MMA: genetically heterogenous disorder Incidence: MSUD, 1 in 185 000; 1 in 200 in Old Order Mennonite population; PA, 1 in 100 000; more common in Saudi Arabians; MMA, 1 in 100 000 Presentation: first presentation in adults very unusual, although there have been a few reported cases in the literature Acute decompensation: encephalopathy, acidosis, hyperammonaemia, developmental delay, abnormal neurology, cardiomyopathy (PA > MMA), pancreatitis (MMA, PA), renal failure (more often in MMA; occasionally, adults with milder forms of MMA may present with just renal impairment) Diagnosis and biochemistry: metabolic acidosis with high anion gap; hyperammonaemia during acute decompensation Characteristic abnormalities on AA or OA and AC profiles Management: natural protein restriction, avoid catabolism liver transplantation (MSUD, PA), renal or combined liver and renal transplantation (MMA), carnitine supplementation (PA, MMA) BCAA-free supplements (MSUD) Prognosis: variable depending on severity; acute decompensations can be fatal

[a]Acute presentation.

A, Adult presentation reported (including first presentation in adulthood); *AA*, amino acid; *AC*, acylcarnitine profile; *BCAA*, branched-chain amino acids; *OA*, organic acid; *OTC*, ornithine transcarbamylase.

Box E36.2 Lysosomal storage disorders

Gaucher (A)	Beta-glucocerebrosidase (GBA) deficiency Incidence: panethnic, 1 in 75 000; AJ population, 1 in 850 Presentation: 94% of patients have non-neuronal disease (type 1) characterized by hepatosplenomegaly, bone marrow involvement, cytopaenia, and skeletal symptoms such as severe bone pain and avascular necrosis. Presentation can occur in adulthood, frequently with asymptomatic splenomegaly. The neuronopathic form is much rarer (6% of patients). Acute (1%) (type 2) presentation almost always at younger than 1 year of age and includes severe progressive neurological deficit leading to early death. Chronic (5%) (type 3) presentation includes mild to severe neurological and skeletal symptoms and organomegaly. Intermediate forms occur, as can late neurological presentations in adulthood with neurological symptoms including a supranuclear horizontal gaze palsy Acute complications: avascular necrosis of bone and splenic infarcts Heterozygous mutations may increase risk of Parkinson's syndrome Diagnosis and biochemistry: deficient GBA activity in leukocytes or fibroblasts, 'Gaucher cells' on biopsy; genetic conformation Management: supportive, enzyme replacement therapy (ERT) using recombinant GBA
Fabry (A, X)	Alpha-galactosidase A (α-GLA) deficiency Incidence: 1 in 40 000 Presentation: late onset of milder forms is much more common; acroparesthaesia (burning or numbness in the hands and feet), cornea verticillate, angiokeratomas, gastrointestinal symptoms (classical phenotype, most common presenting symptoms in paediatric population); adults more often present with end-organ disease; arrhythmia, cardiac fibrosis, hypertrophic cardiomyopathy; proteinuria, renal impairment; stroke; decreased exercise tolerance, fatigue, depression, and acroparesthesia worsened by stress, heat, or exercise Females can also be symptomatic because of lyonization of the X chromosome but often experience end-organ symptoms usually a decade later than males Wide phenotypic spectrum in both males and females based on residual enzyme activity levels Diagnosis and biochemistry: deficient (α-GLA) activity in white blood cells; activity can be normal in 60% of affected females; genetic confirmation is therefore needed Management: supportive, ERT
Niemann-Pick disease (A)	Acid sphingomyelinase deficiency Incidence: 1 in 250 000; 1 in 40 000 in AJ Presentation: age of presentation and severity are variable; can present for the first time in adulthood; splenomegaly, hypersplenism; hepatomegaly, hepatic fibrosis, cirrhosis; dyslipidaemia (high low-density lipoprotein cholesterol and triglycerides); osteoporosis; storage facies; cherry-red macular spot (25%); lung fibrosis causing restrictive lung disease Niemann-Pick A: severe neurodegenerative form; usually fatal in early childhood Diagnosis and biochemistry: deficient enzyme activity in leukocytes, genetic confirmation Management: supportive, ERT
Tay–Sachs disease (A)	Beta-hexosaminidase (isoform A) deficiency Incidence:1 in 250 000; 1 in 25 (vs 1 in 250 in the general population) carrier frequency in AJ; actual incidence in Jewish populations has now significantly decreased in many countries because of effective counselling Presentation: childhood disease is fatal with rapidly progressive neurology, visceromegaly, dysmorphia, and increased startle response Late-onset adult disease: age of onset and severity vary; cerebellar and motor neurone dysfunction causing tremor, dysarthria, chorea, and dystonia Diagnosis and biochemistry: deficient enzyme activity in leukocytes, genetic confirmation Management: supportive
Pompe's disease (GSD type 2) (A)	Acid alpha-glucosidase (acid maltase) deficiency Incidence: 1 in 40 000 Presentation: adult with Pompe's disease are usually symptomatic in the third to fourth decades of life with limb girdle muscle weakness and respiratory insufficiency secondary to diaphragm weakness, which is often major cause of death acute decompensation related to respiratory failure Classical infantile form: more severe presentation in the first few months of life Diagnosis and biochemistry: deficient enzyme activity in leukocytes; vacuolated lymphocytes; vacuolar myopathy on muscle biopsy; pulmonary function test (>20% change in vital capacity between sitting and supine position) Management: supportive, ERT

A, Adult presentations reported; *AJ,* Ashkenazi Jewish population; *X,* X-linked inheritance.

i Box E36.3 Common glycogen storage disorders

GSDs affecting mainly muscle GSD 2 (Pompe's disease) (A) GSD 5 (McArdle's disease) (A)[a]	See **Box E36.2** Myophosphorylase deficiency Incidence: 1 in 100 000 Presentation: first presentation in adults not unusual; exercise intolerance; pain within 5–10 min of starting exercise ('second wind' phenomenon, i.e., symptoms improve within a few minutes if able to continue with exercise); rhabdomyolysis Diagnosis and biochemistry: raised CK, flat lactate with an exaggerated rise in ammonia on non-ischaemic exercise test, genetic testing, PAS staining vacuolar myopathy (although muscle biopsy should not normally be needed for diagnosis) Management: supportive with aerobic conditioning exercise programmes and dietary advice to improve exercise tolerance and avoidance of unaccustomed strenuous exercise; oral glucose (for e.g., glucose polymer drink with 25% glucose) 5 min before exercise
1a (glucose-6-phosphatase deficiency) and 1b (glucose-6-phosphatase translocase deficiency)[a] GSD IIIA[a]	Incidence: 1 in 100 000 Presentation: hypoglycaemia, ketosis, and lactic acidosis in infancy; growth retardation massive hepatomegaly; renal impairment; hepatic adenomas; and hepatocellular carcinoma (GSD 1b can present with neutropenia and features similar to inflammatory bowel disease); no cognitive impairment if hypoglycaemia is avoided; with improved treatment, most transition into adulthood; unlikely to present for the first time in adulthood; isolated case reports of first presentation in adults; adults are much more likely to present with complications related to cardiac or liver involvement rather than hypoglycaemia Diagnosis and biochemistry: enzymology and genetic testing for diagnosis; secondary abnormalities include hyperuricemia, ketosis, lactic acidosis, and hypertriglyceridaemia Management: dietary therapy with regular doses of slow-release cornstarch to maintain normoglycaemia, ER; most severe cases may need continuous feed through gastrostomy; patients with GSD 1b need management of neutropenia and inflammatory disease; SGLT2 inhibitors decrease the incidence and severity of neutropenia in patients with GSD 1b Debrancher enzyme deficiency Incidence: 1 in 100 000; 1 in 5400 in North African Jewish ancestry Presentation: less severe than GSD 1; fasting hypoglycaemia less severe than GSD 1; ketosis but no lactic acidosis; mild hyperlipidaemia; childhood presentation with hypoglycaemia and hepatomegaly; kidneys not involved; hepatic adenomas less common; muscle weakness with raised CK may occur; cardiac complications, including ventricular hypertrophy Diagnosis and biochemistry: enzymology and genetic testing for diagnosis; may have raised CK Management: dietary therapy with regular doses of slow-release cornstarch to maintain normoglycaemia; ER

[a]Emergency presentations reported.

A, Adult presentations reported; *AR,* autosomal recessive; *CK,* creatine kinase; *ER,* emergency regimen; *PAS,* periodic acid-Schiff; *SGLT2,* sodium–glucose transporter 2.

i Box E36.4 Other inherited metabolic diseases affecting carbohydrate metabolism

Galactosaemia	Most commonly galactose-1-phophate uridyltransferase deficiency Incidence: 1 in 40 000 Presentation: neonatal presentation on ingestion of galactose (milk or formula) with failure to thrive, liver disease, cataract, and renal tubular dysfunction; treated adults can have developmental delay, neurological deficits, or premature ovarian failure Adult presentation rare; however, exclude the diagnosis in patients with premature ovarian failure ± cataracts ± neurological disease Diagnosis and biochemistry: elevated levels of galactose-1-phosphate in RBCs Management: immediate removal of galactose from the diet as soon as suspected without awaiting diagnostic confirmation; galactose restriction for life; supportive therapy, including options to freeze eggs before onset of premature ovarian failure
Glucose transporter 1 (GLUT-1) deficiency (AD)	GLUT-1 deficiency Incidence: 1 in 90 000 Presentation: symptoms secondary to decreased facilitated glucose transport into brain via GLUT-1; symptoms include seizures, developmental delay, and complex movement disorder; symptoms worse a few hours postprandial Diagnostic and biochemistry: lumbar puncture after a 4- to 6-hour fast; paired CSF-to-blood glucose ratio <0.45 suggestive; confirmation with genetics and RBC glucose uptake studies Management: ketogenic diet imitates metabolic state of fasting so brain uses ketones as alternative fuel; ketogenic diet for life; antiepileptic medication
Hereditary fructose intolerance	Aldolase B deficiency Incidence:1 in 20 000 Presentation: ingestion of fructose, sucrose, or sorbitol causes acute illness (within minutes) of abdominal pain, vomiting, and diarrhoea More severe symptoms, including hypoglycaemia, seizures, and coma, can occur Persistent exposure leads to liver steatosis, fibrosis, or failure and renal tubular disease or Fanconi's syndrome; patients may go undiagnosed or have the diagnosis delayed until adulthood because of self-restricted fructose intake and thus symptom prevention Diagnostic and biochemistry: enzymology now replaced by genetic testing Management: withdrawal of all enteral and parenteral sources of fructose medications, infusions, and supplements should be checked for fructose; supplementation with vitamin C and other vitamins Prognosis good if all fructose is avoided

AD, Autosomal dominant inheritance; *CSF,* cerebrospinal fluid; *RBC,* red blood cell.

Box E36.5 Fatty acid oxidation disorders

Medium-chain acyl-CoA dehydrogenase deficiency (MCADD)	Incidence: 1 in 17 000; more common in people of north European descent Diagnosis and biochemistry: characteristic profile on AC or OA, genetics; newborn screening in the UK and a number of countries Presentation: in adults, first presentation with rhabdomyolysis, most often triggered by excess alcohol intake; paediatric presentation: hypoketotic hypoglycaemia, encephalopathy, liver dysfunction, cot death Management: avoid fasting, regular meals No specific dietary modifications needed other than avoiding foods that have very high medium-chain fat content; ER
Carnitine palmitoyltransferase 2 (CPT2) deficiency	Incidence: exact frequency not known (>300 cases reported in the literature) Presentation: as above + myopathy in paediatric presentation Diagnosis and biochemistry: characteristic profile on AC and OA, genetics Management: avoid fasting, regular meals No specific dietary modifications needed in milder adult presentation; more severe paediatric presentations may require restriction of long-chain fats in the diet; medium-chain fat supplements may be used to supplement calories; ER
VLCADD: very long-chain acyl-CoA dehydrogenase deficiency	Incidence: 1 in 40 000 Presentation: as in CPT2 + neuromuscular weakness, cardiomyopathy in paediatric presentation Diagnosis and biochemistry: characteristic profile on AC/OA, genetics Management: as for CPT2 deficiency, need cardiac surveillance and restriction of long-chain fatty acids depending on severity both in adult and paediatric presentations; medium-chain fat supplements may be used to supplement calories; ER

AC, Acylcarnitine profile; *ER*, emergency regimen; *OA*, organic acid profile.

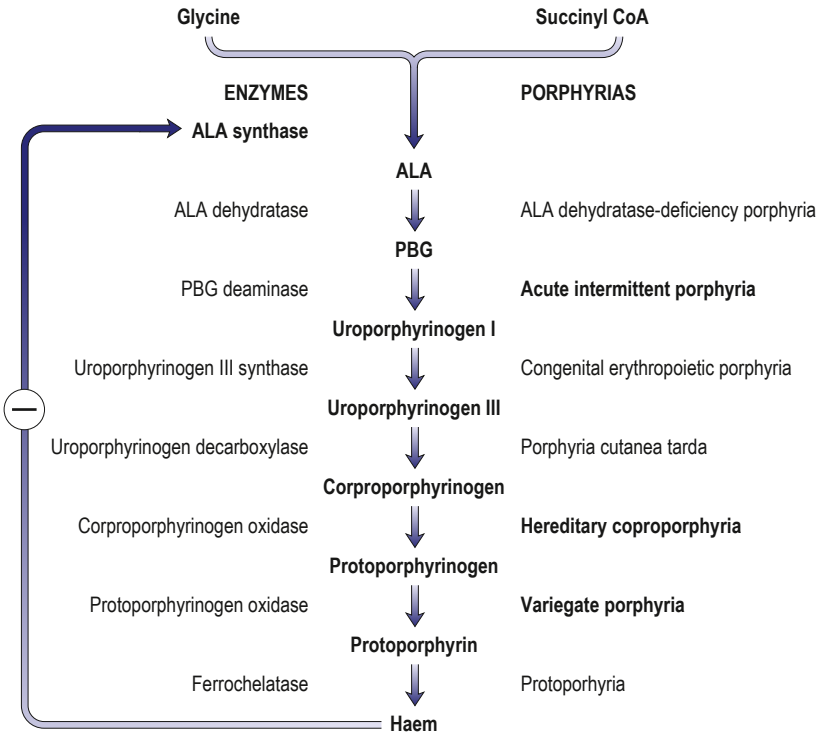


Fig. 36.3 Haem synthesis involves eight enzymes and occurs mainly in the bone marrow and liver. Inherited deficiencies of each of the enzymes result in distinct disease phenotypes collectively known as the porphyrias. Acute porphyrias include acute intermittent porphyria, variegate porphyria, hereditary coproporphyria, and delta-aminolevulinic acid (ALA) dehydratase deficiency porphyria. *PBG,* Porphobilinogen.

Peroxisomal diseases

Peroxisomes have a role in lipid metabolism, processing reactive oxygen species, and oxidation of specific biomolecules.

X-Adrenoleukodystrophy

Being X-linked, presentation is more severe in males, who present with the cerebral form. Symptoms usually begin between 4 and 10 years of age with behavioural and neuropsychiatric problems; cognitive decline; and variable progressive neurological decline, which is fatal. The presentation may be delayed into adulthood and even into the third or fourth decade of life. Patients also often develop adrenal insufficiency, which may precede development neurological symptoms. Although no treatment exists for symptomatic patients, presymptomatic patients may be amenable to treatment with bone marrow transplant or a clinical trial of gene therapy. Therefore, X-adrenoleukodystrophy (ALD) should be excluded in all patients with adrenal insufficiency in whom no other cause is found. Female carriers and adult males may also present with adrenomyeloneuropathy, resulting in spastic paraparesis of the lower limbs and bladder disturbances. Cerebral ALD and adrenal insufficiency almost never occur in female carriers. Adult physicians should exclude ALD in patients with suggestive symptoms because a positive diagnosis will initiate cascade testing to find any presymptomatic relatives.

Adult Refsum's disease

Adult Refsum's disease is a recessive disease that results in phytanic acid accumulation in tissues and plasma secondary to mutations in the *PHYH* or *PEX7* gene. The diagnosis should be excluded in patients presenting with retinitis pigmentosa, which (along with anosmia) is present in almost all patients with adult Refsum's disease. These two features often precede development of other features such as ataxia, peripheral neuropathy, and cataract by several years. One-third of patients may have bony abnormalities such as shortened fourth metatarsals. Cardiac arrhythmias may occur, particularly when phytanic acid levels are high. Treatment includes a phytanic acid–restricted diet and avoidance of fasting or rapid weight loss. Plasmapheresis may be needed at times of acute decompensation. Without treatment, there is gradual progressive deterioration. Elevated plasma phytanic acid levels with low pristanic acid levels are diagnostic.

Further reading

Gorman GS, Chinnery PF, DiMauro S et al. Mitochondrial diseases. *Nat Rev Dis Primers*. 2016; 2:16080.
Hollack CEM, Lachmann RH. *Inherited Metabolic Disease In Adults: A Clinical Guide*. Oxford: Oxford University Press; 2016.
Saudubray JM, Baumgartner MR, Walter J, Garcia-Cazorla A. *Inborn Metabolic Diseases: Diagnosis and Treatment*. 7th edn. Cham: Springer; 2022.
Zschocke J, Hoffmann GF. *Vademecum Metabolicum: Diagnosis and Treatment of Inborn Errors of Metabolism Paperback—13 Jan*. 2021.
www.bimdg.org.uk *British Inherited Metabolic Diseases Group*.

Questions for this chapter can be found on Elsevier eBooks+.

37 Obesity

Anjali Zalin, Ritwika Mallik, and Beenish Masood

Background 1248
Prevalence 1248
Aetiology 1248
Clinical approach to obesity 1251
Motivational interviewing 1252
History 1252
Examination 1252
Anthropometric measurements 1252
Body mass index and obesity diagnosis 1253
Investigations 1255
Complications of obesity 1257
Pathophysiology of obesity complications 1259
Management 1260
General principles of management 1260
Lifestyle interventions 1261
Medical therapy 1265
Bariatric surgery 1267

CORE SKILLS AND KNOWLEDGE

Obesity is a complex, chronic disease that adversely impacts the patient's quality of life and life expectancy. Significant contributors to death and disability in people living with obesity include cardiovascular disease, cancer, chronic kidney disease, diabetes mellitus, and musculoskeletal disorders.

Global obesity rates are rising sharply. There are many biopsychosocial drivers, including increased intake of energy-dense foods and sedentary time, in modern societies. Addressing these challenges requires highly skilled, empathic clinicians working in partnership with people living with obesity. On a broader societal level, effective policies across sectors such as urban planning, food marketing, education, and others are vital.

A substantial body of evidence demonstrates that effective obesity care can make a difference. Across a population, it has the potential to reduce the burden of diabetes, hypertension, cardiovascular disease, liver disease, and obstructive sleep apnoea (OSA). A variety of lifestyle, pharmacological, and surgical interventions have been shown to be effective in reducing weight. For individuals, communities, and the local health economy, the effects of good obesity care are transformative.

Key skills within obesity include:

- Developing a compassionate, non-judgemental consulting style; negotiating with patients about the best interventions; and using motivational interviewing techniques
- Understanding the physiology of weight gain and its effects on multiple body systems
- Working with multidisciplinary teams to address all aspects of obesity care
- Understanding the social determinants of health and developing leadership skills to advocate for appropriate public health measures and effective obesity services.

BACKGROUND

Prevalence

Obesity is a complex, chronic disease with multisystem impact. It is defined by the World Health Organization (WHO) as abnormal, excessive fat accumulation that presents a risk to health. Worldwide obesity rates have nearly tripled since 1975 and continue to rise. Currently, 2 billion adults are living with overweight, and of them, 650 million are living with obesity (body mass index [BMI] ≥30 kg/m^2).

Obesity does not affect a population equally. For example, adults in the most deprived regions have almost double the prevalence of obesity than those living in more affluent regions, a trend that begins in childhood in poorer populations. There are also sex disparities: females are twice more likely to gain weight and develop obesity-associated complications.

Obesity is responsible for 4.7 million premature deaths each year, and life expectancy decreases as BMI increases. Definitive and coordinated action is needed to address the obesity epidemic. A key step is to recognize obesity as a disease and work to challenge the stigma and bias experienced by patients.

Obesity was recognized as a disease as far back as 1948 in the WHO's International Classification of Disease, although this has not been consistently visible in clinical practice. The concept of obesity as a disease is now endorsed by many global organizations, supporting delivery of effective, responsive care for people living with obesity and working to reduce bias.

Aetiology

The aetiology of obesity is complex, and previous assertions including 'eat less, move more' have failed to address globally rising obesity rates. This simple narrative fails to explain why people respond

differently to obesogenic environments and respond differently to different interventions for obesity.

Current thinking acknowledges the principle of energy balance but also presents evidence for multiple other factors that promote and sustain obesity.

The landmark Foresight report published by the UK government in 2007 combined expertise from across society to map the complex factors contributing to obesity. This report comments that 'causes of obesity are extremely complex encompassing biology and behaviour, but set within a cultural, environmental and social framework' (**Box 37.1**).

Obesogenic environment

The obesogenic environment has strong sociocultural, economic, and policy drivers. Food environment includes the immediate eating environment (distractions, atmosphere, time of day) as well as the food itself (presentation, portion size, serving style).

People make an estimated 200 food choices daily, including food type and quantity. A host of neurobiological processes plays a role in this decision-making, which extend beyond simple 'willpower'. Most choices are subconscious and strongly responsive to environmental influences.

Persuasive food marketing plays a key role in promoting highly processed foods, typically with high sugar, high fat, and low fibre content. These foods are usually affordable and highly palatable. Many people also eat from widely distributed and easily accessible food outlets where food has higher caloric value and fewer micronutrients. These outlets often open for extended hours and may offer a food delivery service. This is further compounded by increasing sedentary behaviour linked to limited green spaces; unsafe, polluted public areas; and increased opportunities for working from home.

Biology: genetics and obesity

There are marked variations in how individuals respond to the obesogenic environment described. It is widely accepted that genetics and related biological differences have a powerful influence, interacting with environmental factors to result in the obesity phenotype. Twin studies demonstrate that the heritability of body weight (the percentage attributed to genetics) is similar to that of height.

Initially, candidate genes were identified in mice displaying severe obesity and hyperphagia. Reverse genetic techniques uncovered that mutations in the *ob* or obese gene (encoding for adipocyte-produced hormone leptin) resulted in severe obesity. Leptin is a hormone that provides the sensation of satiety and regulates appetite. This seminal body of work has broadened thinking and enabled the biological basis of obesity to gain acceptance.

Further research identified the central role of leptin-related pathways in food intake, reward, and preference. In addition to congenital leptin deficiency, mutations were also uncovered in genes encoding the leptin receptor (LEPR) and other components of the melanocortin pathway (**Fig. 37.1**). Genome-wide association studies have now highlighted more than 1000 other genes and hundreds of single nucleotide polymorphisms that affect feeding behaviour and contribute to obesity.

Epigenetics in obesity

Although crucial in our understanding of obesity, genetics does not fully explain the variance observed in the population. A solely genetic explanation also does not account for rising obesity rates over the timespan observed. Here, 'epigenetic modifications' have emerged as an important phenomenon. This refers to changes in gene expression mediated by mechanisms such as DNA methylation, histone modification, and non-coding RNA-associated gene silencing. Epigenetic modifications change how DNA is transcribed without altering the actual DNA sequence. These changes may be physiological but also arise in response to environmental stimuli.

An example of epigenetic change resulting in obesity is the hypomethylation of the agouti viable yellow *(Avy)* gene in mice. As a result, there are reduced methylation and aberrant expression of a protein binding to the melanocortin-4 receptor (MC4R) in the hypothalamus, a critical regulator of energy homeostasis and appetite. This results in hyperphagia (insatiable appetite). Feeding affected mice (*Avy* gene mutation) with food rich in folate and methionine (both acting as methyl donors for DNA) improved their body weight and insulin resistance.

Box 37.1 Summary of aetiological factors in obesity

Category	Aetiological factor (examples)
Biology	• Genetic predisposition • Level of satiety • Low resting metabolic rate • Ethnicity • Parental obesity
Early life	• Intrauterine environment (diet, weight, glucose tolerance) • Breast-feeding (reduces childhood obesity)
Individual activity	• Sedentary behaviour (e.g., television viewing) • Reduced outdoor activity in childhood • Sedentary occupation • Reductions in walking and cycling • Reduced fitness impairing ability to engage in exercise
Environmental factors	• Financial costs of physical exercise (e.g., gym membership fees) • Perceived danger in the environment • Walkability of the living environment • Cultural values associated with activity patterns • Lack of outdoor activities (reduced green spaces, parks, and playgrounds) • Endocrine disrupting chemicals: molecules present in food and cosmetics (e.g., bisphenol A and phyto-oestrogens) that interfere with the endogenous hormonal milieu and affect weight • Reduced diversity in the gut microbiome
Individual psychology	• Mixed feelings and beliefs about health lifestyle choices • Self-esteem • Stress, distractions • Level of food literacy • Level of parental control and family dynamics • Difficulty translating intention into action
Food consumption	• Increased variety and low cost of food • Surplus of energy-dense food • Larger portion sizes • Easy availability of low-cost ultra-processed foods
Food production	• Pressure for increased palatability of food • Low market prices of ultra-processed foods • Higher cost of ingredients for fresh cooking
Societal influences	• Social acceptability of weight • Importance of ideal body size image • Food education • Food advertising

Adapted from Tackling obesities: Future choices report.

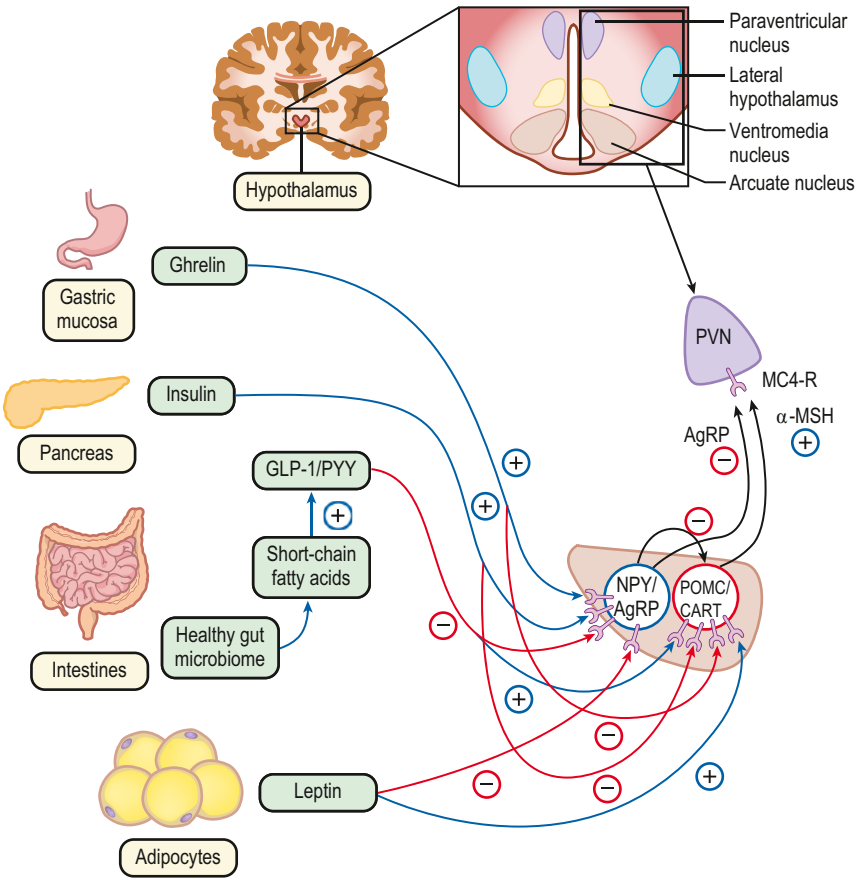


Fig. 37.1 Appetite is regulated by various peripheral peptides which act on nuclei in the hypothalamus. The main peripheral peptides which act on these neurons [neuropeptide Y (NPY)/agouti-related protein (AgRP)] and [pro-opiomelanocortin (POMC)/CART (cocaine and amphetamine regulated transcript)] to regulate appetite include: (a) ghrelin: secreted by the stomach, acts on GHS-R1a receptors on NPY/AgRP to stimulate appetite and inhibit POMC/CART neurons; (b) insulin: secreted by endocrine pancreas, acts on insulin receptors on NPY/AgRP and POMC/CART to inhibit appetite; (c) GLP-1 (glucagon like peptide 1): secreted by the distal ileum and colon, acts on GLP-1 receptors; delay gastric emptying, stimulate insulin and somatostatin secretion and inhibition glucagon secretion. They inhibit NPY/AgRP neurons and stimulate POMC/CART neurons to suppress appetite. A healthy gut microbiome produces more short chain fatty acids which increase secretion of GLP-1 and PYY. (d) PYY (peptide YY): secreted by the ileum, colon, and rectum, acts on Y2 receptors on NPY neurons to inhibit them and Y2 receptors on POMC/CART neurons to stimulate them and suppress appetite; (e) leptin: secreted by adipocytes, acts on leptin receptor (LEPR) on NPY/AgRP to inhibit them and on POMC/CART neurons to stimulate them thereby suppressing appetite. Disorders with leptin deficiency and leptin resistance leads to hyperphagia (insatiable appetite) and obesity. The hypothalamic nuclei involved in appetite regulation include the paraventricular, ventromedial, and arcuate nuclei. The arcuate nucleus consists of two main types of neurons mentioned above regulating food intake and energy expenditure. These include: (1) NPY/AgRP neurons which **increase appetite** (orexigenic) via agouti-related protein (AgRP). (2) POMC/CART neurons which **reduce appetite** (anorexigenic) via α-melanocyte stimulating hormone (α-MSH) and increase energy expenditure. Upon activation, NPY/AgRP neurons secrete AgRP which inhibit MC4-R receptors on the paraventricular nucleus (PVN) and increase appetite. In contrast, activated POMC/CART neurons produce α-MSH which stimulate MC4-R receptors on the PVN and reduces appetite. MC4-R mutations also lead to hyperphagia and obesity as with leptin deficiency and resistance. These pathways are actively regulated via conscious and subconscious signals from the brain as food intake results in activation of reward pathways in the brain. *(Diagram by Beenish Masood.)*

Epigenetic modifications are part of 'developmental programming' that occur *in utero* or during early life in response to an adverse environment (e.g., starvation or food scarcity). These adaptations prepare the fetus or child to survive in the presence of food scarcity later in life. If, however, these individuals grow up in a food-replete environment, these epigenetic modifications function as a catalyst for weight gain and promote obesity. Most epigenetic markers are erased during early development, although some of these markers persist and are passed on to offspring.

Physiological mediators of obesity

Energy balance is a key principle. There are two interrelated yet distinct mechanisms:

- *Continuous positive energy balance* in which energy intake exceeds energy expenditure
- *Reset of the body weight* to a higher setpoint so that a higher body weight is actively defended by the body itself through a variety of mechanisms

In humans, energy intake is from food and drink ingested. Energy expenditure is the sum of the energy expended for:

1. **Basal metabolic rate:** This is the energy required for sustaining basic functions of life at rest, which is determined by body size and lean body weight. This accounts for 50% to 70% of total energy expenditure and varies significantly between individuals with obesity despite similar weight. People with lower metabolic rates struggle to lose weight.

2. **Physical activity and digestion:** This is the most variable component of the daily expenditure, depending on patterns of behaviour.
3. **Nutrient absorption (diet-induced thermogenesis):** This is the amount of energy required to digest, absorb, and store ingested nutrients and is much higher for certain food constituents (e.g., proteins) than for others (e.g., dietary fats and carbohydrates). It accounts for 10% of total daily energy expenditure.

Fig. 37.1 presents a model of energy flow within the human body. The brain is the primary organ regulating our weight. It integrates external and internal signals to influence food intake and energy balance (subconsciously). The initial stimulatory input *(green arrow)* is from the environment and food intake. From there, hormones (including insulin) relay signals that have an impact on food intake. Weight gain causes ectopic lipid distribution within organs.

Control of appetite

The control of appetite is complex but elegantly orchestrated. Feeding behaviour and energy intake are driven by the metabolic need to eat to alleviate discomfort (drive reduction) or by food reward (fulfilment by eating).

Certain hypothalamic neurons are crucial in controlling feeding behaviour (Fig. 37.1). After a meal, gastric and duodenal distension and the release of many substances (peripherally and centrally) cause satiety and signal to stop eating.

These substances include cholecystokinin, bombesin, glucagon-like peptide 1 (GLP-1), enterostatin and somatostatin (released from the small intestine), and glucagon and insulin (from the pancreas). The lateral hypothalamic area, and the paraventricular and arcuate nuclei play a key role in integrating signals involved in appetite and body weight regulation. The control of appetite is illustrated in Fig. 37.1, and key peripheral hormones are described below.

Peptide YY (PYY)* is a rapid and short-acting signal produced by the L cells of the large bowel and distal small bowel shortly after food intake. It is released in proportion to the energy ingested. PYY reduces appetite through inhibition of the appetite-stimulating pathway (NPY/AgRP-expressing neurones). There are numerous other peripheral appetite-suppressing signals, including GLP-1 and oxyntomodulin, which, like PYY, are produced by the gut in a nutrient-dependent manner.

Ghrelin* is a 28-amino-acetylated peptide produced by the oxyntic cells of the fundus of the stomach that activates the central appetite-stimulating pathway. Levels are high before a meal and reduce rapidly after ingestion of a meal or glucose (in contrast to PYY, which increases after a meal). It also acts as a long-term signal because its circulating concentration in weight-stable individuals is inversely related to BMI. Ghrelin is also increased in several situations in which there is a negative energy balance, such as long-term exercise, very-low-calorie diets, anorexia nervosa, and both cancer and cardiac cachexia. (An exception is vertical banded gastric bypass surgery, in which its concentration is low rather than high.) Obestatin is a peptide produced by the same gene that encodes ghrelin and counteracts the increase in food intake induced by ghrelin.

The endocannabinoid system is also involved in the control of food intake and energy balance. There are two receptors, endocannabinoid (CB1) in the brain and CB2 in the periphery. CB1 receptors are in the cerebral cortex, cerebellum, and hippocampus.

Tumor necrosis factor (TNF) and interleukin-2 are cytokines that are elevated in inflammatory and traumatic conditions and also suppress appetite; exact pathways are unclear. Appetite inhibitors include dopamine, serotonin, and γ-aminobutyric acid, and appetite stimulators include opioids.

The gut–brain axis

The gastrointestinal (GI) tract and the central nervous system communicate via immune, neural, and endocrine pathways. Relevant messengers include bacterial metabolites, gut hormones, and neuroactive substances.

The gut microbiome impacts nutrient metabolism and energy expenditure. Its role in obesity is via its effects on the permeability of the GI mucosa and fermentation and absorption of polysaccharides. Animal studies report that the gut microbiome in obesity is less diverse than in lean individuals and impacts appetite and energy consumption.

Hedonic hunger

This is a nonhomeostatic hunger caused by excess hedonic drive (to feel better) rather than caused by a true metabolic need (as in starvation). Thus, individuals eat without being hungry despite large fat reserves. Neuroimaging revealed that when fasting humans viewed pictures of high-calorie and low-calorie foods, the former increased neural activity in the orbitofrontal cortex, ventral striatum, amygdala, and anterior insula (i.e., brain reward areas). Hence, specific foods activate brain reward and determine food preference. Eating is stimulated by a metabolic need, hedonic hunger, or a combination of the two (Fig. 37.2).

Further reading

Bray GA, Kim KK, Wilding JPH; World Obesity Federation. Obesity: a chronic relapsing progressive disease process. A position statement of the World Obesity Federation. *Obes Rev.* 2017;18(7):715–723.
Loos RJF, Yeo GSH. The genetics of obesity: from discovery to biology. *Nat Rev Genet.* 2022;23(2):120–133.
Schwartz MW, Seeley RJ, Zeltser LM, Drewnowski A, Ravussin E, Redman LM, Leibel RL. Obesity pathogenesis: an endocrine society scientific statement. *Endocr Rev.* 2017;38(4):267–296.
https://www.worldobesity.org/resources/resource-library/world-esity-atlas-2023. *World Obesity Atlas*.
https://www.gov.uk/government/publications/reducing-obesity-obesity-system-map. *Tackling Obesity: Future Choices, Foresight report*.

CLINICAL APPROACH TO OBESITY

The clinical assessment of obesity should focus on establishing *why* a person has obesity and *how* it is affecting them and lay the foundations to consider the *best treatment plan*. Assessment should be done is a sensitive and non-judgemental way and undertaken when the individual is ready (Box 37.2).

The 'making every contact count' approached promoted in the UK encourages all clinicians to take opportunities presented in every clinical encounter to prompt patients towards positive health behaviours. There may be opportunities in an obesity clinic to suggest non–weight-related lifestyle measures (e.g., smoking cessation). Clinicians in other settings should look for opportunities to talk about weight reduction in people living with overweight or obesity in an appropriate way even if this is not the main focus of the clinical encounter.

Clinicians should be mindful that many people will have struggled with their weight for years and encountered stigma and bias both in society and within healthcare settings. It is not uncommon to hear that people have relayed negative messages to themselves internally. In this context, **language matters** and can have a significant impact on how an individual reacts to the discussion and

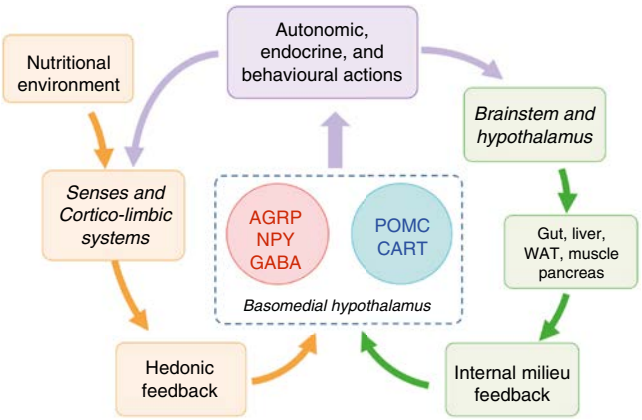


Fig. 37.2 Internal and external appetite regulation. Agouti-related protein (AgRP), neuropeptide Y (NPY), and gamma-aminobutyric acid (GABA) neurons (hunger neurons) integrate metabolic and hedonic signals and translate them, so a person feels hungry (similar to the accelerator of a car). Proopiomelanocortin (POMC) and cocaine- and amphetamine-regulated transcript (CART) neurons control this hunger, like a brake pedal. *WAT,* White adipose tissue. (*From Blaming the brain for obesity: Integration of hedonic and homeostatic mechanisms*.)

Box 37.2 Principals of caring for patient with obesity

In caring for people living with obesity, it is critical to recognize:

- The person living with obesity is the expert in their condition and best understands personal barriers and potential solutions to treatment.
- Stigma impacts quality of life and reduces engagement.
- One of the most important roles of an obesity specialist is to create an empathic environment where a patient's story can be heard without judgement.
- A deeper understanding of an individual's obesity phenotype supports personalized care.
- Multidisciplinary working is key to effective care delivery.

advice provided. To build rapport and minimize stigma, it is especially important to use clear, person-first language and adopt a neutral, non-judgemental tone (**Box 37.3**).

Motivational interviewing

Motivational interviewing (MI) is a valuable tool to support empowerment and elicit an individual's intrinsic motivation and is a key approach developed to help encourage behaviour change and new habit forming. People often have a strong desire to change, but complex biopsychosocial factors can impair their progress. In this context, directive advice, 'education', and messages such as 'eat less and move more' are largely unhelpful.

MI uses a variety of evidence-based strategies to develop an empathic environment and encourage rapport between the patient and clinician. Within this space, individuals are guided to harness their own intrinsic motivation and drive change to support improvement. The ethos of MI or 'spirit' recognizes individual autonomy and expertise while emphasizing strengths rather than deficits. Advice is offered but not imposed. Clinical staff require initial training as well as ongoing coaching and supervision in the skills and strategies of MI to increase their competency in the method. The 'OARS' nemonic summarizes the principles and basic skills for MI (**Box 37.4**).

History

A comprehensive medical history for a person living with obesity is a key step. It enables a deeper understanding and supports the development of a personalized treatment strategy, aligning with personal goals. A comprehensive obesity-focused history is described in **Box 37.5**.

This may begin by trying to understand changes in a patient's weight throughout their life and in particular in relation to different life events and previous efforts to lose weight (**Fig. 37.3**). Medication may be a key driver of weight gain and may inhibit attempts to lose weight (**Box 37.6**).

Validated questionnaires may be used to complement the history. Examples include the Binge Eating Scale or the Food Frequency questionnaire for assessing eating habits, the generalized anxiety disorder screening (GAD-7) or patient health questionnaire-9 (PHQ-9) scores for assessing mood, and the Epworth and STOP-BANG scores for assessing sleep hygiene and risk of OSA, see Chapter 20 ('Obstructive sleep apnoea' section).

Examination

Physical examination of a person living with obesity has four main aims:

1. Identify obesity-related co-morbidities.
2. Consider the impact of obesity on existing co-morbidities.
3. Consider secondary causes of obesity.
4. Obesity phenotyping and profiling.

To do this effectively, it is important to ensure the dignity and comfort of the individual. There should be privacy and space and specialist settings equipped with appropriately sized chairs and weighing and monitoring equipment. For example, larger cuffs should be available to measure blood pressure (BP) because standard cuffs may overestimate the BP of a person with obesity. Key aspects of the physical examination in a patient with obesity are included in **Fig. 37.4**.

Anthropometric measurements

Waist-to-height ratio

Central fat distribution increases morbidity and mortality risk. Measuring waist-to-height ratio (**Fig. 37.5**) can identify individuals at greater risk.

Skinfold measurements

This is useful for individuals who may be housebound or unable to step onto scales. It is measured with callipers over the middle of the

Box 37.3 Examples of preferred language and language to avoid when consulting with people living with obesity

Seeking to be more…	Examples
Empathic	It sounds as though your weight is really hard to manage at the moment.
Empowering and inclusive	What changes do you believe are needed right now?
Respectful	I appreciate you being open to talk about this.
Trust building	I will have a conversation with the team and feed back to you.
Person centred	What thoughts have you had yourself about your eating?
Encouraging and positive	I can see the effort you are putting in; keep up the great work!
Clear	Yes, your weight has not come down yet.
Reassuring	It is not unusual for people not to lose weight immediately; it's hard work that your body is fighting.
Understanding	Now doesn't sound the best time to be concentrating on your weight.
Exploring	What makes you say, 'I feel like a failure'?
Environmentally aware	Make sure that your clinic and waiting room are appropriate for people who are of a larger size.
Collaborative	Let's talk together about the different options and then see what you think would suit you best.
Helpful and supportive	Offer specific help when appropriate, such as a referral, or a dedicated appointment to set goals and make plans.
Congruent words and behaviours	Looking at the person when welcoming or asking questions.
Culturally competent	Explore individuals' cultural, religious or faith, and spiritual beliefs.
Seeking to be less…	**Examples**
Stigmatizing	You're in denial.
Shaming or blaming	It's being so overweight that is causing you to have all these problems.
Authoritarian	You must not eat any of these foods.
Demanding	Before you come to see me, I want you to complete this food chart so I can check what's going wrong.
Disapproving	You aren't meant to eat that.
Discriminating	I don't think they'd get much from the exercise class.
Stereotyping	People from this background often prefer high-fat foods.
Assumptive	I think you just need to go for an operation; you won't cope with the other options.
Prejudging	No one in that family has ever taken much notice of their weight; they will be the same.
Judgemental	That's the wrong choice; it'll never work.
Threatening	If you don't lose weight, you will end up with your leg chopped off or just plain dead

Obesity UK. Summary of language matters guide.

Box 37.4 The OARS approach in motivational interviewing

O: open-ended questions	• How do you feel about your health right now? • How would you like things to be different?
A: affirmations	• You handled yourself really well in that situation. • I know how challenging this is, and you have made a lot of positive changes already.
R: reflections	• I see that you understand a lot about how weight is related to your health, and you really want to be in more control. • You're wondering if…
S: summarizing	• Here is what I've heard. Tell me if I've missed anything. • Let me see if I understand so far. I am wondering what you think your next step should be.

triceps muscle. Normal values are less than 20 mm in male adults and less than 30 mm in female adults.

Neck circumference

This is measured with a standard tape measure by starting at the Adam's apple and then bringing the tape around to the starting point (with care taken not to pull too tight) (Fig. 37.5). More than 43 cm in male adults and 41 cm in female adults indicates a high risk for sleep apnoea.

Bioelectrical impedance analysis

Some centres have access to bioelectrical impedance analysis technology, which can provide information about body composition (i.e., body fat mass, segmental body fat, visceral fat, muscle mass, total body water, basal metabolic rate, and metabolic age). These deeper insights into a body type can help guide conversations and management. Serial measurements are useful for motivation.

Body mass index and obesity diagnosis

BMI defines and classifies obesity severity (Box 37.7). It is calculated by dividing an adult's weight in kilograms by their height in metres squared. A lower cut-off is used for certain ethnicities to reflect higher visceral adiposity and increased cardiometabolic risk.

BMI has many advantages on the population level: it is inexpensive and simple to measure (without highly specialized equipment and training). However, at an individual level, it may not always provide an accurate picture or fully define risk. For example, because muscle is denser than fat, muscular individuals may be classified by BMI as 'obese'.

BMI also does not account for gender, age, or fat distribution and does not necessarily correlate with health risk. Certain individuals may be of an obese range BMI but considered 'metabolically healthy', but others may develop type 2 diabetes mellitus at a non-obese range of BMI. In case of the latter, a popular hypothesis is that everyone has a 'personal fat threshold' that determines their personal capacity to accumulate fat. After this threshold, they deposit visceral fat (e.g., liver, pancreas). This then triggers insulin resistance, type 2 diabetes, and other complications.

In view of these complexities and to obtain a more personalized profile of an individual living with obesity, measurements such as body composition and waist circumference are useful

Box 37.5 Components of an obesity-focused medical history

Age	• Consider physiological age and frailty • Digital literacy (relevant to care delivery) may vary with age • Specific factors relevant to young adults (see 'Obesity through the life course' section)
Sex	• Sex differences in obesity prevalence and risk • Sex-specific variations relating to obesity such as bone health, reproductive health, and body composition • Sex specific co-morbidities (e.g., polycystic ovarian syndrome, erectile dysfunction) • Transgender considerations (with specialist support)
Ethnicity	• Ethnic-specific criteria exist for access to certain interventions • Perspectives on weight management and dietary changes may also vary with ethnicity
Occupation	• Employed people spend one-third of their day in the workplace • Working environment influences sedentary time, food environment, and stress levels • Insufficient sleep and circadian misalignment secondary to shift work contribute to weight gain
Weight history	• Use open questioning and active listening to identify the reason why someone is seeking help • BMI is used to classify severity, but this may not be the main reason for a person to seek help • Assess the person's view of their weight and possible reasons for weight gain • Ask about weight at different stages of life (including birth weight) • Ask about minimum and maximum weight through life • Identify life events that may have resulted in weight changes; a visual representation (e.g., lifestyle events body weight graph) can support reflection and insights (see **Fig. 37.6**) • Identify previous attempts at weight loss and modalities used, as well as outcomes and learning from these experiences • Type and perception of hunger • Describe weight pattern (e.g., rapid onset versus gradual gain over years); rapid onset of weight gain points to a secondary cause • Adaptive behaviours relating to weight, e.g., excessive exercise, insulin misuse, or purging (vomiting or laxative use); it is a sensitive area, and clinicians should obtain relevant training and gain local awareness of support agencies for signposting (e.g., eating disorder unit) • Consider pointers toward monogenic genetic obesity: early-onset severe obesity, hyperphagia, learning disability, syndromic features, central endocrine abnormalities, increased severity; resistance to treatment
Past medical history	• Enquire specifically about obesity-related co-morbidities including diabetes, heart disease, OSA, renal impairment, and hypertension • For mental health diagnoses and eating disorders, clarify treatment modality and stability; restrictive diets can lead to a flare-up of binge-eating disorder • For diabetes mellitus, ask about duration of disease, glycaemic control, and related complications; diabetes onset <5 years has higher likelihood of remission with significant weight loss; ensure their diabetes treatment supports weight loss (see **Box 37.6**) • Enquire about suitability for bariatric surgery: previous general anaesthetic, reflux, hiatus hernia, previous abdominal surgery • Consider diagnoses that hinder weight loss (e.g., untreated OSA, which can lead to stress, fatigue and reduced exercise tolerance) • Ask about secondary causes for obesity such as Cushing's syndrome, hypothyroidism, or hypothalamic disorder or injury (this may be accompanied by hyperphagia)
Drug history	• Identify previous obesity pharmacotherapy and outcomes • Enquire specifically about drugs causing weight gain (**Box 37.6**) and corticosteroids
Family history	• Review if weight consistent with family trends
Social history	• Physical activity: type of occupation, day-to-day mobility, sedentary time, barriers to mobility (physical, psychosocial), structured activity (e.g., exercise classes, swimming) and level of exertion, experience and interest in digital or wearable technology, self-perceived level of fitness • Diet history, including food diary, speed of eating and eating environment, autonomy in meal preparation, and external influences • Sleep hygiene and quality • Functional status: see staging above • Economic status: deprivation is strongly associated with obesity, and food poverty or cost of living challenges can have an impact on food choices and quality • Smoking history • Alcohol intake (remember the calorie intake) • Support network • Who in the household prepares meals and will support with weight loss attempts? • Identify social isolation, which maybe a barrier to change

BMI, Body mass index; *OSA*, obstructive sleep apnoea.

to measure in addition to BMI. Adiposity-based chronic disease (ABCD), a newer diagnostic term, has also been proposed (American Association of Clinical Endocrinology guidance). This goes beyond simple BMI measurements and identifies obesity as a chronic disease, outlines its pathophysiology, and reduces stigma.

Staging of obesity

Because of the limitations of a simple BMI-based approach, newer staging systems have been developed that include an assessment of the impact of obesity on multiple areas of function and lifestyle (**Fig. 37.6**). These staging systems have been shown to correlate with mortality rate, as well as quality of life and other health indicators

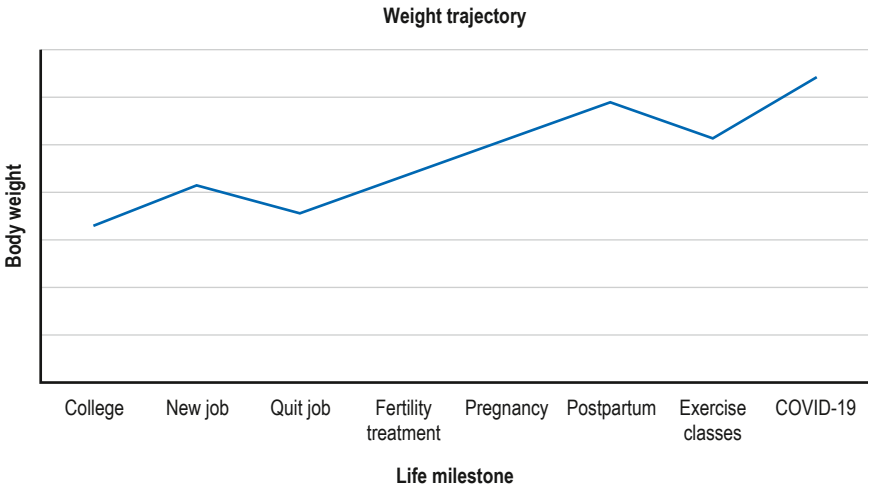


Fig. 37.3 **Lifestyle events: body weight graph.**

Box 37.6 **Common prescriptions and their impacts on weight**

Drug class	Weight loss	Weight neutral	Weight gain
Steroids			+ (e.g., prednisolone)
Hormonal contraceptives			+ (e.g., OCPs, DMPA)
Antihistamines			+ (e.g., cetirizine, fexofenadine, desloratadine)
β2-blockers			+ (e.g., atenolol, metoprolol)
α2-agonists			+ (e.g., clonidine)
Antidepressants		+ (e.g., escitalopram, duloxetine)	+ (e.g., amitriptyline, citalopram, fluoxetine, sertraline)
Atypical antipsychotics		+ (e.g., aripiprazole)	+ (e.g., olanzapine, clozapine, risperidone, quetiapine)
Anticonvulsants	+ (e.g., topiramate, zonisamide)	+ (e.g., levetiracetam, lamotrigine)	+ (e.g., valproate, carbamazepine, pregabalin)
Mood stabilizers			+ (e.g., lithium)
Antidiabetic medications			
Insulin			+
Biguanides	+		
SGLT2 inhibitors	+		
GLP-1RA	+		
DPP-4 inhibitors		+	
Sulfonylureas			+
Thiazolidinediones			+
Meglitinides	+		
α-glucosidase inhibitors		+	

DMPA, Depot-medroxyprogesterone acetate; *DPP-4*, dipeptidyl peptidase 4; *GLP-1 RA*, glucagon-like peptide 1 receptor agonist; *OCP*, oral contraceptive pill; *SGLT2*, sodium-glucose cotransporter-2.

and can be used to discuss with patients about weight reduction strategies.

Investigations

Investigations should consider:

- Underlying or secondary causes of obesity
- Severity of obesity and staging
- Associated co-morbidities of obesity
- Cardiovascular risk and metabolic syndrome
- Nutritional status
- Obesity profiling to guide future therapeutic options.

It is not uncommon to uncover silent conditions that may have a significant impact on the patient's health. These conditions may have unrecognized symptoms or be asymptomatic. Examples include liver disease, atrial fibrillation, OSA, and renal impairment. The investigation phase therefore presents an opportunity to 'make every contact count' and develop a clearer picture of the individual and their risk.

As with investigations in all areas of medicine, there should be careful thought. Excessive or unwarranted investigation carries the risk of unnecessary anxiety; expense; and in some cases, harm from radiation exposure. It is recommended that the clinical history and examination guide specialized investigations. Centres may

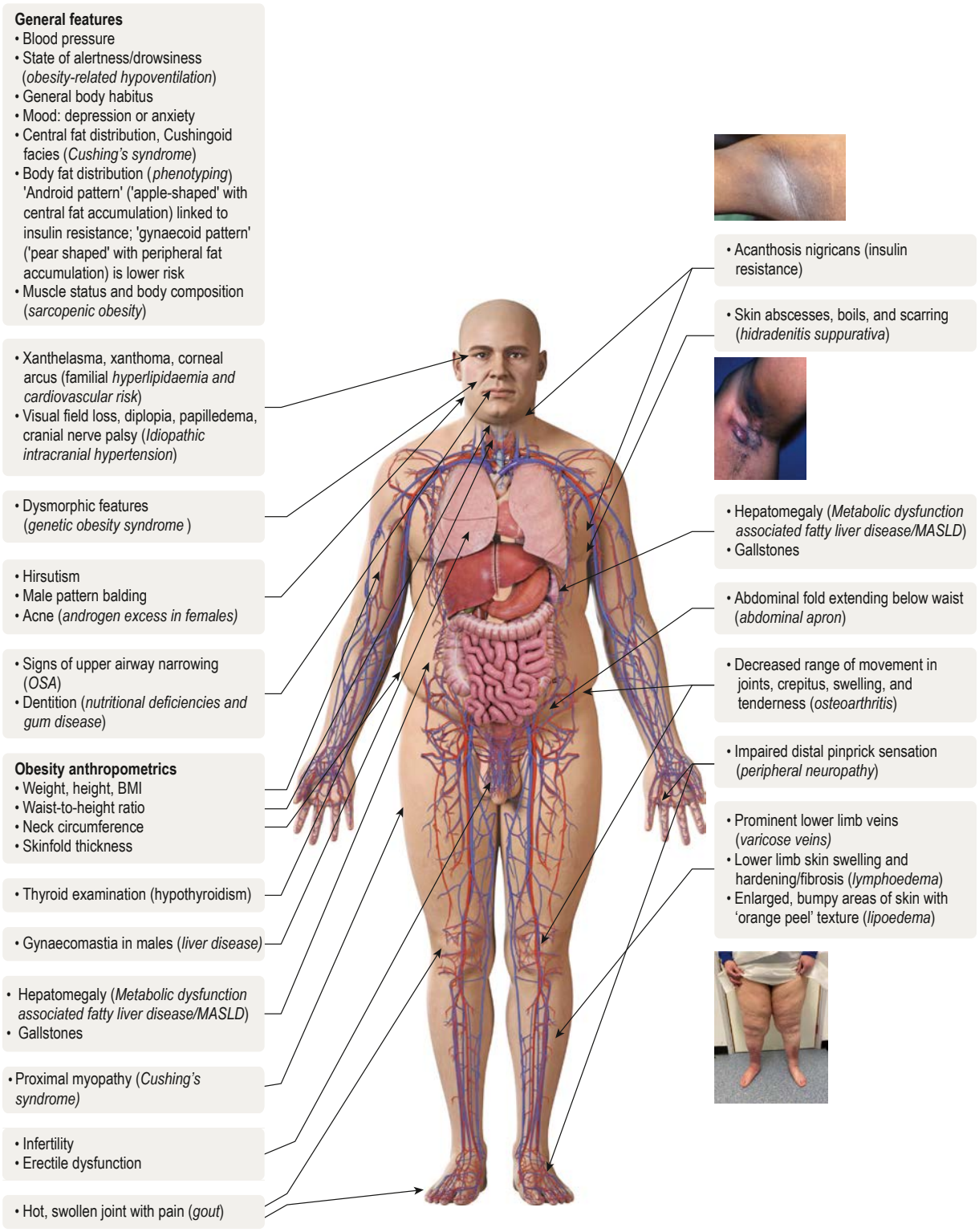


Fig. 37.4 Features to consider during examination of a person living with obesity. *BMI,* Body mass index; *MASLD,* metabolic dysfunction–associated steatotic liver disease; *OSA,* obstructive sleep apnoea.

adopt a 'standard panel' (**Box 37.8**) that is agreed on locally to reflect the local population and health service.

Specialist services should anticipate technical challenges and be well prepared with equipment such as larger scanners and beds. If this is not considered, it can lead to significant distress and embarrassment for both parties. Limitations of certain investigations (e.g., transthoracic echocardiogram that report 'limited views due to body habitus') should be addressed sensitively. It should also be anticipated that certain investigations may uncover unexpected or new diagnoses as well as create additional referral needs. Specialist pathways (e.g., to respiratory services) may need to be discussed and agreed.

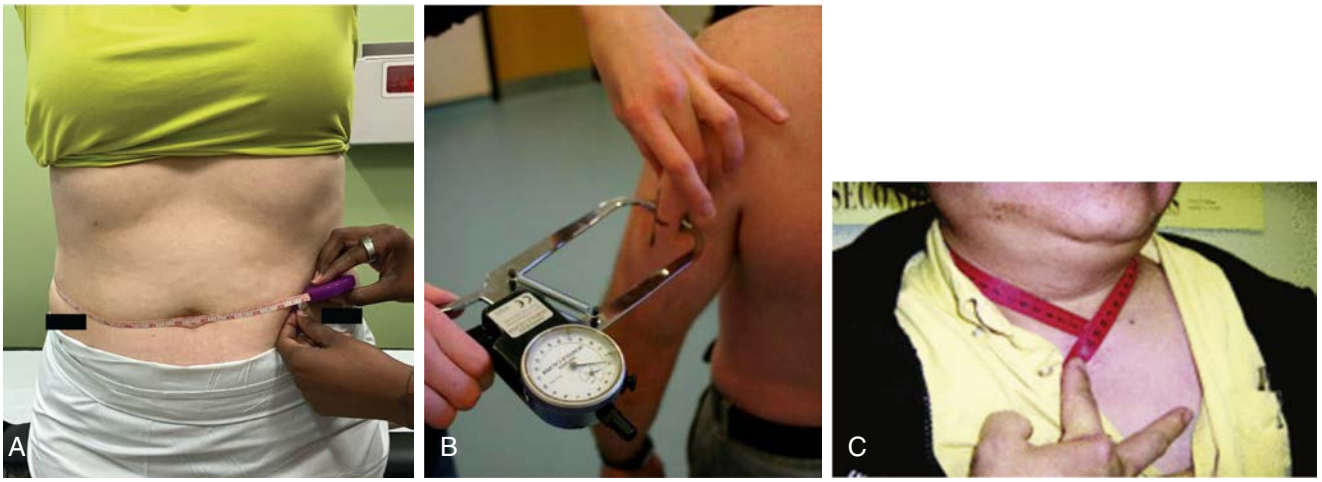


Fig. 37.5 Anthropometric measurement in obesity. (A) Waist-to-height ratio: To measure waist circumference, find a point midway between the lowest point of the ribcage and the iliac crest. Place a tape measure around the individual at this level, passing over the navel. Ask the patient to breath in and out, and measure at the end of expiration. **(B)** Skinfold: This is useful for individuals who may be housebound or unable to step onto scales. It is measured with calipers over the middle of the triceps muscle. Normal values are <20 mm in males and <30 mm in females. **(C)** Neck circumference: This is measured with a standard tape measure by starting at the Adam's apple and then bringing the tape around to the starting point (take care not to pull too tight). More than 43 cm in males and 41 cm in females indicates high risk for sleep apnoea.

Box 37.7 Classification using body mass index, including ethnic specific criteria

	General population BMI criteria (kg/m²)	Ethnicity-specific BMI criteria[a] (kg/m²)
Healthy weight	18.5–24.9	18.5–22.9
Overweight	25–29.9	23–27.4
Obesity class 1	30–34.9	27.5–32.4
Obesity class 2	35–39.9	32.5–37.4
Obesity class 3	>40	>37.5

[a]South Asian, Chinese, other Asian, Middle Eastern, Black African and African-Caribbean.
BMI, Body mass index.
(NICE guidance, CG189.)

Furthermore, specialist investigations maybe indicated depending on the clinical evaluation and proposed treatment plan. These may be considered when there is suspicion of a secondary cause of obesity (**Box 37.9**) or when additional tests are required to determine eligibility for specific treatment modalities (e.g., bariatric surgery) (**Box 37.10**).

Genetic investigations

Heritability is a feature in many patients with obesity; however, in some, the pattern of genetic association is the dominant aetiological factor. Features suggesting a genetic aetiology include early-onset obesity (aged 3–5 years); hyperphagia (extreme hunger); learning disabilities; hypothyroidism; difficulty losing weight; and, in some cases, weight regain after bariatric surgery. There is increasing awareness of monogenic obesity with the development of the novel licensed treatment setmelanotide, an MC4R agonist. Genetic testing is now accessible and performed in selected candidates. Testing requires training not only to take samples but also in genetic counselling and interpretation of the test results. It is best practice to link in with a centre of expertise and work with relevant specialists (e.g., learning disability teams, rare disease support groups).

Further reading

Aasheim ET, Aylwin SJ, Radhakrishnan ST et al. Assessment of obesity beyond body mass index to determine benefit of treatment. *Clinical Obes*. 2011;1(2–3):77–84.
Sharma AM. M, M, M & M: a mnemonic for assessing obesity. *Obes Rev*. 2010;11(11):808–809.
Vallis M, Piccinini-Vallis H, Sharma AM, Freedhoff Y. Clinical review: modified 5 As: minimal intervention for obesity counseling in primary care. *Can Fam Physician*. 2013;59(1):27–31.
https://www.nice.org.uk/guidance/cg189. *Obesity: identification, assessment and management. CG189.*
https://www.worldobesity.org/training-and-events/scope/e-learning. *SCOPE learning*.

COMPLICATIONS OF OBESITY

Obesity is a multisystem disease, and a full obesity assessment includes consideration of related conditions in multiple different organ systems.

Cardiovascular disease

Cardiovascular diseases are the main cause of death in people with obesity. Deposition of lipid-laden macrophages in vessel walls is accelerated by obesity and mediated via increased insulin resistance and inflammation. This promotes atherosclerosis, ***coronary artery disease, and structural heart changes (via fibrosis)***. Higher adiposity also results in increased total blood volume and increased cardiac output. This in turn results in ventricular hypertrophy and arrhythmias. Accompanying hypertension, dyslipidaemia and hyperglycaemia (metabolic syndrome) contribute to a higher prevalence of ***heart failure with preserved ejection***

	Stage 0	Stage 1	Stage 2	Stage 3
	'Normal health'	*'At risk of disease'*	*'Established disease'*	*'Advanced disease'*
Airways	Normal	Snoring	Requires CPAP	Cor pulmonale
Body mass index	<35	35–40	40–60	>60
Cardiovascular	<10% risk	10%–20% risk	Heart disease	Heart failure
Diabetes	Normal	Impaired fasting glycaemia	Type 2 diabetes	Uncontrolled type 2 diabetes
Economic	Normal	Expensive travel/clothes	Workplace discrimination	Unemployed due to obesity
Functional	Can manage three flights of stairs	Manages one or two flights of stairs	Requires walking aids or wheel chair	Housebound
Gonadal	Normal	PCOS	Infertility	Sexual dysfunction
Health perceived	Normal	Low mood or QoL	Depression or poor QoL	Severe depression
Body image	Normal	Dislikes body	Body image dysphoria	Eating disorder with purging (drug use or vomiting)

Fig. 37.6 Modified King's obesity staging criteria. *CPAP,* Continuous positive airway pressure; *PCOS,* polycystic ovary syndrome; *QoL,* quality of life.

Box 37.8 A typical panel of diagnostic tests for use in an obesity service

Standard obesity panel example:

- Full blood count
- Urea, creatinine, and electrolytes
- Glycated haemoglobin or HbA_{1c},
- Lipid profile
- Liver biochemistry (with calculation of a 'fibrosis-4' index based on age, ALT, AST, and platelet count to assess the likelihood of MASLD)
- Thyroid function tests
- Ferritin folate, vitamin B_{12}, vitamin D
- Urinalysis for protein

ALT, Alanine transaminase; *AST,* aspartate transaminase; *MASLD,* metabolic dysfunction–associated steatotic liver disease.

fraction. Any of these cardiac conditions in isolation or combination may lead to sudden cardiac death. Other complications include ***obesity-associated cardiomyopathy, left ventricular hypertrophy, cor pulmonale, pulmonary hypertension of obesity, and lymphoedema***.

Respiratory disease

Anatomic and neuromuscular changes in obesity are key factors in sleep-related breathing disorders. ***OSA*** (Chapter 20, 'Obstructive sleep apnoea' section) is caused by progressive expiratory narrowing in the retropharyngeal area, which results in a negative collapsing pressure during inspiration. ***Obesity hypoventilation syndrome*** is associated with chronic type 2 respiratory failure caused by diminished respiratory drive: maximal inspiratory and expiratory pressures are reduced, and there are reduced ventilation in the lower pulmonary lobes and hypercapnia. Possible mechanisms for this are leptin resistance, genetic predisposition, and sleep-disordered breathing. There is also a higher incidence of respiratory infections, asthma exacerbations, and pulmonary emboli in people living with obesity.

A striking association was seen between obesity and ***COVID-19–related death*** during the pandemic. Large-scale studies were conducted in England, the USA, and other countries revealing that people with obesity were more severely affected by COVID-19. One study showed that 77% of people hospitalized were either overweight (29%) or obese (48%). This is likely due to immune dysregulation and the low-grade inflammation of obesity predisposing to severe infection.

Box 37.9 Secondary causes of obesity

Obesity is a prominent feature of the following conditions and syndromes:

Polycystic ovarian syndrome: weight gain, irregular periods, hirsutism, subfertility

Hypothyroidism: tiredness, hair loss, feeling cold, constipation, dyslipidaemia, bradycardia, mood changes, dry skin and hair, menstrual irregularity

Cushing's syndrome: proximal muscle weakness, wasting of the extremities, wide purplish striae, easy bruising, supraclavicular fat; hypertension and glucose intolerance may also be present (high suspicion, sensitivity)

Prader-Willi syndrome: infants and toddlers present with hypotonia and failure to thrive; older children and adults manifest hyperphagia, intellectual disability, behavioural concerns, short stature, and hypogonadism

Bardet-Beidl syndrome: intellectual disability, retinitis pigmentosa, polydactyly, hypogonadism, deafness, and renal disease

Albright hereditary osteodystrophy (pseudohypoparathyroidism type 1a): short stature, short metacarpals and metatarsals, hypocalcaemia, precocious puberty, cognitive deficit

Alström syndrome: blindness, deafness, acanthosis nigricans, type 2 diabetes mellitus, cirrhosis, hypogonadism, renal disease

Hypothalamic obesity: abnormal weight gain caused by structural destruction of the hypothalamus characterized by rapid weight gain, hyperphagia, decreased energy expenditure, hypopituitarism, reduced quality of life

Liver disease

Metabolic dysfunction–associated steatotic liver disease (MASLD) is a newer term for non-alcoholic fatty liver disease. It is particularly prevalent in people with obesity. MASLD is related to chronic caloric excess, leading to fat accumulation in liver, especially of triglycerides. This is now one of the leading causes of cirrhosis in developed societies (Chapter 23, 'Metabolic dysfunction–associated steatotic liver disease' section).

Gastrointestinal disease

Obesity is linked to various disorders, including reflux disease, erosive oesophagitis, Barrett's esophagitis, oesophageal adenocarcinoma, erosive gastritis, gastric cancer, diarrhoea, colonic diverticular disease, colonic polyps and cancer, cholelithiasis, cholecystitis, acute pancreatitis, and pancreatic cancer. Aetiological factors include altered peristalsis, changes in sphincter pressures, chronic intestinal inflammation, increased colon transit time, increased permeability of the GI mucosa, changes in GI hormone

Box 37.10 Specialist tests considered for the investigation of obesity as directed by clinical assessment

Specialist tests considered for the investigation of obesity as directed by clinical assessment
Testosterone
Luteinizing hormone
Follicle-stimulating hormone
Sex hormone–binding globulin
B-type natriuretic peptide (false-negative results are more frequent in patients with obesity)
Investigations for Cushing's syndrome
Genetic obesity screening (next-generation sequencing)
Spirometry
Polysomnography or overnight sleep oximetry
Venous bicarbonate
Continuous positive airway pressure compliance report
Liver elastography
Electrocardiography
Echocardiography
Ultrasonography of the abdomen
Upper gastrointestinal endoscopy
Transvaginal ultrasonography
Dual-energy X-ray absorptiometry (bone density scan)

secretion (increased resistin and visfatin, reduced leptin), and altered bile acid absorption.

Endocrine disease

Obesity-associated endocrine dysfunction is linked to pro-inflammatory adipokines secreted by adipose tissue. Disorders include ***insulin resistance, prediabetes, type 2 diabetes mellitus, dyslipidaemia, hypothyroidism, and hypercortisolemia***.

In people with central obesity, there is ***functional hypercortisolism***. 11-β steroid dehydrogenase type 1 present in adipose tissue converts cortisone to cortisol. Therefore, an increase in adipose tissue results in increase in cortisol. This is accompanied by increased nocturnal Adrenocorticotropic hormone and cortisol levels, increased urinary free cortisol, and adrenal gland enlargement. The type of disturbance in the hypothalamic–pituitary–adrenal axis depends on adipose tissue distribution and individual stress response.

Reproductive system disease

Reproductive dysfunction occurs in both males and females living with obesity as a result of changes in sex hormone metabolism. Adipose tissue has increased aromatise cytochrome 450 activity, which converts testosterone to oestradiol. Higher oestradiol downregulates Glucose transporter type 4 receptors and worsens insulin resistance. This in turn reduces sex hormone–binding globulin and results in raised free testosterone. High oestradiol also reduces gonadotropin release from the pituitary, which reduces ovulation in females and spermatogenesis in males. In females, there are higher incidences of ***precocious puberty, polycystic ovarian syndrome, anovulation, infertility, sexual dysfunction, and hyperandrogenism***. Males with obesity experience a higher rate of ***low testosterone, decreased libido, and subfertility***.

Neurological disease

Idiopathic intracranial hypertension is caused by increased brain water content and excess cerebrospinal fluid production. It is triggered by weight gain, hormonal changes, increased cerebral venous pressure, and other factors. Inflammation in the brain may contribute to the higher rates of ***cognitive impairment*** and ***dementia*** seen in those with obesity. Other neurological complications include higher incidences of ***stroke, dementia,*** and ***polyneuropathy***.

Renal disease

Obesity predisposes to ***chronic kidney disease*** via associated conditions such as hypertension or diabetes as well as directly through glomerulosclerosis. Stress incontinence is another complication commonly seen in females with obesity.

Musculoskeletal disease

Musculoskeletal symptoms are common in patients living with obesity. These may relate to ***osteoarthritis*** and mechanical factors or muscle inflammation and impaired tissue healing in the context of chronic inflammation. Obesity is a risk factor for ***osteoporosis*** with contributions from inactivity, sarcopenia, and vitamin D deficiency.

Surgical

It is well recognized that obesity predisposes to increased perioperative risk. Weight reduction is often advised before procedures such as joint replacement and organ transplantation to support access and outcomes.

Psychological

There is a complex, bidirectional relationship between obesity and mental health. People living with obesity report a high prevalence (between 20% and 60%) of conditions such as depression and anxiety. These conditions are often exacerbated by factors that include weight bias, social isolation, and medication side-effects.

Obstetric and perinatal

Females who conceive when at a higher BMI are more likely to develop ***pregnancy-related hypertension, pre-eclampsia, postpartum infections, and gestational diabetes*** and their babies are more likely to be born preterm and via ***caesarean section***.

Skin

Dermatological complications associated with obesity include hidradenitis suppurativa, hirsutism, acanthosis nigricans, psoriasis, cellulitis, venous stasis ulcers, and cellulitis.

Pathophysiology of obesity complications

Inflammation

Obesity is a state of chronic low-level inflammation. Abundant immune cells in adipose tissue and macrophage aggregates increase proportionally with the degree of obesity.

Around 50 pro-inflammatory adipokines, including TNF-α, interleukin-6, adiponectin, resistin, and visfatin, are secreted by visceral adipose tissue. This causes disruptions of endocrine and nervous signals. Changes may be seen from early childhood.

Inflammation is proposed to be a central mechanism linking obesity to metabolic and vascular complications and increased risks of cancer and infectious disease. Evidence suggests that effective obesity care can reduce this inflammation.

The Krebs cycle and obesity

Abnormalities of the Krebs cycle and its components are implicated in inflammation. This is also emerging as an important concept in obesity linking metabolic stress to inflammation.

Fumarate hydratase deficiency has been linked to elevated levels of hypoxia-inducible factor-1 and other cytotoxic and oncogenic substrates. Tumours deficient in fumarate dehydrogenase have very high growth-differentiation factor 15 levels, linked to cellular stress and conditions such as obesity, diabetes mellitus, and liver disease. This discovery has led to novel therapeutic targets and is promising for future obesity research.

Further reading

Dandgey S, Patten E. Psychological considerations for the holistic management of obesity. *Clin Med (Lond).* 2023;23(4):318–322.
Jacob P, McCafferty K. Assessment and management of chronic kidney disease in people living with obesity. *Clin Med (Lond).* 2023;23(4):353–356.
Mourougavelou V, Chowdhury TA. Management of hyperglycaemia in people with obesity. *Clin Med (Lond).* 2023;23(4):364–371.
Young R, Benjamin A. The assessment and management of obstructive sleep apnoea–hypopnoea syndrome and obesity hypoventilation syndrome in obesity. *Clin Med (Lond).* 2023;23(4):372–379.
Zaromytidou M, Savvatis K. The weight of obesity in hypertrophic cardiomyopathy. *Clin Med (Lond).* 2023;23(4):357–363.

MANAGEMENT

General principles of management

An effective treatment strategy should be:

- Tailored to specific phenotype (based on clinical evaluation)
- Mindful of personal goals and barriers
- Effective for the severity of obesity and complications
- Sustainable within the context of lifestyle, support network, and affordability
- Evidence based and in line with local guidance and resource.

In designing a management strategy, a personalized care plan document (digital or paper) can add value and offer a reference point. The 'five As of obesity care' (**Box 37.11**) also offers useful framework.

Sensitive exploration of previous weight loss attempts and what has or has not worked provides valuable information. For example, a specific dietary strategy may have been classed as a 'failure', but this could be related to insufficient explanation around the weight loss trajectory or unrecognized binge eating disorder impairing progress. Information gleaned from these conversations can be used in the design of the current management strategy. This can enhance the chance of success.

Interventions can range from 'doing nothing' to bariatric surgery. It is helpful to describe the severity of obesity and associated risk (using staging systems as needed) to an individual to support their decision-making. Ideally, a range of choices would be presented. Suitability as well as eligibility should be considered by the individual and the multidisciplinary team. For example, a patient may meet criteria for bariatric surgery but live within a chaotic situation and may be actively self-harming. In this case, there is a significant risk associated with a bariatric intervention, and care for the person's mental health should take precedence.

Other considerations

Ethos

- Obesity is a chronic disease; discussions should cover relapse and potential strategies.
- It is important to remember that obesity is a complex, heterogeneous disorder. There is no single solution. Identifying the root cause of obesity by considering 'why' rather than simply 'how' will enable clinicians to plan an effective strategy.

Box 37.11 The five A's of obesity care

Ask	Ask for permission to explore weight. Explore readiness to change.
Assess	Assess body mass index class and stage. Assess for drivers, complications, and barriers.
Advise	Advise on obesity risks. Discuss the benefits of modest weight loss. Discuss treatment options.
Agree	Agree of realistic weight loss expectations. Focus on behavioural goals (Specific, Measurable, Attainable, Relevant, Time-bound [SMART]) and health outcomes. Agree on a treatment plan.
Assist	Assist in addressing drivers and barriers. Assist in accessing resources and programmes. Refer to appropriate provider. Arrange follow-up.

Adapted from Elaine A. Sherson, Elizabeth Yakes Jimenez, Nikki Katalanos, A review of the use of the 5 A's model for weight loss counselling: differences between physician practice and patient demand, Family Practice, *Volume 31, Issue 4, August 2014, Pages 389–398, https://doi.org/10.1093/fampra/cmu020.*

Approach

- Adopt a holistic approach and consider biopsychosocial factors. In preparing an individual to change, questions such as 'What would make you feel more ready?' can prompt self-reflection and ignite intrinsic motivation.

Communication

- There is heterogeneity in response to all strategies. Individuals should be counselled for nonresponse as well as response and appreciate that this is not a personal failure.
- Information should be provided on the fact that weight loss is rarely linear.
- Management strategies should be presented with information on predicted (average) weight loss. This will support managing expectations in line with goals.
- Repercussions of significant weight loss (i.e., loose skin, psychological impact on body image, relationships) should be considered and discussed.

Principles of management

- Multidisciplinary delivery is central to obesity care.
- Interventions should be considered as tools rather than solutions.
- Ineffective strategies should be promptly reviewed and adjusted.
- Identify high-risk individuals who require expedited care (e.g., those with declining renal function).
- The most appropriate management strategy may be watchful waiting. For example, for some it may not be the right time to engage with an obesity service (e.g., after recent bereavement). For these patients, gently guide them to return for support when able and if they would wish to.
- Management may include signposting to internal or external services and treatment of conditions within the expertise of the team, typically including metabolic conditions and nutritional deficiencies.
- The goal of management is to achieve clinically meaningful weight loss; evidence shows this starts from 5%.

Weight management goals

It is important to ascertain an individual's weight loss goals before embarking on a treatment strategy. If these have been

suggested from an external source (e.g., a surgical team to access another procedure), then it is important to review this in line with available and suitable interventions and individual circumstances. Realistic and personalized goal setting is key. A reasonable strategy is to:

- Prevent further weight gain
- Reduce body weight
- Maintain this lower weight for the long term.

Potential therapeutic strategies

- Lifestyle interventions are first-line treatment but generally result only in modest weight loss (5%–10%). Weight maintenance can be challenging because of physiological compensations.
- Remission of co-morbidities and cardiovascular risk reduction requires greater weight loss (10%–25%) (Fig. 37.7). This may be achievable with newer medical therapies.
- Bariatric surgery currently offers the greatest sustainable weight loss (for suitable candidates), in the region of more than 25%.
- It is important to map the weight loss against predicted and appreciate that individuals will respond differently. A poor or no response should prompt an alternative strategy.
- After successful weight loss, there is an expected shift in appetite-related hormones, increased appetite for palatable food, and a decrease in resting energy expenditure, which are thought to contribute to weight regain. This physiological response should not be attributed to lack of motivation, and adequate support should be provided.

The multidisciplinary team in obesity care

It is best practice to work with a multidisciplinary team to deliver care to people living with severe and complex obesity (Fig. 37.8). High-quality training in obesity care is available to allow professionals in different fields to become specialists in this area.

In the UK, a tiered approach is used in which basic obesity care is delivered in primary care, and more specialist interventions such as surgery is offered in secondary care in tier 3 and 4 services (Fig. 37.9).

Obesity through the life course

Obesity impacts individuals at every stage of their lives, including early childhood. Tailoring the approach and resources to the individuals' age and maturity is likely to add value:

- **Pregnancy and fetal outcomes:** Strong evidence links obesity to adverse maternal and fetal outcomes. Prenatal counselling is valuable. Fertility often improves as weight falls and should be considered in counselling for new drugs and bariatric surgery.
- **Young adults:** Leaving home, starting a new job, marriage, and parenthood are key transition points and may combine cost of living pressures with increased responsibility, triggering weight gain. The psychological burden on a young adult with obesity should be considered. Often this has persisted through childhood and involved adverse childhood events (including abuse and bullying). It may have resulted in social isolation. Psychology assessment is important. E-health strategies may support engagement. Complications begin silently at this stage, so vigilance is required to maintain awareness of genetic obesity.
- **Adults and middle age:** This is a key phase in the disease because co-morbidities usually begin to manifest. Care requires coordination with a range of specialists. People often have tried many strategies by this stage so require specific attention to novel strategies, rapport, and engagement.
- **Later life and older adults:** Sarcopenic obesity occurs with age because of low skeletal mass in combination with increased fat mass. This can exacerbate the age-related decline in physical function and contribute to frailty. Weight maintenance may be an appropriate strategy rather than weight loss.

Lifestyle interventions

Lifestyle intervention is the foundation of all obesity management strategies. It includes diet; physical activity; and other parameters such as sleep hygiene, alcohol moderation, and stress management. Also linked are measures to reduce cardiovascular risk in obesity such as smoking cessation. Lifestyle management should be underpinned by robust behaviour change theory and delivered in a manner that resonates with the individual, such as face-to-face consults, digital options, and group sessions. Outcomes vary, and studies suggest many of the effects of lifestyle intervention are short-lived, with weight rising after the intervention period has finished.

Dietary strategies

Obesity has been described as 'a normal condition in an abnormal environment'. There are many challenges to consuming a healthy, balanced diet within an obesogenic environment. Highly palatable foods high in fat, salt, and sugar with large portion sizes have become increasingly prevalent and accessible across the world. Food insecurity, economic crises, and many differing consumer messages create a landscape that can be extremely challenging to navigate.

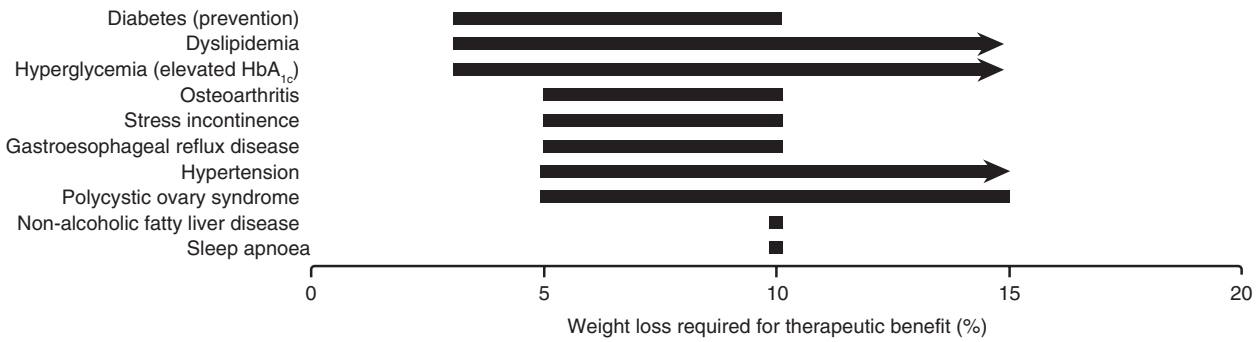


Fig. 37.7 Weight loss required for therapeutic benefit. *(Adapted from Cefalu, W. T., Bray, G. A., Home, P. D., Garvey, W. T., Klein, S., Pi-Sunyer, F. X., … Ryan, D. H., et al. [2015]. Advances in the science, treatment, and prevention of the disease of obesity: Reflections from a diabetes care editors' expert forum.* Diabetes Care, *38[8], 1567–1582.)*

Dietetic advice for obesity care is specialized. Inappropriate or generic advice with insensitive delivery can be harmful and lead to attrition from services.

When planning a dietary approach with a patient:

- Gain understanding about previous dietary strategies and outcomes, as well as the patient's food knowledge.
- Explain that 'fad' diets typically lead to rapid weight loss (often lean body mass) but lack a sound evidence base and are often unsustainable. Aim to understand if a cycle of weight loss and regain has been experienced and promote alternative, evidence-based approaches.

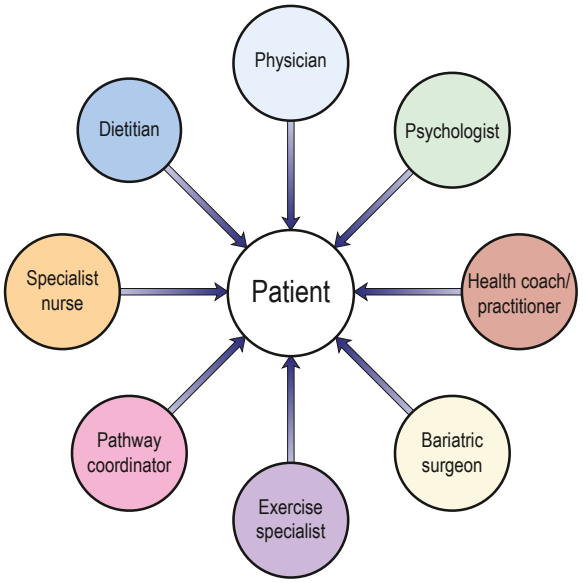


Fig. 37.8 **Multidisciplinary team example.**

- Perform a comprehensive dietetic assessment using validated tools (e.g., food diaries and food frequency questionnaires). Aim to understand typical dietary intake, meal timings, take-away consumption, and snack and fluid intake as well as drivers for these patterns, including mood.
- Understand if any co-morbidities may be having an impact on eating behaviour (e.g., hyperphagia in genetic obesity, preference for certain foods and patterns of eating in autism or attention-deficit hyperactivity disorder, trauma).
- Understand lifestyle factors, such as shift working, that have an impact on eating.
- Screen for disordered eating, which can be highly prevalent in obesity (up to 15%). This includes binge-eating disorder and night eating syndrome. Counsel and signpost accordingly.
- Before the implementation of a specific strategy, work on stabilizing eating behaviour, meal planning, and managing portion sizes. This can reduce unplanned, high-energy consumption and compensatory eating.
- Ask about consumption of sugar-sweetened beverages ('fizzy drinks'), which have been linked to disrupted physiological responses to feeding and reduced satiety, as well as poor dental health.
- Identify and track nutritional deficiency, which is common despite obesity.
- Encourage self-monitoring and weight tracking to support weight maintenance after weight loss. For example, a patient may be encouraged to monitor their weight at home and check in with the clinical team if their weight hits certain thresholds after an intervention (e.g., 2 kg as a red flag and 4 kg to return for intervention).

Specific dietary strategies

A range of specific dietary strategies can promote weight loss and other health benefits. Overarching aims should be to reduce

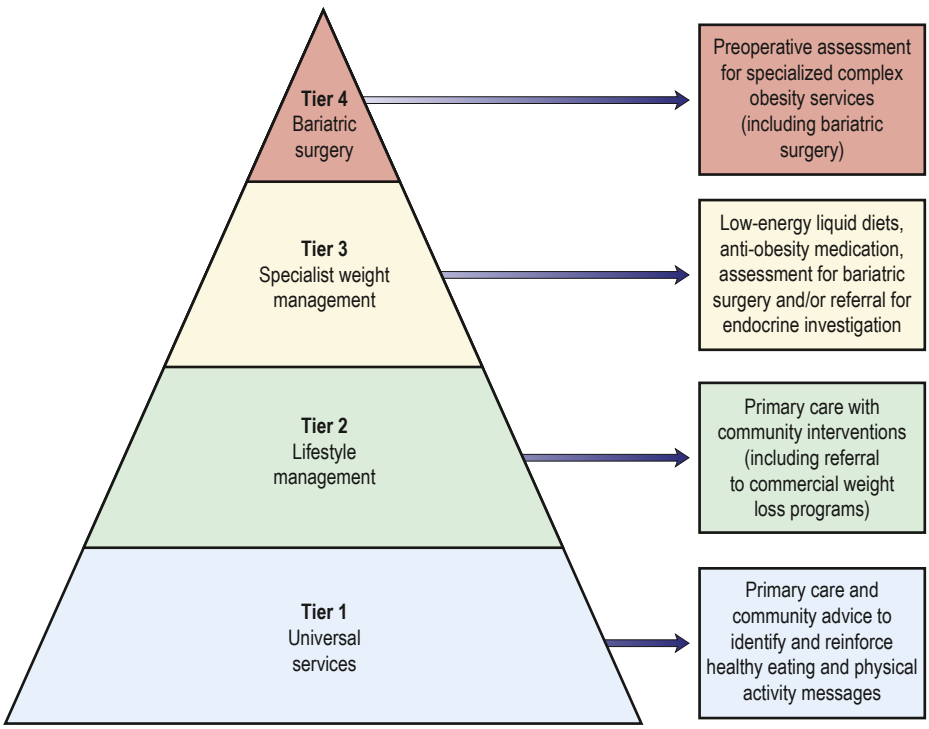


Fig. 37.9 **The tiered weight management system in England.**

energy intake while maintaining nutritional quality in a manner that is practical, sustainable, and evidence based. There is no 'right or wrong diet' as such, and a decision on strategy should be a shared decision.

A significant evidence base also demonstrates enhanced outcomes when a dietary intervention is combined with an effective lifestyle intervention. This is useful to convey in consultation. The Look AHEAD Trial study compared a 4-year intensive programme (including lifestyle counsellor, dietary interventions, portion-controlled meal plans, physical activity, and behavioural modification techniques) with usual medical care in people with diabetes. The intensive intervention group showed mean weight loss at 1 year of –8.6% versus –0.7%. This was sustained over 4 years with a mean weight loss of –6.15% in the intervention group. **Box 37.12** summarizes the main evidence-based dietary approaches for weight management.

Physical activity in obesity care

People living with obesity report lower levels of physical activity than those of normal body weight. They are also more likely to be sedentary (expending <1.5 metabolic equivalents in a sitting or lying position). Physical activity increases energy expenditure.

A strong body of evidence links increased physical activity to health benefit including cardiovascular disease. This guides national recommendations (**Fig. 37.10**).

There are, however, many reasons why these goals may not be easily attainable in people with obesity:

- Co-morbidities affecting mobility or fitness and lack of tailored support to be active
- Limited knowledge on exercise intensity and how this is perceived (e.g., breathlessness is to be expected with moderate intensity activity but may be experienced as pathology in those with co-morbid obesity who have not received a graduated programme with counselling)
- Stigma and bias having an impact on confidence to be active in public spaces
- Fear that gym equipment may not be the appropriate size
- Deprivation and health inequalities linked with obesity and physical inactivity.

With these factors, the national recommendations may seem unattainable and demoralizing. Yet people living with obesity who are sedentary may, in fact, have the most to gain from becoming active (**Fig. 37.11**).

Substantial evidence exists for the benefits of physical activity in obesity care:

- Increased efficacy and sustainability for other weight loss interventions
- Preservation of lean body mass during weight loss, particularly when rapid
- Improvements in cardiovascular risk factors and lower all-cause mortality
- Improved cardiorespiratory function in preparation for bariatric surgery
- Improvements in depression and anxiety (various meta-analyses).

Box 37.12 Dietary approaches for weight management and related evidence

Dietary strategy	Description	Evidence/phenotype
DASH (Dietary Approaches to Stop Hypertension) diet	Foods including vegetables, fruits, wholegrains, and low saturated fat. The diet limits sodium intake but is rich in potassium, calcium, and magnesium.	Evidence for hypertension, CVD risk, lipid profile Impact on body weight: 1.4 kg (systematic review).
Mediterranean diet	Emphasis on monounsaturated fats (extra-virgin olive oil, vegetables, nuts) and unrefined cereals. Reduced intake of processed meat and refined sugars.	Substantial body of evidence for improved CVD risk (e.g., PREDIMED study). Mean weight loss, 1.75 kg, enhanced to 4 kg when combined with lifestyle measures.
Low- calorie and very-low-calorie diets	Low-calorie diets are those that restrict daily caloric intake to 1000–1200 kcal. Very-low-calorie diets restrict daily caloric intake to 800 kcal/day or less.	Usually closely supported within a structured programme. Selection and readiness are important. Valuable for rapid weight loss such as before bariatric surgery (e.g., the 'liver shrinkage' diet) or in preparation for a specific intervention (e.g., transplant). Landmark evidence from national studies (DIRECT, DROPLET).
Low-GI diet	GI describes the impact of carbohydrate on blood glucose levels. Lower GI foods tend to be higher in fibre (e.g., vegetables, pulses) and can improve satiety.	Weight reduction of 1.8 kg. Reduced cancer and mortality rate in association with higher fibre content (WHO).
Low-carbohydrate diet	Ranges between very low levels (ketogenic or 20–50 g/day) to higher levels of up to 230 g/day. In the context of significant consumption, may still equate to a reduced approach.	Improvements in diabetes mellitus with evidence for PCOS and MASLD. Improves triglycerides but may have an adverse impact on LDL. Up to 9-kg weight loss.
Intermittent fasting	Intermittent fasting is voluntarily refraining from eating during a certain time period and limiting the number of calories eaten during this period.	An umbrella review of meta-analyses of randomized clinical trials reported that there is moderate- to high-quality evidence of a positive association between intermittent fasting and 'anthropometric and cardiovascular' outcomes.
Time-restricted eating	Time-restricted eating similarly refers to abstinence from food during a specific period of time without limiting the number of calories eaten during this period.	The TREAT randomized clinical trial shows there was a considerable decrease in weight for the group with time-restricted eating compared with the group with consistent meal timing.

CVD, Cardiovascular; *GI,* glycaemic index; *LDL,* low-density lipoprotein; *MASLD,* metabolic dysfunction–associated steatotic liver disease; *PCOS,* polycystic ovarian syndrome; *WHO,* World Health Organization.

Fig. 37.10 National recommendations on physical activity. *(UK Chief Medical Officers' Physical Activity Guidelines 2019.)*

Implementation of physical activity in obesity care

In counselling, it is important to communicate that physical activity alone is unlikely to have a major impact on weight loss. This can set expectations. Additional benefits should be emphasized as well as the synergistic effect of activity with other interventions. In terms of modalities, aerobic, high intensity-interval training, and resistance programmes all have evidence for benefit. 'E-health' approaches (see next section) may help patients engage with regular exercise programmes.

E-health in obesity care

Information and communication technologies used to support healthcare are broadly referred to as e-health. Examples include smartphone applications or apps that provide lifestyle support and promote behaviour change. Some offer clinical facing portals and links to wearable technology. Benefits include flexibility, ease of access to care, personalization, and support for motivation and planning. Challenges include attrition rates, digital exclusion, and integration with clinical services.

Given the rising obesity rates and accumulating evidence, it seems likely that there will be strong interest in promoting e-health delivery at scale, for example, to support patients monitoring their weight, calculating the calory content of food, or recording and engaging with exercise programmes. As with any other intervention in obesity, it is worthwhile considering this as a tool rather than a solution. Given the complex psychological burden of obesity, it should not be a substitute for face-to-face delivery because often people with obesity benefit hugely from and value the rapport and support offered from face-to-face delivery. For people who cannot or prefer not to access e-health, alternatives should be presented.

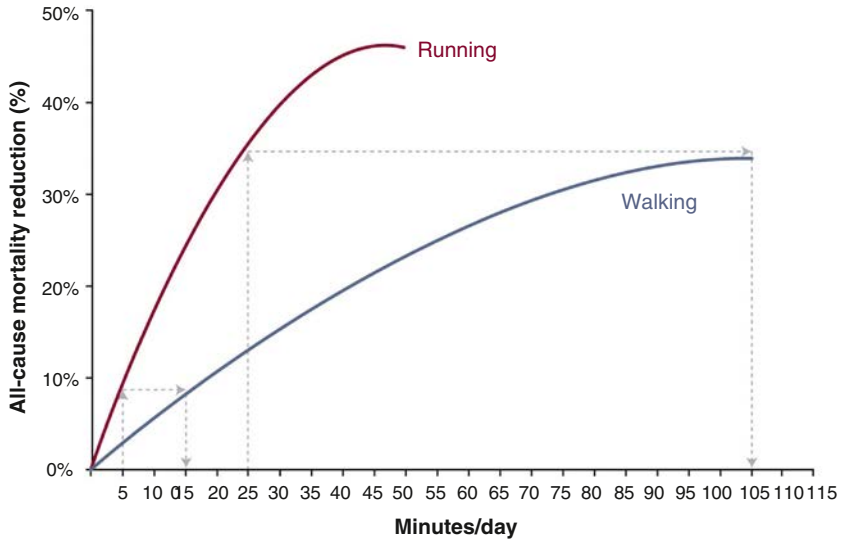


Fig. 37.11 Comparison of benefits between walking and running. A 5-min run generates the same benefits as a 15-min walk, and a 25-min run is equivalent to a 105-min walk. *(Minimal Amount of Exercise to Prolong Life: To Walk, to Run, or Just Mix It Up? August 2014* Journal of the American College of Cardiology *64[5]:482–484. https://doi.org/10.1016/j.jacc.2014.05.026.)*

Medical therapy

Compared with other disease areas, there has, until recently, been a limited range of options for the medical treatment of patients with obesity. Several agents have been withdrawn over the years for safety concerns (e.g., the cannabinoid-1 receptor blocker rimonabant). We are now in a promising new era.

As the pathophysiology of obesity has been elucidated in more detail, a range of therapeutic targets have emerged. Medical treatment for people with obesity is now one of the most highly publicized and exciting areas of drug development. This has led to increased social media traffic, media reports, and drug shortages. In this context, it is especially important to emphasize the medical intervention as a 'tool, not a solution'. Alongside medical therapy, patients should ideally be thoroughly reviewed by different members of the multidisciplinary team and undergo education about monitoring and likely side-effects, a detailed medication review, psychological support, and lifestyle coaching.

A medication review includes:

- Reviewing all existing medications and trying to remove those associated with weight gain, if appropriate and endorsed by specialist colleagues
- Reviewing drugs that will need to be deprescribed in the context of significant weight loss (e.g., bariatric surgery) such as antidiabetic or antihypertensive medications.

Vitamin D

There is strong evidence linking vitamin D deficiency with obesity and its associated diseases. However, whether this is a cause or a consequence remains to be determined. Whereas some studies suggest that vitamin D accumulates in adipose tissue, which leads to lower plasma levels, others suggest impaired 25-hydroxylation in the liver as the cause. Other studies suggest that low vitamin D predisposes to obesity by increasing pre-adipocyte differentiation into adipocytes. In either case, there appears to be a role for vitamin D replacement in obesity management.

Orlistat

Orlistat is an inhibitor of pancreatic and gastric lipases. It reduces dietary fat absorption by up to 30% and aids weight loss. In the UK, it is incorporated into obesity guidance and licensed for people with a BMI of 28 kg/m^2 or greater and obesity-related comorbidities or with a BMI of 30 kg/m^2 or greater. Because of its mechanism of action, orlistat can cause unpleasant side-effects (e.g., oily stools and faecal incontinence). This is in the context of relatively modest weight loss. For these reasons, it is frequently not tolerated. Counselling is important before prescription to support retention. Advice on lower fat diets may be beneficial. Orlistat is typically prescribed for 12 months but may be used for longer with support. A double-blind, randomized, controlled study (Xendos) reported 10.6-kg weight loss with the addition of lifestyle modification. Orlistat also has evidence in diabetes prevention (37% relative risk reduction), cardiovascular prevention, and renal failure.

Glucagon-like peptide 1 receptor agonists

Endogenous GLP-1 is a peptide consisting of 31 amino acids. It is secreted mainly by L-cells in the gut and ileum but also from the brain (nucleus tractus solitarius) in response to food intake. Receptors for GLP-1 are widely expressed across the body, resulting in a range of physiological effects (Fig. 37.12). In the context of weight management, effects such as increased satiety and reduced gastric emptying are clinically important.

The evidence base for GLP-1 receptor agonists is extensive and includes global, large-scale, randomized controlled studies (Fig. 37.13). In the SCALE trials of liraglutide, weight loss exceeded 5% of previous body weight in half of patients and exceeded 10% of previous body weight in 25%. There was also a significant reduction in the number of treated patients who developed diabetes. Semaglutide may offer even greater weight loss (10% weight loss in >70% of patients and 15% weight loss in >50% in the STEP clinical trial programme).

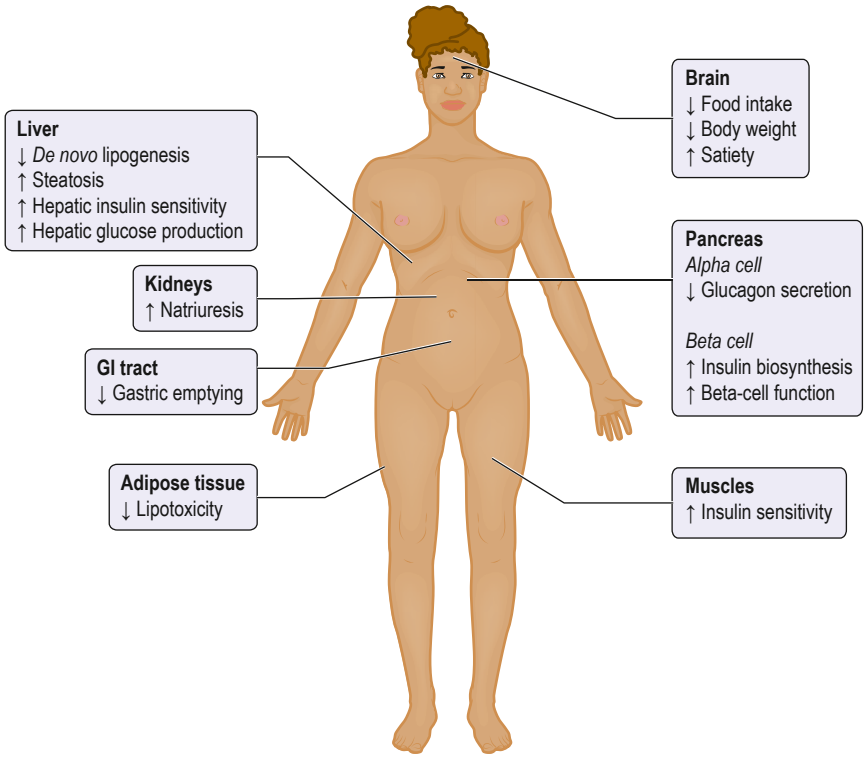


Fig. 37.12 Systemic effects of GLP-1 receptor agonists. *GI,* Gastrointestinal.

Clinical delivery of a glucagon-like peptide 1 agonist programme

GLP-1 agonists are given by daily or weekly subcutaneous injection. Phone prompts may be helpful in reminding patients when doses are due. Patient education and peer support are valuable. It is especially important to counsel people on common side-effects (e.g., headache, GI disturbance), how to manage these side-effects, and how the side-effects typically settle with time. This minimizes attrition.

Baseline and serial measurements of selected parameters (e.g., body composition, BP) can support with evaluation and programme retention. For example, scales may not immediately show weight loss, but body composition may change favourably. This insight can support motivation and retention.

Counselling should also cover:

- Indication
- Administration (important to check comfort, competence, and dexterity for injections)
- Titration schedule
- Side-effects and strategies for management (e.g., smaller meals, food diary)
- Clinical features of severe side-effects (e.g., pancreatitis, cholecystitis)
- Monitoring
- Cautions
- Expected benefits (note that weight loss may be less pronounced in certain groups)
- Review points and contact for the clinical team for support and questions
- Conception
- Concurrent lifestyle changes
- Variability in response.

Bupropion and naltrexone combination therapy

Bupropion is an antidepressant, and naltrexone is an opioid antagonist. The combination works synergistically to support weight loss as part of a multicomponent plan. The treatment works centrally with bupropion activating proopiomelanocortin (POMC) neurones in the hypothalamus, an effect is augmented by naltrexone. The overall impact is reduced appetite and cravings. There are numerous cautions, contraindications, and side-effects for this agent, and it should be used in centres with experience.

Phentermine and topiramate combination therapy

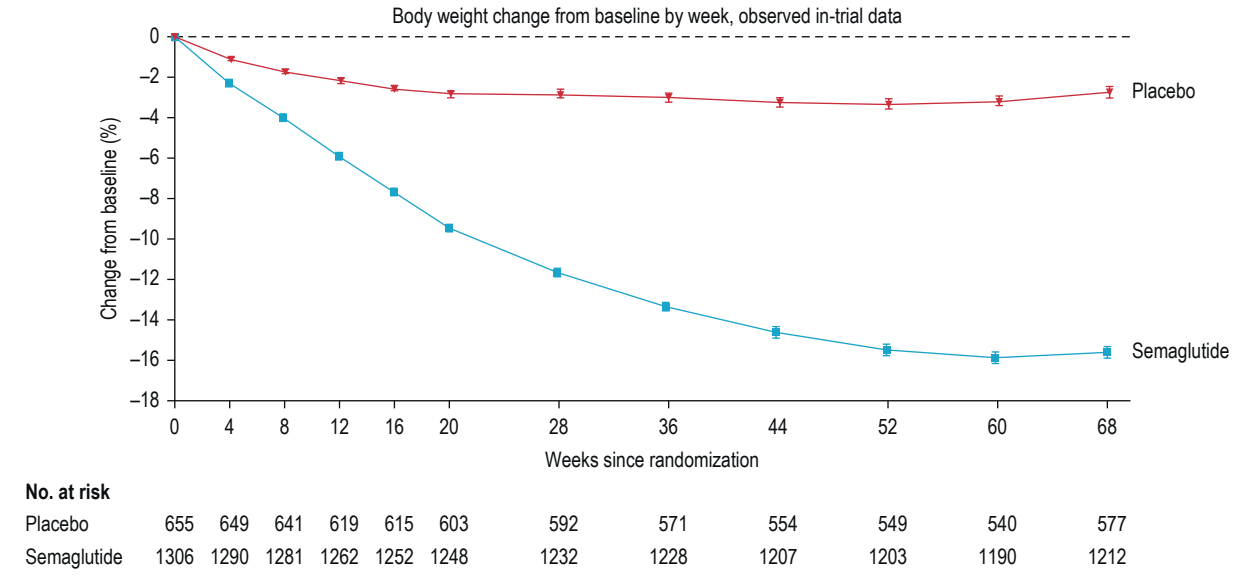
Phentermine is a sympathomimetic, and topiramate is a carbonic anhydrase inhibitor (used as an anticonvulsant). The combination acts to suppress appetite and increase satiety. This treatment is not recommended in patients with uncontrolled hypertension and significant cardiac disease. Careful monitoring is required.

Setmelanotide

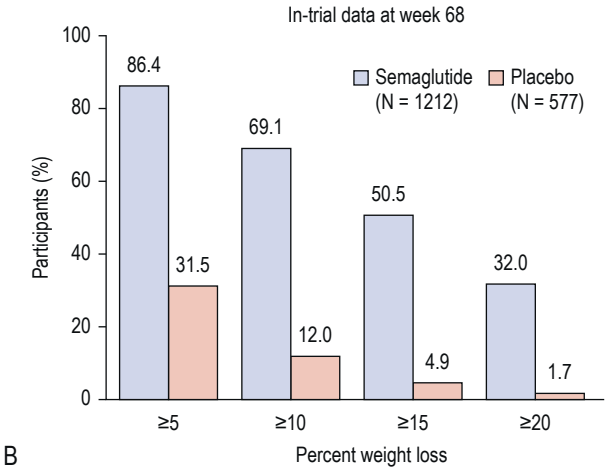
Setmelanotide is a highly specialist drug indicated for a few specific monogenic forms of obesity only. Treatment is via subcutaneous injection and is typically available through selected centres with relevant genetic expertise. It is an MC4R agonist approved for monogenic obesity caused by POMC, proprotein convertase subtilisin/kexin type 1, or LEPR deficiency in individuals ages 6 years of age and older. The most common side-effects are GI effects, hyperpigmentation, and injection site reactions. Clinical trials are underway for further indications.

Diabetes medications

Type 2 diabetes and obesity often (but not always) coexist. International guidance supports the preferential use of agents that support weight loss in people living with obesity and diabetes (see Chapter 35).



A



B

Fig. 37.13 Effect of once-weekly semaglutide, as compared with placebo, on body weight.
(From The New England Journal of Medicine, Wilding J.P.H., Batterham, R.L., Calanna S,. et al., Once-Weekly Semaglutide in Adults with Overweight or Obesity, N Engl J Med *2021;384:989-1002. Massachusetts Medical Society. Reprinted with permission from Massachusetts Medical Society.)*

Future therapies for obesity and areas of interest

Observations from bariatric surgery (in which the primary effects are thought to be mediated via altered concentrations of multiple gut hormones) have led to the development and study of novel combination therapies.

- ***Tirzeptide*** is a novel dual (glucagon-like peptide 1 [GLP-1] and gastric inhibitory polypeptide [GIP] receptor) agonist and first in class. It is currently licensed for type 2 diabetes in the USA, and significant weight loss outcomes have been reported (with 40%–69% achieving ≥10% weight loss in the SURPASS clinical trial programme). Tirzepatide is currently awaiting approval for obesity care.
- ***Retatrutide,*** an agonist of GLP-1, GIP, and glucagon receptors, causes substantial reductions in body weight.
- ***Orforglipron,*** an oral once-daily non-peptide GLP-1 receptor agonist, achieved a mean weight loss of 14.7% at 36 weeks in adults with obesity or overweight in phase 2 clinical trial.
- Use of a GLP-1 agonist with **glucagon**, which in addition to the well-recognized impact of increasing glucose levels also has favourable effects on energy balance, which supports weight loss. Additional areas of interest are also emerging from trial data, such as benefits on liver fibrosis.
- Use of a GLP-1 agonist with **amylin** (e.g., cagrilintide). Amylin is co-secreted with insulin and has a role in appetite regulation as well as reduced gastric emptying.
- Monoclonal antibody against the activin type II receptor. Activin is a protein complex with various biological effects, including regulation of the menstrual cycle and energy metabolism.

Bariatric surgery

Bariatric surgery (also known as metabolic surgery) is an effective and durable treatment for people with severe obesity, linked to improvements in obesity-related co-morbid conditions as well as all-cause mortality. A systematic review and

meta-analysis of 621 studies with 135 246 patients undergoing a range of bariatric procedures reported an overall average weight loss of 38.5 kg per patient, or 55.9% of excess body weight. This currently makes bariatric surgery the most potent treatment for obesity in appropriate candidates. A range of surgical options are available, and novel approaches show particular promise. In carefully selected and prepared individuals, surgery has a low mortality rate, minimal complication rate, and short lengths of stay. Some centres offer day case surgery or enhanced recovery protocols.

Remission of co-morbidities such as type 2 diabetes and OSA are realistic goals, but it is key to counsel patients that there are no guarantees. In the case of diabetes, validated scores such as DiaRem, which consider diabetes duration and treatment, can offer prediction scores for remission. In the postoperative period, glycaemic control often improves dramatically within a few days and before any significant weight loss.

Patient selection

The keys to the success of bariatric interventions are appropriate selection, work-up, and follow-up of individuals. People who wish to access bariatric surgery should commit to lifelong follow-up, modifications to their diet and lifestyle, and engagement with the service multidisciplinary team. Different models exist according to region, but typically, individuals are required to complete a comprehensive weight management programme before surgery, including medical and lifestyle elements. This is to ensure adequate preparation, selection, and understanding.

Unfortunately, uptake of this effective treatment can be limited, and in the UK, for example, it is estimated that fewer than 1% of eligible patients receive this intervention. There are various reasons for this, including cultural, economic, societal, and healthcare factors. In certain cases, individuals may choose to self-pay for surgery or travel overseas.

Preparation for surgery

In terms of **counselling and preparation** for bariatric surgery, the following are important:

- Identification of appropriate individuals based on individual preference, local criteria, and clinical phenotype. For example, people with recent-onset diabetes are more likely to achieve remission with surgery and should be considered for this approach if they are otherwise suitable and interested.
- Robust multidisciplinary assessment and combined discussion. Key members of the team include a specialist dietitian, psychologist, nurse, surgeon, physician, and anaesthetist. All should be happy to proceed as well as the individual. Certain high-risk cases may need dedicated meetings and more detailed discussion. Some individuals may need to be expedited (e.g., when bariatric surgery is required as a bridge to access organ transplantation).
- Counselling on risks of surgery. This includes standard risks as for all surgery (e.g., deep vein thrombosis) but also includes dumping syndrome, postbariatric hypoglycaemia, excess skin, anastomotic leak, strictures, internal hernias, and nutritional deficiencies.
- Psychological evaluation and readiness. Psychological instability may lead to postoperative deterioration (e.g., a switch of comfort eating to alcoholism or complications such as weight regain). It is important to explore body image, coping strategies, and partner and social support.
- Anaesthetic preparation. Relevant investigations include cardiac, metabolic, and sleep studies, as well as optimization of risk factors, including diabetes mellitus, hypertension, OSA, smoking, diet (a 'liver shrinkage diet'), and physical fitness using a 'prehabilitation' approach.
- Preconception counselling. Fertility is typically improved after bariatric surgery, but because of the dramatic weight and metabolic shifts in the postoperative period, females are advised to wait for at least 1 year before attempting conception.

Selection of bariatric procedure

Although bariatric surgery is an effective intervention with a strong evidence base, it is not without complications and is an arduous and risky process for those who are not effectively selected and prepared.

The type of procedure performed is a shared decision but should be guided by the views of the operating bariatric surgeon. Procedures may be classified as restrictive (reducing volume of digestive tract and reducing oral intake) or malabsorptive (reducing absorption of calories) and include the following.

Roux-en-Y gastric bypass

The Roux-en-Y gastric bypass (RYGB) is one of the most effective and widely performed procedures worldwide, typically performed laparoscopically under general anaesthetic. The stomach is reduced to a small pouch by stapling and then directly connected to the jejunum, thereby bypassing the small bowel (**Fig. 37.14**). It is a combined restrictive and malabsorptive procedure. After the procedure, the patient is provided with a tailored diet of thin fluids, gradually building up to solid food. They may be back at work in a matter of weeks.

Sleeve gastrectomy

This is a popular procedure and usually quicker to perform than the RYGB. The majority (85%) of the lateral aspect or curvature of the stomach is resected, leaving a tubular 'sleeve', which is typically less than 2 cm in diameter. It is important to identify people with reflux because higher intraluminal pressures can exacerbate symptoms.

Adjustable gastric banding

Adjustable gastric banding (AGB) is a restrictive procedure created through the placement of a silicone band at the upper aspect of the stomach. This band is connected to the skin via a filling port, which may then be adjusted as required. AGB gained popularity, particularly by those who perceived it as a reversible and less invasive procedure compared with other options. But trends have changed, and it is less commonly practised now because of comparatively lower weight loss outcomes, complication rate (e.g., band slippage, port infection, erosion), and follow-up burden (band adjustments). It is no longer offered within the UK National Health Service.

Other surgical techniques

'One anastomosis gastric bypass' is a newer, simplified form of the RYGB procedure reporting significant weight loss. The 'duodenal switch' procedure is a complex malabsorptive procedure with increased complications that is sometimes used as a second-step 'rescue' procedure.

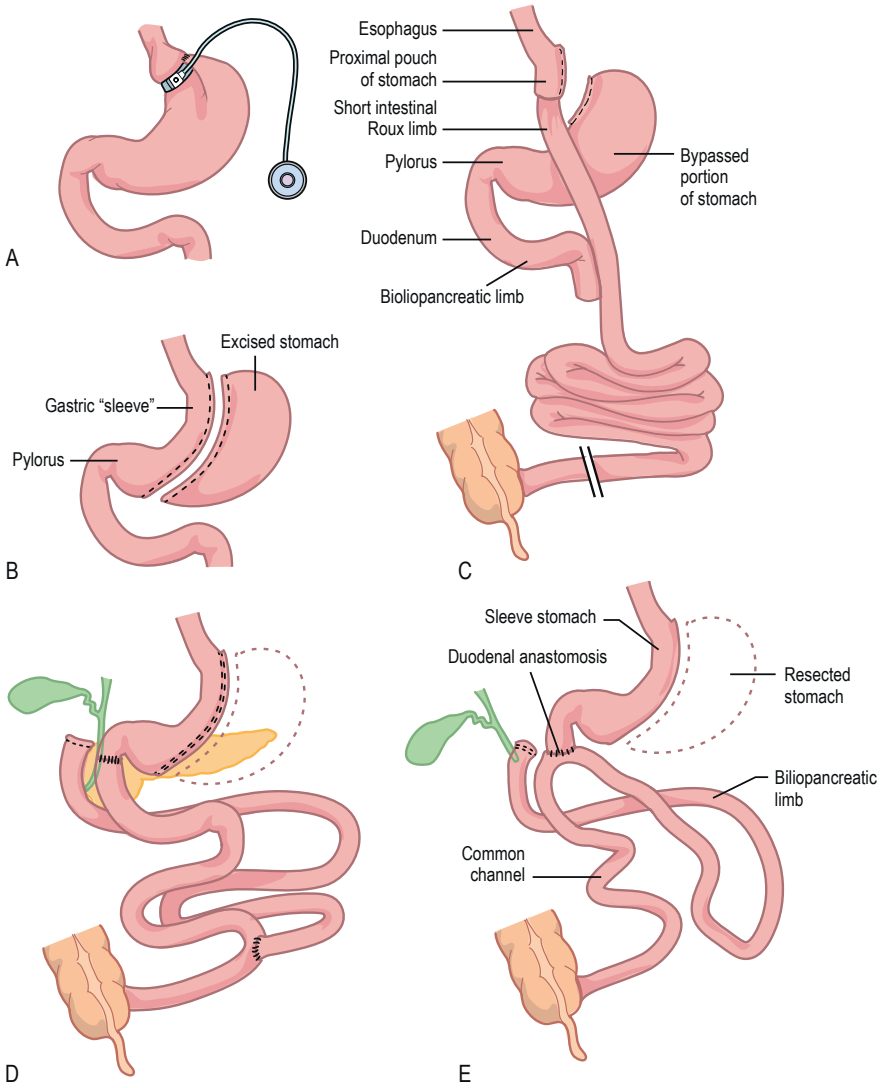


Fig. 37.14 **(A)** Adjustable gastric banding. **(B)** Sleeve gastrectomy. **(C)** Roux-en-Y gastric bypass. **(D)** Duodenal switch. **(E)** Single anastomosis duodeno-ileal bypass with sleeve gastrectomy.

Endoscopic procedures

A number of novel endoscopic procedures have been developed that tend to have a lower morbidity than surgery and that can, in some cases, be performed without anaesthetic. Long-term outcome data are awaited. Examples include:

- **Gastric balloon:** A small balloon (silicone or rubber) is inserted within the stomach. With inflation (500–750 mL of water), it offers restriction. There is a risk of deflation and migration of the balloon and minimal weight loss.
- **Endoscopic sleeve gastrectomy:** Using endoscopic access, suturing is placed inside the stomach to reduce the volume by up to 70%. Weight loss outcomes equivalent to the surgical sleeve gastrectomy have been reported.
- **Transoral outlet reduction endoscopy** is used in patients who have undergone gastric bypass surgery but have then regained weight. Endoscopically delivered suturing reduces the aperture of the anastomosis and aims to slow transit. It is also used to address postoperative complications, including dumping.

Further reading

Askari A, Jambulingam P, Gurprashad R et al. The surgical management of obesity. *Clin Med (Lond).* 2023;23(4):330–336.
Flint S, Batterham RL. The need to personalise approaches for the prevention and management of obesity. *EClinicalMedicine*. 2023;58; 101944.
Lee V. Introduction to the dietary management of obesity in adults. *Clin Med (Lond).* 2023;23(4):304–310.
McGowan B et al. The bariatric-metabolic physician's role in managing clinically severe obesity. *Curr Obes Rep*. 2021;10(3):263–273.
Melson E, Miras AD, Papamargaritis D. Future therapies for obesity. *Clin Med (Lond).* 2023;23(4):337–346.

Significant websites

https://www.scope-elearning.org. *SCOPE e-learning*.
oen.org.uk *OEN: Obesity Empowerment Network—'Nothing about us without us'*.
https://movingmedicine.ac.uk. *Moving Medicine*.
https://easo.org/talking-about-obesity-obesityuk-language-matters-guide/. *Talking about obesity: Obesity UK Language Matters Guide—EASO*.
https://bomss.org. *British Obesity and Metabolic Surgery Society: BOMSS—British Obesity & Metabolic Surgery Society*.
who.int *World Health Organization, Health Services Delivery Framework: Health service delivery framework for prevention and management of obesity*.

Questions for this chapter can be found on Elsevier eBooks+.

38 Haematology

Michael F. Murphy, K. John Pasi, and Noemi Roy

Clinical skills 1271
History 1271
Examination 1271
Investigations 1271
Background 1271
The formation of blood cells (haemopoiesis) 1271
Haemopoietic growth factors 1271
Peripheral blood 1273
Red blood cells 1276
Anaemia 1277
Microcytic anaemia 1279
Normocytic anaemia 1283
Macrocytic anaemias 1283
Megaloblastic anaemia 1283
Vitamin B12 (cobalamin) 1284
Folic acid 1285
Management and prevention of megaloblastic anaemia 1286
Macrocytosis without megaloblastic changes 1286
Anaemia caused by marrow failure (aplastic anaemia) 1287
Haemolytic anaemias 1288
Inherited haemolytic anaemia 1289
Red blood cell membrane defects 1289
Haemoglobin abnormalities 1291
The thalassaemias 1292
Sickle syndromes 1295
Metabolic disorders of the red blood cell 1299
Acquired haemolytic anaemia 1301
Autoimmune haemolytic anaemias 1301
Alloimmune haemolytic anaemia 1304
Non-immune haemolytic anaemia 1305
Polycythaemia 1306
The spleen 1306
Blood transfusion 1307
Blood groups 1308
Procedure for blood transfusion in hospitals 1308
Pre-transfusion compatibility testing 1308
Blood ordering 1309
Complications of blood transfusion 1309
Immunological complications 1310
Non-immunological complications 1312
Strategies for the avoidance of unnecessary transfusion 1312
Blood, blood components, and blood products 1313
White blood cells 1314
Neutrophils 1314
Eosinophils 1315
Basophils 1315
Monocytes 1315
Lymphocytes 1315
Haemostasis 1316
Vessel wall 1316
Platelets 1316
Coagulation and fibrinolysis 1316
Investigation of bleeding disorders 1320
Vascular disorders 1321
Platelet disorders 1321
Thrombocytopenia 1321
Platelet function disorders 1323
Thrombocytosis 1323
Inherited coagulation disorders 1324
Acquired coagulation disorders 1325

CORE SKILLS AND KNOWLEDGE

Haematologists treat patients with disorders of the blood. This chapter focuses on general haematology, including the investigation of anaemia and non-malignant white blood cell (WBC) and bleeding disorders, alongside the safe management of blood transfusion. Haematological malignancies (see Chapter 39) and thrombotic disease (see Chapter 40) are dealt with elsewhere.

The skills and responsibilities of haematologists extend beyond caring for the patients with specific haematological disorders and include ensuring that blood results reported by the laboratory analysers are accurate, establishing that blood used for transfusions is safe and available when and where it is needed, helping other specialties interpret results of laboratory tests, and guiding the use of blood products.

Key learning objectives for general haematology at undergraduate level include:

- Understanding anaemia: causes, investigation, and management
- Assessing whether disorders of WBCs and platelets are caused by primary haematological disease or secondary to other causes
- Understanding the way that blood transfusions are carried out and the mechanisms by which safety is ensured
- Understanding the mechanisms of disorders of haemostasis and thrombosis, as well as their investigation and management.

Specific opportunities for learning include attending ward rounds on inpatient haematology units, attending haematology clinics to understand the problems that cause patients to be referred from primary care, seeking out patients on a haematology day unit or medical admissions ward to understand the immediate and long-term management of haematological disorders, and attending laboratory 'morphology' meetings when laboratory science is correlated with individual patient care.

CLINICAL SKILLS

History

Symptoms of haematological disease tend to be caused by abnormalities in individual cell lines (Fig. 38.1).

Examination

Key clinical signs of haematological disease are included in Fig. 38.2. Most can be attributed to the effects of over- or under-production of different cell lineages or are features of underlying haematological disease processes (e.g., splenomegaly).

Investigations

Specific investigation of haematological disease generally begins with examination of the peripheral blood (Box 38.1). Bone marrow aspiration, with or without biopsy (Fig. 38.3), is a crucial investigation in the diagnosis and monitoring of many haematological disorders. These procedures allow determination of cellularity (increased or decreased?) as well as morphology (do the blood cell precursor cells look as they should?). The procedure and the cell lineages present in bone marrow are detailed in Fig. 38.3 and Box 38.2.

BACKGROUND

The formation of blood cells (haemopoiesis)

Blood consists of:

- Red blood cells (RBCs)
- White blood cells (WBCs)
- Platelets
- Plasma, in which the above elements are suspended.

Plasma is the liquid component of unclotted blood, which contains soluble fibrinogen. Serum is what remains after formation of the fibrin clot.

The haemopoietic system includes the bone marrow, liver, spleen, lymph nodes, and thymus. There is a huge turnover of cells, with RBCs surviving for approximately 120 days, platelets approximately 7 days, and granulocytes only 7 hours. The production of as many as 10^{13} new myeloid cells (all blood cells except for lymphocytes) per day in the normal, healthy state requires tight regulation according to the needs of the body.

Blood islands are formed in the yolk sac in the third week of gestation and produce primitive blood cells, which migrate to the liver and spleen. These organs are the chief sites of haemopoiesis from 6 weeks' to 7 months' gestation, when the bone marrow becomes the main source of blood cells. In childhood and adult life, the bone marrow is the only normal source of blood cells, apart from T lymphocytes, which are made in the thymus.

At birth, haemopoiesis takes place in the marrow of nearly every bone. As a child grows, the active red marrow is gradually replaced by fat (yellow marrow) so that haemopoiesis in the adult becomes confined to the central skeleton and the proximal ends of the long bones. Only if the demand for blood cells increases and persists do areas of red marrow extend. Pathological processes interfering with normal haemopoiesis may result in resumption of haemopoietic activity in the liver and spleen, which is referred to as ***extramedullary haemopoiesis***.

All blood cells are derived from pluripotent stem cells. These stem cells have two key properties:

1. ***Self-renewal:*** production of more stem cells
2. ***Proliferation and differentiation*** into progenitor cells, committed to one specific cell line

Pluripotent stem cells differentiate into mature blood cells through intermediate progenitor cells, which have lost the ability to self-renew but have high proliferative capacity. These progenitor cells can be broadly classified according to the blood cells that they are programmed to make (Fig. 38.3F). For example, the common lymphoid progenitor gives rise to T- and B-lymphoid cells; whereas the former are produced in the thymus, B-cell production primarily occurs in the bone marrow. The common myeloid progenitor gives rise to all non-lymphoid cells in the bone marrow via a series of intermediate progenitor cells. These cells cannot be identified in bone marrow biopsies but are recognized by their ability to form colonies (reflecting their high proliferative capacity) when haemopoietic cells are immobilized in a soft gel matrix, the so-called colony-forming unit (CFU).

These haemopoietic stem and progenitor cells interact closely with several components contained within the specialized microenvironment of the bone marrow, including non-haemopoietic stromal cells, blood vessels, and extracellular matrix, together creating a nurturing microenvironment for blood cell development. Central to this microenvironment are haemopoietic growth factors produced by stromal cells in the bone marrow and elsewhere, which are key regulators of blood cell production.

Haemopoietic growth factors

Haemopoietic growth factors are glycoproteins (GPs), which regulate differentiation and proliferation of haemopoietic progenitor cells and the function of mature blood cells. Some growth factors are present in the circulation, and other growth factors are produced within the bone marrow microenvironment by stromal cells such as fibroblasts, osteoblasts, or endothelial cells or at sites of inflammation by activated T cells, monocytes, and macrophages. Growth factors act on their respective cell surface receptors, expressed on haemopoietic cells at various stages of development, to maintain the haemopoietic progenitor cells and stimulate increased production of one or more cell types in response to stresses such as blood loss and infection (Fig. 38.3F).

These haemopoietic growth factors, including erythropoietin (EPO), interleukin (IL)-3, IL-6, IL-7, IL-11, IL-12, β-catenin, stem cell factor (Steel factor or C-kit ligand) and Fms-like tyrosine kinase 3, act

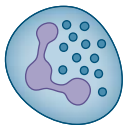
	Too many	Too few
Red cells	**Too many red blood cells** • Headache • Blurred vision • Could it be secondary? • Any hypoxia or dehydration? • Is there associated high white count, especially basophils? • Check JAK2 for primary polycythaemia	**Too few red blood cells** • Shortness of breath, fatigue • Chest pain: severe anaemia can exacerbate ischaemic heart disease and needs rapid treatment • Think about bone marrow failure • Are white cells low as well? • Look at trends over time – is this new or ongoing? • Ask about diet with haematinic deficiencies: vegans have poor intake of iron and B_{12}
White cells 	**Too many white cells** • Any recent infection? • High neutrophils are usually reactive • Is there a history of allergy, asthma, eczema? This raises eosinophil count • Remember to request a blood film so the morphology can be assessed • Immature white cells indicate malignant haematological disorders	**Too few white cells** • Any recent infection? This can be the cause or the effect of low white cells • Ask about mouth ulcers and family history • Benign neutropenia is seen in young females • Take a drug history: low white cells are a common side-effect • Make sure patients with neutrophils $<1 \times 10^9$ know who to call if they have a fever • Neutropenic sepsis is fatal if not treated rapidly with i.v. antibiotics
Platelets	**Too many platelets** • Headache, itching • Bleeding: very high platelet counts increase risk of bleeding • Is there evidence of infection or malignancy? • Secondary thrombocytosis is more common than primary • Look for splenomegaly: all myeloproliferative disorders are associated with this	**Too few platelets/abnormal clotting** • Bleeding: look for petechiae and blood blisters • Recent infection: immune thrombocytopenia often follows viral infection • Ask about rheumatological disorders • Take a drug history: low platelets are a common side-effect • Understand what platelet counts are safe for procedures • Platelet transfusions are rarely needed for immune thrombocytopenia • Bleeding history: ask about previous operations, labour, dental extractions, menstruation

Fig. 38.1 Key questions from the history and clinical signs are highlighted to guide the investigation of abnormal blood counts and common haematological presentations. *i.v.,* Intravenous; *JAK,* Janus kinase.

via specific receptors on cell surfaces to stimulate downstream signalling within the cell, such as the cytoplasmic Janus kinase pathway. This in turn activates signalling cascades within the cell, ultimately altering gene expression in the cell nucleus and thus influencing cell behaviour (e.g., to promote proliferation). Colony-stimulating factors (CSFs, with a prefix that indicates the cell type; Fig. 38.1), as well as ILs and EPO, regulate the lineage-committed progenitor cells.

Thrombopoietin (TPO) is produced in the kidneys, liver, and certain bone marrow stromal cells and acts to control platelet production, along with IL-6 and IL-11.

In addition to these factors, which stimulate haemopoiesis, there are other factors that inhibit the process including tumour necrosis factor (TNF) and transforming growth factor beta (TGF-β).

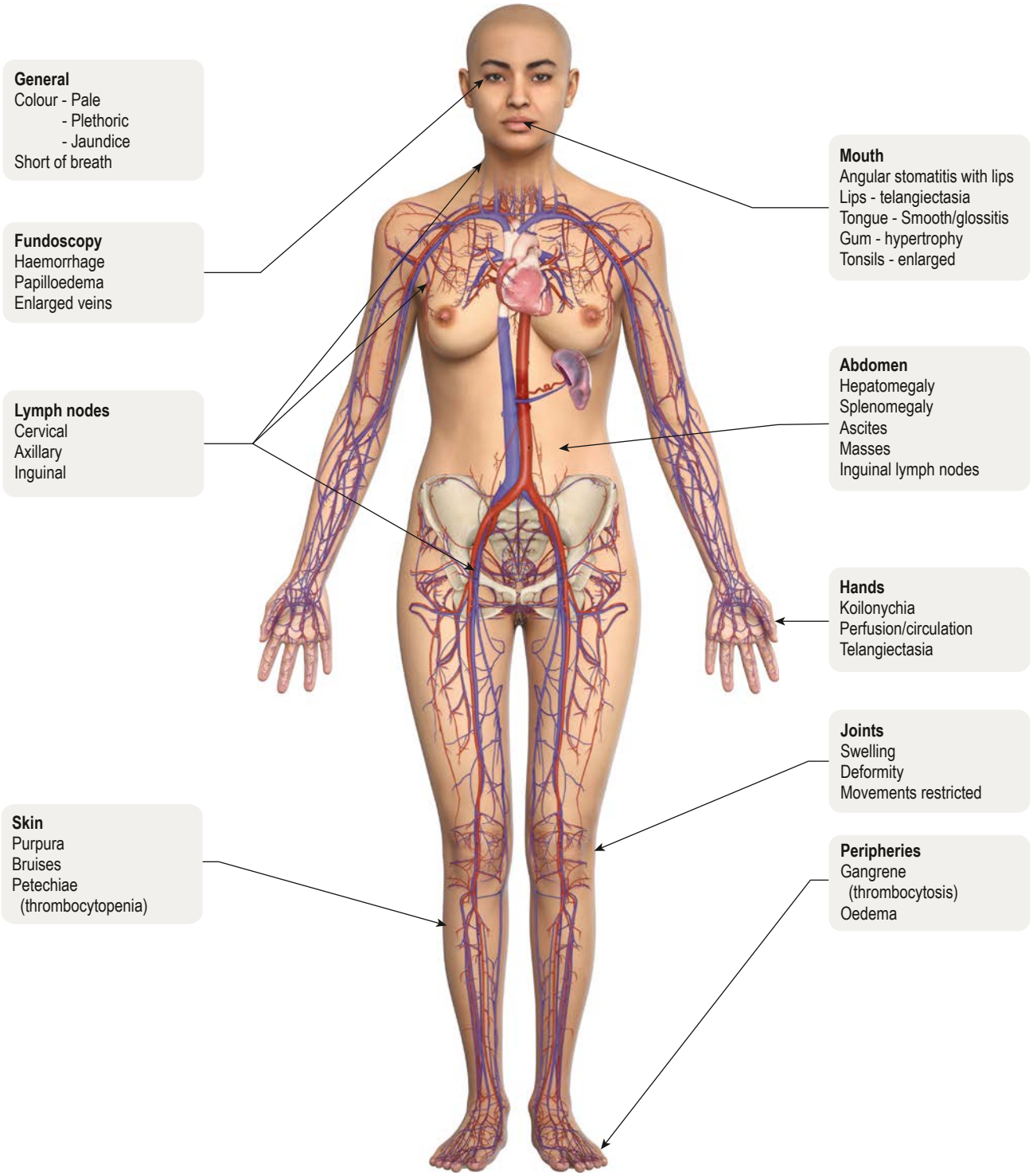


Fig. 38.2 **Clinical features of haematological disease**

Uses in treatment

Many growth factors have been produced by recombinant DNA techniques and are used clinically. Examples include:

- Granulocyte-colony-stimulating factor (G-CSF), used to accelerate haemopoietic recovery after chemotherapy and haemopoietic cell transplantation
- EPO, used to treat anaemia in patients with chronic kidney disease (CKD) and in patients with certain blood cancers
- TPO receptor agonists, used to treat patients with immune thrombocytopenic purpura (ITP).

Peripheral blood

Automated cell counters are used to measure the haemoglobin (Hb) concentration and numbers and sizes of RBCs, WBCs, and platelets (**Box 38.1**). Other indices can be derived from these values. A blood film examined by a professional experienced in blood cell morphology is an essential adjunct because definitive abnormalities of cells can be seen. This is particularly important in some haematological emergencies such as acute promyelocytic leukaemia (see Chapter 39) and thrombotic thrombocytopenic purpura (TTP;

Box 38.1 Reference values for peripheral blood

	Males	Females
Haemoglobin (g/L)	135–175	115–155
Packed cell volume (haematocrit; L/L)	0.4–0.54	0.37–0.47
Red blood cell count (10^{12}/L)	4.5–6.0	3.9–5.0
Mean corpuscular volume of red blood cell (fL)	80–96	
Mean corpuscular Hb (pg)	27–32	
Mean corpuscular Hb concentration (g/L)	320–360	
Red blood cell distribution width (%)	11–15	
White cell count (10^9/L)	4.0–11.0	
Platelets (10^9/L)	150–400	
Erythrocyte sedimentation rate (mm/h)	<20	
Reticulocytes	0.5%–2.5% (50–100 × 10^9/L)	

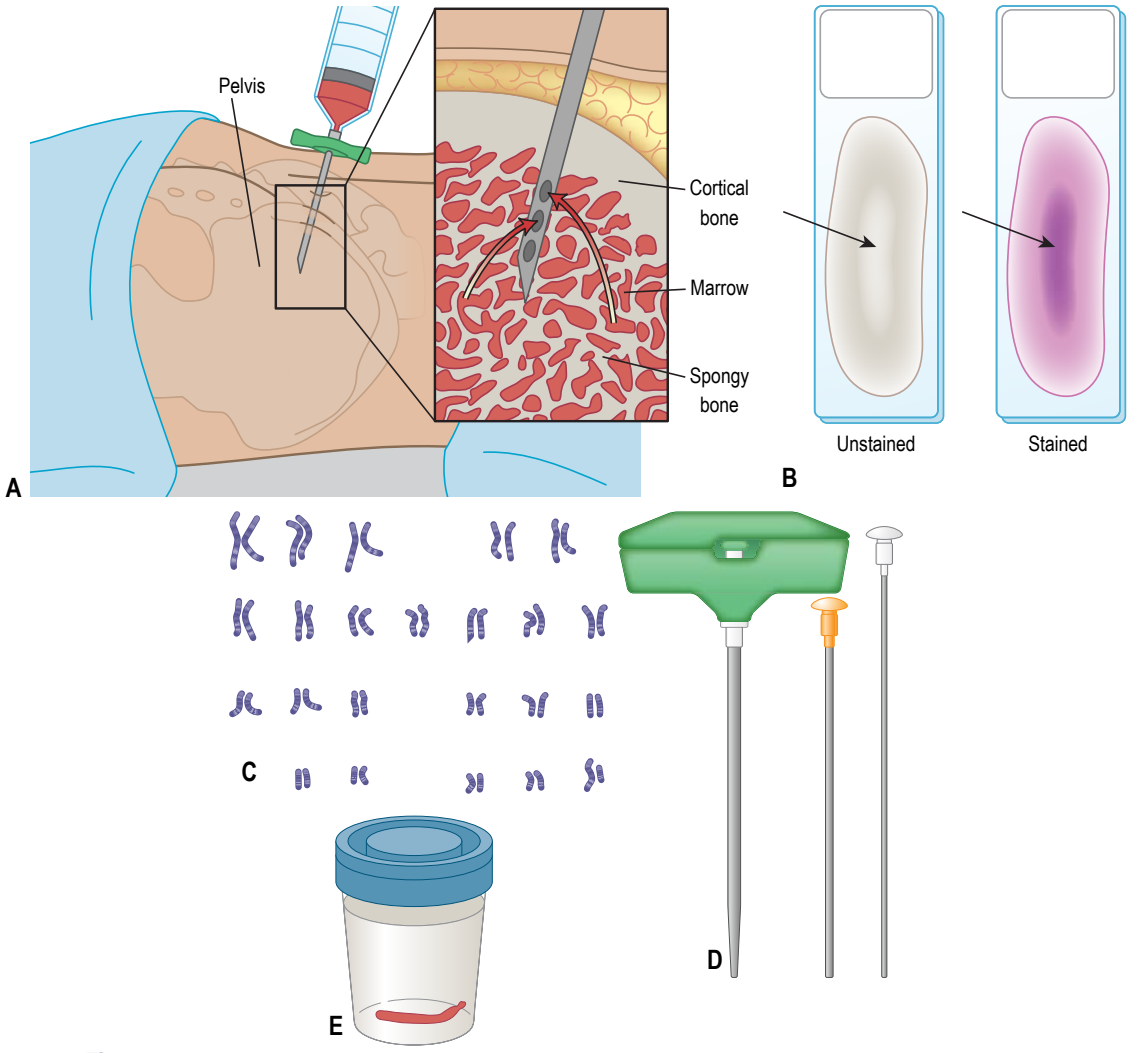


Fig. 38.3 Bone marrow aspiration and biopsy (trephine). **(A)** This is generally carried out as an outpatient procedure under local anaesthetic using aseptic technique. Informed consent is obtained. The skin overlying the superior posterior iliac crest is cleaned, and local anaesthetic is injected subcutaneously and around the periosteum, which is extremely well innervated. When the area is numb, an aspirate needle is used to obtain a liquid sample. **(B)** The sample is spread on a slide, stained, and examined under a microscope for morphology. **(C)** A sample is also sent for cytogenetic analysis, and a karyotype is generated, allowing identification of gross chromosomal rearrangements such as gain or loss of chromosomes or translocations. **(D)** A trephine needle is needed to obtain a bone biopsy. **(E)** The biopsy is a hard core of bone, which must be decalcified in chelating solution before being embedded in paraffin. Sections are then cut and stained, permitting the architecture of the marrow cells to be examined, and immunohistochemistry is carried out to detect abnormal cell populations.

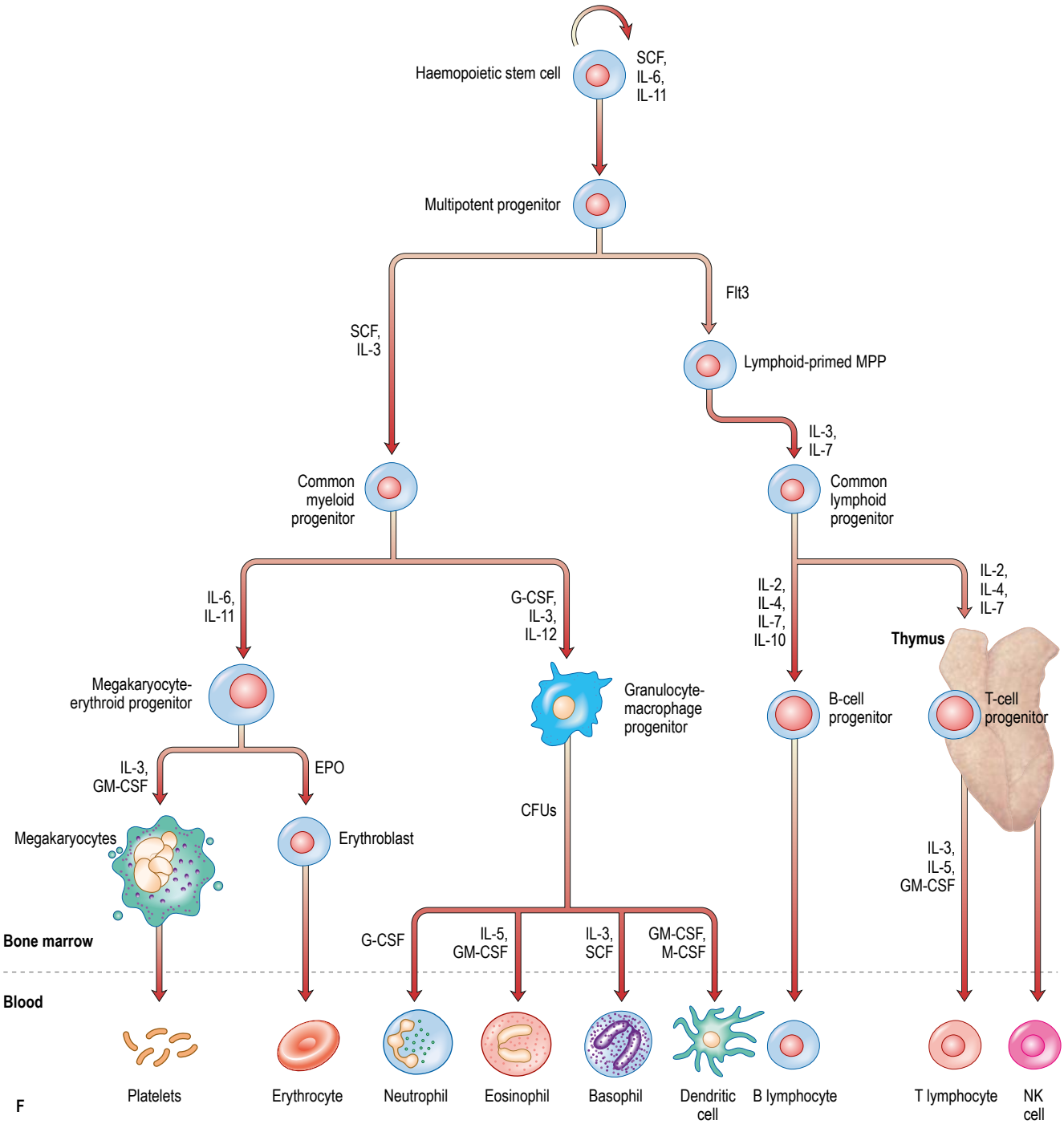


Fig. 38.3—Cont'd **(F)** Haemopoiesis arises from multipotent stem cells, which give rise to all different blood types according to the presence of internal and external regulatory molecules, hormones, and cytokines. As such, cell numbers are closely coordinated to respond rapidly to changes in requirements. *EPO,* Erythropoietin; *Flt3,* Fms-like tyrosine kinase 3; *G-CSF,* granulocyte-colony-stimulating factor; *GM-CSF,* granulocyte-macrophage-colony-stimulating factor; *IL,* interleukin; *MPP,* multipotent progenitor; *NK,* natural killer; *SCF,* stem cell factor; *TPO,* thrombopoietin.

see 'Thrombotic thrombocytopenic purpura' section), which require specific therapy within hours of diagnosis to prevent mortality.

- ***The mean corpuscular volume (MCV)*** of RBCs is a useful index in classification of anaemia (Fig. 38.4).
- ***The red cell distribution width (RDW)*** is a measure of variability in size of RBCs. An elevated RDW suggests increased variation in RBC size, e.g., anisocytosis in iron deficiency.
- ***The white cell count (WCC)*** (or white blood count [WBC]) gives the total number of circulating leucocytes. Differential counts specifically enumerate neutrophils, monocytes, lymphocytes, eosinophils, and basophils.
- ***Reticulocytes*** are young RBCs and usually make up less than 2.5% of all RBCs. The reticulocyte count gives a guide to erythroid activity in the bone marrow. An increased count is seen with

increased marrow activity (e.g., after haemorrhage or haemolysis) and during response to treatment with a specific haematinic. A low count in the presence of anaemia indicates an inappropriate response by the bone marrow and may be seen in bone marrow failure or deficiency of a haematinic.

- ***The erythrocyte sedimentation rate (ESR)*** is the rate of fall of RBCs in a column of blood and is a measure of the acute-phase response. A raised ESR reflects an increase in the plasma concentration of large proteins, such as fibrinogen and immunoglobulins. These proteins cause rouleaux formation, with RBCs clumping together and therefore falling more rapidly (Fig. 38.5). The ESR increases with age and is higher in females than males. Underlying pathology may be immunological, infective, ischaemic, malignant, or traumatic.
- ***Plasma viscosity*** is a measurement used instead of the ESR in some laboratories. It is also dependent on the concentration of large molecules such as fibrinogen and immunoglobulins and not affected by the concentration of Hb.
- ***C-reactive protein (CRP)*** is a protein produced in the acute-phase response. It is synthesized exclusively in the liver and rises within 6 hours of an acute event. CRP concentrations increase with fever (possibly triggered by IL-1, IL-6, TNF-α, and other cytokines), in inflammatory conditions, and after trauma. It follows the clinical state of the patient more rapidly than the ESR and is unaffected by the concentration of Hb but is less helpful than ESR or plasma viscosity in monitoring chronic inflammatory diseases such as systemic lupus erythematosus (SLE).

RED BLOOD CELLS

Erythropoiesis

RBC precursors pass through several stages in the bone marrow. The earliest morphologically recognizable cells are proerythroblasts. Smaller erythroblasts result from cell divisions, and precursors at each stage progressively contain less RNA and more Hb in the cytoplasm. The nucleus becomes more condensed and is eventually extruded from the late normoblast in the bone marrow when the cell becomes a reticulocyte.

- ***Reticulocytes*** contain residual ribosomal RNA and are still able to synthesize Hb. They remain in the marrow for about 1 or 2 days before release into the circulation, where they lose their RNA and become mature RBCs (erythrocytes) after another 1 or 2 days. Mature RBCs are non-nucleated biconcave discs.
- ***Late erythroblasts*** (nucleated RBCs) are not normally present in peripheral blood but are present in extramedullary haemopoiesis, cases of severe stress (e.g., severe infection, cardiac arrest), and some marrow disorders. About 10% of ***erythroblasts*** die in the bone marrow, even during normal erythropoiesis. Survival is dependent on erythropoietin (EPO) cell signalling. This 10% excess production of erythroblasts (ineffective erythropoiesis) ensures that a rapid drop in haemoglobin (e.g., acute haemorrhage or haemolysis) can lead to a rapid increase in terminally differentiated RBCs. Ineffective erythropoiesis is substantially increased in some anaemias, such as thalassaemia major and megaloblastic anaemia.
- ***EPO*** is a hormone that controls erythropoiesis. It is a heavily glycosylated polypeptide with a molecular weight of 30 400 Da produced in renal peritubular cells (90%) and in the liver (10%). Production is regulated mainly by tissue oxygen tension, increasing production in hypoxia (e.g., anaemia, cardiac or pulmonary disease). The *EPO* gene is one of several genes regulated by the hypoxic sensor pathway. The 3′-flanking region has a hypoxic response element necessary for induction of gene transcription in hypoxic cells. Hypoxia-inducible factor 1 is a transcription factor that binds to the hypoxia response element and acts as a master regulator of several genes responsive to hypoxia. EPO stimulates an increase in the proportion of bone marrow precursor cells committed to erythropoiesis by allowing a greater number of erythroid precursors to survive, and Colony Forming Unit-Erythroid (CFU-E) are stimulated to proliferate and differentiate. Increased production of EPO occurs in certain tumours such as renal cell carcinoma and when there is hypoxia, resulting in secondary erythrocytosis.

Haemoglobin synthesis

Hb performs the main functions of RBCs, carrying oxygen to the tissues and returning carbon dioxide from the tissues to the lungs. Each normal adult Hb molecule (HbA) has a molecular weight of 68 kDa and consists of two α- and two β-globin polypeptide chains ($\alpha_2\beta_2$). HbA makes up about 97% of the Hb in adults. Two other Hb types, HbA_2 ($\alpha_2\delta_2$) and HbF ($\alpha_2\gamma_2$), are found in adults in small amounts (1.5%–3.2% and <1%, respectively) (see Box 38.10).

Haem synthesis occurs in the mitochondria of the developing RBC (Fig. 38.6). The major rate-limiting step is the conversion of glycine and succinic acid to δ-aminolevulinic acid (ALA) by ALA synthase. Vitamin B_6 is a coenzyme for this reaction, which is inhibited by haem and stimulated by EPO. Two molecules of δ-ALA condense to form a pyrrole ring (porphobilinogen). These rings are then grouped in fours to produce protoporphyrins, and with the addition of iron, haem is formed. Haem is then inserted into the globin chains to form an Hb molecule. The structure of Hb is shown in Fig. 38.7.

Haemoglobin function

The biconcave shape of RBCs provides a large surface area for the uptake and release of oxygen and carbon dioxide. Hb becomes saturated with oxygen in the pulmonary capillaries, where the partial pressure of oxygen is high, and Hb has a high affinity for oxygen. Oxygen is released in the tissues where the partial pressure of oxygen is low, and Hb has a low affinity for oxygen.

In adult Hb, a haem group is bound to each of the four globin chains. Each haem group has a porphyrin ring with a ferrous atom, which can reversibly bind one oxygen molecule. The Hb molecule exists in two conformations, relaxed (R) and taut (T), corresponding to oxyhaemoglobin and deoxyhaemoglobin, respectively. The binding of one oxygen molecule to deoxyhaemoglobin increases the oxygen affinity of the remaining binding sites; this property is known as 'cooperativity' and is the reason for the sigmoid shape of the oxygen dissociation curve (Fig. 10.3). The binding of oxygen can also be influenced by secondary effectors such as hydrogen ions (acidity), carbon dioxide, and RBC bisphosphoglycerate (2,3-BPG). RBC metabolism produces 2,3-BPG from glycolysis. Binding of 2,3-BPG stabilizes the T conformation, reduces affinity for oxygen, and increases release of oxygen in the tissues. The P_{50} is the partial pressure of oxygen at which the Hb is 50% saturated with oxygen. When the primary limitation to oxygen transport is in the lungs (e.g., lung disease, high-altitude exposure), P_{50} is reduced to enhance oxygen loading. When the primary limitation is in the periphery (e.g., heavy exercise, anaemia) P_{50} is increased to enhance oxygen

unloading. Hydrogen ions and carbon dioxide added to blood also cause a reduction in the oxygen-binding affinity of Hb (Bohr effect), helping the exchange of carbon dioxide and oxygen in the tissues. A summary of normal RBC production and destruction is given in Fig. 38.8.

ANAEMIA

Anaemia is present when there is a decrease in Hb in the blood below age- and sex-specific reference limits (Box 38.1). Alterations in Hb concentration may occur because of changes in the plasma volume, as shown in Fig. 38.9. A reduction in plasma volume (e.g., dehydration) leads to a high Hb. An increase in plasma volume lowers the Hb concentration, even when this combined with a small increase in RBC volume, as occurs in pregnancy.

Anaemia can be classified in a variety of ways:

1. Decreased production or increased destruction
2. Inherited or acquired
3. By MCV (Fig. 38.4):
   - Hypochromic microcytic with a low MCV
   - Normochromic normocytic with a normal MCV
   - Macrocytic with a high MCV

Although there is some biological rationale for classification, some types of anaemia can present in more than one category.

Box 38.2 Techniques for obtaining bone marrow

Explain the technique to the patient and obtain consent.

Aspiration

- Site: usually the iliac crest
- Give local anaesthetic injection
- Use a special bone marrow needle (e.g., Salah)
- Aspirate marrow
- Make smear with glass slide
- Stain with:
  - Romanowsky technique
  - Perls' reaction (acid ferrocyanide) for iron

Trephine

Indications

- Better assessment of cellularity (e.g., aplastic anaemia)
- Better assessment of presence of infiltration or fibrosis

Technique

- Site: usually the posterior iliac crest
- Give local anaesthetic injection
- Use special needle (e.g., Jamshidi, which is longer and wider than for aspiration)
- Obtain core of bone
- Fix in formalin; decalcify (this takes a few days)
- Stain with:
  - Haematoxylin and eosin
  - Reticulin stain

Clinical features

Patients with anaemia may be asymptomatic. A slowly falling concentration of Hb allows for haemodynamic compensation and enhancement of the oxygen-carrying capacity of the blood. A rise in 2,3-BPG causes a shift of the oxygen dissociation curve to the right, so that oxygen is more readily given up to tissues. Where blood loss is rapid, more severe symptoms occur, particularly in older adults.

Symptoms

The symptoms of anaemia are non-specific and include breathlessness, fatigue, headaches, palpitations, and faintness (Fig. 38.1). Anaemia exacerbates cardiorespiratory problems (e.g., angina or intermittent claudication precipitated by anaemia). A good way to assess the effects of anaemia is to ask about breathlessness in relation to different levels of exertion such as walking on the flat or climbing a flight of stairs.

Signs

- Pallor
- Tachycardia
- Systolic flow murmur
- Cardiac failure

Specific signs seen in different types of anaemia include:

- Koilonychia: spoon-shaped nails seen in longstanding iron-deficiency anaemia
- Jaundice: haemolytic anaemia
- Bone deformities: thalassaemia major
- Leg ulcers: sickle cell disease

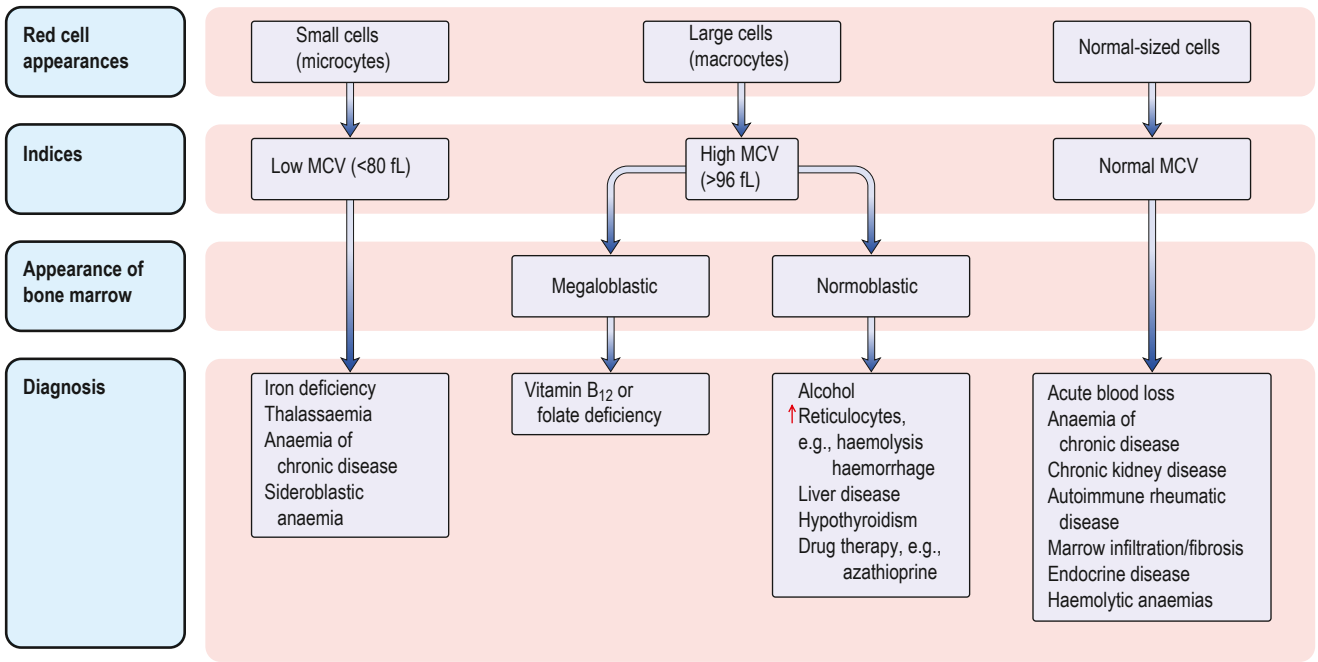


Fig. 38.4 Classification of anaemia. *fL,* Femtolitres; *MCV,* mean corpuscular volume; *RBC,* red blood cell.

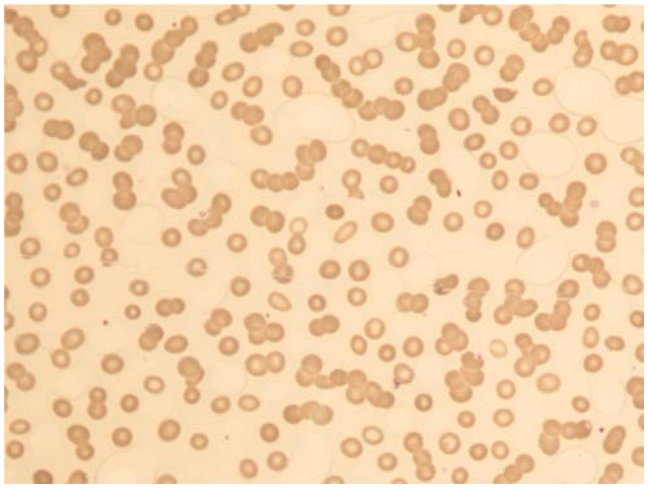


Fig. 38.5 Peripheral blood film (total magnification 400×) showing rouleaux formation in the red blood cells. In states of increased plasma protein concentration (e.g., infection, inflammation, paraprotein), the discoid red blood cells stack together like coins.

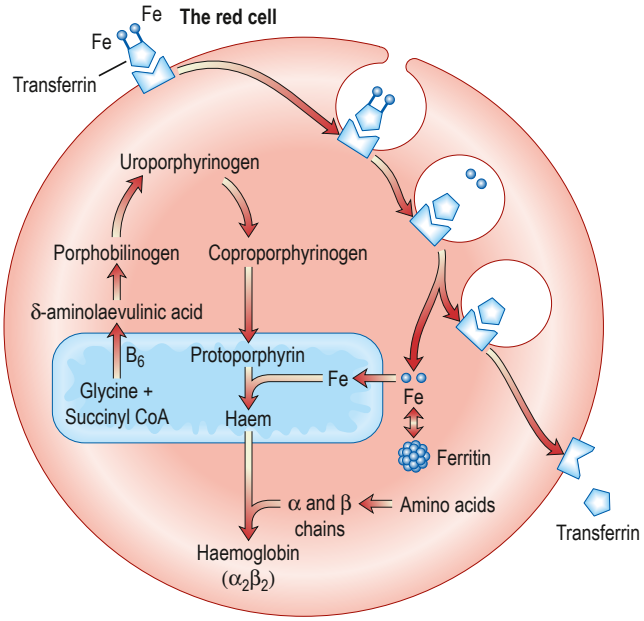


Fig. 38.6 Haemoglobin (Hb) synthesis. Transferrin attaches to a surface receptor on developing red blood cells. Iron is released and transported to the mitochondria, where it combines with protoporphyrin to form haem. Protoporphyrin itself is manufactured from glycine and succinyl CoA. Haem combines with α and β chains (formed on ribosomes) to make Hb.

Anaemia is not a final diagnosis. A cause should always be sought.

Investigations

Peripheral blood

A low Hb should be evaluated with:

- RBC indices
- WCC
- Platelet count
- Reticulocyte count (indicates marrow activity)
- Blood film: abnormal RBC morphology may indicate the diagnosis (Fig. 38.10).

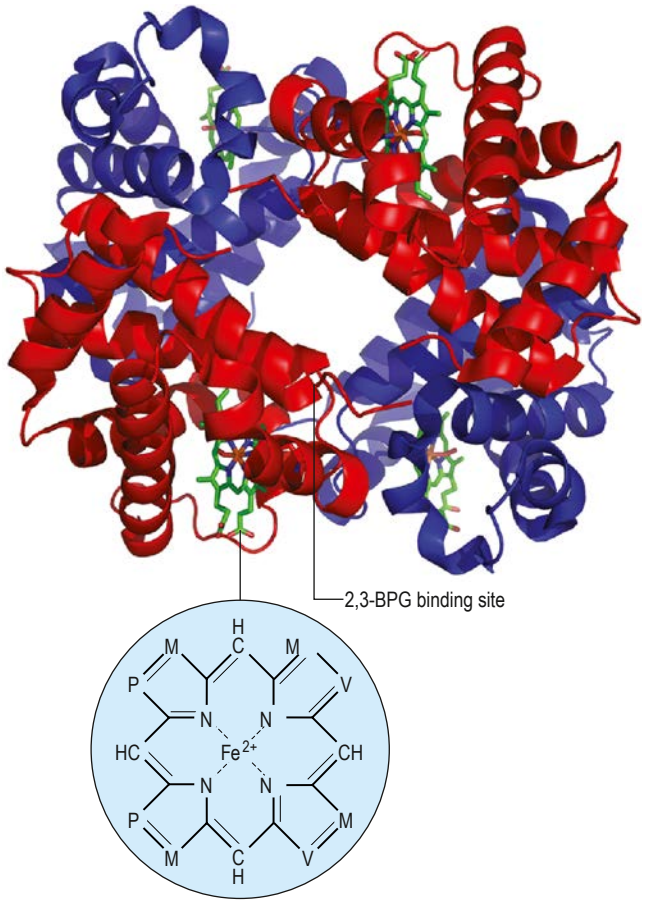


Fig. 38.7 Model of the haemoglobin molecule showing α *(red)* and β *(blue)* chains. Bisphosphoglycerate (2,3-BPG) binds in the centre of the molecule and stabilizes the deoxygenated form by cross-linking the β chains. *M,* Methyl; *P,* propionic acid; *V,* vinyl.

A dimorphic blood film is when two populations of RBCs are seen:

- 'Double deficiencies': e.g., combined iron and folate deficiency in coeliac disease or after treatment of anaemia with the appropriate haematinic when old RBCs exhibit the deficiency and new RBCs have a normal morphology
- After blood transfusion
- Inherited congenital sideroblastic anaemia (rare)

Bone marrow

Techniques for obtaining bone marrow are shown in Box 38.2. Bone marrow examination is performed to further investigate peripheral blood abnormalities:

- ***Aspiration:*** provides a film for microscopic examination of developing haemopoietic cell morphology. The liquid sample can also be used for immunophenotyping (assessment of the presence or absence of antigens on the surface of cells), cytogenetic investigations, assessment for molecular markers, or microbiological culture.
- ***Trephine:*** provides a core of bone that is processed as a histological specimen. It gives an overall view of bone marrow architecture, cellularity, and presence or absence of abnormal infiltrates. It can be also used for immunohistochemical investigations.

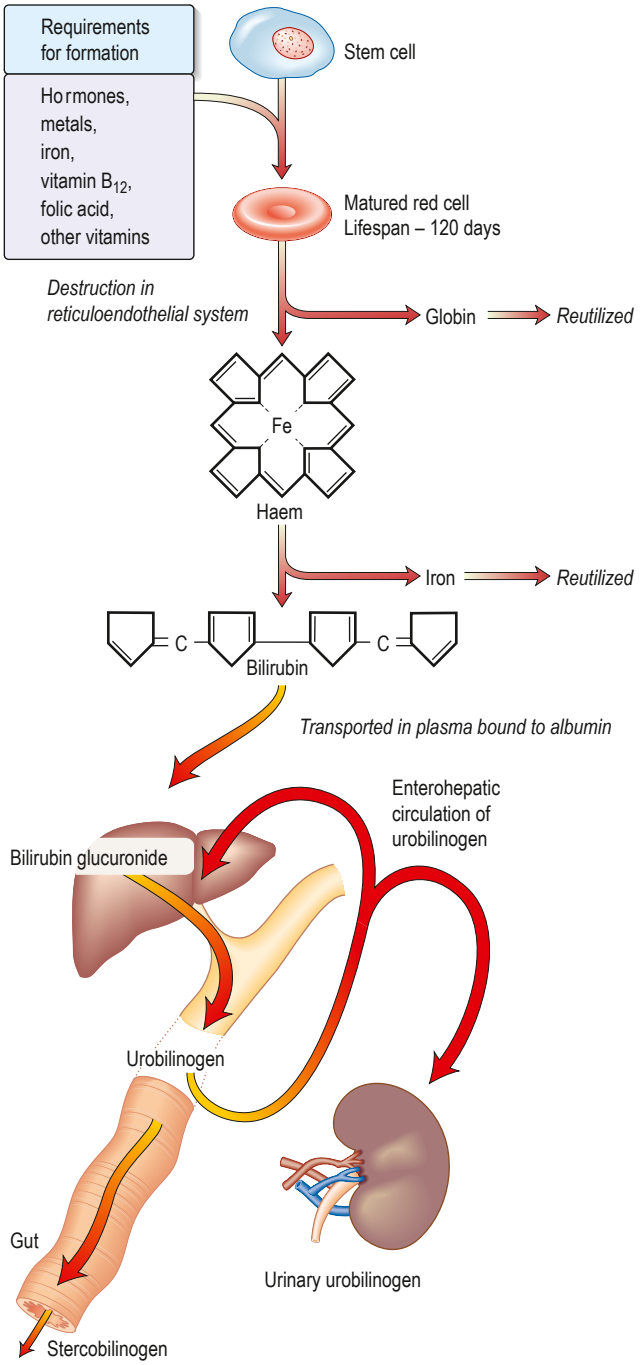


Fig. 38.8 **Red blood cell production and breakdown.**

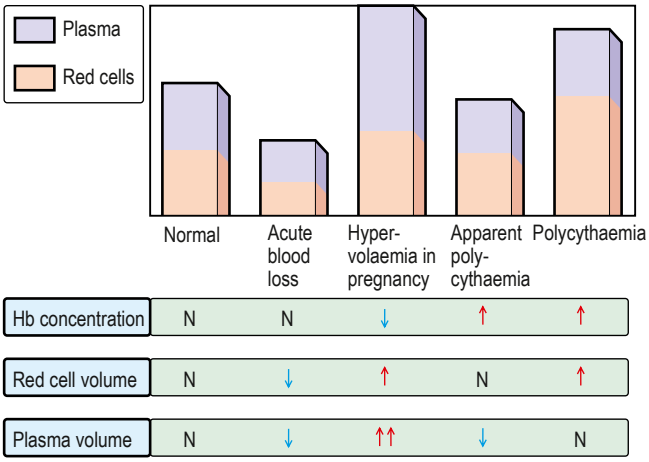


Fig. 38.9 **Alterations of haemoglobin (Hb) in relation to plasma.**

Fig. 38.10 **Hypochromic microcytic cells *(arrow)* on a blood film.** Poikilocytosis (variation in shape) and anisocytosis (variation in size) are seen.

The following are assessed:

- Cellularity of the marrow
- Type of erythropoiesis (e.g., normoblastic or megaloblastic)
- Morphology of all cell lines
- Cellularity of cell lines
- Infiltration of the marrow, e.g., presence of non-haemopoietic cells such as cancer cells
- Iron stores

MICROCYTIC ANAEMIA

Microcytic anaemia most commonly results from iron deficiency, the most common cause of anaemia globally, affecting 30% of the world's population. This occurs because of the body's limited ability to absorb iron and frequent loss of iron in bleeding. Although iron is abundant, most is in the insoluble ferric (Fe^{3+}) form, which has poor bioavailability. Ferrous (Fe^{2+}) iron is more readily absorbed. When anaemia of chronic disease is long-standing, iron restriction leads to microcytosis. Other causes of a microcytic hypochromic anaemia (Fig. 38.4) include defects in globin production (thalassaemia) or defects in haem synthesis, termed 'sideroblastic anaemia'.

Iron

Dietary intake

The average daily diet in the UK contains 15 to 20 mg of iron, although normally only 10% of this is absorbed. Absorption may be increased to 20% to 30% in iron deficiency and pregnancy. The main dietary iron is non-haem iron derived mainly from cereals commonly fortified with iron. Availability is affected by other dietary constituents such as calcium and tannins preventing absorption. Haem iron is derived from Hb and myoglobin in red or organ meats.

Absorption

Factors influencing iron and haem iron absorption are shown in Fig. 38.11 and Box 38.3.

Dietary haem iron is more rapidly absorbed than non-haem iron derived from vegetables and grains. Most haem is absorbed

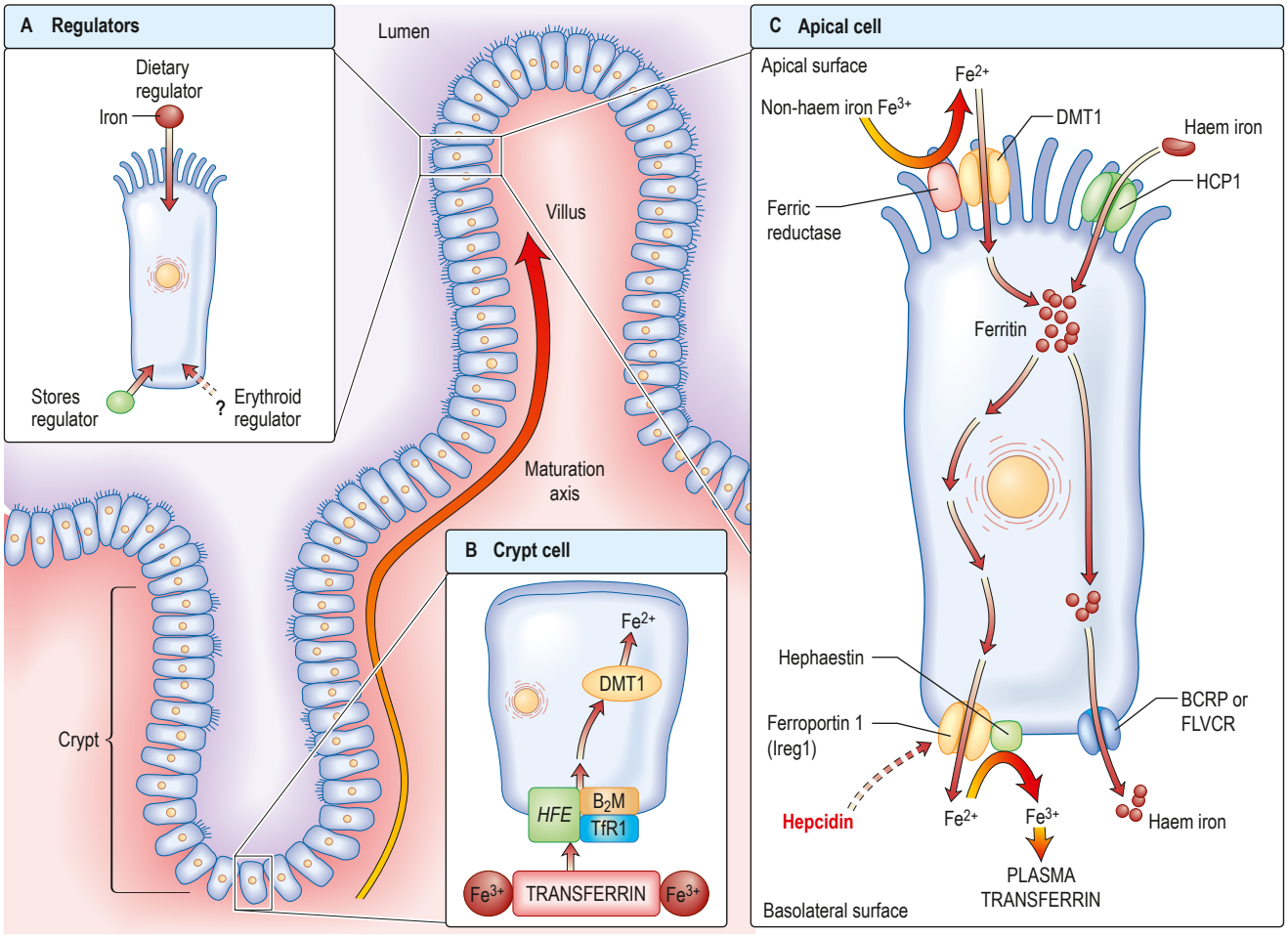


Fig. 38.11 Iron absorption. (A) *Regulation of the absorption of intestinal iron*. The iron-absorbing cells of the duodenal epithelium originate in the intestinal crypts and migrate toward the tip of the villus as they differentiate (maturation axis). Absorption of intestinal iron is regulated by at least three independent mechanisms. First, iron absorption is influenced by recent dietary iron intake *(dietary regulator)*. After a large dietary bolus, absorptive cells are resistant to iron uptake for several days. Second, iron absorption can be modulated considerably in response to body iron stores *(stores regulator)*. Third, a signal communicates the state of bone marrow erythropoiesis to the intestine *(erythroid regulator)*. **(B)** ***Duodenal crypt cells*** sense body iron status through the binding of transferrin to the *HFE/B_2M/TfR1* gene complex. Cytosolic enzymes change the oxidative state of iron from ferric (Fe^{3+}) to ferrous (Fe^{2+}). A decrease in crypt cell iron concentration upregulates the divalent metal transporter (DMT1). This increases as crypt cells migrate up the villus and become mature absorptive cells. **(C)** ***Apical cell***. Dietary iron is reduced from the ferric to the ferrous state by the brush border ferric reductase. ***DMT1*** facilitates iron absorption from the intestinal lumen. The export proteins (e.g., ferroportin 1, hephaestin) transfer iron from the enterocyte into the circulation, depending on hepcidin concentration. A second pathway absorbs intact haem iron into the circulation via breast cancer resistant protein (BCRP) and feline leukaemia virus subgroup C (FLVCR). *B_2M,* β_2-Microglobulin; *HCP1,* haem carrier protein-1; *HFE,* hereditary haemochromatosis gene; *TfR1,* transferrin receptor.

in the proximal intestine, with absorptive capacity decreasing distally. The intestinal haem transporter haem carrier protein 1 is highly expressed in the duodenum and is upregulated by hypoxia and iron deficiency. Some haem iron may be absorbed intact into the circulation via enterocytes by two exporter proteins, breast cancer resistant protein and feline leukaemia virus subgroup C (Fig. 38.11).

Non-haem iron absorption occurs primarily in the duodenum. Non-haem iron is dissolved in the low pH of the stomach and reduced from ferric to ferrous form by brush border ferric reductase. Cells in duodenal crypts sense the body's iron requirements and retain this information as they mature into cells capable of absorbing iron at the tips of the villi. Divalent metal transporter 1 (previously termed natural resistance-associated macrophage protein) transports iron and other metals across the apical (luminal) surface of the mucosal cells in the small intestine.

Inside the mucosal cell, iron may be transferred across the cell to the plasma or stored as ferritin. The body's iron status at the time the absorptive cell developed from the crypt cell is probably the crucial deciding factor. Iron stored as ferritin is lost into the gut lumen when mucosal cells are shed; this regulates iron homeostasis. Transport of iron across the basolateral surface of mucosal cells involves a transporter protein, ferroportin 1 through its iron-responsive element. This transporter protein requires hephaestin, an accessory multicopper protein (Fig. 38.11).

Box 38.3 Factors that influence iron absorption

- Haem iron is absorbed better than non-haem iron.
- Ferrous iron is absorbed better than ferric iron.
- Gastric acidity helps to keep iron in the ferrous state and soluble in the upper gut.
- Formation of insoluble complexes with phytate or phosphate decreases iron absorption.
- Iron absorption is increased with low iron stores and increased erythropoietic activity (e.g., bleeding, haemolysis, high altitude).
- Absorption is decreased in iron overload, except in hereditary haemochromatosis, in which it is increased.

The body iron content is closely regulated by control of iron absorption. There is no physiological mechanism for eliminating excess iron. The key molecule regulating iron absorption is ***hepcidin***, a 25-amino acid peptide synthesized in the liver. Hepcidin regulates the iron-exporting activity of ferroportin by binding to it causing internalization and degradation, thereby decreasing iron efflux from iron-exporting tissues into plasma. High concentrations of hepcidin that occur in inflammation states via inflammatory cytokines (e.g., IL-6) destroy ferroportin and limit iron absorption. Low concentrations of hepcidin in anaemia, low iron stores, and hypoxia encourage iron absorption. In haemochromatosis (Fig. 23.27), mutations in the genes *HFE, HJV,* and *TfR2* interrupt hepcidin synthesis. A deficiency of hepcidin in intestinal cells leads to less bound ferroportin and more iron release into the plasma, which accumulates in an uncontrolled fashion into target organs such as the liver, pancreas, heart, and endocrine glands.

Anaemias characterized by ineffective erythropoiesis (e.g., thalassaemia) are associated with excessive and inappropriate iron absorption. This is caused by down-regulation of hepcidin and up-regulation of ferroportin, mediated by a newly identified hormone, erythroferrone, produced by developing erythroblasts.

Transport in the blood

Normal serum iron concentration is about 13 to 32 μmol/L; there is diurnal variation, with higher concentrations in the morning. Serum iron is not used clinically for assessing iron status. Iron is transported in the plasma bound to transferrin, a protein synthesized in the liver. Each transferrin molecule binds two atoms of ferric iron and is normally one-third saturated. Most iron bound to transferrin comes from macrophages in the reticuloendothelial system and not from iron absorbed by the intestine. These recycle the iron obtained from senescent RBCs when Hb is broken down, a process that mostly occurs in the spleen. Transferrin-bound iron attaches by specific receptors to erythroblasts and reticulocytes in the marrow and the iron is removed (Fig. 38.6). In an average adult male, 20 mg of iron, chiefly obtained from RBC breakdown in the macrophages of the reticuloendothelial system, is incorporated into Hb every day.

Iron stores

About two-thirds of total body iron is in the circulation as Hb (2.5–3.0 g in a normal adult male). The rest is stored in reticuloendothelial cells, hepatocytes, and skeletal muscle cells (0.5–1.5 g). About two-thirds of this is stored as ferritin and one-third as haemosiderin. Small amounts of iron are found in plasma (~4 mg bound to transferrin), with some in myoglobin and enzymes.

Ferritin is a water-soluble complex of iron and protein. It is present in small amounts in plasma, proportional to the degree of iron loading. It is also an acute-phase protein and rises with infection and inflammation. It is more easily mobilized than haemosiderin for Hb formation.

Haemosiderin is an insoluble iron–protein complex found in macrophages in the bone marrow, liver, and spleen. Unlike ferritin, it is visible on light microscopy in tissue sections and bone marrow films after staining by Perls' reaction.

Requirements

Each day, 0.5 to 1.0 mg of iron is lost in faeces, urine, and sweat. Menstruation results in the loss of 30 to 40 mL of blood per cycle, an average of about 0.5 to 0.7 mg of iron per day. Menstrual blood loss more than 100 mL usually results in iron deficiency because increased iron absorption from the gut cannot compensate. Demand for iron increases during growth (~0.6 mg/day) and pregnancy (1–2 mg/day).

Increases in the body iron content (haemochromatosis) are classified into:

- ***Hereditary haemochromatosis:*** a mutation in the *HFE* gene or other iron-controlling proteins causes uncontrolled increased iron absorption.
- ***Secondary haemochromatosis*** (transfusion siderosis): iron overload in conditions treated by regular blood transfusion.
- ***Non-transfusional iron overload:*** ineffective erythropoiesis drives uncontrolled iron absorption from the gut via erythroferrone, e.g., thalassaemia, sideroblastic anaemia, congenital dyserythropoietic anaemia.

Iron deficiency

Iron-deficiency anaemia develops when there is inadequate iron for Hb synthesis. The causes are:

- Increased loss: blood loss
- Increased demands (e.g., growth and pregnancy)
- Decreased absorption (e.g., post-gastrectomy, poor dietary intake).

Most iron deficiency is caused by blood loss, usually from the uterus or gastrointestinal (GI) tract. A common cause of iron deficiency worldwide is blood loss from the GI tract resulting from parasites (e.g., hookworm). A poor diet, predominantly containing vegetables, contributes to the high prevalence of iron deficiency in low-resource settings. Iron deficiency is not uncommon in infancy, when iron intake is insufficient for the demands of growth. It is more prevalent in preterm infants and when introduction of solid food is delayed.

Clinical features

Symptoms of anaemia are described earlier (see 'Anaemia' section). Features of longstanding iron deficiency include:

- Brittle nails
- Spoon-shaped nails (koilonychia)
- Atrophy of the papillae of the tongue
- Angular stomatitis
- Brittle hair
- A syndrome of dysphagia and glossitis (Plummer–Vinson or Paterson–Brown–Kelly syndrome).

The diagnosis of iron-deficiency anaemia requires a clinical history, including dietary intake, use of non-steroidal anti-inflammatory drugs (NSAIDs) associated with GI bleeding, and presence of blood in the faeces. In females, careful enquiry should be made about the duration of periods, the occurrence of clots, and the number of sanitary towels or tampons used (three to five per day is normal).

Box 38.4 Microcytic anaemia: differential diagnosis

	Iron deficiency	Anaemia of chronic disease	Thalassaemia trait (α or β)	Sideroblastic anaemia
MCV	Decreased	Low normal or normal	Very low for degree of anaemia	Decreased in inherited type but often increased in acquired type
Serum iron	Decreased	Decreased	Normal	Increased
Serum TIBC	Increased	Decreased	Normal	Normal
Serum ferritin	Decreased	Normal or increased	Normal	Increased
Serum soluble transfer receptors	Increased	Normal	Normal or increased	Normal or increased
Iron in marrow	Absent	Present	Present	Present
Iron in erythroblasts	Absent	Absent or decreased	Present	Ring forms

MCV, Mean corpuscular volume of red blood cells; *TIBC,* total iron-binding capacity.

Investigations

- ***Blood count and film*** (Fig. 38.10). RBCs are microcytic (MCV <80 fL) and hypochromic (mean corpuscular haemoglobin (MCH) <27 pg). There are poikilocytosis (variation in shape) and anisocytosis (variation in size).
- ***Transferrin saturation.*** Serum iron is not helpful in the assessment of clinical iron status. Transferrin saturation is a more accurate measure: iron deficiency is present when this decreases below 19% (Box 38.4).
- ***Serum ferritin*** concentration reflects stored iron. Reference values are 30 to 300 μg/L (11.6–144 nmol/L) in males and 30 to 200 μg/L (5.8–96 nmol/L) in premenopausal females. In simple iron deficiency, a low serum ferritin confirms the diagnosis. However, ferritin is an acute-phase reactant and concentrations increase in inflammatory or malignant diseases, and in liver damage. These can result in a normal or raised ferritin concentration in iron deficiency. Very high concentrations of ferritin may occur in hepatitis and haemophagocytic lymphohistiocytosis.
- ***Serum soluble transferrin receptors.*** The number of transferrin receptors released into the serum from bone marrow erythroblasts increases in iron deficiency and correlate with estimation of iron stores from bone marrow aspiration. This can help distinguish iron deficiency from anaemia of chronic disease (Box 38.4) and may avoid the need for bone marrow examination. It may be helpful in the investigation of complicated causes of anaemia. However, it is not readily available in routine clinical practice.
- ***Serum hepcidin concentrations*** are possible in research settings but are not routinely available because of a lack of standardization of assays.
- ***Other investigations*** as indicated by clinical history and examination. Investigations of the GI tract are often required to determine the cause of iron deficiency.

Differential diagnosis

Anaemia with microcytosis and hypochromia does not always indicate iron deficiency. The most common other causes are thalassaemia, sideroblastic anaemia, and anaemia of chronic disease. In these disorders, iron stores are normal or increased. The differential diagnosis of microcytic anaemia is shown in Box 38.4.

Management

The correct management of iron deficiency is to find and treat the underlying cause and to give iron to correct anaemia and replace iron stores. Taking iron increases the Hb concentration of patients with iron deficiency by approximately 10 g/L per week unless other factors, such as bleeding, are present.

Oral iron is all that is required in most cases. The best preparation is ferrous sulphate. A 200-mg tablet provides 60 mg of elemental iron. Emerging evidence supports a quick rise in serum hepcidin concentrations after an oral iron dose, suppressing the absorption of subsequent doses. Prescription of oral iron replacement should therefore be no more than once daily. Optimizing absorption includes taking the tablet 30 minutes before food with a small amount of orange juice. (Vitamin C helps convert ferric to ferrous iron, aiding absorption.) Tannins in tea inhibit iron absorption and should be avoided.

In low-income countries, distribution of iron tablets and fortification of food are the main approaches. However, iron supplementation programmes have been ineffective, mainly because of poor adherence. There are also concerns regarding an increased risk of infection in areas where malaria and other parasites are endemic.

Oral iron should be given for long enough to correct the Hb concentration and to replenish iron stores; this can take 6 months. The most common causes of failure to respond to oral iron are:

- Reduced adherence
- Ongoing blood loss
- Incorrect diagnosis e.g., thalassaemia trait.

These possibilities should be considered before parenteral iron is used. Parenteral iron is indicated in those with intolerance to oral preparations, severe malabsorption, and chronic disease (e.g., inflammatory bowel disease). Iron stores are replaced faster with parenteral iron than with oral iron, but the haematological response is no quicker. Parenteral iron can be given by slow intravenous infusion of low-molecular-weight iron dextran (requires a test dose because of a risk of anaphylactoid reactions), iron sucrose, ferric carboxymaltose, or iron isomaltoside 1000; oral iron should be discontinued when intravenous treatment is given.

Anaemia of chronic disease

One of the most common types of anaemia, particularly in hospitalized patients, is anaemia of chronic disease, occurring in chronic inflammatory diseases such as Crohn's disease, rheumatoid arthritis, SLE, polymyalgia rheumatica, and malignancy or with chronic infections such as tuberculosis (TB). There are decreased release of iron from the bone marrow to developing erythroblasts, an inadequate EPO response to anaemia, and decreased RBC survival.

Serum iron and total iron-binding capacity are low. Serum ferritin is normal or raised because of the inflammatory process. Serum soluble transferrin receptor concentration is normal

(Box 38.4). Stainable iron is present in the bone marrow, but iron is not seen in developing erythroblasts. Patients do not respond to oral iron therapy. Treatment is, in general, that of the underlying disorder, though intravenous iron may help. Recombinant EPO therapy is used in renal anaemia (see Chapter 26) and occasionally in inflammatory disease (rheumatoid arthritis, inflammatory bowel disease).

High concentrations of hepcidin expression play a key role (see earlier). Because numerous pathogens are dependent on iron for multiplication, responding to infection and inflammation by increasing hepcidin concentrations and sequestering iron from the circulation is evolutionarily advantageous; IL-6 production during infection is a potent driver of hepcidin expression. However, hepcidin concentrations lead to poor delivery of iron to the marrow and hence anaemia. Measurement of hepcidin concentrations is emerging as a useful test to help distinguish anaemia of chronic disease from iron-deficiency anaemia.

Sideroblastic anaemia

Sideroblastic anaemias are inherited or acquired disorders characterized by a refractory anaemia (that does not respond to haematinic supplementation), a variable number of hypochromic cells in the peripheral blood, and excess iron and ring sideroblasts in the bone marrow. The presence of ring sideroblasts is the diagnostic feature of sideroblastic anaemia. Accumulation of iron in the mitochondria of erythroblasts is caused by disordered haem synthesis, forming a ring of iron granules around the nucleus that can be seen with Perls' reaction. The blood film is often dimorphic; ineffective haem synthesis is responsible for the microcytic hypochromic cells. Sideroblastic anaemias can be inherited as an X-linked or autosomal recessive forms. A structural defect in δ-ALA synthase, the pyridoxine-dependent enzyme responsible for the first step in haem synthesis (Fig. 38.6), is responsible for the X-linked form. Acquired causes include myeloproliferative disorders and myeloid leukaemia. Primary acquired sideroblastic anaemia is one of the myelodysplastic syndromes and is responsible for most cases of sideroblastic anaemia in adults. It can also be caused by drugs (e.g., isoniazid), alcohol misuse, or lead toxicity and can occur in other disorders such as rheumatoid arthritis, carcinomas, and megaloblastic and haemolytic anaemias.

Management

Some patients respond when drugs or alcohol are withdrawn if these are the causative agents. In occasional cases, there is a response to pyridoxine, particularly in the X-linked inherited form. Treatment with folic acid may be required to treat accompanying folate deficiency.

Further reading

Adams PC, Jeffrey G, Ryan J. Haemochromatosis. *Lancet* 2023; 401:1811–1821.
Camaschella C. Iron-deficiency. *Blood* 2019; 133:30–39.
Cullis JO, Fitzsimons EJ, Griffiths WJ, Tsochatzis E, Thomas DW. Investigation and management of a raised serum ferritin. *Br J Haematol* 2018; 181:331–340.
Pasricha SR, Tye-Din J, Muckenthaler MU, Swinkels DW. Iron deficiency. *Lancet* 2021; 397:233–248.
Pavord S, Daru J, Prasannan N et al. UK guidelines on the management of iron deficiency in pregnancy. *Br J Haematol* 2020; 188:819–830.
Stoffel NU, Cercamondi CI, Brittenham G et al. Iron absorption from oral iron supplements given on consecutive versus alternate days and as single morning doses versus twice-daily split dosing in iron-depleted women: two open-label, randomised controlled trials. *Lancet Haematol* 2017; 4:e524–e533.
Weiss G, Ganz T, Goodnough LT. Anaemia of inflammation. *Blood* 2019; 133:40–50.

NORMOCYTIC ANAEMIA

Normocytic, normochromic anaemia (Fig. 38.4) is seen in:

- Anaemia of chronic disease
- Endocrine disorders, e.g., hypopituitarism, hypothyroidism, and hypoadrenalism
- Haematological disorders, e.g., aplastic anaemia and some haemolytic anaemias
- Early blood loss
- Inherited causes, e.g., Diamond-Blackfan anaemia, which usually presents in the first year of life and is associated with an extremely low reticulocyte count.

MACROCYTIC ANAEMIAS

These can be divided into megaloblastic and non-megaloblastic types, depending on bone marrow findings.

Megaloblastic anaemia

Megaloblastic anaemia is characterized by the presence erythroblasts in the bone marrow with delayed nuclear maturation because of defective DNA synthesis (megaloblasts). Megaloblasts are large and have large immature nuclei. The nuclear chromatin is more finely dispersed than normal and has an open, stippled appearance (Fig. 38.12). In addition, giant metamyelocytes are frequently seen. These cells are about twice the size of normal cells and often have twisted nuclei. Megaloblastic changes occur in:

- Vitamin B_{12} deficiency and abnormal vitamin B_{12} metabolism
- Folic acid deficiency and abnormal folate metabolism
- Other defects of DNA synthesis
  - Congenital enzyme deficiencies in DNA synthesis, e.g., orotic aciduria
  - Drugs interfering with DNA synthesis, e.g., hydroxycarbamide (hydroxyurea), azathioprine
- Myelodysplasia caused by dyserythropoiesis.

Haematological findings

- The MCV is characteristically greater than 96 fL unless there is a coexisting cause of microcytosis, in which case there may be a dimorphic picture with a normal or low average MCV.

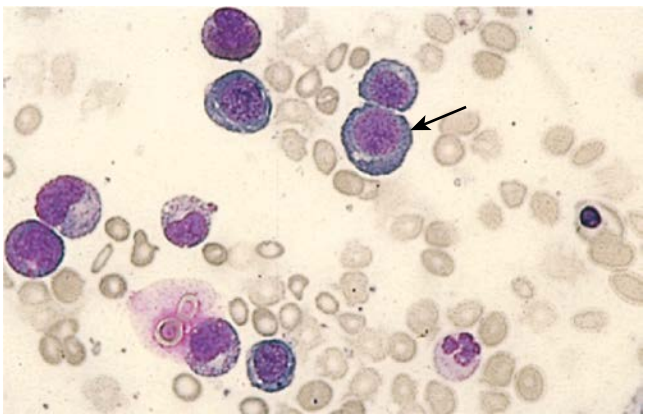


Fig. 38.12 Megaloblasts *(arrow)* in the bone marrow.

- The peripheral blood film shows oval macrocytes with hypersegmented polymorphs with six or more lobes in the nucleus (Fig. 38.13).
- If severe, there may be leucopenia and thrombocytopenia.
- Lactate dehydrogenase (LDH) is typically elevated, reflecting ineffective erythropoiesis.

Biochemical basis of megaloblastic anaemia

The key biochemical problem common to both vitamin B_{12} and folate deficiencies is a block in DNA synthesis owing to an inability to methylate deoxyuridine monophosphate to deoxythymidine monophosphate, which is used to build DNA (Fig. 38.14). The methyl group is supplied by the folate coenzyme methylene tetrahydrofolate (THF). Deficiency of folate reduces the supply of this coenzyme; deficiency of vitamin B_{12} slows the demethylation of methyltetrahydrofolate (methyl THF) and prevents cells from receiving THF for synthesis of methylene TFH polyglutamate.

Other congenital and acquired forms of megaloblastic anaemia are caused by interference with purine or pyrimidine synthesis, inhibiting DNA synthesis.

Vitamin B_{12} (cobalamin)

Vitamin B_{12} is synthesized by certain microorganisms. Humans are dependent on animal sources. It is found in meat, fish, eggs and milk but not in plants. Vitamin B_{12} is not usually destroyed by cooking. The average daily diet contains 5 to 30 μg of vitamin B_{12}, of which 2 to 3 μg is absorbed. The average adult stores 2 to 3 mg, mainly in the liver. It may take 2 years or more after absorptive failure before vitamin B_{12} deficiency develops because daily losses are small (1–2 μg).

Vitamin B_{12} consists of a small group of compounds, the cobalamins, which are composed of a planar group with a central cobalt atom (corrin ring) and a nucleotide set at right angles. Vitamin B_{12} was first crystallized as cyanocobalamin, but the main natural cobalamins have deoxyadenosyl-, methyl-, and hydroxocobalamin groups attached to the cobalt atom. Deoxyadenosylcobalamin is a coenzyme for the conversion of methylmalonyl CoA to succinyl CoA. Hence, measurement of methylmalonic acid was once used as a test for vitamin B_{12} deficiency, but this is no longer routine.

Absorption and transport

Vitamin B_{12} is liberated from protein complexes in food by gastric enzymes. It then binds to a vitamin B_{12}–binding protein ('R' binder) related to plasma transcobalamin I (TCI) and derived from saliva.

Fig. 38.13 Macrocytes and a hypersegmented neutrophil *(arrow)* on a peripheral blood film.

Vitamin B_{12} is released from the 'R' binder by pancreatic enzymes and then becomes bound to intrinsic factor. Intrinsic factor is a GP with a molecular weight of 45 kDa secreted by gastric parietal cells along with H^+ ions. It combines with vitamin B_{12} and carries it to a specific receptor on the surface of the mucosa of the ileum, cubilin. Vitamin B_{12} enters the ileal cells and intrinsic factor remains in the lumen and is excreted. Vitamin B_{12} is transported from the enterocytes to the bone marrow and other tissues by the GP transcobalamin II (TCII). Vitamin B_{12} bound to TCII is known as holotranscobalamin or 'active B_{12}' because this is the form taken up by cells. Although TCII is the essential carrier protein for vitamin B_{12}, the amount of B_{12} on TCII is low. Vitamin B_{12} in plasma is mainly bound to TCI (70%–90%). Only about 1% of an oral dose of B_{12} is absorbed 'passively' without the need for intrinsic factor.

Vitamin B_{12} deficiency

There are several causes of B_{12} deficiency and abnormal B_{12} metabolism (Box 38.5). The most common cause of vitamin B_{12} deficiency in adults is pernicious anaemia (PA). Malabsorption of vitamin B_{12} because of pancreatitis, coeliac disease, or treatment with metformin is mild and does not usually result in significant deficiency.

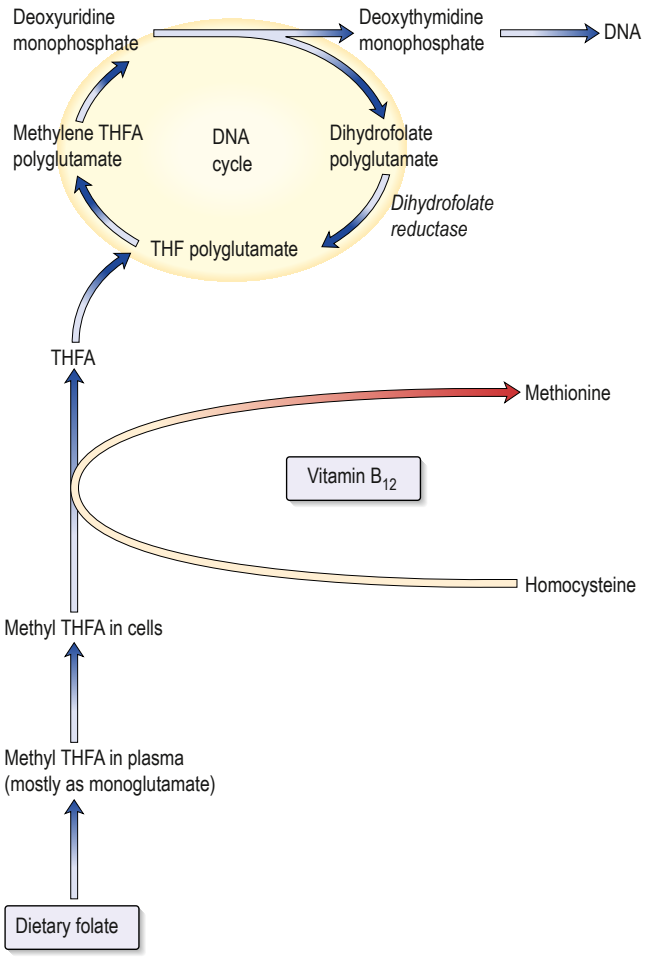


Fig. 38.14 Biochemical basis of megaloblastic anaemia. The metabolic relationship between vitamin B_{12} and folate and their role in DNA synthesis. Megaloblastic anaemia is characterized by erythroblasts with delayed nuclear maturation caused by defective DNA synthesis. *THFA,* Tetrahydrofolate.

Box 38.5 Vitamin B_{12} deficiency and abnormal B_{12} utilization

Low dietary intake
- Veganism

Impaired absorption

Stomach
- Pernicious anaemia
- Gastrectomy or bariatric surgery
- Congenital deficiency of intrinsic factor

Small bowel
- Ileal disease or resection
- Bacterial overgrowth
- Tropical sprue
- Fish tapeworm *(Diphyllobothrium latum)*

Abnormal utilization
- Congenital transcobalamin II deficiency
- Nitrous oxide (inactivates vitamin B_{12})

Pernicious anaemia

Pernicious anaemia (PA) is an autoimmune disorder with atrophic gastritis, loss of parietal cells in the gastric mucosa, consequent failure of intrinsic factor production, and vitamin B_{12} malabsorption. This disease is common in older adults: about 1 in 8000 aged older than 60 years are affected in the UK. It is seen in all ethnicities and is more common in females than males.

Pathogenesis

There is an association with other autoimmune diseases, particularly thyroid disease, Addison's disease, and vitiligo. Approximately 50% of all patients with PA have thyroid antibodies. There is a higher incidence of gastric carcinoma with PA (1%–3%) than in the general population.

Parietal cell antibodies are present in the serum in 90% of patients with PA but are also found in 10% of normal individuals. Conversely, intrinsic factor antibodies, although found in only 50% of patients with PA, are specific for this diagnosis.

Pathology

Fundal autoimmune gastritis is present, with plasma cell and lymphoid infiltration. The parietal and chief cells are replaced by mucin-secreting cells. Affected patients have achlorhydria and absent secretion of intrinsic factor. The histological abnormality can be improved by corticosteroid therapy, which supports an autoimmune basis for the disease.

Clinical features

The onset of PA is insidious, with progressively increasing symptoms of anaemia. Patients are sometimes said to have a lemon-yellow colour owing to a combination of pallor and mild jaundice caused by excess breakdown of Hb. A red, sore tongue (glossitis) and angular stomatitis are sometimes present.

Neurological changes, if left untreated for a long time, can be irreversible. Neurological abnormalities occur only with very low concentrations of serum B_{12} (<60 ng/L or 50 pmol/L) and occasionally in patients who are not clinically anaemic. The classical neurological features are those of a polyneuropathy progressively involving peripheral nerves and the posterior and eventually lateral columns of the spinal cord (subacute combined degeneration). Patients present with symmetrical paraesthesiae in the fingers and toes, early loss of vibration sense and proprioception, progressive weakness, and ataxia. Paraplegia may result. Dementia, psychiatric problems, hallucinations, delusions, and optic atrophy may also occur.

Investigations

- ***Haematological findings*** show features of a megaloblastic anaemia.
- ***Bone marrow*** shows typical features of megaloblastic erythropoiesis (Fig. 38.12), although it is frequently not sampled in cases of straightforward macrocytic anaemia with a low serum vitamin B_{12}.
- ***Serum bilirubin and LDH*** may be raised because of ineffective erythropoiesis. Normally, a minor fraction of serum bilirubin results from premature breakdown of newly formed RBCs in the bone marrow. In many megaloblastic anaemias, in which destruction of developing RBCs increased, serum bilirubin can be elevated.
- ***Serum vitamin B_{12}*** is usually well below 160 ng/L, which is the lower end of the reference range. Serum vitamin B_{12} can be assayed using radio-isotope dilution or immunological assays. Numerous factors can affect B_{12} concentrations, such as pregnancy and the oral contraceptive pill, which lower it.
- ***Holotranscobalamin*** is the 'active' fraction of cobalamin. Measurement may be a better marker for vitamin B_{12} deficiency than serum vitamin B_{12} concentration.
- ***Serum methylmalonic acid and homocysteine*** are raised in B_{12} deficiency, but testing is recommended only in complex cases (e.g., a strong suspicion of vitamin B_{12} deficiency with normal vitamin B_{12} concentration).

Absorption tests

Vitamin B_{12} absorption tests are no longer performed in the UK because radioactive vitamin B_{12} is not available.

Gastrointestinal investigations

In PA, there is achlorhydria. Intubation studies can be performed to confirm this but are rarely carried out in routine practice. Endoscopy or barium meal examination of the stomach is performed only if gastric symptoms are present.

Differential diagnosis

Vitamin B_{12} deficiency must be differentiated from other causes of megaloblastic anaemia, principally folate deficiency, but usually this is quite clear from blood concentrations.

PA should be distinguished from other causes of vitamin B_{12} deficiency by testing for intrinsic factor antibodies (Box 38.5). Patients negative for intrinsic factor antibodies with no other cause of vitamin B_{12} deficiency may still have PA.

Folic acid

Folic acid monoglutamate is not present in nature but occurs as polyglutamates. Folates present in food are polyglutamates in reduced dihydrofolate or THF forms. Polyglutamates are broken down to monoglutamates in the upper GI tract. During the absorptive process, these are converted to methyl THF monoglutamate, which is the main form in the serum. The first step in which methyl THF entering cells from the plasma is converted into folate polyglutamates involves methylation of homocysteine to methionine and requires both methylcobalamin and methyl THF as coenzymes. Intracellular polyglutamates are the active forms of folate and act as coenzymes in the transfer of single carbon units in amino acid metabolism and DNA synthesis (Fig. 38.14).

Dietary intake

Folate is found in green vegetables (e.g., spinach, broccoli) and offal (e.g., liver, kidney). Cooking loses 60% to 90% of folate. The minimal daily requirement is about 100 μg.

Folate deficiency

Causes of folate deficiency are shown in **Box 38.6**. The main cause is poor intake, which may occur alone or in combination with excessive utilization or malabsorption. The body's reserves of folate are about 10 mg. In a deficient diet, folate deficiency develops over the course of about 4 months, but folate deficiency may develop rapidly in patients who have both a poor intake and excess utilization (e.g., critical illness).

Folic acid supplements at the time of conception and in the first 12 weeks of pregnancy reduce the incidence of neural tube defects. In the USA and Canada, mandatory fortification of grain products, such as bread, flour, and rice, has substantially improved folate status and been associated with a significant reduction in neural tube defects. There is controversy about the role of folate supplementation in the reduction of cardiovascular and cerebrovascular disease by lowering homocysteine concentrations.

Clinical features

Patients with folate deficiency may be asymptomatic or may present with symptoms of anaemia or the underlying cause. Glossitis can occur. Unlike with vitamin B_{12} deficiency, neuropathy does not occur.

Investigations

The haematological findings are those of a megaloblastic anaemia.

Blood measurements

Serum folate reflects recent folate status and intake. It is usually measured by immunological methods; a concentration below 3 μg/L (7 nmol/L) is indicative of deficiency. The amount of folate in the RBCs is a measure of tissue folate over the lifetime of RBCs; a concentration below 150 μg/L (340 nmol/L) is consistent with folate deficiency, however, serum folate concentration, rather than red cell folate, is usually sufficient for the diagnosis of folate deficiency.

Further investigations

In many cases, the cause is not obvious from the clinical picture or dietary history. Occult GI disease should then be suspected, and appropriate investigations, such as small bowel biopsy, may be performed.

Box 38.6 Causes of folate deficiency

Nutritional (major cause)

Poor intake
- Poor social conditions
- Starvation
- Alcohol excess (also causes impaired utilization)
- Cancer

Gastrointestinal disease
- Partial gastrectomy or bariatric surgery
- Coeliac disease
- Crohn's disease

Antifolate drugs
- Anticonvulsants
  - Phenytoin
  - Primidone
- Methotrexate
- Pyrimethamine
- Trimethoprim

Excess utilization

Physiological
- Pregnancy
- Lactation

Pathological
- Haematological disease with excess red blood cell production (e.g., haemolysis)
- Malignant disease with increased cell turnover
- Inflammatory disease
- Metabolic disease (e.g., homocystinuria)
- Haemodialysis or peritoneal dialysis

Malabsorption
- Small bowel disease (minor effect compared with inadequate intake)

Management and prevention of megaloblastic anaemia

Treatment depends on the type of deficiency. Blood transfusion is not usually indicated in patients with chronic anaemia; indeed, it can be dangerous to transfuse older adult patients because heart failure may be precipitated. Folic acid may produce a haematological response in vitamin B_{12} deficiency but may aggravate the neuropathy. Large doses of folic acid alone should therefore not be used to treat megaloblastic anaemia unless the serum vitamin B_{12} concentration is known to be normal. In severely ill patients, it may be necessary to treat with both folic acid and vitamin B_{12} while awaiting serum concentrations.

Management of vitamin B_{12} deficiency

Hydroxocobalamin 1000 μg is given intramuscularly to a total of 5 to 6 mg over the course of 2 weeks followed by 1000 μg every 3 months for the rest of the patient's life. Clinical improvement may occur within 48 hours. A reticulocytosis can be seen 2 to 3 days after starting therapy, peaking at 5 to 7 days. Improvement of the polyneuropathy may occur over 6 to 12 months, but longstanding spinal cord damage is irreversible. Hypokalaemia can occur, and if severe, supplements should be given. Iron deficiency often develops in the first few weeks of therapy. Hyperuricaemia also occurs, but clinical gout is uncommon. In patients who have had a total gastrectomy or an ileal resection, vitamin B_{12} should be monitored; if concentrations are low, prophylactic intramuscular vitamin B_{12} should be given. Oral treatment with vitamin B_{12} may be used (e.g., in vegans), but good adherence is required.

Management of folate deficiency

Folate deficiency can be corrected by giving 5 mg of folic acid daily. The same haematological response occurs as with treatment of vitamin B_{12} deficiency. Treatment should be given for about 4 months to replace body stores. Any underlying cause, such as coeliac disease, should be treated.

Prophylactic folic acid (400 μg daily) is recommended for all females planning a pregnancy and in early pregnancy to reduce neural tube defects. Females who have had a child with a neural tube defect should take higher dose (5 mg) folic acid daily before and during subsequent pregnancies.

Prophylactic folic acid (5 mg/day or weekly), is given to patients with chronic haematological disorders in which there is rapid cell turnover and to those undergoing renal dialysis.

Macrocytosis without megaloblastic changes

A raised MCV with macrocytosis on the peripheral blood film can occur with a normoblastic rather than a megaloblastic bone marrow.

A common ***physiological*** cause of macrocytosis is pregnancy. Macrocytosis may also occur in the newborn. Common ***pathological*** causes are:

- Alcohol excess
- Liver disease
- Reticulocytosis (e.g., caused by haemolysis)
- Hypothyroidism
- Some haematological disorders (e.g., aplastic anaemia, myelodysplasia, pure RBC aplasia, multiple myeloma)
- Drugs (e.g., hydroxycarbamide, azathioprine)
- Cold agglutinins caused by autoagglutination of RBCs (the MCV decreases to normal with warming of the sample to 37°C).

In all these conditions, concentrations of vitamin B_{12} and folate are normal. Exact mechanisms in each case are uncertain, but in some, there is increased lipid deposition in the RBC membrane. An increased number of reticulocytes also leads to a raised MCV because they are large cells. High alcohol consumption is a frequent cause of a raised MCV and can be used as a surrogate marker for monitoring excess alcohol consumption. Megaloblastic anaemia can also occur in alcohol excess because of a toxic effect of alcohol on erythropoiesis or dietary folate deficiency.

Further reading

British Committee for Standards in Haematology. Guidelines for the diagnosis and treatment of cobalamin and folate disorders. *Br J Haematol* 2014; 166:496–513.

ANAEMIA CAUSED BY MARROW FAILURE (APLASTIC ANAEMIA)

Aplastic anaemia is defined as pancytopenia with hypocellularity (aplasia) of the bone marrow; there are no leukaemic, cancerous, or other abnormal cells in the peripheral blood or bone marrow. It is usually an acquired condition but may rarely be inherited.

Aplastic anaemia is caused by a reduction in the number of pluripotent stem cells (**Fig. 38.3F**), together with either a fault in those remaining or an immune reaction against them so they are unable to repopulate the bone marrow. Failure of only one cell line may also occur, resulting in isolated deficiencies such as the absence of RBC precursors in pure RBC aplasia. Evolution to myelodysplasia, paroxysmal nocturnal haemoglobinuria (PNH), or acute myeloid leukaemia occurs in some cases, probably owing to the emergence of an abnormal clone of haemopoietic cells.

Aetiology

Causes of aplasia are shown in **Box 38.7**. Immune mechanisms are probably responsible for most cases of idiopathic acquired aplastic anaemia and play a part in the persistence of many secondary cases. Activated cytotoxic T cells in the blood and bone marrow are responsible for the bone marrow failure.

Many drugs may cause marrow aplasia. Cytotoxic drugs such as busulfan and doxorubicin are expected to cause transient aplasia with therapeutic use. However, some individuals develop aplasia due to sensitivity to non-cytotoxic drugs such as chloramphenicol, gold, carbimazole, chlorpromazine, phenytoin, ribavirin, tolbutamide, NSAIDs, and many others for which aplasia is reported as a rare side-effect.

? Box 38.7 Causes of aplastic anaemia

Primary
- Inherited e.g., Fanconi anaemia
- Idiopathic (67% of cases)

Secondary
- Chemicals e.g., benzene, toluene, glue sniffing
- Drugs
  - Chemotherapy
  - Antibiotics e.g., chloramphenicol, gold, penicillamine, phenytoin, carbamazepine, carbimazole, azathioprine
- Insecticides
- Ionizing radiation
- Infections
  - Viral e.g., hepatitis, Epstein–Barr virus, human immunodeficiency virus, erythrovirus
  - Other e.g., tuberculosis
- Paroxysmal nocturnal haemoglobinuria
- Miscellaneous e.g., pregnancy

Inherited aplastic anaemias are rare. Multiple gene mutations have been identified in genes involved in DNA repair, ribosomal processing, and telomere maintenance. Fanconi anaemia is an autosomal recessive condition with skeletal, skin, eye, renal, and central nervous system abnormalities. It usually presents between the ages of 5 and 10 years and is associated with an increased risk of malignancies.

Clinical features

The clinical manifestations of marrow failure from any cause are anaemia, bleeding, and infection. Bleeding is often the predominant initial presentation of aplastic anaemia with bruising on minimal trauma or blood blisters in the mouth. Physical findings include ecchymoses, bleeding gums, and epistaxis. Mouth infections are common. Lymphadenopathy and hepatosplenomegaly are rare in aplastic anaemia.

Investigations

- Pancytopenia
- The virtual absence of reticulocytes
- A hypocellular or aplastic bone marrow with increased fat spaces (**Fig. 38.15**)

Differential diagnosis

Aplastic anaemia must be differentiated from other causes of pancytopenia (**Box 38.8**). A bone marrow trephine is essential for assessment of the bone marrow cellularity.

Management and prognosis

Treatment of patients with aplastic anaemia depends on the underlying cause and the likelihood of spontaneous recovery of blood counts. Careful attention to supportive care is essential while awaiting bone marrow recovery. In some cases, specific treatment can be used to accelerate marrow recovery.

The main risk is infection and stringent measures should be undertaken to avoid this. Any suspicion of infection in a severely neutropenic patient (neutrophil count $<0.5 \times 10^9$/L) should receive broad-spectrum parenteral antibiotics. Supportive care, including transfusions of RBCs and platelets, should be given as necessary. The cause of the aplastic anaemia must be eliminated if possible.

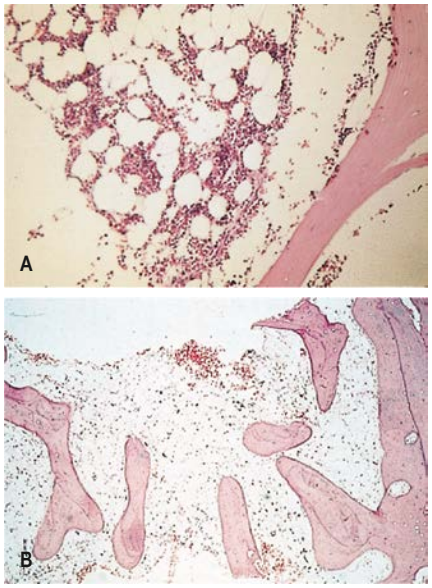


Fig. 38.15 Bone marrow trephine biopsies in low-power view. **(A)** Normal cellularity. **(B)** Hypocellularity in aplastic anaemia.

Box 38.8 Causes of pancytopenia

- Aplastic anaemia (Box 38.7)
- Drugs
- Megaloblastic anaemia
- Bone marrow infiltration or replacement
  - Hodgkin and non-Hodgkin lymphoma
  - Acute leukaemia
  - Myeloma
  - Bone metastases
  - Myelofibrosis
- Hypersplenism
- Systemic lupus erythematosus
- Disseminated tuberculosis
- Paroxysmal nocturnal haemoglobinuria
- Overwhelming sepsis

The course of aplastic anaemia can be variable, ranging from a rapid spontaneous remission to persistent, increasingly severe pancytopenia, which may lead to death through haemorrhage or infection. The most reliable prognostic factors are the number of neutrophils, reticulocytes, and platelets and the cellularity of the bone marrow.

Severe aplastic anaemia and a poor prognosis is associated with the presence of two of the following three features:

- Neutrophil count less than 0.5×10^9/L
- Platelet count less than 20×10^9/L
- Reticulocyte count less than 40×10^9/L.

Haemopoietic stem cells (allogeneic bone marrow transplantation) are the treatment of choice for patients with severe aplastic anaemia younger than the age of 40 years who have a human leucocyte antigen (HLA)-identical sibling donor, in whom it gives a 75% to 90% chance of long-term survival.

Immunosuppressive therapy is recommended for:

- Patients with severe disease older than the age of 40 years
- Younger patients with severe disease who do not have an HLA-identical sibling donor
- Patients who do not have severe disease but who are transfusion dependent.

Standard immunosuppressive treatment is antithymocyte globulin (ATG) and ciclosporin, which results in response rates of 60% to 80% and 5-year survival rates of 75% to 85%.

Stem cell transplantation using matched unrelated donors is an option for patients younger than the age of 50 years who have no matched sibling donor and who have failed to respond to immunosuppression. Results are improving (5-year survival rate, 65%–73%). The main problems are graft rejection, graft-versus-host disease, and viral infections.

Concentrations of haemopoietic growth factors (Fig. 38.3F) are normal or increased in most patients with aplastic anaemia, so these factors are ineffective as primary treatment.

Steroids should not be used to treat patients with severe aplastic anaemia except for those with serum sickness caused by ATG.

Adult pure RBC aplasia is associated with a thymoma in 5% to 15% of cases, and thymectomy occasionally induces a remission. It may also be associated with autoimmune disease or be idiopathic. Steroids, cyclophosphamide, azathioprine, and ciclosporin are effective treatments in some cases.

Further reading

British Society of Haematology. Guidelines for the diagnosis and management of adult aplastic anaemia. *Br J Haematol* 2016; 172:187–207.

HAEMOLYTIC ANAEMIAS

Haemolytic anaemias are caused by increased destruction of RBCs. RBCs normally survive about 120 days, but in patients with haemolytic anaemias, survival times are considerably shortened. Breakdown of normal RBCs occurs in the macrophages of the bone marrow, liver, and spleen (Fig. 38.8).

Consequences of haemolysis

Shortening RBC survival does not always cause anaemia because there is a compensatory increase in RBC production by the bone marrow. If the RBC loss can be contained within the marrow's capacity for increased output, then haemolysis can exist without anaemia ***(compensated haemolytic disease)***. The bone marrow can raise its output by six to eight times by increasing the proportion of cells committed to erythropoiesis ***(erythroid hyperplasia)*** and by expanding the volume of active marrow. In addition, immature RBCs ***(reticulocytes)*** are released prematurely. These cells are larger than mature cells and stain with a light blue tinge on a peripheral blood film (termed ***polychromasia***) because of the presence of residual ribosomal RNA. Reticulocytes may be counted accurately as a percentage of all RBCs on a blood film using a supravital stain for residual RNA (e.g., new methylene blue).

Sites of haemolysis

Extravascular haemolysis

In most haemolytic conditions, RBC destruction is extravascular. The RBCs are removed from the circulation by macrophages in the reticuloendothelial system, particularly the spleen.

Intravascular haemolysis

When RBCs are rapidly destroyed within the circulation, Hb is liberated. This is initially bound to plasma haptoglobins, but these soon become saturated. Excess free plasma Hb is filtered by the

renal glomerulus and enters the urine, although small amounts are reabsorbed by the renal tubules, broken down, and deposited as ***haemosiderin***. This can be detected in a spun sediment of urine using Perls' reaction. Some free plasma Hb is oxidized to ***methaemoglobin***, which dissociates into ***ferrihaem*** and globin. ***Plasma haemopexin*** binds ferrihaem, but if its binding capacity is exceeded, ferrihaem becomes attached to albumin, forming ***methaemalbumin***. On spectrophotometry of the plasma, methaemalbumin forms a characteristic band; this is the basis of ***Schumm's test***.

The ***liver*** plays a major role in removing Hb bound to haptoglobin and haemopexin and any remaining free Hb. Free Hb is toxic and scavenges nitric oxide, which is a critical signalling molecule for smooth muscle relaxation and control of pulmonary pressure.

Evidence for haemolysis

- Increased RBC production accompanies increased RBC breakdown (Fig. 38.16).
- Demonstration of shortened RBC lifespan: RBC survival can be estimated from ^{51}Cr-labelled RBCs given intravenously, but this test is rarely performed.
- Intravascular haemolysis: haemosiderinuria, very low or absent haptoglobins, and the presence of methaemalbumin (positive Schumm's test).

Various laboratory studies will be necessary to determine the exact type of haemolytic anaemia present. Causes of haemolytic anaemias are shown in Box 38.9.

INHERITED HAEMOLYTIC ANAEMIA

Red blood cell membrane defects

The normal RBC membrane consists of a lipid bilayer crossed by integral proteins with an underlying lattice of proteins (cytoskeleton), including spectrin, actin, ankyrin, and protein 4.1, attached to integral proteins (Fig. 38.17).

Hereditary spherocytosis

Hereditary spherocytosis (HS) is the most common inherited haemolytic anaemia in Northern Europeans, affecting 1 in 5000. It can be inherited in autosomal dominant or recessive manners and by spontaneous *(de novo)* mutation. In the absence of a family history,

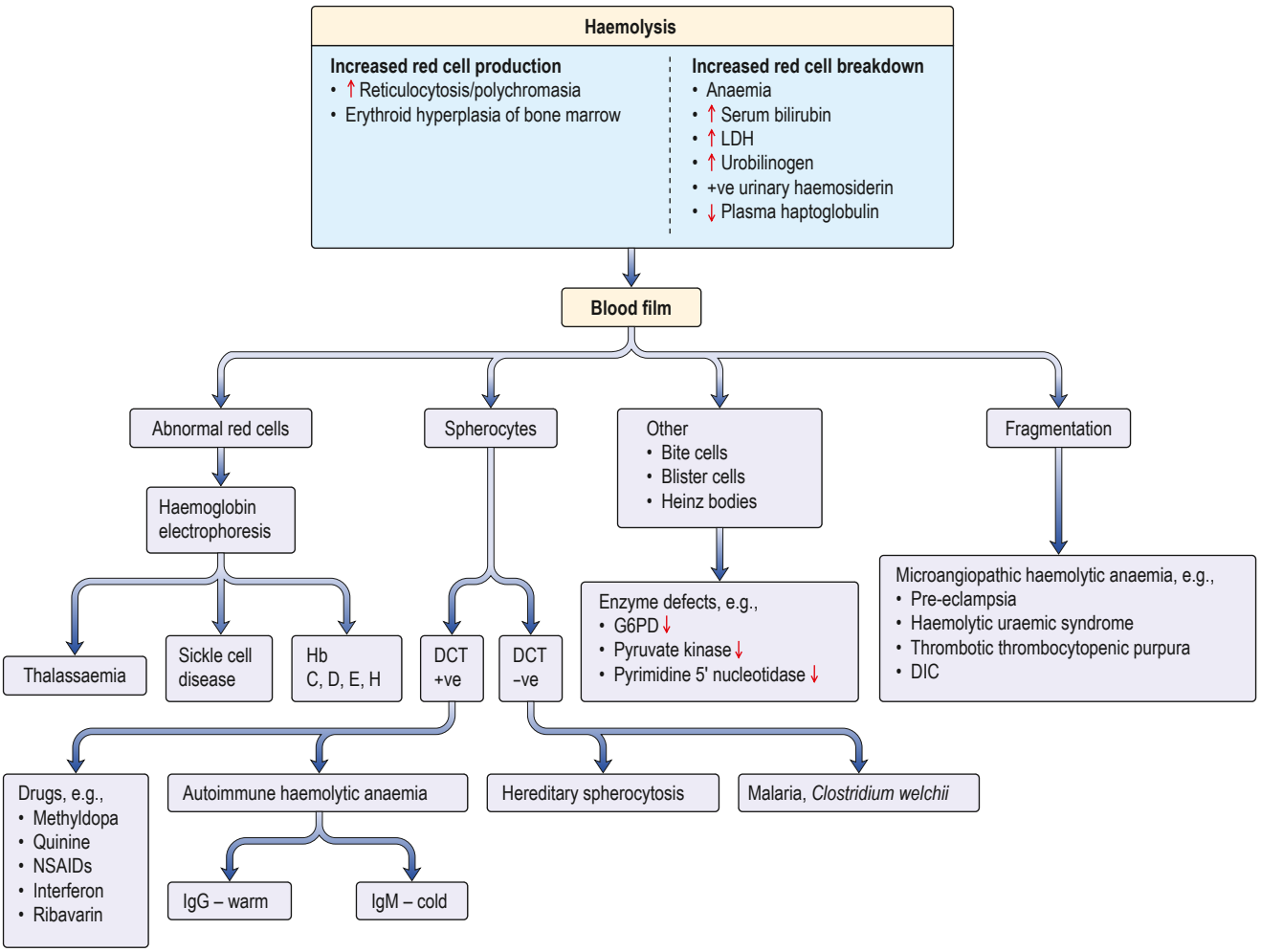


Fig. 38.16 Haemolysis: diagnostic and haematological features. *DCT,* Direct Coombs' test; *DIC,* disseminated intravascular coagulation; *G6PD,* glucose-6-phosphate dehydrogenase; *Hb,* haemoglobin; *Ig,* immunoglobulin; *LDH,* lactate dehydrogenase; *NSAIDs,* non-steroidal anti-inflammatory drugs; *RBC,* red blood cell.

the most important differential diagnosis is autoimmune haemolytic anaemia, which can usually be excluded by a negative direct antiglobulin test (DAT) result.

HS is caused by defects in RBC membrane proteins or the cytoskeleton, resulting in the cells losing part of the cell membrane as they pass through the spleen, possibly because the lipid bilayer is inadequately supported by the membrane skeleton. The best-characterized defect is a deficiency in the structural protein spectrin, but quantitative defects in other membrane proteins have been identified (Fig. 38.17), with ankyrin defects being the most common. The abnormal RBC membrane in HS is associated with an increased permeability to sodium. This requires an increased rate of active transport of sodium out of the cells, which is dependent on adenosine triphosphate (ATP), produced by glycolysis. The surface-to-volume ratio decreases, and cells become spherocytic. Spherocytes are more rigid and less deformable than normal RBCs. They are unable to pass through the splenic microcirculation so have a shortened lifespan.

? Box 38.9 Causes of haemolytic anaemia

Inherited

Red blood cell membrane defect
- Hereditary spherocytosis
- Hereditary elliptocytosis

Haemoglobin abnormalities
- Thalassaemia
- Sickle cell disease

Metabolic defects
- Glucose-6-phosphate dehydrogenase deficiency
- Pyruvate kinase deficiency
- Pyrimidine kinase deficiency

Acquired

Immune
- Autoimmune
  - Warm
  - Cold
- Alloimmune
  - Haemolytic transfusion reactions
  - Haemolytic disease of the newborn
  - After allogeneic bone marrow or organ transplantation
- Drug induced

Non-immune
- Acquired membrane defect: paroxysmal nocturnal haemoglobinuria
- Mechanical
  - Microangiopathic haemolytic anaemia
  - Valve prosthesis
  - March haemoglobinuria
- Secondary to systemic disease: kidney and liver failure
- Miscellaneous
  - Infections (e.g., malaria, mycoplasma)
  - *Clostridium perfringens (welchii)*, generalized sepsis
  - Drugs and chemicals causing damage to red blood cell membrane or oxidative haemolysis
  - Hypersplenism
  - Burns

Clinical features

The condition may present with jaundice at birth. However, jaundice can be delayed for many years. Some patients may go through life with no symptoms and are detected only during family studies. Patients may eventually develop anaemia; gallstones; splenomegaly; or, rarely, leg ulcers. As in many haemolytic anaemias, the disease course may be interrupted by aplastic, haemolytic, and megaloblastic crises. Aplastic anaemia usually occurs after infections, particularly with parvovirus, which infects developing erythroblasts and causes temporary abolition of RBC production, often requiring a blood transfusion for the patient to recover. Megaloblastic anaemia is the result of folate depletion caused by hyperactivity of the bone marrow. Chronic haemolysis leads to the formation of pigment gallstones.

Investigations

- ***Anaemia:*** usually mild but occasionally can be severe
- ***Blood film:*** spherocytes (Fig. 38.18) and polychromasia
- ***Haemolysis:*** raised serum bilirubin, urinary urobilinogen, and LDH (released by the destruction of RBCs)
- ***Osmotic fragility:*** RBCs placed in solutions of increasing hypotonicity take in water, swell, and eventually lyse. Spherocytes tolerate hypotonic solutions less well than normal biconcave RBCs. Osmotic fragility tests are infrequently carried out in routine practice but may be useful to confirm suspicion of spherocytosis on a blood film.

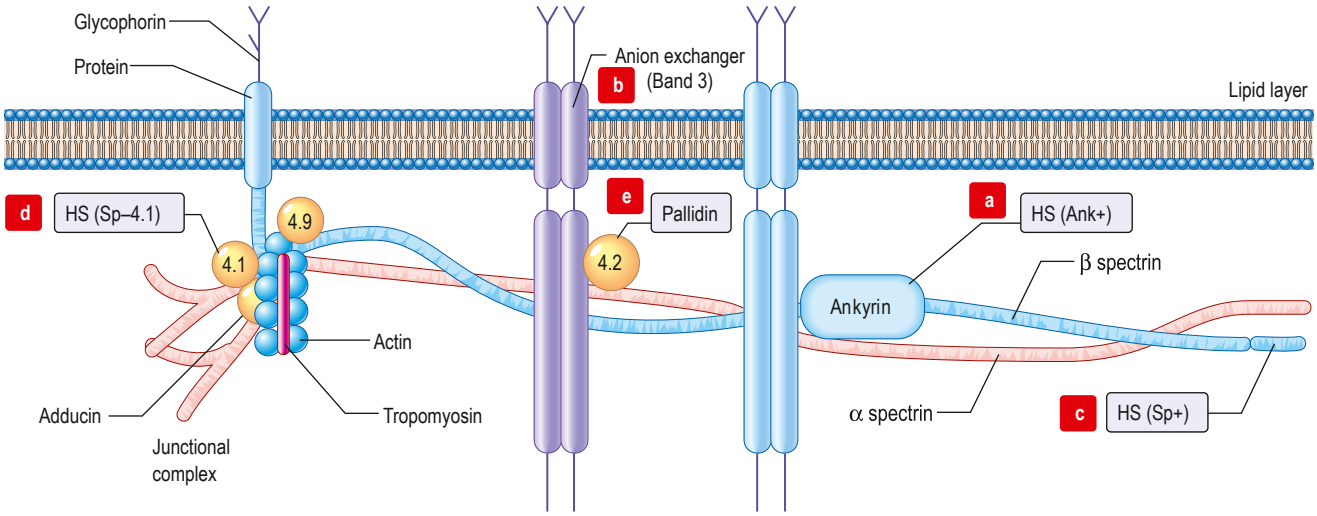


Fig. 38.17 Hereditary spherocytosis (HS) and hereditary elliptocytosis (HE): red blood cell membrane show the sites of the principal defects. Vertical interactions producing HS: **(a)** ***ankyrin mutation***, HS (Ank$^+$) producing deficiency (45% of cases); **(b)** HS ***band 3 deficiency*** (20%); **(c)** ***β-spectrin deficiency***, HS (Sp$^+$) (<20%); **(d)** ***abnormal spectrin/protein 4.1 binding***, HS (Sp–4.1); and **(e)** ***protein 4.2 (pallidin) deficiency***. These produce various autosomal dominant and recessive forms of the disease. Horizontal interactions producing HE: α-spectrin (80%), protein 4.1 (15%), and β spectrin (5%).

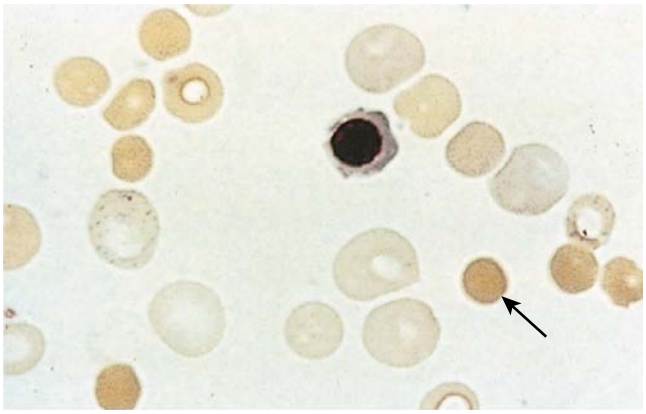


Fig. 38.18 Spherocytes *(arrow)*. This blood film also shows reticulocytes, polychromasia, and a nucleated erythroblast.

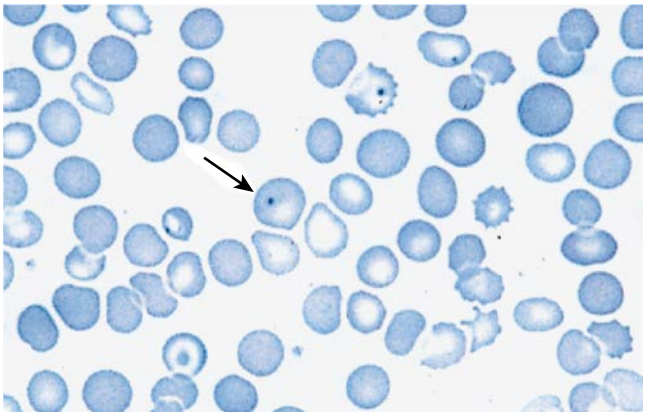


Fig. 38.19 Post-splenectomy film with Howell–Jolly bodies *(arrow)*. Target cells and irregularly contracted cells are also shown.

- ***Direct antiglobulin (Coombs') test:*** negative in HS, virtually ruling out autoimmune haemolytic anaemia in which spherocytes are also commonly present
- ***EMA dye binding.*** The dye eosin-5-maleimide binds to band 3, a RBC membrane protein. When assessed by flow cytometry, more than 20% reduced fluorescence is diagnostic of HS, providing a rapid diagnostic test.

Management

Most patients require no specific treatment. They may need occasional transfusions during pregnancy or intercurrent infections. Rarely, patients are transfusion dependent. For the remainder, splenectomy is indicated to relieve symptoms from anaemia or splenomegaly, reverse growth failure, and prevent recurrent gallstones. It is best to postpone splenectomy until after childhood because sudden overwhelming fatal infections, usually caused by encapsulated organisms such as pneumococci, may occur. Splenectomy should be preceded by appropriate immunization and followed by life-long penicillin prophylaxis. In addition to the well-known risk of bacterial infection, there is also a significant risk of adverse arterial and venous thromboembolic events after splenectomy, and this needs to be considered when deciding whether to proceed to splenectomy.

After splenectomy, spherocytes persist, but the Hb usually returns to normal because the RBCs are no longer destroyed. Other changes after a splenectomy include a raised WBC and platelet count, reticulocytosis, and the presence of Howell-Jolly bodies (nuclear remnants) on a blood film (Fig. 38.19).

Hereditary elliptocytosis

This disorder of the RBC membrane is inherited in an autosomal dominant manner and has a prevalence of 1 in 2500 in Whites. RBCs are elliptical because of deficiencies of protein 4.1 or the spectrin–actin–4.1 complex, which leads to weakness of horizontal protein interaction and the membrane defect (Fig. 38.17). Clinically, it is a similar condition to HS but milder. Only a minority of patients have anaemia, and only occasional individuals require splenectomy.

Rarely, HS or elliptocytosis may be inherited in a homozygous fashion, giving rise to a severe haemolytic anaemia sometimes necessitating splenectomy in early childhood (hereditary pyropoikilocytosis).

Hereditary stomatocytosis

Stomatocytes are RBCs in which the pale central area appears slit-like. Their presence in large numbers may occur in a hereditary haemolytic anaemia associated with a membrane defect, but excess alcohol intake is also a common cause. Although these hereditary conditions are very rare, a correct diagnosis is required; splenectomy is contraindicated because it may result in fatal thromboembolic events.

Further reading

Bolton-Maggs PHB, Langer JC, Iolascon A et al. *Guidelines for the Diagnosis and Management of Hereditary Spherocytosis*. London: British Committee for Standards in Haematology; 2011. https://b-s-h.org.uk/guidelines/.
Mohandas N. Inherited hemolytic anemia: a possessive beginner's guide. *Hematology* 2018; 2018:377–381.

Haemoglobin abnormalities

In early embryonic life, haemoglobins Gower 1, Gower 2, and Portland predominate. Later, fetal haemoglobin (HbF), which has two α and two γ chains, is produced (Fig. 38.20). There is increasing synthesis of β chains from 13 weeks' gestation, and at term, there is 80% HbF and 20% HbA. Completion of the Hb switch from HbF to HbA occurs after birth, when the genes for γ chain production are further suppressed and there is rapid increase in synthesis of β chains. BCL IIA, a zinc finger protein, is one of several proteins that suppress γ gene expression and is a key genome editing target in the effort to reactivate HbF in adults with haemoglobinopathies. There is little HbF produced (normally <1%) by 6 months of age. The δ chain is synthesized just before birth, and HbA_2 ($\alpha_2\delta_2$) remains at a concentration of about 2% throughout adult life (Box 38.10).

A normal individual has four α-globin chain genes (Fig. 38.20), with two α-globin genes on each haploid genome (genes derived from one parent). These are situated close together on chromosome 16. The genes controlling the production of ε, γ, δ, and β chains are close together on chromosome 11. The globin genes are arranged on chromosomes 16 and 11 in the order in which they are expressed and combine to give different haemoglobins. Normal Hb synthesis is shown in Fig. 38.6. Understanding how globin chain production changes through fetal life and birth forms the basis of understanding how different genetic abnormalities lead to particular phenotypes.

Haemoglobinopathies

Abnormalities can occur in:

- Globin chain production: quantitative defect (e.g., thalassaemia)
- Structure of the globin chain: qualitative defect (e.g., sickle cell disease)
- Combined defects of globin chain production and structure (e.g., sickle cell β-thalassaemia, HbE).

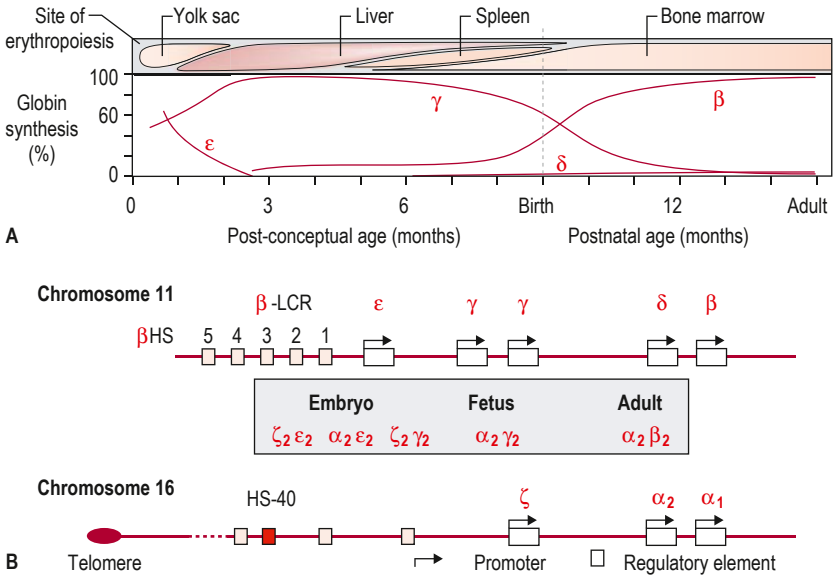


Fig. 38.20 Normal development switches in globin expression. **(A)** Sites of haemopoiesis and globin changes at various gestational ages. **(B)** Structure of α-like and β-like globin gene clusters. *HS,* Major upstream regulatory element; *LCR,* locus control region. *(From Higgs DR, Engel JD, Stamatoyannopoulos G. Thalassaemia.* Lancet *2012; 379:373–383, with permission.)*

Box 38.10 Types of haemoglobin (Hb)

	Hb	Structure	Comments
Normal	A	$\alpha_2\beta_2$	Makes up 97% of adult Hb
	A_2	$\alpha_2\delta_2$	Makes up 2% of adult Hb Elevated in β-thalassaemia
	F	$\alpha_2\gamma_2$	Normal Hb in fetuses from 3rd month of pregnancy to delivery
			Increased in β-thalassaemia
			Makes up <1% of Hb in adult
Abnormal chain production	H	β_4	Found in α-thalassaemia Biologically useless
	Barts	γ_4	Makes up 100% of Hb in homozygous α-thalassaemia Biologically useless
Abnormal chain structure	SC	$\alpha_2\beta_2{}^S\alpha_2\beta_2{}^C$	Substitution of valine for glutamine acid in position 6 of β chain Substitution of lysine for glutamic acid in position 6 of β chain

Further reading

Forget BG. Progress in understanding the hemoglobin switch. *N Engl J Med* 2011; 365:852–854.
Lu HY, Orkin SH, Sankaran VG. Fetal hemoglobin regulation in beta-thalassemia. *Hematol Oncol Clin North Am* 2023; 37(2):301–312.

The thalassaemias

The thalassaemias affect people throughout the world (Fig. 38.21), and at least 60 000 severely affected individuals are born every year. Normally, there is balanced (1:1) production of α and β chains. The defective synthesis of globin chains in thalassaemia leads to 'imbalanced' globin chain production, causing precipitation of the excess globin chains within the RBC precursors, resulting in ineffective erythropoiesis. Precipitation of globin chains in mature RBCs leads to haemolysis. This concept of globin chain imbalance is critical in understanding the relationship between genotype and phenotype (the greater the imbalance, the worse the phenotype), as well as understanding how novel therapies are used to ameliorate disease.

Beta-thalassaemia

In ***homozygous*** (identical mutations on both alleles) or ***compound heterozygous*** (different damaging mutations on both alleles) β-thalassaemia, either no normal β chains are produced (β^0) or β-chain production is very reduced (β^+). This leaves an excess of α chains, which precipitate in erythroblasts and RBCs, causing ineffective erythropoiesis and haemolysis.

In the carrier (heterozygous) state, excess α chains combine with δ and result in modestly increased quantities of HbA_2, which is used to screen for β-thalassaemia carrier status. Heterozygous β-thalassaemia carriers exhibit asymptomatic microcytosis with or without mild anaemia. Box 38.11 shows findings in the homozygote and heterozygote for the common types of β-thalassaemia.

Genetics

Molecular errors accounting for more than 200 genetic defects leading to β-thalassaemia have been characterized. Unlike in α-thalassaemia, the defects are mainly point mutations rather than gene deletions. The mutations result in defects in transcription, RNA splicing and modification, translation via frame shifts, and nonsense codons producing highly unstable β-globin that cannot be used.

Clinical syndromes

The terminology of clinical syndromes in thalassaemia have changed in recent years, moving away from the 'major' and 'minor' descriptions and now focusing on whether they are 'transfusion-dependent thalassaemias' (TDTs) or 'non–transfusion-dependent thalassaemias' (NTDTs) (Fig. 38.22). Patients who have NTDTs may require occasional blood transfusions (e.g., with infection, in pregnancy). Patients with NTDTs may be transferred to the TDT category if it is decided that regular transfusions would improve their growth, development, and wellbeing. 'Thalassaemia trait' is the term still used to describe the symptomless heterozygous carrier state.

Thalassaemia trait (carrier)

Heterozygous β-thalassaemia is asymptomatic. Anaemia is mild or absent. RBCs are hypochromic and microcytic with a low MCV and MCH, and the condition may be confused with iron deficiency. However, the two are easily distinguished because serum ferritin and iron stores are normal in thalassaemia trait (Box 38.4). The RDW is usually normal. Hb electrophoresis usually shows a raised HbA_2 and often a slightly raised HbF (Fig. 38.23). Iron should not be given in the absence of proven coincidental iron deficiency.

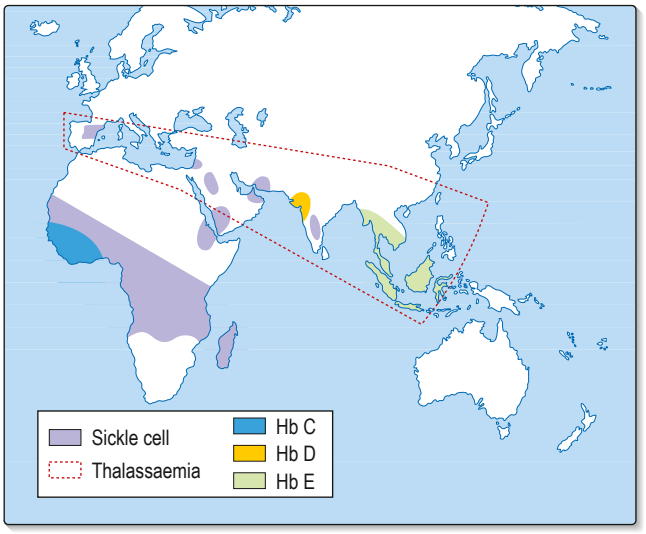


Fig. 38.21 Major haemoglobin (Hb) abnormalities: geographical distribution. Map lines delineate study areas and do not necessarily depict accepted national boundaries.

Box 38.11 Usual findings in β-thalassaemia

Type of thalassaemia	Homozygous or compound heterozygous phenotype	Heterozygous phenotype
β^+ (very reduced β-chain production)	**Transfusion-dependent anaemia** HbA (10%–30%) HbF (70%–90%) HbA_2 variable	**Mild reduction in Hb** HbA (90%–95%) Increased HbA_2 (3.6%–4.2%)
β^0 (no β-chain production)	**Transfusion-dependent anaemia** HbF (95%–98%) HbA_2 variable	**Mild reduction in Hb** HbA (90%–95%) Increased HbA_2 (4%–5%)

Hb, Haemoglobin.

Non–transfusion-dependent thalassaemia

This condition was previously referred to as 'thalassaemia intermedia'. Patients are symptomatic with moderate anaemia (Hb 70–100 g/L), but do not require regular transfusions.

NTDT occurs in patients in whom the effects of the excess α-globin chains are attenuated. This can be caused by a reduced production (co-inheritance of α-thalassaemia) or the α-chains being 'mopped up' by γ chains. Although there is usually very little γ-chain production after birth, some patients may co-inherit 'hereditary persistence of fetal Hb', in which γ-chain production is increased, sometimes to substantial amounts. By reducing α-chain production or increasing that of γ chains, there is less excess α-chain precipitation, thus helping to reduce the degree of globin imbalance and causing a consequent reduction in ineffective erythropoiesis and haemolysis.

NTDT may also arise from the inheritance of a heterozygous β^+ mutation with an excess of α-chain genes (e.g., triplicated α-chain genes). The more α-chain genes inherited, the greater the imbalance and the more severe the phenotype.

Patients with NTDTs may have splenomegaly and bone deformities. Recurrent leg ulcers, gallstones, and infections are also seen. Patients may be iron overloaded despite a lack of regular blood transfusions caused by excessive iron absorption resulting from the underlying dyserythropoiesis that drives Erfe production, suppressing hepcidin to undetectable concentrations.

Transfusion-dependent thalassaemia

This condition was previously referred to as 'thalassaemia major'. Most children affected by homozygous or compound heterozygous β-thalassaemia present during the first year of life with:

- Failure to thrive and recurrent bacterial infections
- Severe anaemia from 3 to 6 months, when the switch from γ- to β-chain production should normally occur
- Extramedullary haemopoiesis that leads to hepatosplenomegaly and bone expansion, giving rise to the classical thalassaemic facies (Fig. 38.24A).

Skull X-rays show the characteristic 'hair-on-end' appearance of bony trabeculation because of expansion of the bone marrow into cortical bone (Fig. 38.24B). Expansion of bone marrow can also be seen in plain films of the hand (Fig. 38.24C).

The classic features of untreated thalassaemia major are generally observed only when there is an absence of good blood transfusion support.

Management of symptomatic thalassaemia

The aims of treatment are to suppress ineffective erythropoiesis; prevent bony deformities; and allow normal activity, growth, and development.

- ***Long-term folic acid*** supplements are required.
- ***Regular transfusions*** should be given to keep the Hb above 100 g/L. Blood transfusions may be required every 4 to 6 weeks.
- ***Splenectomy*** was previously used to reduce transfusion requirements. However, these are associated with the development of

pulmonary hypertension and should be used only rarely and after discussion with a wide multidisciplinary team to ensure that the likely benefits outweigh risks. Splenectomy is usually delayed until after the age of 6 years because of the risk of infection; prophylaxis against infection is required (see 'The spleen' section).

- ***Iron overload*** caused by repeated transfusions (transfusion haemosiderosis) may lead to damage to the endocrine glands, liver, pancreas, and myocardium by adolescence if inadequate iron chelation results from a lack of adherence or access to chelators. Cardiomyopathy and associated cardiac tachyarrhythmias are the leading causes of morbidity and mortality. Magnetic resonance imaging (MRI) is useful for monitoring iron overload; both the heart (myocardial T_2-relaxation time) and the liver can be monitored. Iron chelation using the once-daily oral iron chelator deferasirox is now standard where it is available and economically accessible. Elsewhere, desferrioxamine remains the standard chelator, although it must be administered parenterally as an overnight subcutaneous infusion on 5 to 7 nights each week. Ascorbic acid 200 mg/day increases urinary excretion of iron in response to desferrioxamine. Deferiprone, an oral iron chelator, is usually used for cardiac iron overload or in combination with another chelator.
- ***Intensive treatment with desferrioxamine*** has been reported to reverse damage to the heart in patients with severe iron overload, but excessive doses of desferrioxamine may cause cataracts, retinal damage, and sensorineural deafness. Infection with *Yersinia enterocolitica* occurs in iron-loaded patients treated with desferrioxamine. Iron overload should be periodically assessed by measuring serum ferritin and assessing hepatic iron stores by MRI.
- ***Bone marrow transplantation*** has been used in young patients with HLA-matched siblings. It has been successful in suitable patients with a 3-year mortality rate of less than 5%, but there is a high mortality (>50%) in patients in poor condition with iron overload and liver dysfunction. Only 25% of patients have a sibling-matched donor.
- ***Prenatal diagnosis*** is discussed in the 'Prenatal screening and diagnosis of severe haemoglobin abnormalities' section.
- ***Gene therapy*** is detailed in Chapter 5.

Thalassaemia carrier	NTDT	TDT
Thalassaemia trait	'Thalassaemia intermedia'	'Thalassaemia major'
No transfusions	Intermittent transfusions	Transfusion dependence
Heterozygous for one of: • 1 or 2 deletions of α globin • β^0-globin mutation (no β globin produced) • β^+-globin mutation (small amount of β globin produced)	• HbH disease (3 α gene deletions) • HbE/β-thalassaemia • β-thalassaemia trait with co-inherited multiple copies of α globin • β-thalassaemia major with co-inherited hereditary persistence of fetal haemoglobin or α-thalassaemia	Homozygous or compound heterozygous for either: • β^0-globin mutation (no β globin produced) • β^+-globin mutation (small amount of β globin produced)

Fig. 38.22 Clinical syndromes in thalassaemia. The terms 'thalassaemia intermedia' and 'thalassaemia major' have been replaced by 'non–transfusion-dependent thalassaemia' (NTDT) and 'transfusion-dependent thalassaemia' (TDT). Patients can fluctuate between these categories, depending on their life circumstances. The phenotype of affected individuals is influenced by modifying factors such as the reactivation of γ globin or the co-inheritance of changes in the number of α-globin genes.

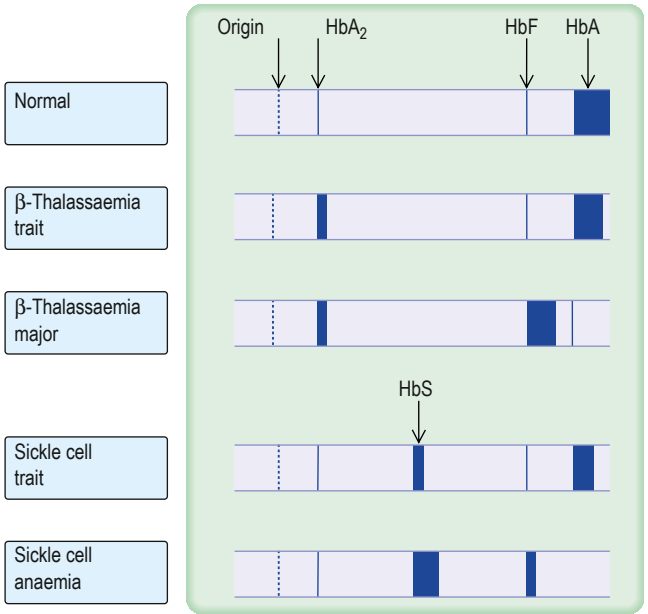


Fig. 38.23 Patterns of haemoglobin (Hb) electrophoresis.

Alpha-thalassaemia

Genetics

In contrast to β-thalassaemia, α-thalassaemia is often caused by gene deletions, although mutations of the α-globin genes may also occur. The gene for α-globin chains is duplicated on chromosome 16, meaning that a normal person has a total of four α-globin genes. Deletion of one α-chain gene (α^+) or both α-chain genes (α^0) on each chromosome 16 may occur (**Box 38.12**). The former is most common. Deletions arise during meiosis because of unequal crossover during non-homologous recombination, one of the main ways in which genetic variability arises. Although one of the gametes has a missing alpha gene, the other will have an excess of alpha genes, a term called 'triplicated alpha' (**Fig. 38.25A**). The risk to subsequent generations is more significant in α^0 than α^+ as depicted in **Fig. 38.25B**.

- ***Four-gene deletion*** (deletion of both genes on both chromosomes). There is no α-chain synthesis, and only Hb Barts (γ4) is present. Hb Barts cannot carry oxygen. This is incompatible with life (**Boxes 38.10 and 38.12**). Fetuses are pale and oedematous with enormous livers and spleens, a condition called 'hydrops fetalis'. If this is detected *in utero*, intrauterine transfusions can be given so that the fetus can grow and survive until term, at which point the baby will have TDT.
- ***Three-gene deletion.*** The severe reduction in α-chain synthesis results in HbH disease, which is common in parts of Asia. HbH has four β-chains, although patients may also have low concentrations of HbA and Hb Barts. HbA_2 is normal or reduced. HbH does not transport oxygen and precipitates in erythroblasts and erythrocytes. These patients have moderate anaemia (Hb 70–100 g/L) and splenomegaly (previously referred to as thalassaemia intermedia). The patients usually have NTDT.
- ***Two-gene deletion*** (α-thalassaemia trait). There is microcytosis with or without mild anaemia. HbH bodies may be seen on blood film staining with brilliant cresyl blue.

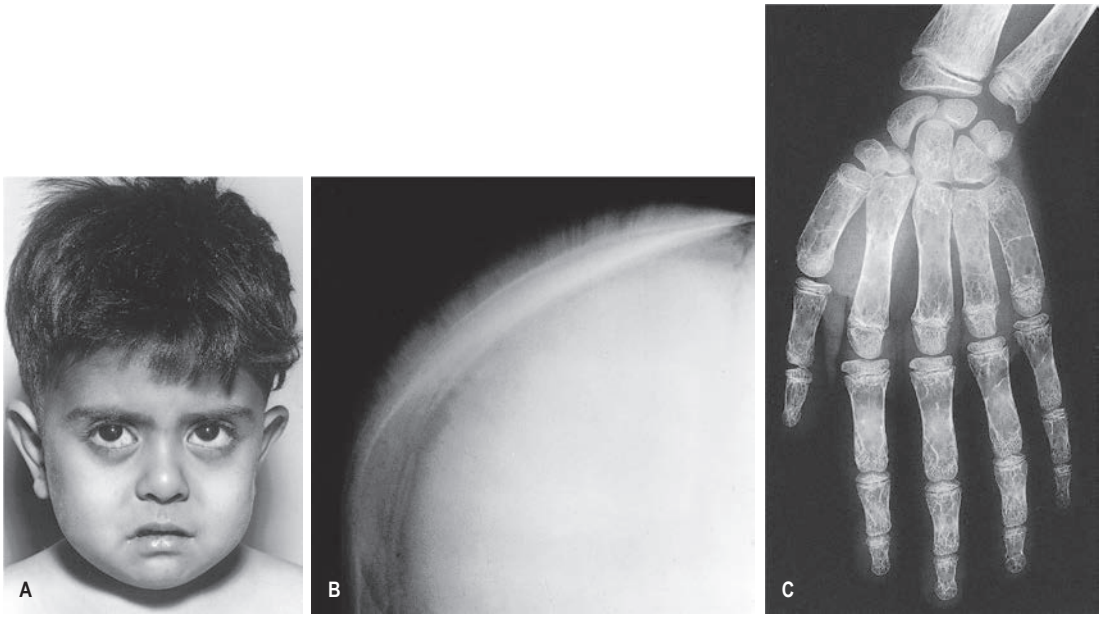


Fig. 38.24 Thalassaemia. (A) ***Child with thalassaemia*** showing the typical facial features. (B) ***Skull radiograph*** of a child with β-thalassaemia showing a 'hair-on-end' appearance. (C) ***Radiograph of a hand*** showing expansion of marrow and a thinned cortex.

Box 38.12 The α-thalassaemias

Number of α-globin genes deleted	Genotype[a]	Hb type	Clinical picture
4	−−/−−	Hb Barts (γ4)	Hydrops fetalis
3	−−/−α	HbH (β4)	Moderately severe anaemia Splenomegaly
2	−α/−α	HbA	Mild anaemia
	Or	Some HbH bodies	
	−−/αα		
1	αα/−α	HbA	Very mild/no anaemia

[a]The normal α-globin genotype is αα/αα (i.e., four α-globin genes present).
Hb, Haemoglobin.

- ***One-gene deletion.*** The blood picture is usually normal, but there microcytosis may be present.

Globin chain synthesis studies for the detection of a reduced ratio of α to β chains may be necessary for the definitive diagnosis of α-thalassaemia trait, although standard DNA analysis using gene sequencing can detect the most common deletions that lead to α-thalassaemia trait.

Further reading

Guerra A, Parhiz H, Rivella S. Novel potential therapeutics to modify iron metabolism and red cell synthesis in diseases associated with defective erythropoiesis. *Haematologica* 2023.
Higgs DR. The molecular basis of α-thalassemia. *Cold Spring Harb Perspect Med* 2013; 3(1):a011718.
Piel FB, Weatherall DJ. The α thalassaemias. *N Engl J Med* 2014; 371:1908–1916.
Saeidnia M, Fazeli P, Farzi A et al. An expert overview on therapies in non-transfusion-dependent thalassemia: classical to cutting edge in treatment. *Hemoglobin* 2023; 47(2):56–70.

Sickle syndromes

Sickle cell Hb (HbS) results from a single-base mutation of adenine to thymine, which produces a substitution of valine for glutamic acid at the sixth codon of the β-globin chain ($\alpha_2\beta_2^{6glu\rightarrow val}$). In the homozygous state ***(sickle cell anaemia)***, both genes are abnormal (HbSS), whereas in the heterozygous state (***sickle cell trait***, HbAS), only one chromosome carries the abnormal gene. Because the synthesis of HbF is normal, the disease usually does not manifest itself until the HbF decreases to adult concentrations at about 6 months of age.

The sickle gene is most common in Africans (up to 25% gene frequency in some populations) but is also found in India, the Middle East, and Southern Europe. Its distribution mirrors that of malaria infestation.

Pathogenesis

Deoxygenated HbS molecules are insoluble and polymerize. The flexibility of the cells is decreased; they become rigid and take up their characteristic sickle appearance (Fig. 38.26). This process is initially reversible, but with repeated sickling, cells eventually lose their membrane flexibility and become irreversibly sickled. This is due to dehydration, partly caused by potassium leaving the RBCs via a calcium-activated potassium channel called the Gardos channel. These irreversibly sickled cells are dehydrated and dense and do not return to normal when oxygenated. Sickling can produce:

- shortened RBC survival (haemolysis)
- impaired passage of cells through the microcirculation, leading to obstruction of small vessels and tissue infarction.

Sickling is precipitated by infection, dehydration, cold, stress, acidosis, and hypoxia. In many cases, the cause is unknown, but adhesion proteins on activated endothelial cells (vascular cell adhesion molecule 1) may play a causal role, particularly in vaso-occlusion when rigid cells are trapped, facilitating polymerization. HbS releases its oxygen to the tissues more easily than does normal

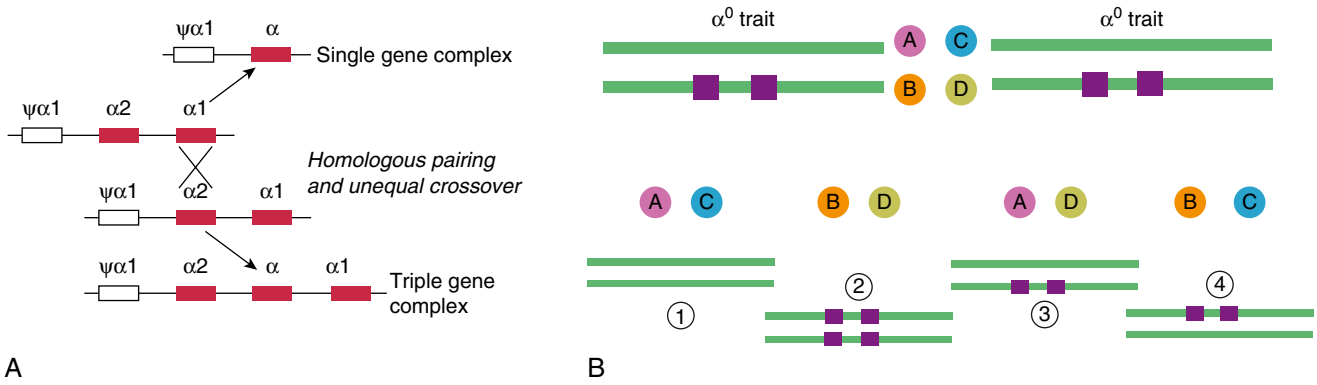


Fig. 38.25 Alpha-thalassaemia. **(A)** Deletions of α globin result from unequal crossover in the homologous recombination that occurs during meiosis. This also results in one allele with a triplicated α gene, which is of no clinical consequence in healthy individuals but can exacerbate the globin chain imbalance in individuals with reduced or abnormal β globin. **(B)** Possible outcomes from a pairing of two individuals with α^0 thalassaemia. Antenatal screening is needed because of a 25% risk of hydrops fetalis in the fetus. There is a 50% chance that the offspring will be an α^0 carrier and a 25% chance that the offspring will have all four α-globin chains.

Fig. 38.26 Sickle cells *(arrow)* and target cells.

Hb, and patients therefore feel well despite being anaemic (except, of course, during crises or complications).

Depending on the type of Hb chain combinations, three clinical syndromes occur:

- ***Homozygous HbSS*** patients have the most severe disease. Patients with co-inheritance of HbS and β-thalassaemia can have equally severe disease.
- ***Combined heterozygosity (HbSC) for HbS and C*** (see later) causes intermediate symptoms.
- ***Heterozygous HbAS (sickle cell trait)*** has no symptoms.

Sickle cell anaemia

Clinical features

Vaso-occlusive crises

An early presentation may be acute pain in the hands and feet (dactylitis) owing to vaso-occlusion of the small vessels. Severe pain in other bones, including the femur, humerus, vertebrae, ribs, and pelvis, occurs in older children and adults. Attacks vary greatly in frequency from patient to patient and sometimes in the same patient from year to year. Patients can be moderately asymptomatic or require multiple hospital admissions because of painful vaso-occlusive crises. Overall, 25% of patients with sickle cell disease report experiencing daily pain. Fever often accompanies pain.

Acute chest syndrome

This occurs in up to 30%, and pulmonary hypertension and chronic lung disease are the most common causes of death in adults with sickle cell disease. The acute chest syndrome is caused by infection, fat embolism from necrotic bone marrow in a vaso-occlusive crisis, or pulmonary infarction caused by sequestration of sickle cells. It consists of shortness of breath, chest pain, pyrexia, hypoxia, and new chest radiograph changes because of consolidation. The presentation may be gradual or very rapid, carrying a risk of death in a few hours. Management is with pain relief, high-flow supplemental oxygen, and antibiotics (cefotaxime and clarithromycin), which should be started immediately. If there is no improvement, exchange transfusion will reduce the amount of HbS to less than 30%. Continuous positive airways pressure may be necessary. Infections can be caused by *Chlamydia* and *Mycoplasma* spp., as well as *Streptococcus pneumoniae*. Use of an incentive spirometer to encourage deep breathing in patients admitted with vaso-occlusive crises reduces the risk of developing the acute chest syndrome.

Pulmonary hypertension

Pulmonary hypertension (defined as mean pulmonary artery pressure of >25 mmHg by right heart catheterization) is a known consequence of sickle cell anaemia, occurring in approximately 10% of patients. Its mechanism is not understood, but hypothesized contributory factors include sickle cell vasculopathy, damage from repeated chest crises, repeated thromboembolism, and intravascular haemolysis, which affect reactive oxygen species and concentrations of scavenging nitric oxide. Development of pulmonary hypertension increases the risk of hypoxaemia and worsening sickle cell crises and is an independent risk factor for death in patients with sickle cell anaemia. Diagnosis of pulmonary hypertension in patients with sickle cell anaemia should prompt consideration of therapies to reduce concentrations of HbS (see later) and referral to a pulmonary hypertension specialist.

Anaemia

Chronic haemolysis produces a stable Hb concentration, usually in the 60 to 80 g/L range, but an acute fall in Hb can occur owing to:

- Splenic sequestration
- Bone marrow aplasia
- Further haemolysis caused by drugs, acute infection, or associated glucose-6-phosphate dehydrogenase (G6PD) deficiency.

Splenic sequestration

Vaso-occlusion produces an acute painful enlargement of the spleen. There are splenic pooling of RBCs and hypovolaemia, leading in some to circulatory collapse and death. The condition supervenes in childhood before multiple splenic infarctions have occurred, eventually leading to a fibrotic, non-functioning spleen. After an episode of sequestration has occurred, it is likely to recur and be life-threatening. In some cases, surgical splenectomy is required to prevent recurrence. Liver sequestration can also occur.

Bone marrow aplasia

This most commonly follows infection with parvovirus B19, which invades proliferating erythroid progenitors. There is a rapid fall in Hb with no reticulocytes in the peripheral blood because of the failure of erythropoiesis in the marrow. Patients may require transfusion until their bone marrow recovers.

Long-term problems

- ***Growth and development.*** Young children are short but regain their height by adulthood. However, they remain below the normal weight. Sexual maturation is often delayed and may require hormone therapy.
- ***Bones.*** Bones are a common site for vaso-occlusive episodes, leading to chronic infarcts. Avascular necrosis of the hips and shoulders, compression of vertebrae, and shortening of bones in the hands and feet occur. Osteomyelitis is more frequent in sickle cell disease and is caused by *Staphylococcus aureus*, *Streptococcus pneumoniae*, and salmonella. Occasionally, hip or knee joint replacement may be required at a young age.
- ***Infections.*** Infections are common in tissues susceptible to vaso-occlusion, such as the bones, lungs, and kidneys.
- ***Leg ulcers.*** These occur spontaneously (vaso-occlusive episodes) or follow trauma. They are usually found over medial or lateral malleoli. They often become infected and are quite resistant to treatment; blood transfusion may facilitate ulcer healing.
- ***Cardiac problems include*** cardiomegaly, arrhythmias, and iron overload cardiomyopathy. Myocardial infarctions are caused by thrombotic episodes, not secondary to atheroma.
- ***Neurological complications*** occur in 25% of patients and include transient ischaemic attacks, fits, cerebral infarction, cerebral haemorrhage, and coma. About 11% of patients younger than 20 years of age have strokes without a stroke prevention programme. The most common finding is obstruction of a distal intracranial internal carotid artery or a proximal middle cerebral artery. About 10% of children without neurological signs or symptoms have abnormal blood-flow velocity indicative of clinically significant arterial stenosis; such patients have a very high risk of stroke. It has been demonstrated that if children with stenotic cranial artery lesions, as demonstrated on transcranial Doppler ultrasonography, are maintained on a regular RBC transfusion exchange programme designed to suppress erythropoiesis so that no more than 30% of the circulating RBCs are their own, about 90% of strokes in such children are prevented. Whether they need to remain on these transfusions for life remains to be determined. Silent infarcts, seen on MRI but not manifesting clinically, are common in children with sickle cell disease and associated with poor educational attainment.
- ***Cholelithiasis.*** Pigment stones occur as a result of chronic haemolysis.
- ***Liver complications.*** Chronic hepatomegaly and liver dysfunction are caused by trapping of sickle cells.
- ***Renal complications.*** Chronic tubulointerstitial nephritis occurs (see Chapter 26), and 10% of patients develop CKD requiring long-term kidney replacement therapy.
- ***Priapism.*** An unwanted painful erection occurs because of vaso-occlusion and can be recurrent. This requires urgent treatment because it may result in impotence. Treatment is with an α-adrenergic blocking drug, analgesia, and hydration. When this is not successful, aspiration or shunts may be required.
- ***Eye complications.*** Background retinopathy, proliferative retinopathy, vitreous haemorrhages, and retinal detachments all occur. Annual eye checks are required.
- ***Pregnancy.*** Impaired placental blood flow causes spontaneous abortion, intrauterine growth restriction, pre-eclampsia, and fetal death. Painful episodes, infections, and severe anaemia occur. Pregnancy requires monitoring by a combination of haematologists and high-risk obstetricians.

Investigations

- ***Blood count.*** The concentration of Hb is in the range of 60 to 80 g/L, with a high reticulocyte count (10%–20%).
- ***Blood film.*** This shows features of hyposplenism (Fig. 38.19) and sickling (Fig. 38.26).
- ***Sickle solubility test.*** A mixture of HbS in a reducing solution such as sodium dithionite gives a turbid appearance because of precipitation of HbS, whereas normal Hb gives a clear solution. Commercial kits, such as Sickledex, are available for rapid screening for the presence of HbS (e.g., before surgery and in the emergency department).
- ***Hb electrophoresis*** is needed to confirm the diagnosis. There is no HbA, 80% to 95% HbS, and 2% to 20% HbF (Fig. 38.23).
- ***Parents.*** The parents of affected children show features of sickle cell trait.

Management

Precipitating factors (see earlier) should be avoided or treated quickly. Complications that require inpatient management are shown in Box 38.13.

Acute painful attacks require supportive therapy with intravenous fluids and adequate analgesia. Oxygen and antibiotics are given if indicated. Crises can be extremely painful and require strong, usually narcotic, analgesia. Morphine is the drug of choice. Milder pain can sometimes be relieved by codeine, paracetamol, or NSAIDs. Use of incentive spirometry is recommended for patients hospitalized for a vaso-occlusive crisis.

Infection prophylaxis includes penicillin 500 mg/day and vaccination with polyvalent pneumococcal, meningococcal, and *Haemophilus influenzae* type b vaccines because of functional hyposplenism.

Folic acid is given to all patients with haemolysis.

Anaemia

Blood transfusions should be given only when there are clear indications. Patients with steady-state anaemia, those having minor surgery, and those having painful episodes without complications should not be transfused. Transfusions (top-up or exchange) should be given acutely for acute chest syndrome and acute anaemia associated with splenic sequestration or an aplastic crisis, aiming for an Hb of 100 g/L and HbS below

Box 38.13 Complications of sickle cell disease requiring inpatient management

- Pain uncontrolled by non-opiate analgesia
- Swollen, painful joint(s)
- Acute sickle chest syndrome or pneumonia
- Mesenteric sickling, bowel ischaemia
- Splenic or hepatic sequestration
- Central nervous system deficit
- Cholecystitis (pigment stones)
- Cardiac arrhythmias
- Renal papillary necrosis resulting in colic or haematuria
- Hyphaema, retinal detachment
- Priapism

30%. Before elective operations, top-up transfusions are given to increase the Hb to 100 g/L and reduce risk of perioperative complications. During pregnancy, repeated transfusions may be used to reduce the proportion of circulating HbS to below 30% to prevent sickling and improve placental function. Regular exchange transfusions are given to children with a history of stroke (secondary prevention) or a high risk of stroke as assessed by transcranial Doppler (primary prevention). Exchange transfusions may be necessary in patients with severe or recurrent crises and before emergency surgery. RBCs for transfusion should be matched for Rhesus (Rh) and Kell blood groups to minimize the risk of alloimmunization, which arises from the mismatch between blood donors and recipients. Ideally, patients requiring long-term transfusion, such as those with sickle cell anaemia and thalassaemia, should undergo molecular blood group typing before their first transfusion.

Hydroxycarbamide (hydroxyurea) is widely used as therapy for sickle cell anaemia. It acts, at least in part, by increasing HbF concentrations. Hydroxycarbamide has been shown in trials to reduce episodes of pain, acute chest syndrome, and the need for blood transfusions, as well as to increase lifespan. It has been shown to be safe and effective and probably reduces long-term complications such as kidney dysfunction. It is now offered to almost all patients with sickle cell disease from the age of 12 months. Blood counts must be carefully monitored for neutropenia, which is a side-effect of treatment.

Novel therapies include crizanlizumab (anti–P-selectin antibody administered as a monthly intravenous infusion) and Voxelotor (a tablet that stabilizes oxygenated HbS). Their use is currently limited because not enough clinical trials have shown benefit, but they provide new avenues for treatment.

Stem cell transplantation has been used to treat patients with sickle cell anaemia but in lower numbers than for thalassaemia. Children and adolescents younger than 16 years of age who have severe complications (strokes, recurrent chest syndrome, or refractory pain) and an HLA-matched donor are the best candidates for transplantation.

Counselling and psychological support are important parts of improving long-term outcomes for people living with chronic illnesses. A multidisciplinary team should be involved, with regular clinic appointments to build up relationships. Adolescents require careful counselling including psychosocial aspects of care and contraception.

Prognosis

Some patients with HbSS die in the first few years of life from either infection or episodes of sequestration. However, there is marked variation in disease severity, and some patients have a relatively normal lifespan with few complications. In most, the length and quality of life are adversely affected, and regular monitoring is important to prevent long-term complications.

Sickle cell trait

These individuals have no symptoms unless extreme circumstances cause anoxia (e.g., flying in a non-pressurized aircraft). At a population level, sickle cell trait confers some protection against *Plasmodium falciparum* malaria, and the sickle cell gene has been seen as an example of a balanced polymorphism: the advantage of malaria protection in the heterozygote is balanced by the mortality of the homozygous condition. However, for an individual, the presence of sickle cell trait does not obviate the need for antimalarial prophylaxis. Typically, there is 60% HbA and 40% HbS. It should be emphasized that unlike thalassaemia trait, the blood count and film of a person with sickle cell trait are normal. The diagnosis is made by a positive sickle test or Hb electrophoresis (Fig. 38.23).

Other structural globin chain defects

There are many Hb variants, and most are not associated with any clinical manifestations. However, some Hb variants may interact with HbS; for example, compound heterozygosity for HbC and HbS gives rise to HbSC disease. The clinical course of HbSC disease is generally somewhat milder than that of HbSS disease, but there is an increased likelihood of retinopathy and thrombosis, which may lead to thrombosis in pregnancy.

Combined defects of globin chain production and structure

Abnormalities of Hb structure (e.g., HbS, HbC) can occur in combination with thalassaemia. The combination of β-thalassaemia trait and sickle cell trait (sickle cell β-thalassaemia) resembles sickle cell anaemia (HbSS) clinically.

HbE ($\alpha_2\beta_2^{26glu\rightarrow lys}$) is the most common Hb variant in South-east Asia and the second most prevalent Hb variant worldwide. HbE heterozygotes are asymptomatic; Hb concentration is normal, but RBCs are microcytic. Homozygous HbE causes a mild microcytic anaemia, but the combination of heterozygosity for HbE and β-thalassaemia produces a variable anaemia, which can be as severe as TDT (β-thalassaemia major).

Prenatal screening and diagnosis of severe haemoglobin abnormalities

Of the offspring of parents who both have either β-thalassaemia or sickle cell trait, 25% have TDT or sickle cell anaemia, respectively. Recognition of heterozygous states in parents and family counselling provide a basis for antenatal screening and diagnosis. Pregnant persons with either sickle cell trait or thalassaemia trait are identified at antenatal booking either by selective screening of high-risk groups or by universal screening of all pregnancies. Beta-thalassaemia trait can be detected by a low MCV and MCH and confirmed by Hb electrophoresis. However, sickle cell trait is undetectable from a blood count and needs a specific request to screen for sickle cell trait.

If a Hb defect is detected in pregnancy, the partner should be tested. Antenatal diagnosis is offered if both are affected, and there is a risk of a severe fetal Hb defect, particularly TDT. Fetal DNA analysis can be carried out using amniotic fluid, chorionic villus sampling, or fetal blood samples. Termination of pregnancy

is discussed if the fetus is found to be severely affected. Chorionic villus biopsy has the advantage that it can be carried out in the first trimester, thus avoiding the need for second-trimester termination.

Further reading

Piel FB et al. Defining global strategies to improve outcomes in sickle cell disease: a Lancet Haematology Commission. *Lancet Haematol* 2023.
Qureshi A, Kaya B, Pancham S et al. Guidelines for the use of hydroxycarbamide in children and adults with sickle cell disease. *Br J Haematol* 2018; 181:460–475.
Ribeil JA, Pollock G, Frangoul H, Steinberg MH. An integrated therapeutic approach to sickle cell disease management beyond infancy. *Am J Hematol* 2023; 98(7):1087–1096.

Metabolic disorders of the red blood cell

Red blood cell metabolism

A mature RBC has no nucleus, mitochondria, or ribosomes and is therefore unable to synthesize proteins. RBC enzymes maintain the viability and function of the cells, but the lifespan of these enzymes is limited because there is no nucleus to replenish them after they become defective. Energy is required in the form of ATP to maintain flexibility of the membrane and the biconcave shape to allow passage through small vessels and for regulation of sodium and potassium pumps to ensure osmotic equilibrium. In addition, it is essential for Hb to be maintained in the reduced state.

The enzyme systems responsible for producing energy and maintaining a reduced state (Fig. 38.27) are:

- ***The glycolytic (Embden–Meyerhof) pathway:*** Glucose is metabolized to pyruvate and lactic acid with production of ATP.
- ***The hexose monophosphate (pentose phosphate) pathway:*** This provides reducing power for the RBCs in the form of NADPH (the reduced form of nicotinamide adenine dinucleotide phosphate).

These biochemical processes are understandably complex; however, it is possible to focus on relevant aspects that lead to a clinical phenotype. For example, pyruvate kinase (Fig. 38.27, highlighted in *orange*) is the last enzyme in the glycolytic pathway. When this is defective, less ATP is formed, and less energy is available to the RBCs for their needs. RBCs use ATP to maintain the correct concentrations of solutes inside the cytoplasm by driving ion pumps; without enough ATP, these ionic gradients cannot be maintained, and cell survival is shortened.

About 90% of glucose is metabolized by the glycolytic pathway, and 10% goes through the hexose monophosphate shunt whose main function is to maintain glutathione in a reduced state. Glutathione is necessary to combat oxidative stress to the RBCs, and failure of this mechanism may result in:

- ***Rigidity*** caused by cross-linking of spectrin, which decreases membrane flexibility (Fig. 38.17) and causes 'leakiness' of the RBC membrane
- ***Oxidation of the Hb molecule***, producing methaemoglobin and precipitation of globin chains as Heinz bodies localized on the inside of the membrane; these bodies are removed from circulating RBCs by the spleen.

2,3-BPG is formed from a side-arm of the glycolytic pathway labelled the 'Rappaport-Luebering shunt' in Fig. 38.27. It binds to the central part of the Hb tetramer, fixing it in the low-affinity state. A decreased affinity with a shift in the oxygen dissociation curve to the right enables more oxygen to be delivered to the tissues. 2,3-BPG therefore has a role in modulating the severity of these disorders.

Glucose-6-phosphate dehydrogenase deficiency

The G6PD enzyme occupies a vital position in the hexose monophosphate shunt highlighted in *yellow* in Fig. 38.27. It oxidizes glucose-6-phosphate to 6-phosphoglycerate with the reduction of NADP to NADPH. The reaction is necessary in RBCs, where it is the only source of NADPH, which is used via glutathione to protect the RBCs from oxidative damage. G6PD deficiency is a common condition that presents with a haemolytic anaemia and affects millions of people throughout the world, particularly in Africa, the Mediterranean, the Middle East (~20%), and South-east Asia (up to 40% in some regions).

The gene for G6PD is localized to chromosome Xq28 near the factor VIII gene. Deficiency is therefore more common in males than females. However, female heterozygotes can also have clinical problems because of lyonization whereby, because of random X-chromosome inactivation, female heterozygotes have two populations of RBCs: a normal one and a G6PD-deficient one.

There are more than 400 structural types of G6PD, and mutations are mostly single amino acid substitutions (missense point mutations). The World Health Organization has classified variants by the degree of enzyme deficiency and severity of haemolysis. The most common ones with normal activity are called type B^+ (present in almost all Whites and ~70% of Black Africans) and type A^+ (present in about 20% of Black Africans). There are many variants with reduced activity but, only two are common:

- African (A–type). Deficiency is mild and more marked in older cells. Haemolysis is self-limiting because young RBCs newly produced by the bone marrow have nearly normal enzyme activity.
- Mediterranean type. Both young and old RBCs have very low enzyme activity. After an oxidant shock, the Hb concentration may fall precipitously; death may follow unless the condition is recognized and the patient is transfused urgently.

Clinical syndromes

- Acute drug-induced haemolysis (Box 38.14): usually dose related
- Favism: haemolysis after ingestion of fava beans
- Chronic haemolytic anaemia
- Neonatal jaundice
- Infections and acute illnesses precipitating haemolysis
- Mothballs containing naphthalene can also cause haemolysis

Clinical features are caused by rapid intravascular haemolysis with symptoms of anaemia, jaundice, and haemoglobinuria.

Investigations

- The blood count is normal between attacks.
- ***During an attack***, the blood film may show irregularly contracted cells, bite cells (cells with an indentation in the membrane), blister cells (cells in which the Hb appears to have become partially detached from the cell membrane; Fig. 38.28), Heinz bodies (RBC inclusions composed of denatured Hb, best seen on films stained with methyl violet), and reticulocytosis.
- ***Haemolysis*** is evident (see Fig. 38.16 and 'Evidence for haemolysis' section).
- ***Screening tests*** include demonstration of the decreased ability of G6PD-deficient cells to reduce dyes and direct assay of enzyme concentration. Immediately after an attack, test results

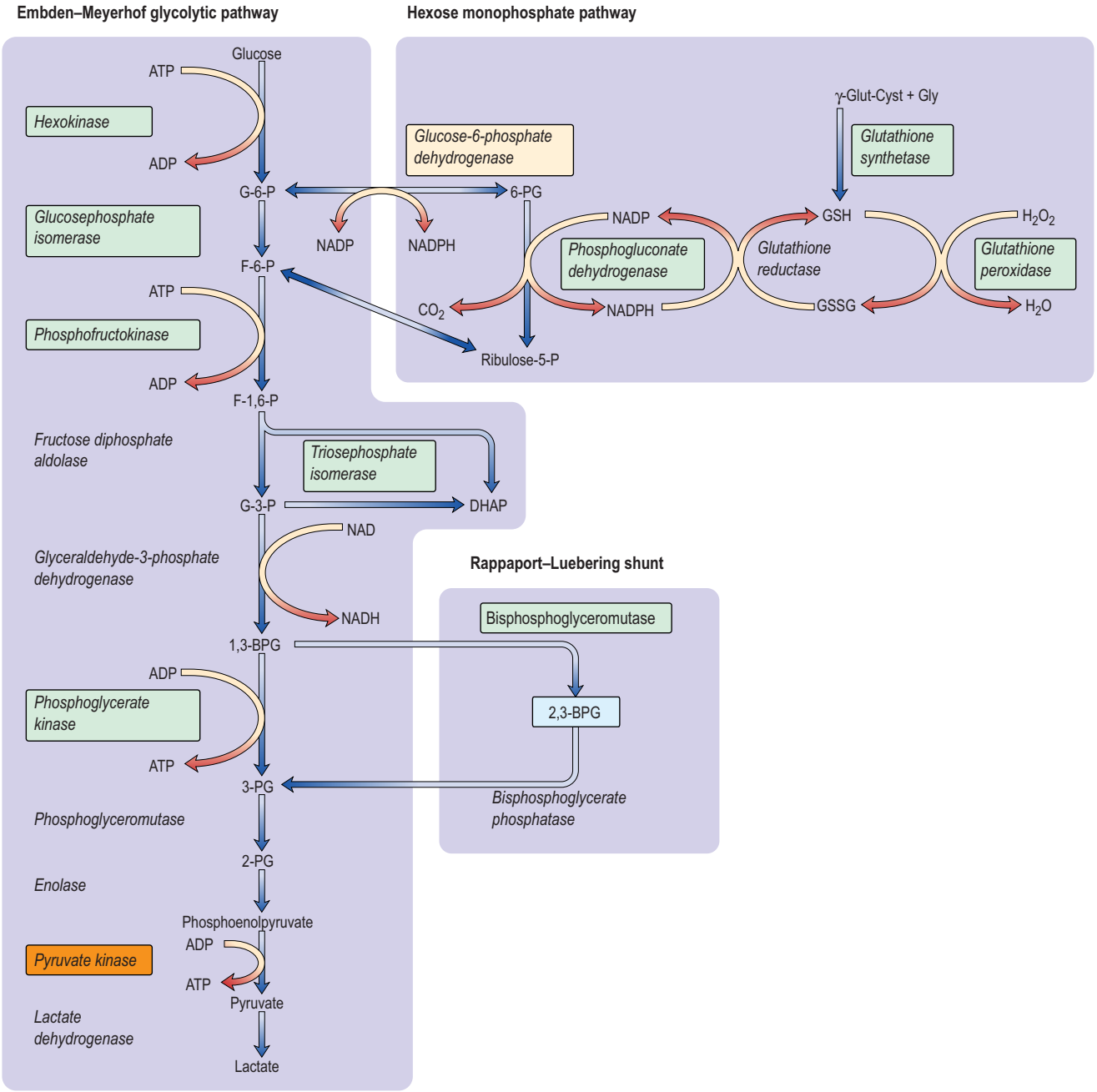


Fig. 38.27 Glucose metabolism pathways. The enzymes in *coloured boxes* indicate documented hereditary deficiency diseases. *ADP,* Adenosine diphosphate; *ATP,* adenosine triphosphate; *BPG,* bisphosphoglycerate; *DHAP,* dihydroxyacetone phosphate; *F,* fructose; *G,* glucose; *GSH,* reduced glutathione; *GSSG,* oxidized glutathione; *NADH,* reduced nicotinamide adenine dinucleotide; *NADP,* nicotinamide adenine dinucleotide phosphate; *NADPH,* reduced NADP; *P,* phosphate; *PG,* phosphoglycerate.

may be normal because the oldest RBCs with least G6PD activity are destroyed selectively. Diagnosis of heterozygous females may be difficult because the enzyme concentration may range from very low to normal, depending on Lyonization. However, the risk of clinically significant haemolysis is minimal in patients with borderline G6PD activity.

Management

- Stop the offending drugs.
- Treat the underlying infection.
- Blood transfusion may be life-saving.
- Splenectomy is not usually helpful.

Pyruvate kinase deficiency

This is the most common defect of RBC metabolism after G6PD deficiency; it affects thousands rather than millions of people worldwide. The site of the defect is shown in *orange* in Fig. 38.27. Production of ATP is reduced, causing rigid RBCs. It is inherited as an autosomal recessive condition. Homozygotes have haemolytic anaemia and splenomegaly. Interestingly, with a dysfunctional pyruvate kinase

Box 38.14 Examples of drugs causing haemolysis in glucose-6-phosphate dehydrogenase deficiency

Analgesics
- Aspirin
- Phenacetin (withdrawn in the UK)

Antimalarials
- Primaquine
- Pyrimethamine
- Quinine
- Chloroquine
- Pamaquine

Antibacterials
- Most sulphonamides
- Nitrofurantoin
- Chloramphenicol
- Quinolones

Miscellaneous
- Dapsone
- Vitamin K
- Probenecid
- Quinidine
- Dimercaprol
- Phenylhydrazine
- Methylene blue

Fig. 38.28 'Blister' cells *(arrow)* in glucose-6-phosphate dehydrogenase deficiency.

enzyme, 2,3-BPG accumulates (highlighted in *blue* in Fig. 38.27), causing a right shift in the oxygen dissociation curve. This reduced Hb affinity for oxygen means that more oxygen is released to the tissues and partly compensates for the degree of anaemia.

Investigations

- ***Anaemia*** of variable severity (Hb 50–100 g/L). The oxygen dissociation curve is shifted to the right because of the rise in intracellular 2,3-BPG, which reduces severity of symptoms caused by anaemia.
- ***Blood films*** show distorted ('prickle') cells and a reticulocytosis.
- ***Pyruvate kinase activity*** is low; affected homozygotes have concentrations of 5% to 20%.

Management

Blood transfusions may be necessary during infections and pregnancy. Splenectomy may improve the clinical condition and is usually advised for patients requiring frequent transfusions. Novel drugs that stabilize the enzyme in RBCs have been shown to reduce haemolysis in clinical trials and may become available for clinical use.

Pyrimidine 5′ nucleotidase deficiency

This autosomal recessive disorder is the third most common RBC enzyme disorder after G6PD and pyruvate kinase, causing a haemolytic anaemia with basophilic stippling of the RBCs. The enzyme degrades pyrimidine nucleotides to cytidine and uridine, which feed into the pathway responsible for degradation of RNA in the reticulocytes. Lack of the enzyme results in accumulation of partially degraded RNA, which shows as basophilic stippling in mature RBCs. The enzyme is also inhibited by lead; therefore, basophilic stippling is also seen in lead poisoning (see Chapter 15). The hereditary form can be diagnosed by measuring the enzyme in erythrocytes. A screening test using the ultraviolet absorption spectrum of RBCs is available.

In addition to the G6PD, pyruvate kinase and pyrimidine 5′ nucleotidase deficiencies described here, specialist investigation of other rarer enzyme deficiencies highlighted in *green* in Fig. 38.27 also occurs.

Further reading

Grace RF, Glader B. Red blood cell enzyme disorders. *Pediatr Clin North Am* 2018; 65:579–595.
Roper D, Layton M, Rees D et al. Laboratory diagnosis of G6PD deficiency. A British Society for Haematology Guideline. *Br J Haematol* 2020; 189:24–38.

ACQUIRED HAEMOLYTIC ANAEMIA

These anaemias may be divided into those that have immune, non-immune, or other causes (Box 38.9).

Aetiology

Causes of immune destruction of red blood cells

- Autoantibodies
- Drug-induced antibodies
- Alloantibodies

Causes of non-immune destruction of red blood cells

- Acquired membrane defects (e.g., PNH, see 'Paroxysmal nocturnal haemoglobinuria' section)
- Mechanical factors (e.g., prosthetic heart valves or microangiopathic haemolytic anaemia [MAHA]; see 'Mechanical haemolytic anaemia' section)
- Secondary to systemic disease (e.g., kidney and liver disease)

Miscellaneous causes

- Various toxic substances (e.g., arsenic, toxins produced by *Clostridium perfringens [welchii]*) can disrupt the RBC membrane and cause haemolysis.
- Malaria frequently causes anaemia owing to the combination of reduced RBC survival and reduced production of RBCs.
- Hypersplenism results in reduced RBC survival, which may also contribute to the anaemia seen in those with malaria.
- Extensive burns lead to denaturation of RBC membrane proteins and reduced RBC survival.
- Some drugs (e.g., dapsone, sulfasalazine) cause oxidative haemolysis with Heinz bodies.
- Some ingested chemicals (e.g., weedkillers such as sodium chlorate) can cause severe oxidative haemolysis, leading to acute kidney injury.

Autoimmune haemolytic anaemias

Autoimmune haemolytic anaemias (AIHAs) are acquired disorders resulting from increased RBC destruction caused by

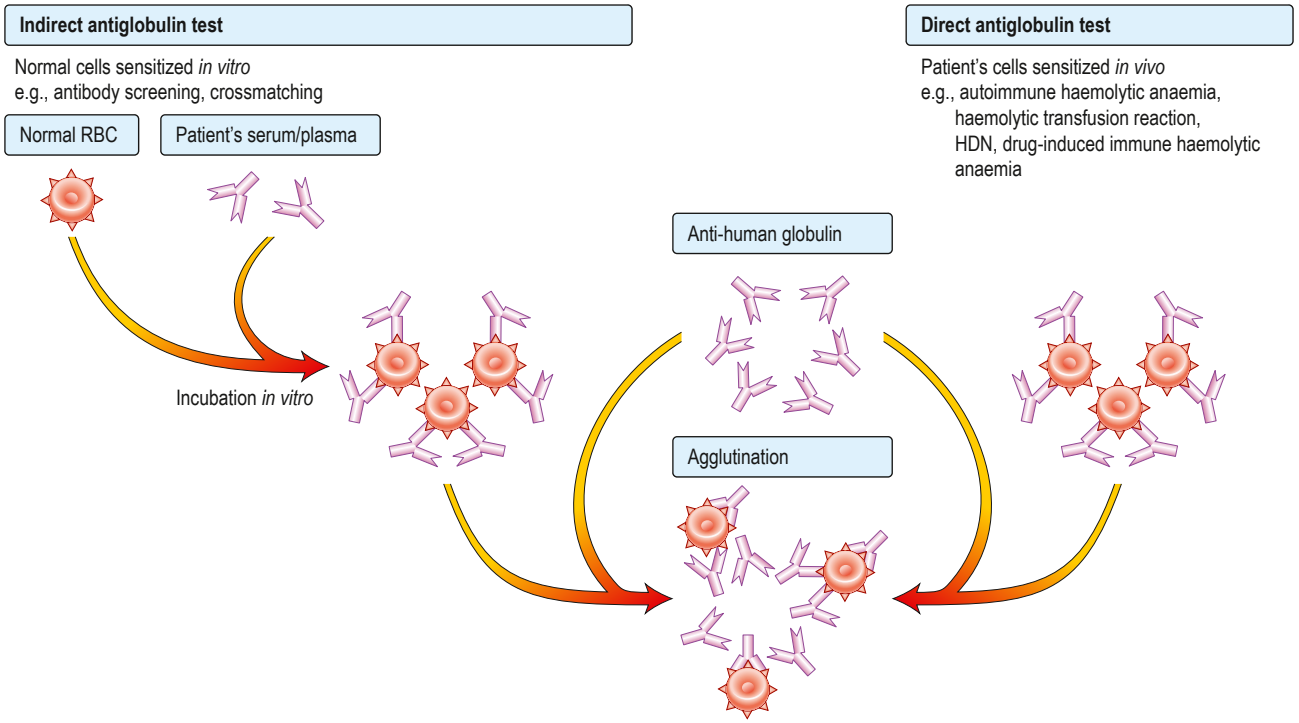


Fig. 38.29 Antiglobulin (Coombs') tests. The anti-human globulin forms bridges between the sensitized cells, causing visible agglutination. The direct test detects patients' cells sensitized *in vivo*. The indirect test detects normal cells sensitized *in vitro*. *HDN,* Haemolytic disease of newborn; *RBC,* red blood cell.

RBC autoantibodies. These anaemias are characterized by the presence of a positive direct antiglobulin (Coombs') test result, which detects evidence of the presence of autoantibody (immunoglobulin [Ig] G, or complement) on the surface of the patient's RBCs (Fig. 38.29).

AIHA is divided into 'warm' (65%), 'cold' (30%), and mixed (5%) types, depending on whether the antibody attaches better to the RBCs at body temperature (37°C) or at lower temperatures. The major features and the causes of these two forms of AIHA are shown in Box 38.15. In warm AIHA, IgG antibodies predominate, and the result of the DAT is positive with IgG alone, IgG and complement, or complement only. In cold AIHA, the antibodies are usually IgM. They easily elute off RBCs, leaving complement, which is detected as C3d.

Immune destruction of red blood cells

IgM and IgG RBC antibodies that fully activate the complement cascade cause lysis of RBCs in the circulation and intravascular haemolysis.

The majority of IgG antibodies do not activate complement, and the coated RBCs undergo extravascular haemolysis (Fig. 38.30). They either are completely phagocytosed in the spleen through an interaction with Fc receptors on macrophages or lose part of the cell membrane through partial phagocytosis and circulate as spherocytes until they become sequestered in the spleen.

Some IgG antibodies may partially activate complement, leading to deposition of C3b on the RBC surface, and this may enhance phagocytosis as macrophages also have receptors for C3b.

IgM antibodies, which are non-complement binding, are rare and have little or no effect on RBC survival. IgM antibodies, which partially rather than fully activate complement, cause adherence of RBCs to C3b receptors on macrophages, particularly in the liver, although this is an ineffective mechanism of haemolysis. Most of the RBCs are released from the macrophages when C3b is cleaved to C3d and then circulate with C3d on their surface, which is not relevant clinically but allows them to be detected by the DAT.

'Warm' autoimmune haemolytic anaemias

Clinical features

These anaemias occur at all ages and in both sexes, although they are most frequent in middle-aged females. They can present as a short episode of anaemia and jaundice, but they often remit and relapse and may progress to an intermittent chronic pattern. The spleen may be palpable. Infection or folate deficiency may provoke a profound fall in Hb concentration. AIHAs are primary (50%) or secondary to an associated disorder (50%). The most common underlying cause is a lymphoproliferative disorder (Box 38.15).

Investigations

- ***Haemolytic anaemia*** is evident (see 'Haemolytic anaemias' section).
- ***Spherocytosis*** is present as a result of RBC damage.
- The ***DAT*** result is positive, with IgG alone (35%), IgG and complement (56%), or complement alone (9%) found on the surface of the RBCs.
- ***Autoantibodies*** may have specificity for the Rh blood group system (e.g., for the e antigen).
- ***Autoimmune thrombocytopenia*** or neutropenia may also be present (***Evans syndrome***).
- ***Abdominal computed tomography (CT)*** should be carried out to detect abdominal lymphoma.

Box 38.15 Causes and major features of autoimmune haemolytic anaemias

	Warm	Cold
Temperature at which antibody attaches best to red blood cells (°C)	37	<37
Type of antibody	IgG	IgM
Direct Coombs' test	Strongly positive	Positive
Causes of primary condition	Idiopathic	Idiopathic
Causes of secondary condition	Autoimmune (e.g., SLE) Chronic lymphocytic leukaemia Lymphomas Hodgkin lymphoma Carcinomas Drugs: many, including methyldopa, penicillins, cephalosporins, NSAIDs, quinine, interferon	Infections (e.g., infectious mononucleosis, *Mycoplasma pneumoniae*, other viral infections [rare]) Lymphomas Paroxysmal cold haemoglobinuria (IgG)

Ig, Immunoglobulin; *NSAID,* non-steroidal anti-inflammatory drug; *SLE,* systemic lupus erythematosus.

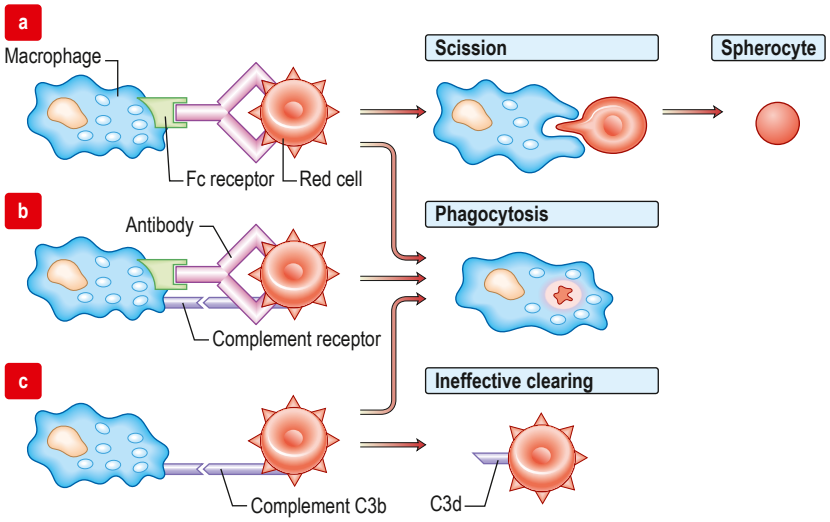


Fig. 38.30 Extravascular haemolysis. This is caused by interaction of antibody-coated cells with cells in the reticuloendothelial system, predominantly in the spleen. **(a)** Spherocytosis results from partial phagocytosis. **(b)** Complete phagocytosis may occur and is enhanced if there is complement as well as antibody on the cell surface. **(c)** Cells coated with complement only are ineffectively removed and circulate with C3d or C3b on their surfaces. *RBC,* Red blood cell.

Management and prognosis

Corticosteroids (e.g., prednisolone 1 mg/kg daily) are effective in inducing a remission in about 80% of patients, although only 30% to 40% have a sustained remission after 1 year. Steroids reduce both production of the RBC autoantibodies and destruction of antibody-coated cells. Rituximab (anti-CD20) is recommended as second-line therapy for patients who fail to respond to steroids or relapse after the dose is reduced. The conventional dosage is 375 mg/m^2 administered as four weekly cycles, but lower doses seem to be equally effective. Response rates of 70% have been reported with a median time to response of 3 to 6 weeks. Splenectomy is effective for patients failing to respond to steroids and rituximab, but many patients prefer to avoid it because of the lack of a guaranteed successful outcome, potential complications, and the need to take antibiotic prophylaxis. Blood transfusion may be necessary if there is severe anaemia, although compatibility testing is complicated by the presence of RBC autoantibodies.

'Cold' autoimmune haemolytic anaemias

Low titres of IgM cold agglutinins reacting at 4°C are normally present in plasma and are harmless. At low temperatures, these antibodies can attach to RBCs and cause their agglutination in the cold peripheries of the body. In addition, activation of complement may cause intravascular haemolysis when the cells return to the higher temperatures in the core of the body. After certain infections, including mycoplasma, cytomegalovirus (CMV), or Epstein–Barr virus (EBV), increased production of *poly*clonal cold agglutinins may occur producing a mild to moderate transient haemolysis.

Chronic cold haemagglutinin disease

Chronic cold haemagglutinin disease usually occurs in older adults, with a gradual onset of haemolytic anaemia owing to the production of *mono*clonal IgM cold agglutinins. After exposure to cold, the

patient develops an acrocyanosis similar to Raynaud's syndrome as a result of RBC autoagglutination.

Investigations

- ***RBCs*** agglutinate in the cold or at room temperature. Agglutination is sometimes seen in the sample tube after cooling but is more easily observed on a peripheral blood film made at room temperature, which is less than body temperature. The agglutination is reversible after warming the sample. Agglutination may cause a spurious increase in MCV.
- ***Cold agglutinin test:*** The titre is elevated to more than 1:512.
- The ***DAT*** result is positive with complement (C3d) alone.
- ***Monoclonal IgM antibodies*** have specificity for the Ii blood group system, usually for the I antigen but occasionally for the i antigen. Developing fetuses and newborns express i antigen until around 13 to 20 months after birth, when I antigen starts to be expressed instead.

Management

Any underlying cause should be treated, if possible. Patients should avoid exposure to cold. Steroids, alkylating agents (chemotherapy), and splenectomy are usually ineffective. Rituximab results in responses in 45% to 60%, but complete responses are unusual. Complement inhibition with sutimlimab, a humanized monoclonal antibody selectively targeting the C1s protein, has been reported to produce rapid responses halting haemolysis with increased Hb concentrations and reduction in fatigue. Blood transfusion may be necessary, and if so, the patient should be in a warm environment; compatibility testing may be difficult because of the cold agglutinin.

Paroxysmal cold haemoglobinuria

Paroxysmal cold haemoglobinuria is a rare condition associated with common childhood infections such as measles, mumps, and chickenpox. Intravascular haemolysis is associated with polyclonal IgG complement-fixing antibodies. These antibodies are biphasic, reacting with RBCs in the cold in the peripheral circulation, with lysis occurring due to complement activation when the cells return to the central circulation. The antibodies have specificity for the P RBC antigen. The lytic reaction is demonstrated *in vitro* by incubating the patient's RBCs and serum at 4°C and then warming the mixture to 37°C (Donath–Landsteiner test). Haemolysis is self-limiting, but RBC transfusions may be necessary. Even though the antibody has specificity for P, transfusion with P-negative RBCs is not necessary.

Drug-induced immune haemolytic anaemia

Drug-induced immune haemolytic anaemias are rare, although more than 100 drugs have been reported to cause immune haemolytic anaemia. Testing for drug-dependent RBC antibodies is not routinely available, and rechallenge with the drug to confirm the diagnosis of drug-induced immune haemolytic anaemia is not advisable. Confirmation of the diagnosis requires:

- A temporal association between administration of a drug and haemolytic anaemia
- A positive DAT result
- Recovery after withdrawal of the drug.

Further reading

Berentsen S, Barcellini W. Autoimmune hemolytic anemias. *NEJM* 2021; 385:1407–1419.
Hill QA, Stamps R, Massey E et al on behalf of the British Society for Haematology. The diagnosis and management of primary autoimmune haemolytic anaemia. *Br J Haem* 2017; 176:395–411.
Hill QA, Stamps R, Massey E et al on behalf of the British Society for Haematology. Guidelines on the management of drug-induced immune and secondary autoimmune haemolytic anaemia. *Br J Haem* 2017; 177:208–220.

Alloimmune haemolytic anaemia

Antibodies produced in one individual react with the RBCs of another. This situation occurs in haemolytic disease of the newborn (HDN); in haemolytic transfusion reactions (see 'Red blood cell complications: haemolytic transfusion reaction' sections); and after allogeneic bone marrow, kidney, liver, cardiac or intestinal transplantation when donor lymphocytes transferred in the allograft ('passenger lymphocytes') produce RBC antibodies against the recipient and cause haemolytic anaemia.

Haemolytic disease of the newborn

HDN is caused by fetomaternal incompatibility for RBC antigens. Maternal alloantibodies against fetal RBC antigens pass from the maternal circulation via the placenta into the fetus, where they destroy fetal RBCs. Only IgG antibodies are capable of transplacental passage from the mother to the fetus. The most common type of HDN is that caused by ABO incompatibility, in which the mother is usually group O and the fetus group A.

HDN caused by ABO incompatibility is usually mild, and exchange transfusion is rarely needed. HDN caused by RhD incompatibility has become much less common in high-income countries since the introduction of anti-D prophylaxis (see later). HDN may be caused by antibodies against antigens in many blood group systems (e.g., other Rh antigens such as c and E, Kell, Duffy, and Kidd).

Sensitization occurs as a result of passage of fetal RBCs into the maternal circulation (which most readily occurs at the time of delivery), so that first pregnancies are rarely affected. However, sensitization may occur at other times (e.g., after a miscarriage, ectopic pregnancy, or blood transfusion) or be caused by episodes during pregnancy that cause transplacental bleeding, such as amniocentesis, chorionic villus sampling, and threatened miscarriage.

Clinical features

These vary from a mild haemolytic anaemia of the newborn to intrauterine death with the characteristic appearance of hydrops fetalis (hepatosplenomegaly, oedema, and cardiac failure). Kernicterus occurs because of severe jaundice in the neonatal period, when the unconjugated (lipid-soluble) bilirubin exceeds 250 μmol/L and bile pigment deposition occurs in the basal ganglia. This can result in permanent brain damage, choreoathetosis, and spasticity. In mild cases, it may present as deafness.

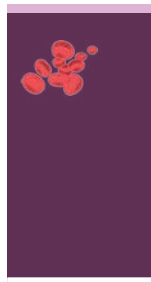
Investigations

All mothers should have their ABO and RhD groups determined and tested for atypical antibodies at antenatal booking and at 28 weeks' gestation. If an antibody is detected, blood group specificity should be determined, and the mother should be retested at least monthly. A rising antibody titre of IgG antibodies or a history of HDN in a previous pregnancy is an indication for referral to a fetal medicine unit.

Antenatal assessment and treatment

If a clinically significant antibody capable of causing HDN, such as anti-D, anti-c, or anti-K, is detected, the father's genotype provides useful information for predicting the likelihood that the fetus carries the relevant RBC antigen. If the father is heterozygous, the genotype of the fetus can be determined from fetal DNA from maternal plasma.

The severity of anaemia is assessed by Doppler flow velocity of the fetal middle cerebral artery; measurement of bile pigments in the amniotic fluid is no longer routine. If the infant appears to have severe anaemia on non-invasive monitoring, ultrasound-guided fetal blood sampling is used to confirm this directly. If necessary, an intrauterine fetal transfusion of RBCs is given.



Birth of an affected infant

A sample of cord blood is obtained at birth. This shows:

- Anaemia with a high reticulocyte count
- Aa positive DAT result
- A raised serum bilirubin

Postnatal management

In mild cases, phototherapy may be used to convert bilirubin to water-soluble biliverdin, which can be excreted by the kidneys. Bilirubin concentrations need to be monitored, and phototherapy may be needed for several days. In more severely affected cases, exchange transfusion may be necessary to replace the infant's antibody-coated RBCs with normal antigen-negative RBCs and to remove bilirubin. Indications for exchange transfusion include severe anaemia or a high or rapidly rising bilirubin concentration.

The blood used for exchange transfusions should:

- Be ABO-compatible with the mother and infant.
- Lack the antigen against which the maternal antibody is directed.
- Be fresh (no more than 5 days from the day of collection).
- Be irradiated to prevent transfusion-associated graft-versus-host disease (TA-GvHD) and be CMV seronegative to prevent transmission of CMV.

Avoidance of exchange transfusion is beneficial because of its risks. Antenatal administration of intravenous immunoglobulin (IVIG) to the mother significantly reduces the need for exchange transfusion.

Prevention of RhD immunization in the mother

Anti-D should be given after delivery when all of the following conditions are fulfilled:

- The mother is RhD negative.
- The fetus is RhD positive.
- There is no maternal anti-D detectable in the mother's serum (i.e., the mother is not already immunized).

The dose is 500 IU of IgG anti-D intramuscularly within 72 hours of delivery. The ***Kleihauer test*** assesses the number of fetal cells in the maternal circulation. A blood film prepared from maternal blood is treated with acid, which elutes HbA. HbF is resistant to this treatment and can be seen when the film is stained with eosin. If large numbers of fetal RBCs are present in the maternal circulation, a higher or additional dose of anti-D is necessary.

Prophylaxis may be required by RhD-negative females at times other than delivery if sensitization may occur, including after an ectopic pregnancy, threatened miscarriage, or termination. The doses of anti-D are 250 IU before 20 weeks' and 500 IU after 20 weeks' gestation. A Kleihauer test should be carried out after 20 weeks' gestation to determine whether more anti-D is required.

Of previously non-immunized RhD-negative females carrying RhD-positive fetuses, 1% to 2% became immunized by the time of delivery. Routine antenatal prophylaxis with anti-D reduces the incidence of immunization during pregnancy and is standard practice in the UK. It is given as two doses of anti-D immunoglobulin of either 500 IU or 1500 IU (one at 28 weeks' and one at 34 weeks' gestation) or as a single dose of 1500 IU either at 28 weeks' or between 28 and 30 weeks' gestation. In the UK in 2016, the National Institute for Care and Health Excellence recommended fetal D typing using maternal plasma to avoid the administration of anti-D to females carrying RhD-negative fetuses.

Further reading

British Committee for Standards in Haematology (BCSH). Guideline for the use of anti-D immunoglobulin for the prevention of haemolytic disease of the fetus and newborn. *Transfus Med* 2014; 24:8–20.
White J, Qureshi H, Massey E et al on behalf of the British Committee for Standards in Haematology. Guideline for blood grouping and red cell antibody testing in pregnancy. *Transfus Med* 2014; 26:246–263.

Non-immune haemolytic anaemia

Paroxysmal nocturnal haemoglobinuria

PNH is a rare form of haemolytic anaemia that results from the clonal expression of haemopoietic stem cells that have mutations in the X-linked gene *PIG-A*. These mutations cause impaired synthesis of glycosylphosphatidylinositol, which anchors many proteins, such as decay accelerating factor (CD55), to the cell surface, and membrane inhibitor of reactive lysis (CD59) to cell membranes. CD55, CD59, and other proteins are involved in complement degradation, and in their absence, the haemolytic action of complement continues.

Clinical features

The major clinical signs are intravascular haemolysis, venous thrombosis, and haemoglobinuria. Haemolysis may be precipitated by infection, iron therapy, or surgery. Characteristically, only the urine voided at night and on waking in the morning is dark, though the reason for this phenomenon is not clear. In severe cases, all urine samples are dark. Urinary iron loss may be sufficient to cause iron deficiency. Some patients present insidiously with signs of anaemia and recurrent abdominal pain.

Venous thrombotic episodes may occur at atypical sites and severe thromboses may be a feature, such as in hepatic (Budd–Chiari syndrome), mesenteric, or cerebral veins. The cause of the increased predisposition to thrombosis is not known but may be:

- Complement-mediated activation of platelets deficient in CD55 and CD59
- Haemolysis releases Hb in the plasma and lowers plasma nitric oxide.

Investigations

- ***Intravascular haemolysis*** is evident (see 'Intravascular haemolysis' section).
- ***Flow cytometric analysis*** of RBCs with anti-CD55 and anti-CD59.
- ***Bone marrow*** is sometimes hypoplastic (or even aplastic) despite haemolysis.

Management and prognosis

PNH is a chronic disorder requiring supportive measures such as blood transfusions, which are necessary for patients with severe anaemia. However, treatment with ***eculizumab*** has revolutionized therapy. The drug is administered intravenously every 7 days for the first 5 weeks and then fortnightly. It is a recombinant humanized monoclonal antibody that prevents cleavage of C5 and therefore formation of the membrane attack complex. It reduces intravascular haemolysis, haemoglobinuria and the need for transfusion, and it improves quality of life. Eculizumab also lowers risk of thrombosis, which was the leading cause of death in patients with PNH. The most serious risk of terminal complement blockade is infection with *Neisseria meningitidis*; thus, vaccination against *N. meningitidis* is recommended 2 weeks before commencing treatment.

PNH may transform into aplastic anaemia or acute leukaemia but may remain stable for many years; the PNH clone may even disappear. With appropriate treatment, the median survival time of patients with PNH is now similar to that of age-matched control participants.

Further reading

Brodsky RA. How I treat paroxysmal nocturnal haemoglobinuria. *Blood* 2021; 137:21304–21309.

Mechanical haemolytic anaemia

RBCs may be injured by physical trauma in the circulation. Direct injury may cause immediate cell lysis or be followed by resealing of the cell membrane with the formation of distorted RBCs or 'fragments'. These cells may circulate for a short period before being destroyed prematurely in the reticuloendothelial system. Causes include:

- Damaged artificial heart valves
- March haemoglobinuria: damage to RBCs in the feet, associated with prolonged marching or running
- MAHA: fragmentation of RBCs occurs in an abnormal microcirculation (e.g., malignant hypertension, eclampsia, haemolytic uraemic syndrome, TTP, vasculitis, disseminated intravascular coagulation [DIC]).

POLYCYTHAEMIA

Polycythaemia (or erythrocytosis) is defined as an increase in Hb or haematocrit (Hct). Hct is a more reliable indicator of polycythaemia than Hb, which may be disproportionately low in those with iron deficiency. Relative erythrocytosis occurs when the RBC volume is normal but there is a decrease in the plasma volume.

Polycythaemia can be classified as primary or secondary. Primary polycythaemia (polycythaemia vera) is a neoplastic process. Secondary polycythaemia can be congenital or acquired. The latter is caused by either an appropriate increase in EPO in response to hypoxia or inappropriate ectopic EPO production by tumours, such as a renal carcinoma. The causes of polycythaemia are given in **Box 38.16**.

Secondary polycythaemias

Serum EPO concentrations are normal or raised in secondary polycythaemia.

Many high-oxygen affinity Hb mutants have been described that lead to increased oxygen affinity but decreased oxygen delivery to the tissues, resulting in compensatory polycythaemia. Chuvash polycythaemia is an autosomal recessive disorder caused by a defect in the oxygen-sensing EPO production pathway caused by a mutation of the von Hippel–Lindau *(VHL)* gene, resulting in increased production of EPO.

Rarely, the discovery of a high EPO concentration may be a clue to the presence of an EPO-secreting tumour.

Management

Treatment is of the precipitating factor (e.g., resection of renal or posterior fossa tumours). The most common cause is heavy smoking, which can produce as much as 10% carboxyhaemoglobin, leading to polycythaemia because of a reduction in blood oxygen-carrying capacity. Heavy smokers may also have respiratory disease.

Complications of secondary polycythaemia are distinct to those seen in polycythaemia vera. Although hyperviscosity symptoms and thromboembolic episodes can occur, management depends on the underlying cause. Complications from myeloproliferative disease, such as progression to myelofibrosis or acute leukaemia, do not develop. Venesection may be symptomatically helpful in certain patients with secondary polycythaemia, particularly if the Hct is above 0.54.

? Box 38.16 Causes of polycythaemia

Primary
- Polycythaemia vera

Congenital
- Mutations in erythropoietin receptor
- High-oxygen-affinity haemoglobins
- Mutations in hypoxia-sensing pathways (e.g., Chuvash polycythaemia mutation in von Hippel–Lindau gene)

Secondary

Increase in erythropoietin caused by hypoxia
- High altitude
- Chronic lung disease
- Cardiovascular disease (right-to-left shunt)
- Sleep apnoea
- Morbid obesity
- Heavy smoking

Caused by an inappropriate increase in erythropoietin
- Renal disease: renal cell carcinoma, Wilms' tumour
- Hepatocellular carcinoma
- Adrenal tumour
- Cerebellar haemangioblastoma
- Massive uterine leiomyoma
- Over-administration of erythropoietin
- Treatment with androgen preparations

Relative
- Stress or spurious polycythaemia
- Dehydration
- Burns

'Relative' or 'apparent' polycythaemia (Gaisböck's syndrome)

The RBC volume is normal, and there is relative polycythaemia because of decreased plasma volume. 'Relative' polycythaemia is more common than polycythaemia vera and typically occurs in middle-aged males, particularly in association with smoking, high alcohol intake, obesity, and high blood pressure. The condition may present with cardiovascular disease, including myocardial or cerebral ischaemia. For this reason, it may be justifiable to venesect selected patients. Management requires modification of lifestyle factors, including smoking and alcohol.

THE SPLEEN

The spleen is the largest lymphoid organ in the body and is situated in the left hypochondrium. There are two anatomical components:

- The red pulp, consisting of sinuses lined by endothelial macrophages and cords (spaces)
- The white pulp, which has a structure like that of lymphoid follicles.

Blood enters via the splenic artery and is delivered to the red and white pulp. During the flow, the blood is 'skimmed', with leucocytes and plasma preferentially passing to white pulp. Some RBCs pass rapidly through into the venous system, and others are held up in the red pulp.

Function

Sequestration and phagocytosis

Normal RBCs, which are flexible, pass through the red pulp into the venous system without difficulty. Old or abnormal cells are damaged by the hypoxia, low glucose, and low pH in the sinuses of the red pulp and are therefore removed by

phagocytosis along with other circulating foreign matter. Howell–Jolly and Heinz bodies and sideroblastic granules have their particles removed by 'pitting' and are then returned to the circulation. IgG-coated RBCs are removed through their Fc receptors by macrophages.

Extramedullary haemopoiesis

Pluripotential stem cells are present in the spleen and proliferate during severe erythroid stress, such as in haemolytic anaemia, thalassaemia major, and myelofibrosis. This occurs because of the presence of anaemia despite every available marrow cavity already being fully engaged in haemopoiesis. It is usually a feature of ineffective erythropoiesis.

Immunological function

About 25% of the body's T lymphocytes and 15% of B lymphocytes are present in the spleen. The spleen shares the function of production of antibodies with other lymphoid tissues.

Blood pooling

Up to one-third of the platelets are sequestrated in the spleen and can be rapidly mobilized. Enlarged spleens pool a significant percentage (up to 40%) of the RBC mass.

Splenomegaly

The spleen enlarges from under the left costal margin inferiorly and medially. It is dull to percussion, and it may be possible to palpate a notch along its leading edge, which further differentiates the mass from an enlarged left kidney. The spleen typically measures 11 cm in its longest dimension. An enlarged spleen is only palpable if it is 1.5 to 2 times the size of a normal spleen. Ultrasound or CT scanning may be used to confirm splenomegaly and may also provide other useful information, such as the presence of abdominal lymphadenopathy.

Aetiology

A clinically palpable spleen can have many causes:

- Infection:
  - Acute (e.g., septic shock, infective endocarditis, typhoid, infectious mononucleosis)
  - Chronic (e.g., TB, brucellosis)
  - Parasitic (e.g., malaria, kala-azar, schistosomiasis)
- ***Inflammation:*** rheumatoid arthritis, sarcoidosis, SLE
- ***Haematological factors:*** haemolytic anaemia, haemoglobinopathies, leukaemia, lymphoma, myeloproliferative disorder
- ***Portal hypertension:*** liver disease
- ***Miscellaneous:*** storage diseases (e.g., Gaucher's disease), amyloidosis, primary and secondary neoplasias, tropical splenomegaly.

Massive splenomegaly is seen in myelofibrosis, chronic myeloid leukaemia, chronic malaria, kala-azar or, rarely, Gaucher's disease.

Hypersplenism

This can result from splenomegaly of any cause. It is commonly seen with splenomegaly from haematological disorders, portal hypertension, rheumatoid arthritis (Felty's syndrome), or lymphoma. Hypersplenism produces:

- Pancytopenia
- Haemolysis caused by sequestration and destruction of RBCs in the spleen
- Increased plasma volume.

Box 38.17 Prophylaxis against infection after splenectomy or in splenic dysfunction

Vaccinate 2–3 weeks before elective splenectomy.

- 23-Valent unconjugated pneumococcal polysaccharide vaccine repeated every 5 years
- Meningococcal polysaccharide vaccine (ACWY)
- Meningococcus group B vaccination
- Annual influenza vaccine

Antibiotic prophylaxis

- Long-term penicillin V 500 mg 12-hourly (if allergic, use erythromycin)

Management

This is often dependent on the underlying cause; splenectomy is sometimes required for severe anaemia or thrombocytopenia. Splenectomy is performed mainly for:

- Trauma
- ITP (see 'Immune thrombocytopenic purpura' section)
- Haemolytic anaemias (see 'Haemolytic anaemias' section)
- Hypersplenism

Problems after splenectomy

An increased platelet count (usually 600–1000 × 10^9/L) may occur for 2 to 3 weeks with risk of thromboembolic phenomena. In the longer term, there is an increased risk of overwhelming infections, particularly pneumococcus.

Prophylaxis against infection after splenectomy or splenic dysfunction

See **Box 38.17** for prophylactic measures. All patients should be educated about the risk of infection and the importance of early recognition and treatment. They should be given an information leaflet and should carry a card, bracelet, or electronic health alert.

Post-splenectomy haematological features

- ***Thrombocytosis*** persists in about 30%.
- The ***WCC*** is usually normal, but there may be a mild lymphocytosis and monocytosis.
- ***Abnormalities in RBC morphology*** are most prominent and include Howell–Jolly bodies (basophilic nuclear remnants), Pappenheimer bodies (sideroblastic granules), target cells, and irregular contracted RBCs (**Fig. 38.19**). Pitted RBCs can be counted.

Splenic atrophy

This is seen in sickle cell disease because of infarction. Hyposplenism is also seen in a wide range of non-haematological diseases, including coeliac disease, dermatitis herpetiformis, and occasionally ulcerative colitis. Post-splenectomy haematological features are seen.

Further reading

Lenti MV, Luu S, Carsetti R et al. Asplenia and spleen hypofunction. *Nat Rev Dis Primers* 2022; 8(1):71.

BLOOD TRANSFUSION

Cells and proteins in the blood express antigens that are controlled by polymorphic genes; a specific antigen may be present in some individuals but not in others. A blood transfusion may

immunize the recipient against donor antigens that the recipient lacks (alloimmunization), and repeated transfusions increase the risk. Similarly, the transplacental passage of fetal blood cells during pregnancy may alloimmunize the mother against fetal antigens inherited from the father. Antibodies stimulated by blood transfusion or pregnancy, such as Rh antibodies, are termed ***immune antibodies***. They are usually IgG, in contrast to ***naturally occurring antibodies*** such as ABO antibodies, which are made in response to environmental antigens present in food and bacteria and are usually IgM.

Blood groups

Blood group antigens are proteins, GPs, or carbohydrates attached to GPs on the surface of RBCs; more than 300 blood groups are recognized. The ABO and Rh systems are the two major blood groups, but incompatibilities involving many other blood groups (e.g., Kell, Duffy, Kidd) may also cause haemolytic transfusion reactions or HDN.

ABO system

This is the most important blood group system in transfusion practice because naturally occurring IgM anti-A and anti-B antibodies can produce rapid and severe intravascular haemolysis of incompatible RBCs.

The ABO system is under the control of three allelic genes, *A*, *B*, and *O*, producing the genotypes and phenotypes shown in **Box 38.18**.

The precursor for A and B antigens is the H antigen, which codes for an enzyme that attaches fucose to galactose on an oligosaccharide structure on the RBC membrane. The *A* and *B* genes control specific enzymes responsible for the addition to H of *N*-acetylgalactosamine for group A and D-galactose for group B (**Fig. 38.31**). The *O* gene is amorphic and does not transform H, so O is not antigenic. The A, B, and H antigens are present in most body cells. They are also found in soluble form in tissue fluids (e.g., saliva, gastric juice) in the 80% of the population who possess secretor genes.

Rh system

Development of IgG RhD antibodies after exposure of RhD-negative individuals to RhD-positive RBCs occurs frequently, causing HDN and haemolytic transfusion reactions.

This system is coded by allelic genes: *C* and *c*; *E* and *e*l and *D* and no *D*, signified as *d*. They are inherited as triplets on each chromosome 1, one from each pair of genes (e.g., *CDE*/*cde*). RhD-negative individuals have no D protein in the RBC membrane, which explains why it is so immunogenic when RhD-negative individuals are exposed to RhD antigen through transfusion or pregnancy. In those of White European ethnicity, the RhD-negative phenotype almost always results from a complete deletion of the *RhD* gene; in those of Black African ethnicity, it can also result from an inactive gene containing stop codons in the reading frame.

Procedure for blood transfusion in hospitals

The safety of blood transfusion in hospitals depends on meticulous attention to detail at each stage before and during transfusion. Avoidance of patient and blood sample identification errors at the time of sample collection for compatibility testing and at the time of transfusion would avoid most serious haemolytic transfusion reactions, almost all of which involve the ABO system.

Pre-transfusion compatibility testing

Blood grouping

The ABO and RhD groups of the patient are determined.

Antibody screening

The patient's serum or plasma is screened for atypical antibodies that may cause a significant reduction in survival of transfused RBCs. The patient's serum or plasma is tested against RBCs from at least two group O donors expressing a wide range of RBC antigens, for detection of IgM RBC alloantibodies (using a direct agglutination test of cells suspended in saline) and IgG antibodies (using an indirect antiglobulin test; **Fig. 38.29**). About 10% of patients have a positive antibody screening result. If this occurs, further testing using a comprehensive panel of typed RBCs to determine the blood group specificity of the antibody is performed. Clinically significant RBC antibodies are detected in about 20% of patients with positive antibody screen results.

Transfusion laboratory testing

Several systems for blood grouping, antibody screening, and cross-matching are available to hospital transfusion laboratories. Not all depend on the traditional agglutination of RBCs in suspension. Others work by differential passage of agglutinated and unagglutinated RBCs through a column of dextran gel matrix (e.g., DiaMed and Ortho Biovue systems) or on the capture of antibodies by RBCs immobilized on the surface of a microplate well (e.g., Capture-R solid phase system).

Selection of donor blood and compatibility testing

Selection procedures

Donor blood of the same ABO and RhD group as the patient is selected. Matching for additional blood groups is carried out for patients with clinically significant RBC antibodies (see later) and for patients who are likely to be multi-transfused and at high risk of developing antibodies (e.g., in sickle cell disease). Many hospitals routinely provide Kell-negative blood for females of childbearing age to minimize the risk of alloimmunization caused by anti-K and subsequent HDN.

Compatibility testing procedures

Timing of sample collection

The timing of sample collection for compatibility testing must take account of pregnancy or recent transfusion because both may

Box 38.18 The ABO system: antigens and antibodies

Phenotype	Genotype	Antigens	Antibodies	Frequency in the UK (%)
O	OO	None	Anti-A and anti-B	43
A	AA or AO	A	Anti-B	45
B	BB or BO	B	Anti-A	9
AB	AB	A and B	None	3

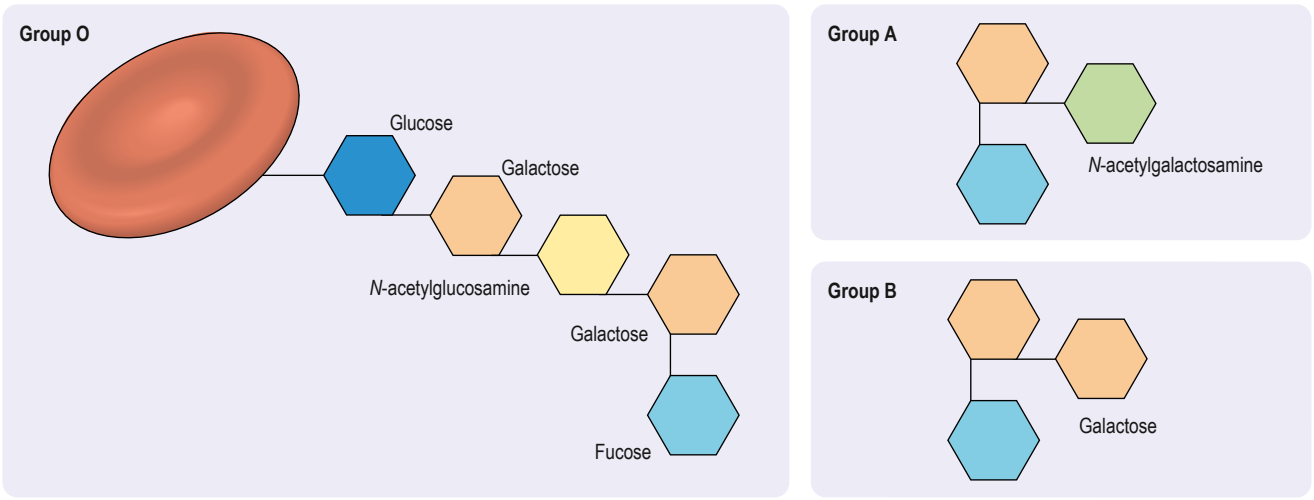


Fig. 38.31 Sugar chains in the ABO blood group system. See text ('ABO system' section).

stimulate antibodies that could cause a haemolytic transfusion reaction. Blood samples should be collected no more than 3 days before transfusion if the patient has been transfused or pregnant within the preceding 3 months.

Patients without atypical red blood cell antibodies

A full crossmatch involves testing the patient's serum or plasma against donor RBCs suspended in saline in a direct agglutination test and using an indirect antiglobulin test. In many hospitals, this serological crossmatch is now omitted because a negative antibody screen makes it highly unlikely that there will be any incompatibility with the donor. Error in collection or labelling of the patient sample or a mix-up of samples in the laboratory are much greater risks. When the antibody screen result is negative, laboratories can use their information system to check the patient records and authorize the release of ABO-compatible donor units ***(computer or electronic crossmatching)***.

Electronic issue can be extended to fridges remote from the main laboratory ***('remote blood issue')*** provided they are electronically linked to the blood transfusion laboratory's information system. The printing of compatibility labels for the blood and its collection are under electronic control, applying the same rules as electronic issue from the main laboratory. Issue of blood using this process reduces the time taken to provide blood for patients, particularly at hospitals without a blood transfusion laboratory, because transport of blood from the central laboratory is not required. There are good examples of centralized transfusion services in cities such as Oxford in the UK and Pittsburgh and Seattle in the USA, and this model is likely to be increasingly employed in the future.

Patients with atypical red blood cell antibodies

Donor blood should be selected that lacks the relevant RBC antigen(s) and is the same ABO and RhD group as the patient. A full crossmatch should always be carried out in case new antibodies have developed.

Blood ordering

Hospitals are increasingly relying on electronic systems for blood ordering; some include algorithms for appropriate use of blood, with 'alerts' if there is non-compliance with agreed blood count triggers for transfusion. Electronic systems also provide data on comparative use of blood for clinical teams and the documentation of consent to transfusion.

Elective surgery

Many hospitals have guidelines for ordering blood for elective surgery (maximum surgical blood ordering schedules). These are aimed at reducing unnecessary crossmatching, which increases the risk of blood eventually becoming outdated and wasted. Many operations in which blood is required only occasionally for unexpected high blood loss can be classified as ***group and save***; this means that when the antibody screen is negative, blood is not reserved but can be made available quickly if necessary—that is, in a few minutes—using the electronic crossmatch procedure described. However, if a patient has atypical antibodies and is undergoing any procedure or surgery that may result in the need for transfusion, compatible blood should be available; this may take several days if the patient has multiple or unusual antibodies.

Emergencies

There may be insufficient time for full pre-transfusion testing. Options include:

- Blood required immediately: 2 units of O RhD-negative blood ***(emergency stock)*** is used to allow additional time for the laboratory to group the patient.
- Blood required in 10 to 15 minutes: Blood of the same ABO and RhD groups as the patient is used ***('group compatible blood')***.
- Blood required in 45 minutes: Most laboratories are able to provide fully crossmatched blood within this time.

Complications of blood transfusion

Reporting schemes under the term ***haemovigilance*** have been established in many countries, including the Serious Hazards of Transfusion (SHOT) scheme in the UK, which produced its first report in 1997. SHOT reported 35 deaths from blood transfusion in its 2022 report of which 18 (51%) were caused by transfusion reaction and 17

Box 38.19 Complications of blood transfusion

Immunological

Alloimmunization and incompatibility

- Red blood cells
  - Immediate haemolytic transfusion reactions
  - Delayed haemolytic transfusion reactions
- Leucocytes and platelets
  - Non-haemolytic (febrile) transfusion reactions
  - TRALI
  - Poor survival of transfused platelets and granulocytes
  - Post-transfusion purpura
  - TA-GvHD
- Plasma proteins
  - Urticarial and anaphylactic reactions

Non-immunological

Transmission of infection

- Viruses
  - Hepatitis: A, B, C, E
  - HIV
  - CMV, HTLV-1, West Nile virus, Dengue, Zika
- Parasites
  - Malaria
  - Trypanosomiasis
  - Babesia
- Bacteria
- Prions

Transfusion-associated circulatory overload
Iron overload caused by multiple transfusions

Bleeding and electrolyte changes: may be caused by massive transfusion of stored blood

CMV, Cytomegalovirus; *HIV,* human immunodeficiency virus; *HTLV-1,* human T-cell leukaemia (or lymphotropic) virus 1; *TA-GvHD,* transfusion-associated graft-versus-host disease; *TRALI,* transfusion-related acute lung injury.

(49%) were caused by error. Transfusion-associated circulatory overload occurred in 8 (23%) patients (Fig. 38.32). Data from the SHOT reports indicating that ***'incorrect blood component transfused'*** is the most frequent type of serious incident with pulmonary complications the most frequent cause of morbidity and mortality. Errors in clinical areas, including collection of the blood sample for compatibility testing, collection of blood from the fridge, or the administration of blood, are more frequent than laboratory errors (Box 38.19).

Prevention of wrong blood transfusions

The potentially serious consequences of errors in transfusion emphasize the need for meticulous procedures. Good training of staff is essential, and regular reviews of practice to ensure compliance with standard procedures are required. New approaches include the use of barcode patient identification and new technology at the bedside. Handheld devices can prompt staff through the key steps and check that the barcode on the patient's wristband matches the barcode on the unit of blood (Box 38.20).

Immunological complications

Alloimmunization

Blood transfusion carries a risk of alloimmunization to the many 'foreign' antigens present on RBCs, leucocytes, platelets, and plasma proteins. Alloimmunization also occurs during pregnancy to fetal antigens inherited from the father and not shared by the mother.

Alloimmunization does not usually cause clinical problems (e.g., haemolytic transfusion reaction) with the first transfusion, but they may occur with subsequent transfusions. There may also be delayed consequences of alloimmunization, such as HDN and rejection of tissue or organ transplants.

Incompatibility

Transfusion of blood cells such as RBCs and platelets into a recipient who has antibodies against antigens on the cells may result in poor survival of the transfused cells and in harmful effects of the antigen–antibody reaction.

Red blood cell complications: haemolytic transfusion reaction (immediate)

This is the most serious and life-threatening complication of blood transfusion and is usually caused by ABO incompatibility. There is complement activation by the antigen–antibody reaction, usually caused by IgM anti-A or anti-B antibodies. Initial symptoms and signs (rigors, lumbar pain, dyspnoea, hypotension, and haemoglobinuria) begin a few minutes after starting the transfusion. Kidney failure may develop. Activation of coagulation occurs and bleeding caused by DIC is a bad prognostic sign. Emergency treatment for patients with shock is needed to maintain blood pressure and kidney function.

Diagnosis

This is confirmed by finding evidence of intravascular haemolysis (e.g., haemoglobinuria) and incompatibility between the donor and recipient. All documentation should be checked for errors such as:

- Failure to check patient identity when taking the sample for compatibility testing (i.e., sample from the wrong patient)
- Mislabelling of the blood sample with the wrong patient's name
- Labelling or handling errors in the laboratory
- Errors in the collection of blood, leading to delivery of the wrong blood to the ward or theatre
- Failure to perform proper identity checks before the blood is transfused (i.e., blood transfused to the wrong patient).

Investigations

To confirm where the error occurred, blood grouping should be carried out on:

- The patient's original sample (used for the compatibility testing)
- A new sample taken from the patient after the reaction
- The donor units.

At the first suspicion of any serious transfusion reaction, the transfusion should always be stopped and the donor units returned to the blood transfusion laboratory with a new blood sample from the patient to exclude a haemolytic transfusion reaction.

Red blood cell complications: haemolytic transfusion reaction (delayed)

This occurs in patients alloimmunized by previous transfusions or pregnancies. The antibody concentration is too low to be detected by pre-transfusion compatibility testing, but a secondary immune response is seen after the transfusion, resulting in destruction of the transfused cells, usually by IgG antibodies.

Haemolysis is usually extravascular because the antibodies are IgG. The patient may develop anaemia and jaundice about 1 week after the transfusion, although most episodes are clinically silent. The blood film shows spherocytosis and reticulocytosis. The DAT result is positive and detection of the antibody is usually straightforward.

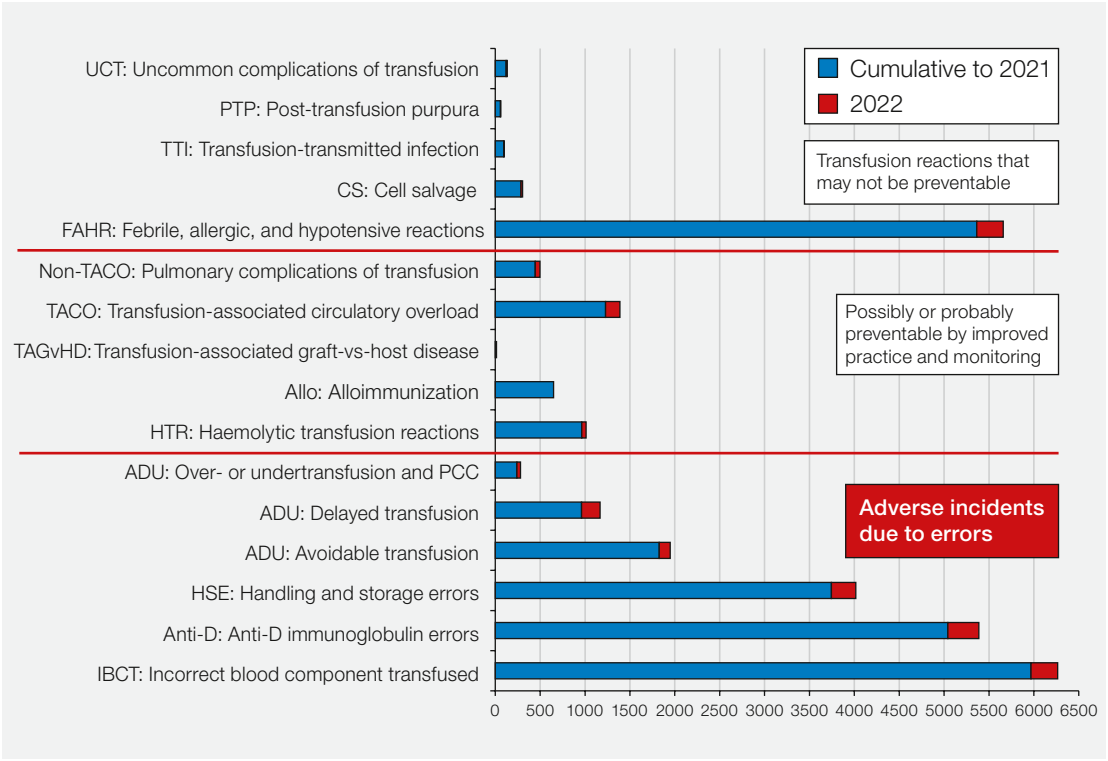


Fig. 38.32 Cumulative data for Serious Hazards of Transfusion categories 1996 to 2022 (*n* = 28877). The adverse events are divided into transfusion reactions that may not be preventable, those that are probably or possibly preventable, and those caused by mistakes. Those reported in 2022 are indicated in red at the *right-hand side* of each *bar*. *PCC,* Prothrombin complex concentrates. *(From Narayan S (ed), Poles D et al. on behalf of the Serious Hazards of Transfusion (SHOT) Steering Group. The 2022 Annual SHOT Report (2023). https://doi.org/10.57911/WZ85-3885. Resources from SHOT are reviewed and updated regularly so please refer to the SHOT website for up-to-date content.)*

Box 38.20 Bedside checks for blood unit compatibility

- Traditional (historical) pre-transfusion bedside checking
  - Two healthcare professionals
  - Multiple checks of written documentation
- Barcode technology
  - Prompts key steps
  - Scans patient's unique barcode ID
  - Including asking patient to state their full name and date of birth
  - Checks patient's details match those on the unit's compatibility label barcode after pre-transfusion testing
  - The unit's compatibility label barcode contains the unique number of the unit, which is matched with the unit number at the top left of the bag ensuring the blood bank has attached the right compatibility label to the bag

Leucocyte complications: non-haemolytic (febrile) transfusion reactions

The typical cause of febrile reactions after RBC transfusions is the presence of leucocyte antibodies in patients who have previously been transfused or pregnant. Antibodies against donor leucocytes lead to the release of pyrogens. The mechanism of febrile reactions is different after platelet transfusions, in which the release of cytokines from donor leucocytes is the most likely cause. Typical signs are flushing, tachycardia, fever (>38°C), chills, and rigors, which are uncomfortable but not life-threatening. Paracetamol may be used to reduce the fever, although it should not be given to patients with thrombocytopenia. The routine introduction of leucocyte-reduced blood in the UK to minimize the risk of transmission of variant Creutzfeldt–Jakob disease (vCJD) by blood transfusion (see later) has reduced the incidence of febrile reactions in the UK.

Transfusion-related acute lung injury

Potent leucocyte antibodies in the plasma of donors (usually multiparous females) may cause transfusion-related acute lung injury (TRALI) characterized by dyspnoea, fever, cough, and shadowing in the perihilar and lower lung fields on a chest radiograph. Prompt respiratory support is essential, and mechanical ventilation is frequently necessary. TRALI usually resolves within 48 to 96 hours but the mortality can occur. The avoidance of female plasma in the preparation of fresh-frozen plasma (FFP) and pooled platelets, and the testing of female plateletpheresis donors for leucocyte antibodies has been implemented in many high-income countries to reduce the risk of TRALI; the number of reported cases has fallen considerably.

Transfusion-associated graft-versus-host disease

Transfused donor lymphocytes that share HLA haplotype with the patient are not rejected and may recognize the patient as 'foreign' causing an acute GvHD reaction including pancytopenia. It is usually fatal. No cases have occurred in the UK for many years, probably because of the routine leucocyte reduction of blood components since 1999. However, certain groups of immunosuppressed patients at particular risk of TA-GvHD should receive irradiated blood to minimize the danger. Such

patients include those with congenital or acquired immunodeficiencies (e.g., individuals who have had treatment with purine analogue chemotherapy, patients with Hodgkin lymphoma, those who have undergone haemopoietic stem cell transplantation, and fetuses and neonates who have received intrauterine transfusions).

Platelet complications: post-transfusion purpura

A rare transfusion reaction characterized by severe thrombocytopenia, which can be life-threatening.

Plasma protein complications: urticaria and anaphylaxis

Urticarial reactions are often attributed to plasma protein incompatibility, but in most cases, they are unexplained. They are common but rarely severe; stopping or slowing the transfusion and administering chlorphenamine 10 mg intravenously are usually sufficient treatment.

Anaphylactic reactions occasionally occur; severe reactions are seen in patients lacking IgA who produce anti-IgA that reacts with IgA in the transfused blood. The transfusion should be stopped and ***adrenaline (epinephrine)*** 0.5 mg intramuscularly and ***chlorphenamine*** 10 mg intravenously given immediately; endotracheal intubation may be required. Patients who have had severe urticarial or anaphylactic reactions should receive washed RBCs. Blood from IgA-deficient donors is required for patients with IgA deficiency.

Immunosuppression

Transfusions are known to have a negative effect on the survival of later kidney allografts caused by transfusion-induced immunomodulation. The precise mechanism is unclear but may be associated with the transfusion of allogeneic leucocytes. Transfusion-induced immunosuppression, including an increase in postoperative infection and tumour recurrence, has been hypothesized but remains unproven.

Non-immunological complications

Transmission of infection

Although stringent measures are being taken to minimize the risk of transfusion-transmitted infection, it may never be possible to guarantee that donor blood is absolutely 'safe'. The current approach to the safety of blood components and plasma in the UK and other high-income countries is cautious, but it is not an absolute guarantee of safety. Clinicians should always consider the patient's requirement for transfusion carefully and only transfuse if clinically appropriate (see later).

Viral transmission

Donor blood in the UK is currently tested for hepatitis B virus (HBV), hepatitis C virus (HCV), hepatitis E virus (HEV), human immunodeficiency virus (HIV)-1, and human T-cell leukaemia virus 1. CMV-seronegative blood is given to fetuses and neonates and during pregnancy. Blood services continue a vigilant search for new infectious agents ('emerging' infections) that may be transmitted by blood transfusion and for methods to prevent their transmission, including donor screening, testing, and pathogen inactivation. Donor questionnaires record recent travel to exclude the possible risk of West Nile virus, the causal agent of meningoencephalitis, which has been transmitted by transfusion and transplantation in the USA, and Zika virus in light of the 2015 to 2016 epidemic.

The risk of transmission of viral infections by blood transfusion varies from country to country, depending on factors such as the underlying prevalence of transfusion-transmitted infections in the population and measures taken to minimize risk of transmission. Viral transmission via blood transfusion is still a major issue in low-income countries. In the UK, the risks of potentially infectious HBV, HCV, or HIV not being detected are very low at less than 1 per million donations tested. Various measures, including treatment with solvents and detergents, are used for inactivating viruses in plasma products such as coagulation factor concentrates and IVIG.

Bacterial transmission

Bacterial contamination of blood components is rare, but when it occurs, it is a frequent cause of death associated with transfusion. Some organisms, such as *Yersinia enterocolitica*, can proliferate in RBC concentrates stored at 4°C. Platelet concentrates stored at 22°C are a more frequent cause of contamination. Measures to avoid bacterial contamination have been implemented in most high-income countries, including strict donor arm cleansing, diversion of the initial blood collection to samples for testing rather than into the collection bag, and bacterial detection systems for platelet concentrate. Such systems were implemented in the UK in 2011, and there have been no cases of bacterial transfusion-transmitted infections since 2015.

Transfusion-transmitted syphilis is very rare. Spirochaetes do not survive for more than 72 hours in blood stored at 4°C, and each donation is tested using the ***Treponema pallidum haemagglutination assay***.

Variant Creutzfeldt–Jakob disease

In the UK, four transmissions of vCJD have occurred after a blood transfusion; the last one was reported in 2004. Measures taken to minimize risk included universal leucocyte reduction of blood components in 1999 because the prion protein was thought to be primarily associated with lymphocytes. Persons who have received a blood transfusion since 1980 are excluded from blood donation in the UK. A ban on the manufacture of blood products using UK donor plasma was lifted in 2020. Imported plasma mostly from the USA had been used to produce blood products such as albumin and immunoglobulins.

Other

- Transfusion associated heart failure due to volume overload
- Iron overload, e.g., from recurrent transfusions in transfusion dependent conditions.

Strategies for the avoidance of unnecessary transfusion

These include:

- Evidence-based criteria for the use of blood components and blood products including:
  - The threshold ('trigger') for transfusion (e.g., Hb concentration or platelet count)
  - The dose. Single-unit transfusions are often sufficient: 'Don't use two when one will do'.
- Discontinuing anticoagulants and antiplatelet drugs whenever possible preoperatively to reduce bleeding in surgical patients
- Identification and treatment of anaemia before surgery and in pregnancy

- Use of antifibrinolytic drugs (e.g., tranexamic acid in major surgery and trauma)
- Use of intraoperative cell salvage of blood in major surgery.

Artificial Hb and other blood substitutes have not yet been developed for routine clinical use. They generally have a short intravascular half-life, and a recent meta-analysis found a significant risk of mortality and myocardial infarction.

Autologous transfusion

An alternative to using blood from volunteer donors is to use the patient's own blood. There are three types of autologous transfusion:

- ***Predeposit.*** The patient donates 2 to 5 units of blood at approximately weekly intervals before elective surgery. There is limited evidence that it reduces blood requirements. In high-income countries, it is only used in rare cases when it is not possible to identify compatible blood because of multiple antibodies.
- ***Preoperative haemodilution.*** Immediately before surgery, 1 or 2 units of blood is removed from the patient and then retransfused to replace operative losses.
- ***Blood salvage.*** Blood lost during or after surgery is collected and retransfused. Several techniques of varying levels of sophistication are available. The operative site must be free of bacteria and bowel contents. Salvage was thought to be contraindicated in cancer surgery because of the risk of facilitating metastases, but many centres are using it with apparent safety. Intraoperative cell salvage is increasingly being used as a way of avoiding the use of donor blood.

In low-income countries, blood from relatives may be used because of a lack of donor blood.

Blood, blood components, and blood products

Most blood collected from donors is processed as follows:

- ***Blood components***, including RBC and platelet concentrates, FFP, and cryoprecipitate, are prepared from a single donation of blood by simple separation (e.g., centrifugation) and are transfused without further processing. Platelet concentrates are also prepared by plateletpheresis (see later).
- ***Blood products***, including coagulation factor concentrates, albumin, and immunoglobulin solutions, are prepared by complex processes using plasma from many donors as the starting material.

In most circumstances, it is preferable to transfuse only the blood component or product required by the patient ***(component therapy)*** rather than using whole blood. This is the most effective way of using donor blood, which is a scarce resource, and it reduces the risk of complications from transfusion of unnecessary blood components. However, there is emerging interest in using whole blood in trauma and other causes of major haemorrhage.

Patients should be provided with information about blood transfusion and alternatives to transfusion whenever possible. This discussion should be documented. Some patients, such as Jehovah's Witnesses, may refuse transfusion and need specialized management when they undergo surgery or receive medical treatments that usually require blood transfusion.

Whole blood

A unit of whole blood consists of 450 mL ± 10% of blood from a suitable donor plus 63 mL of anticoagulant, which is then leucocyte-reduced by filtration. Blood stored at 4°C is given a 'shelf-life' of 5 weeks in the UK (6 weeks in some other countries), when at least 70% of the transfused RBCs should survive normally. Whole blood is rarely used for transfusion, though there is renewed interest in considering its use in major haemorrhage.

Red blood cell concentrates

Virtually all plasma is removed. It is replaced with about 100 mL of an optimal additive solution, such as SAG-M, which contains sodium chloride, adenine, glucose, and mannitol. The mean volume is about 330 mL. The PCV is about 0.57 L/L, but viscosity is low because there are no plasma proteins in the additive solution; this allows fast administration, if necessary. RBC concentrates are used to increase oxygen-carrying capacity in major haemorrhage and to treat non-bleeding patients with severe anaemia. A threshold Hb of 70 g/L for transfusion in anaemia may be increased (e.g., to 80 g/L) in severe cardiovascular disease. Contemporary algorithms for the management of major haemorrhage highlight the need for the early transfusion of FFP and platelets to support haemostasis.

Washed red blood cell concentrates

These are preparations of RBCs suspended in saline, produced by cell separators that remove all but traces of plasma proteins. They are used in patients who have had severe recurrent urticarial or anaphylactic reactions.

Platelet concentrates

These are prepared from single donors either by centrifugation of whole blood or by plateletpheresis of blood using cell separators. They may be stored for up to 7 days at 22°C in the UK after confirmation of a negative bacterial culture result to minimize the risk of bacterial transmission. They are used to treat bleeding in patients with severe thrombocytopenia and prophylactically to prevent bleeding in patients with bone marrow failure when the platelet count is below 10×10^9/L.

Granulocyte concentrates

These are prepared from whole blood as pooled granulocyte products or from blood from single donors using cell separators. They are used for patients with severe neutropenia when there is evidence of bacterial infection. The numbers of granulocytes collected may be increased by treating donors with G-CSF and steroids. The half-life of granulocytes is of the order of a few hours only, so granulocytes need to be collected and transfused daily, making the practicalities of this treatment challenging. There is limited evidence for their effectiveness, and they are commonly associated with febrile and occasional pulmonary reactions.

Fresh-frozen plasma

FFP is prepared by freezing the plasma from 1 unit of blood to –30°C to maintain the concentration of coagulation factors. The volume is approximately 200 mL. FFP contains all the coagulation factors present in fresh plasma and is used mostly for replacement of coagulation factors in acquired coagulation factor deficiencies, such as massive haemorrhage and DIC. It may be treated by a pathogen-inactivation process to minimize the risk of bacterial and blood-borne viral disease transmission.

Prothrombin complex concentrates

These contain factors II, VII, IX, and X and are the recommended treatment for rapid reversal of warfarin toxicity in bleeding patients with an elevated international normalized ratio (INR). The have replaced FFP in this context because of superior efficacy and lower risks of allergic reaction and volume overload.

Cryoprecipitate

This is obtained by allowing the frozen plasma from a single donation to thaw at 4° to 8°C and removing the supernatant. The typical adult dose is two pools each containing 5 donor units. Cryoprecipitate contains factor VIII:C, von Willebrand factor (VWF), and fibrinogen. It may be useful in patients with DIC and other conditions when the fibrinogen concentration is very low. However, fibrinogen concentrates are now available and are increasingly used in preference to cryoprecipitate.

Specific coagulation factor concentrates

Specific factor concentrates are available for factors VII, VIII, IX, X, XI, XIII, and VWF. These are freeze-dried preparations of the specific factor prepared from large pools of plasma. All products come from plasma sources at low risk of vCJD and are virus-inactivated during manufacture with either heat or chemicals. They are used for treating patients with inherited coagulation factor deficiencies when recombinant coagulation factor concentrates are unavailable.

Albumin

There are two preparations:

- ***Human albumin solution 4.5%*** contains 45 g/L of albumin and 160 mmol/L of sodium. It is available in 50-, 100-, 250-, and 500-mL bottles.
- ***Human albumin solution 20%*** contains approximately 200 g/L of albumin and 130 mmol/L of sodium. It is available in 50- and 100-mL bottles.

Human albumin solutions are inappropriate fluids for acute volume replacement or the treatment of shock. However, they are indicated for treatment of management of liver cirrhosis–related complications and as the replacement fluid for plasma exchange.

Normal immunoglobulin

This is prepared from normal plasma. It is used to treat hypogammaglobulinaemia, prevent infections, and in high doses to treat patients with autoimmune disorders, such as ITP (see 'Immune thrombocytopenic purpura' section).

Specific immunoglobulins

These are obtained from donors with high titres of antibodies. Many preparations are available, such as anti-D, anti-hepatitis B, and anti-varicella zoster.

Further reading

Ainley LI, Hewitt PE. Haematology patients and the risk of transfusion transmitted infection. *Br J Haem* 2018; 180:473–483.
Carson JL, Stanworth SJ, Guyatt G et al. Red blood cell transfusion: 2023 AABB International Guidelines. *JAMA* 2023; 330(19):1892–1902.
Delaney M, Wendel S, Bercowitz RS et al for the Biomedical Excellence for Safer Transfusion (BEST) Collaborative. Transfusion reactions: prevention, diagnosis and treatment. *Lancet* 2016; 388:1945–1950.
Mueller MM, Van Remoortel H, Meybohm P et al. on behalf of the ICC PBM Frankfurt 2018 Group. Patient blood management: recommendations from the 2018 frankfurt consensus conference. *JAMA* 2019; 321(10):983–997.
Murphy MF, Roberts DJ, Yazer MH (eds). *Practical Transfusion Medicine*, 6th edn. Oxford: Wiley–Blackwell; 2022.
Narayan S (Ed) Poles D et al. on behalf of the Serious Hazards of Transfusion (SHOT) Steering Group. *The 2021 Annual Shot Report* 2022. http://www.shotuk.org.
Seed CR, Hewitt PE, Dodd RY et al. Creutzfeldt-Jakob disease and blood transfusion safety. *Vox Sang* 2018; 113:220–231.
Shander A, Goodnough LT. From tolerating anemia to treating anemia. *Ann Intern Med* 2019; 170:125–126.
Stanworth SJ, Dowling K, Curry N et al. Haematological management of major haemorrhage: a British Society for Haematology Guideline. *Br J Haematol* 2022; 198(4):654–667.

WHITE BLOOD CELLS

Leucocytes found in peripheral blood are:

- Granulocytes: neutrophils, eosinophils, and basophils (all called granulocytes)
- Non-granulocytes: lymphocytes and monocytes
- The development of these cells is shown in **Fig. 38.3F**.

Neutrophils

Precursors to neutrophils in the bone marrow include:

- Myeloblasts: large cells constituting up to 3% of nucleated cells in the bone marrow. The nucleus is large and contains two to five nucleoli. The cytoplasm is scanty and contains no granules.
- Promyelocytes: similar to myeloblasts but have primary cytoplasmic granules, containing enzymes such as myeloperoxidase.
- Myelocytes: smaller cells without nucleoli. The cytoplasm is more abundant and contains primary and secondary granules.
- Metamyelocyte: the indentation of the nucleus marks the change from myelocyte to metamyelocyte.

The mature neutrophil is a smaller cell with a nucleus that has two to five lobes, with predominantly secondary granules in the cytoplasm, which contain lysozyme, collagenase, and lactoferrin. Peripheral blood neutrophils are equally distributed into a circulating pool and a marginating pool lying along the endothelium of blood vessels. In contrast to the prolonged maturation time of about 10 days for neutrophils in the bone marrow, the half-life in the peripheral blood is extremely short: only 6 to 8 hours. In response to stimuli (e.g., infection, corticosteroid therapy), neutrophils are released into the circulating pool from both the marginating pool and the marrow. Immature WBCs are released from the marrow when a rapid response (within hours) occurs in acute infection (described as a 'shift to the left' on a blood film).

Function

The prime function of neutrophils is to ingest and kill bacteria, fungi, and damaged cells. Neutrophils are attracted to sites of infection or inflammation by chemotaxins. Recognition of foreign or dead material is aided by the coating of particles with immunoglobulin and complement (opsonization) because neutrophils have Fc and C3b receptors. The material is ingested into vacuoles, where it is subjected to enzymic destruction; this is either oxygen dependent with the generation of hydrogen peroxide (myeloperoxidase) or oxygen independent (lysosomal enzymes and lactoferrin). Leucocyte alkaline phosphatase is an enzyme found in leucocytes. It is increased when there is a neutrophilia caused by an acute illness. It is also raised in polycythaemia and myelofibrosis and is reduced in chronic myeloid leukaemia.

Neutrophil leucocytosis

A rise in the number of circulating neutrophils to more than 10×10^9/L occurs in bacterial infections and tissue damage. It may also be seen in pregnancy, during exercise, secondary to smoking, and after corticosteroid administration (**Box 38.21**). With any tissue necrosis, there is a release of various soluble factors, causing leucocytosis. IL-1 is released in tissue necrosis and causes a pyrexia. The pyrexia and leucocytosis that accompany a myocardial infarction are a good example of this and may be wrongly attributed to infection.

Box 38.21 Causes of a neutrophil leucocytosis

- Bacterial infections
- Tissue necrosis (e.g., myocardial infarction, trauma)
- Inflammation (e.g., gout, rheumatoid arthritis)
- Drugs (e.g., corticosteroids, lithium)
- Acute haemorrhage or haemolysis
- Administration of granulocyte colony-stimulating factor
- Haematological
  - Myeloproliferative disease
  - Leucoerythroblastosis
- Physiological (e.g., pregnancy, exercise)
- Malignant disease (e.g., bronchial, breast, gastric)
- Metabolic (e.g., kidney failure, acidosis)
- Congenital (e.g., leucocyte adhesion deficiency, hereditary neutrophilia)
- Smoking

A ***leucoerythroblastic blood film*** (an overproduction of WBCs, with many immature cells and nucleated RBCs) may occur in severe infection, TB, malignant infiltration of the bone marrow, and occasionally after haemorrhage or haemolysis and in severe haemolytic or megaloblastic anaemias.

Neutropenia and agranulocytosis

Neutropenia is defined as a circulatory neutrophil count of less than 1.5 × 10^9/L. The causes are given in **Box 38.22**. A virtual absence of neutrophils is called agranulocytosis. Note that patients from African and Caribbean backgrounds may have lower neutrophil counts, termed 'ethnic neutropenia'. This is caused by a greater proportion of the neutrophils residing close to the vessel walls and not by an absolute reduction in total neutrophil production or function.

Viruses are probably the most common causes of neutropenia. Chemotherapy and radiotherapy predictably produce neutropenia. Other drugs (e.g., anti-thyroid drugs such as carbimazole, some antibiotics) can produce an idiosyncratic cytopenia; a drug cause should always be considered.

Clinical features

Infections may be frequent, are often serious, and become more likely as the neutrophil count falls. An absolute neutrophil count of less than 0.5 × 10^9/L is regarded 'severe' and may be associated with life-threatening infections such as pneumonia and septicaemia. A characteristic glazed mucositis occurs in the mouth, and ulceration is common.

Investigations

The blood film shows marked neutropenia. The appearance of the bone marrow indicates whether the neutropenia is caused by depressed production or increased destruction of neutrophils. Neutrophil antibody studies are performed if an immune mechanism is suspected. Viral serology, including HIV, should be undertaken.

Management

Antibiotics should be given as necessary and according to local guidelines in acute severe neutropenia. Exposure to infections should be minimized.

If the neutropenia seems likely to have been caused by a drug, therapy should be stopped. Recovery of the neutrophil count usually occurs after about 10 days. G-CSF is used to decrease the period of neutropenia after chemotherapy and haemopoietic transplantation. It is also used successfully in the treatment of chronic neutropenia.

Box 38.22 Causes of neutropenia

Acquired

- Viral infection
- Severe bacterial infection (e.g., typhoid)
- Felty's syndrome (rheumatoid arthritis associated)
- Autoimmune neutropenia
- Pancytopenia from any cause, including drug-induced marrow aplasia
- Drug-induced agranulocytosis (e.g., carbimazole)
- Pure white blood cell aplasia

Inherited

- Ethnic (neutropenia is common in African and Caribbean ethnicities)
- Kostmann's syndrome: severe infantile agranulocytosis caused by mutation in the elastane 2 *(ELA2)* gene
- Cyclical: genetic mutation in *ELA2* gene with neutropenia every 2–3 weeks
- Others (e.g., Shwachman–Diamond syndrome, dyskeratosis congenita, Chédiak–Higashi syndrome)

Steroids and high-dose IVIG are used to treat patients with severe autoimmune neutropenia and recurrent infections, and G-CSF has produced responses in some cases.

Eosinophils

Eosinophils are slightly larger than neutrophils and are characterized by a nucleus, usually with two lobes, and large cytoplasmic granules that stain deep red. The eosinophil plays a part in allergic responses and in the defence against helminth and protozoal infections. Eosinophilia is defined as more than 0.4 × 10^9/L eosinophils in the peripheral blood. The causes of eosinophilia are listed in **Box 38.23**.

Basophils

The nucleus of basophils is similar to that of neutrophils, but the cytoplasm is filled with large black granules containing histamine, heparin, and enzymes such as myeloperoxidase. The physiological role of basophils is not known. Binding of IgE causes the cells to degranulate and release histamine and other contents involved in acute hypersensitivity reactions.

Basophils are usually few in number (<1 × 10^9/L) but are significantly increased in some myeloproliferative disorders, such as chronic myeloid leukaemia. Basophilia should prompt further investigations.

Monocytes

Monocytes are slightly larger than neutrophils. The nucleus has a variable shape and may be round, indented, or lobulated. The cytoplasm contains fewer granules than neutrophils. Monocytes are precursors of tissue macrophages and dendritic cells; they spend only a few hours in the blood but can continue to proliferate in the tissues for many years.

A monocytosis (>0.8 × 10^9/L) may be seen in chronic bacterial infections (e.g., TB, infective endocarditis); chronic neutropenia; and myelodysplasia, particularly chronic myelomonocytic leukaemia.

Lymphocytes

Lymphocytes form nearly half of circulating WBCs. They descend from pluripotent stem cells (**Fig. 38.3F**). Circulating lymphocytes are

Box 38.23 Causes of eosinophilia

Parasitic infestations:
- Ascaris
- Hookworm
- Strongyloides

Allergic disorders:
- Hayfever (allergic rhinitis)
- Other hypersensitivity reactions, including drug reactions

Skin disorders:
- Urticaria
- Pemphigus
- Eczema

Pulmonary disorders:
- Bronchial asthma
- Tropical pulmonary eosinophilia
- Allergic bronchopulmonary aspergillosis
- Eosinophilic granulomatosis with polyangiitis

Malignant disorders:
- Hodgkin lymphoma
- T-cell non-Hodgkin lymphoma
- Carcinoma
- Eosinophilic leukaemia
- Acute myeloid leukaemia
- Myeloproliferative diseases associated with platelet-derived growth factor rearrangements

Miscellaneous disorders
- Hypereosinophilic syndrome
- Sarcoidosis
- Hypoadrenalism
- Eosinophilic gastroenteritis

small cells, a little larger than RBCs, with a dark-staining central nucleus. There are two main types: T and B lymphocytes.

Lymphocytosis (lymphocyte count >5 × 10^9/L) occurs in response to viral infections (e.g., EBV, CMV, and HIV) and chronic infections such as TB and toxoplasmosis. It is also found in chronic lymphocytic leukaemia and in some lymphomas.

Further reading

Butt NM, Lambert J, Ali S et al. Guideline for the investigation and management of eosinophilia. *Br J Haematol* 2017; 176:553–572.

HAEMOSTASIS

The integrity of the circulation is maintained by blood flowing through intact vessels lined by endothelial cells. Haemostasis is the host defence mechanism that protects this integrity after injury to the vessel wall and tissue. It is a complex process that depends on interactions between vessel wall, leucocytes, platelets, and coagulation and fibrinolytic mechanisms. Haemostatic systems are normally quiescent, but after tissue injury, become rapidly activated. The formation of the haemostatic plug is shown in Fig. 38.33.

Vessel wall

The vessel wall is lined by endothelium, which in normal conditions prevents platelet adhesion and thrombus formation. This property is caused by:
- Its negative charge
- Thrombomodulin and heparan sulphate expression
- Synthesis of prostacyclin (prostaglandin I_2 [PGI_2]) and nitric oxide, which cause vasodilation and inhibit platelet aggregation
- Production of plasminogen activator.

Injury to vessels causes reflex vasoconstriction, and endothelial damage results in loss of antithrombotic properties, activation of platelets and coagulation, and inhibition of fibrinolysis (Fig. 38.33).

Platelets

Platelet adhesion

When the vessel wall is damaged, the escaping platelets come into contact with and adhere to collagen and subendothelial bound VWF. This adherence is mediated through GP Ib. GP IIb/IIIa is then exposed, forming a second binding site for VWF. Within seconds of adhesion to the vessel wall, platelets begin to undergo a shape change from a disc to a sphere, spread along the subendothelium, and release the contents of their cytoplasmic granules: dense bodies (containing adenosine diphosphate [ADP] and serotonin) and α-granules (containing platelet-derived growth factor, platelet factor 4, β-thromboglobulin, fibrinogen, VWF, fibronectin, thrombospondin, and other factors).

Platelet activation

The release of ADP leads to a conformational change in the GP IIb/IIIa complex (GPIIb/IIIa) on the surfaces of adherent platelets, allowing it to bind to fibrinogen (Fig. 38.34).

Platelet aggregation

Because fibrinogen is a dimer, it can form a direct bridge between platelets and binds platelets into activated aggregates (Fig. 38.33B) and further platelet release of ADP occurs. A self-perpetuating cycle of events is set up, leading to formation of a platelet plug at the site of the injury.

Coagulation

After platelet aggregation and release of ADP, exposed platelet membrane phospholipids are available for assembly of coagulation factor enzyme complexes (tenase and prothrombinase). The presence of thrombin encourages fusion of platelets, and fibrin formation reinforces the stability of the platelet plug. Central to normal platelet function is ***prostaglandin synthesis***, which is induced by platelet activation and leads to the formation of thromboxane A_2 (TXA_2) in platelets (Fig. 38.35). TXA_2 is a powerful vasoconstrictor that reduces cyclic adenosine monophosphate (AMP) concentrations and initiates the platelet release reaction. PGI_2 is synthesized in vascular endothelial cells and opposes the actions of TXA_2. It produces vasodilation and increases the concentration of cyclic AMP, preventing platelet aggregation on the normal vessel wall, as well as limiting the extent of the initial platelet plug after injury.

Coagulation and fibrinolysis

Coagulation involves a series of enzymatic reactions that lead to the conversion of soluble plasma fibrinogen to fibrin-based clot (Fig. 38.36). Roman numerals are used for most of the factors, but I and II are referred to as fibrinogen and prothrombin, respectively; III, IV, and VI are redundant. The active forms are denoted by 'a'. The coagulation factors are primarily synthesized in the liver and are either serine protease enzyme precursors (factors IX, X, and XI and thrombin) or co-factors (V and VIII), except for fibrinogen, which is polymerized to form fibrin.

Coagulation pathway

This enzymatic amplification system was traditionally divided into 'extrinsic' and 'intrinsic' pathways (Fig. 38.37). This concept is useful for the interpretation of clinical laboratory tests such as the prothrombin time (PT) and activated partial thromboplastin time (APTT),

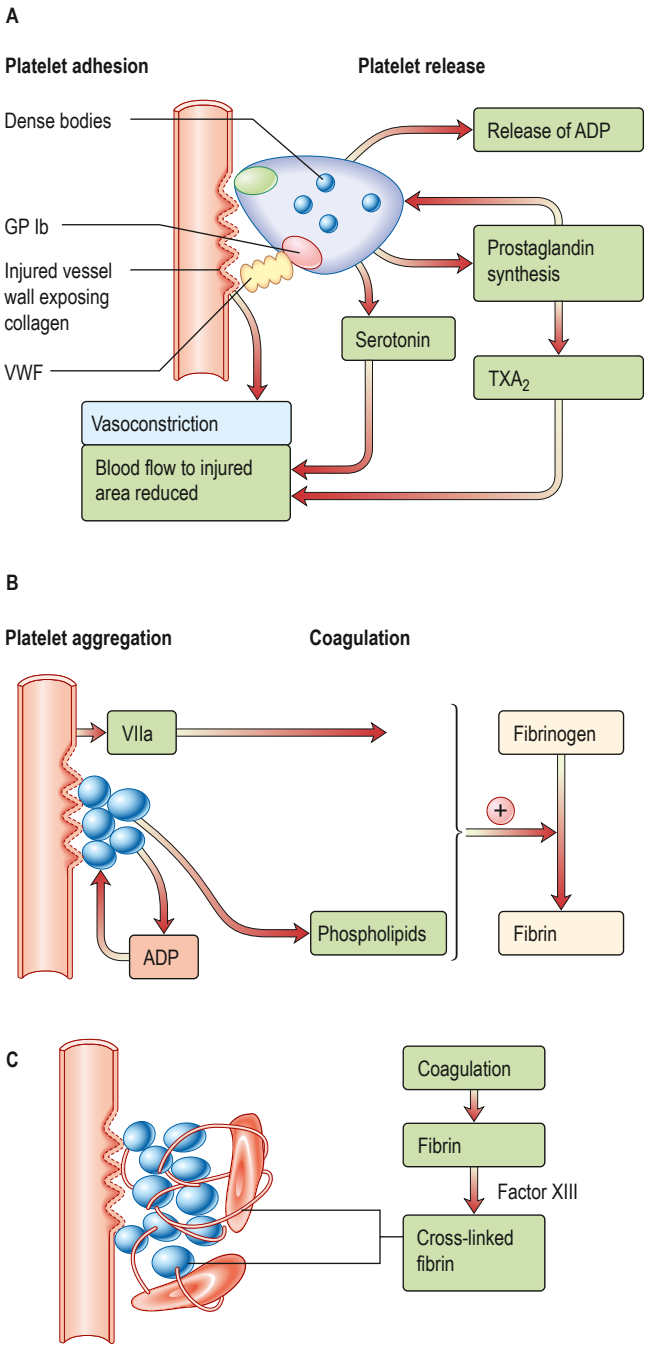


Fig. 38.33 Formation of the haemostatic plug: sequential interactions between the vessel wall, platelets, and coagulation factors. (A) ***Contact of platelets with collagen*** via the platelet receptor glycoprotein (GP) Ib and von Willebrand factor (VWF) activates platelet prostaglandin synthesis, which stimulates release of adenosine diphosphate (ADP) from the dense bodies. Vasoconstriction of the vessel occurs as a reflex and because of release of serotonin and thromboxane A_2 (TXA_2) from platelets. **(B)** ***Release of ADP from platelets*** induces platelet aggregation and formation of the platelet plug. The coagulation pathway is stimulated, leading to the formation of fibrin. **(C)** ***Fibrin strands are cross-linked by factor XIII*** and stabilize the haemostatic plug by binding platelets and red blood cells.

but it is unrepresentative and oversimplifies *in vivo* coagulation. Coagulation is initiated by tissue damage (Fig. 38.36):

- Tissue damage exposes tissue factor (TF), which binds to factor VII.
- The TF–factor VII complex directly converts factor X to active factor Xa and some factor IX to factor IXa.
- In the presence of factor Xa, tissue factor pathway inhibitor (TFPI) inhibits further generation of factor Xa and factor IXa.
- After inhibition by TFPI, the amount of factor Xa produced is insufficient to maintain coagulation. To allow haemostasis to progress to completion, further factor Xa can be generated only by the alternative factor IX–factor VIII pathway. However, enough thrombin exists at this point to activate factor VIII, which dramatically increases the activity of factor IXa (generated by TF–factor VIIa), so further activation of factor X can proceed. Without the amplification and consolidating action of factor VIII–factor IX, bleeding will ensue because generation of factor Xa is insufficient to sustain haemostasis.
- Similarly, thrombin activates factor V, dramatically enhancing conversion of prothrombin to thrombin by factor Xa.
- Thrombin hydrolyses the peptide bonds of fibrinogen, releasing fibrinopeptides A and B and allowing polymerization between fibrinogen molecules to form fibrin. At the same time, thrombin in the presence of calcium ions activates factor XIII, which stabilizes the fibrin clot by cross-linking adjacent fibrin molecules.

Factor VIII

Factor VIII is a key co-factor in the coagulation pathway. It is associated with its carrier protein VWF. When activated, factor VIII increases the conversion of factor X to Xa by factor IXa by approximately 200 000-fold. VWF functions to protect factor VIII from premature proteolytic breakdown and locates it to areas of vascular injury. Factor VIII:C (the antihaemophilic and procoagulant component of the factor VIII complex) has a molecular weight of about 350 kDa.

von Willebrand factor

VWF is a GP with a monomeric molecular weight of about 250 kDa, which readily forms multimers in the circulation with molecular weights of up to 20 000 kDa. It is synthesized by endothelial cells and megakaryocytes and stored in platelet granules and endothelial cells. The high-molecular-weight multimeric forms of VWF are the most biologically active. It has a primary role as an adhesive protein binding to platelets, surface GPs, and constituents of connective tissue. It also acts as a carrier protein for factor VIII (Fig. 38.38).

Factor V

Factor V is similar to factor VIII in that it is a non-enzymatic co-factor. In its thrombin-activated form, factor Va, it significantly enhances conversion of prothrombin to thrombin.

Physiological limitation of coagulation

Without a physiological system to limit blood coagulation, excessive thrombosis could ensue. The natural anticoagulant mechanism regulates and localizes thrombosis to the site of injury.

Antithrombin

Antithrombin, a member of the serine protease inhibitor (serpin) superfamily, is a potent inhibitor of coagulation. It inactivates the serine proteases by forming stable complexes with them and its action is greatly potentiated by heparin.

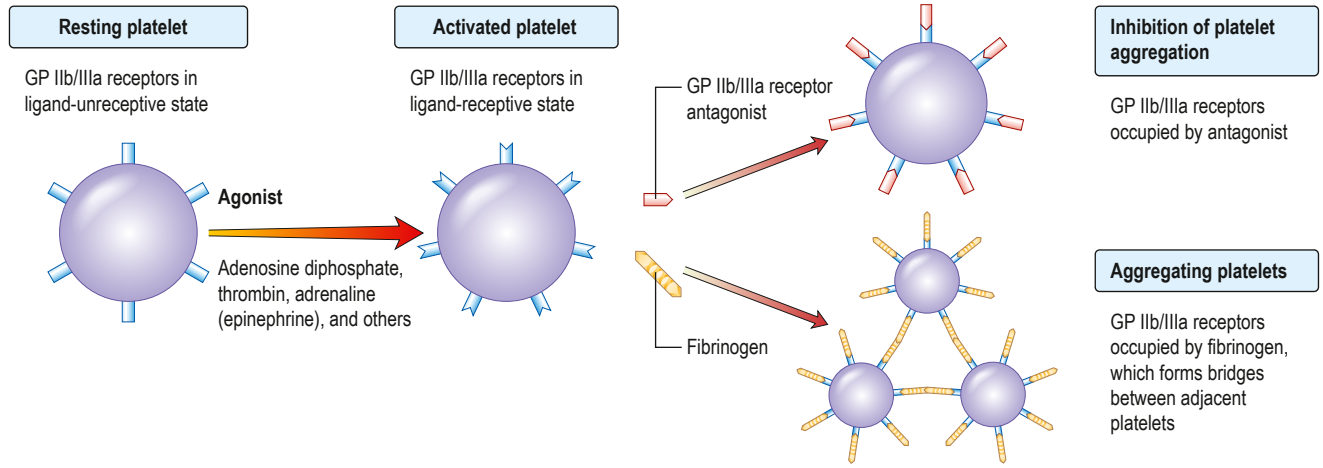


Fig. 38.34 The role of glycoprotein (GP) IIb/IIIa in platelet aggregation and inhibition of platelet aggregation by GP IIb/IIIa receptor antagonists.

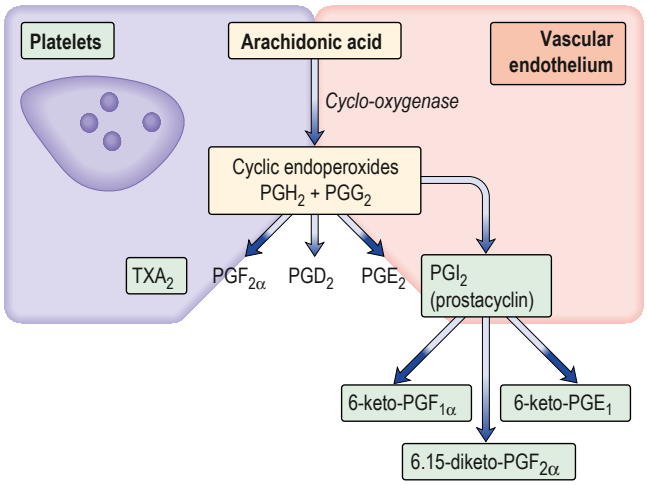


Fig. 38.35 Prostaglandin (PG) synthesis. *TXA_2,* Thromboxane A_2.

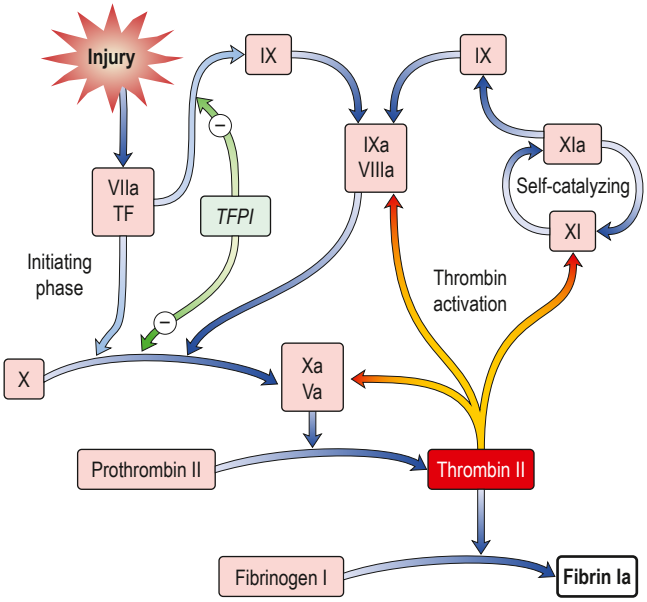


Fig. 38.36 Coagulation mechanism. The *in vivo* pathway begins with the tissue factor–factor VIIa (VIIa–TF) complex activating factor X (Xa) and amplifying factor IX (IXa). Thrombin activates factor XI (XIa), as well as factor IXa, VIIIa, and Va and Xa complexes. *TFPI,* Tissue factor pathway inhibitor.

Activated protein C

This is generated from its vitamin K–dependent precursor, protein C, by thrombin. Thrombin activation of protein C is greatly enhanced when thrombin is bound to thrombomodulin on endothelial cells (Fig. 38.39). Activated protein C inactivates factor Va and factor VIIIa, reducing further thrombin generation.

Protein S

This is a co-factor for protein C, which acts by enhancing binding of activated protein C to the phospholipid surface. It circulates bound to C4b binding protein, but about 30% to 40% remains unbound and active (free protein S).

Other inhibitors

Other natural inhibitors of coagulation include α_2-macroglobulin, α_1-antitrypsin, and α_2-antiplasmin.

Fibrinolysis

Fibrinolysis is a normal haemostatic response that helps to restore vessel patency after vascular damage. The principal component is the enzyme plasmin, which is generated from its inactive precursor plasminogen (Fig. 38.40). This is achieved principally via tissue plasminogen activator (t-PA) released from endothelial cells. Some plasminogen activation may also be promoted by urokinase, produced in the kidneys. Other plasminogen activators (factor XII and prekallikrein) have minor physiological importance.

Plasmin

Plasmin is a serine protease that breaks down fibrinogen and fibrin into fragments X, Y, D, and E, collectively known as fibrin (and fibrinogen) degradation products (FDPs). D-dimer is produced when cross-linked fibrin is degraded. Its presence in the plasma indicates that the coagulation mechanism has been activated.

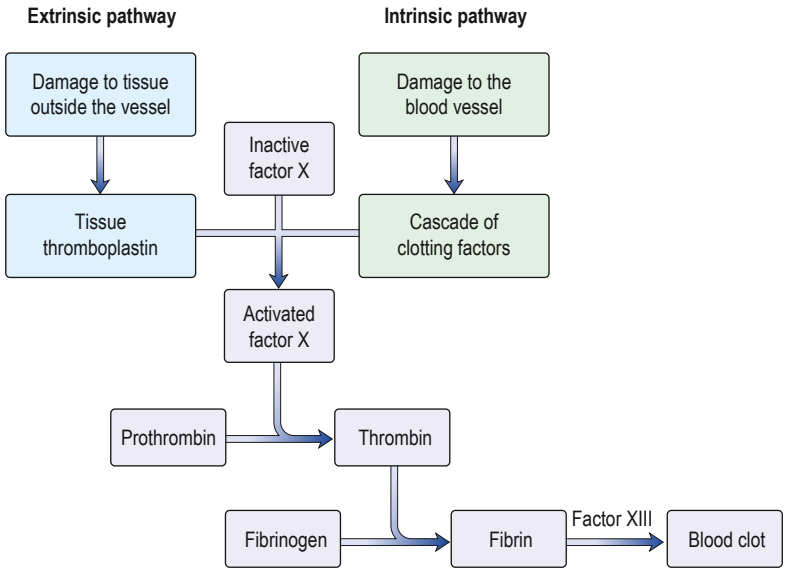


Fig. 38.37 The classic *in vitro* extrinsic and intrinsic coagulation cascade. See also Fig. 38.36. Initiation of coagulation differs significantly *in vitro* from *in vivo*.

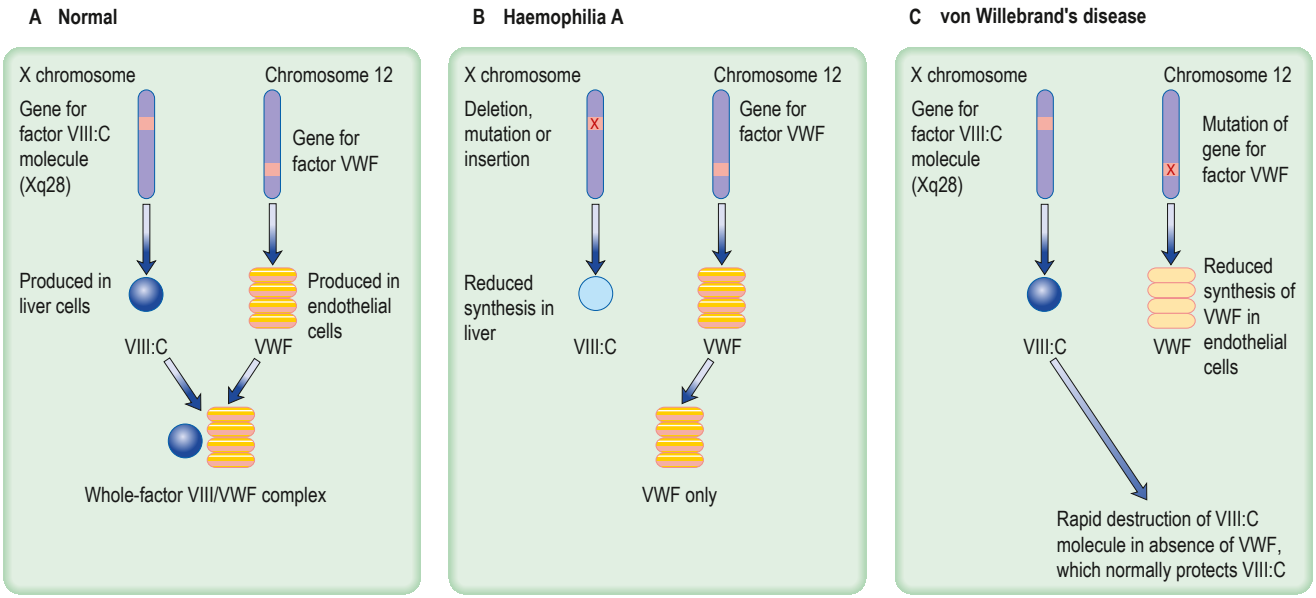


Fig. 38.38 Synthesis of factor VIII and von Willebrand factor (VWF) in inherited coagulation disorders. **(A)** Normal factor VIII synthesis. **(B)** Haemophilia A, showing defective synthesis of factor VIII:C. **(C)** von Willebrand's disease, showing reduced synthesis of VWF.

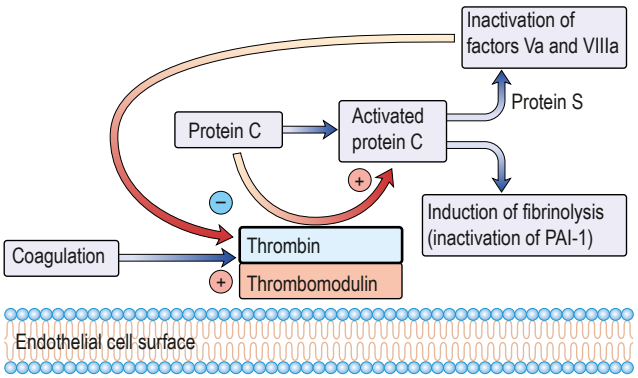


Fig. 38.39 Activation of protein C. *PAI-1,* Plasminogen activator inhibitor 1.

Fibrinolytic system

This is activated by the presence of fibrin. Plasminogen is specifically adsorbed to fibrin and fibrinogen by lysine-binding sites. However, little plasminogen activation occurs in the absence of polymerized fibrin, as fibrin also has a specific binding site for plasminogen activators, but fibrinogen does not (Fig. 38.41).

Tissue plasminogen activator

t-PA is inactivated by plasminogen activator inhibitor-1 (PAI-1). Activated protein C inactivates PAI-1 and therefore induces fibrinolysis (Fig. 38.39). Inactivators of plasmin, such as α_2-antiplasmin (Fig. 38.41) and thrombin-activatable fibrinolysis inhibitor, also contribute to the regulation of fibrinolysis.

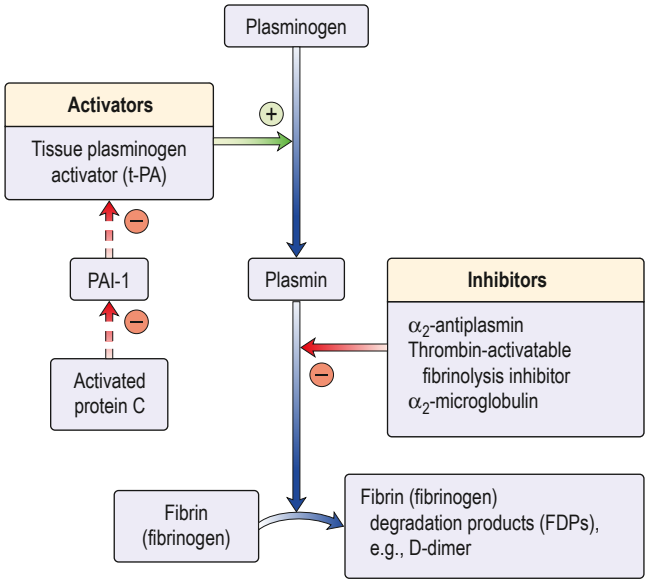


Fig. 38.40 Fibrinolytic system. *PAI-1,* Plasminogen activator inhibitor-1; *t-PA,* tissue plasminogen activator.

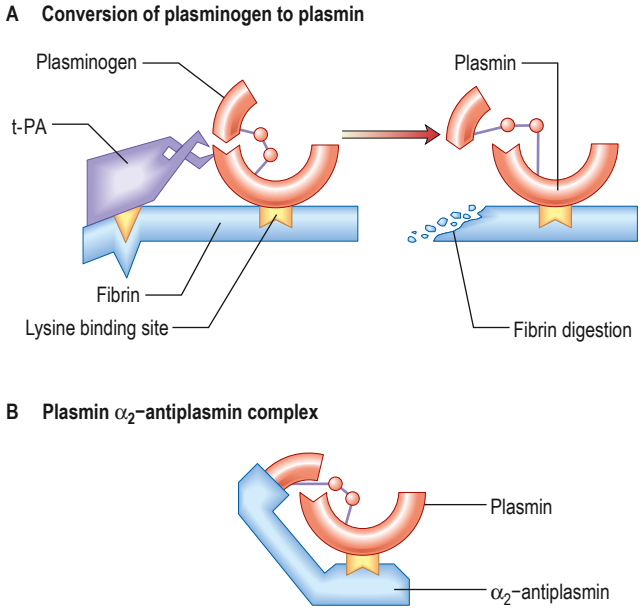


Fig. 38.41 Fibrinolysis. (A) ***The conversion of plasminogen to plasmin*** by tissue plasminogen activator (t-PA) occurs most efficiently on the surface of fibrin, which has binding sites for both plasminogen and t-PA. **(B) *Free plasmin in the blood is rapidly inactivated by α_2-antiplasmin*.** Plasmin generated on the fibrin surface is partially protected from inactivation. The lysine-binding sites on plasminogen are necessary for the interaction between plasmin(ogen) and fibrin and between plasmin and α_2-antiplasmin.

Investigation of bleeding disorders

Although the precise diagnosis of a bleeding disorder depends on laboratory tests, much information is obtained from a structured history and physical examination.

Is there a generalized haemostatic defect?

Supportive evidence includes bleeding from multiple sites, spontaneous bleeding, and excessive bleeding after injury.

Is the defect inherited or acquired?

A family history of a bleeding disorder should be sought. Severe inherited defects usually become apparent in infancy, sometimes at the time of birth with prolonged bleeding from the umbilical cord. Mild inherited defects may only come to attention later in life, such as with excessive bleeding after surgery, childbirth, dental extractions, or trauma. Some milder defects are revealed by routine coagulation screens performed before surgical procedures.

Is the bleeding suggestive of a vascular or platelet defect or a coagulation defect?

Vascular or platelet bleeding

This is characterized by easy bruising and spontaneous bleeding from small vessels. There is often bleeding into the skin. Purpura includes:

- Petechiae: small skin haemorrhages varying from pinpoint size to a few millimetres in diameter that do not blanch on pressure
- Ecchymoses: larger areas of bleeding into the skin

Bleeding also occurs from mucous membranes, especially the nose and mouth.

Coagulation disorders

These are typically associated with bleeding after injury or surgery. In more severe forms, there are haemarthroses and muscle haematomas. There is often a short delay between the precipitating event and overt haemorrhage or haematoma formation.

Laboratory investigations

Blood count and film

These show the number and morphology of platelets and any blood disorder, such as leukaemia or lymphoma. The normal range for the platelet count is 150 to 400 × 10^9/L.

Coagulation tests

These are performed on blood collected into citrate, which neutralizes calcium ions and prevents clotting.

- ***The PT*** is measured by adding tissue factor (thromboplastin) and calcium to patient plasma and establishing the length of time taken for the blood to clot. The normal PT is 11.5 to 13.5 seconds. When used to measure warfarin-based anticoagulation, the PT is expressed as the INR. The PT measures factors VII, X, and V, as well as prothrombin and fibrinogen (the classic 'extrinsic' pathway; Fig. 38.37) and is prolonged if any of these factors is low. It may also be abnormal in patients with liver disease or if the patient is taking oral anticoagulants, including warfarin, and direct oral anticoagulants that target factor Xa.
- ***The APTT*** is also sometimes termed PTT with kaolin. It is measured by adding a surface activator (e.g., kaolin, micronized silica, or ellagic acid), phospholipid (to mimic platelet membrane), and calcium to the patient's plasma. The normal APTT is 26 to 37 seconds and depends on the exact methodology. The APTT measures factors XII, XI, IX, VIII, X, and V, as well as prothrombin and fibrinogen (the classic 'intrinsic' pathway) and is prolonged in deficiencies of one or more of these factors. It is not dependent on factor VII.
- ***The thrombin time (TT)*** is measured by adding thrombin to the patient's plasma. A normal TT is 12 to 14 seconds. It is prolonged in fibrinogen deficiency, in qualitative defects of

Box 38.24 Vascular disorders associated with bleeding/bruising

Congenital
- Hereditary haemorrhagic telangiectasia (Osler–Weber–Rendu disease)
- Connective tissue disorders: Ehlers–Danlos syndrome, osteogenesis imperfecta, pseudoxanthoma elasticum, Marfan's syndrome

Acquired

Severe infections
- Septicaemia
- Meningococcal infections
- Measles
- Typhoid

Autoimmune disorders
- Henoch–Schönlein purpura
- Systemic lupus erythematosus
- Rheumatoid arthritis

Drug-induced disorders
- Corticosteroids
- Sulphonamides

Others
- Senile purpura
- Easy bruising syndrome
- Scurvy
- Factitious purpura

fibrinogen (dysfibrinogenaemia), and in the presence of inhibitors such as heparin or FDPs.

- ***Correction tests*** can be used to differentiate prolonged times in the PT, APTT, and TT caused by various coagulation factor deficiencies and inhibitors of coagulation. Prolonged PT, APTT, or TT caused by coagulation factor deficiencies can be corrected by addition of normal plasma to the patient's plasma. Failure to correct is suggestive of the presence of an inhibitor of coagulation.
- ***Factor assays*** are used to confirm coagulation defects, especially when a single inherited disorder is suspected.
- ***Special tests of coagulation*** will often be required to confirm the precise haemostatic defect. Such tests include estimation of fibrinogen and FDPs, platelet function tests such as platelet aggregation, and platelet granule contents.
- ***Bleeding time*** measures platelet plug formation *in vivo*. Classically, a sphygmomanometer cuff is inflated to 40 mmHg, and standardized incisions are made in the forearm. The time taken for bleeding to stop is recorded. Prolonged times are seen in platelet function defects and when the platelet count is less than 100×10^9/L. Nowadays, this test is very rarely done as it is painful and can scar.

VASCULAR DISORDERS

Vascular disorders can be characterized by easy bruising and bleeding into the skin (Box 38.24). Bleeding from mucous membranes sometimes occurs, but bleeding is rarely severe. Laboratory investigations, including the bleeding time, are normal.

Hereditary haemorrhagic telangiectasia

This is a rare disorder with autosomal dominant inheritance. In most cases, mutations occur in one of three genes—*ENG*, *ALK1*, or *SMAD4*—that encode components of the TGF-β signalling pathway involved in blood vessel development. Dilation of capillaries and small arterioles produces characteristic small red spots that blanch on pressure in the skin and mucous membranes, particularly the nose and GI tract. Recurrent epistaxis and chronic GI bleeding are the major problems that cause chronic iron-deficiency anaemia. Vascular malformations also occur in pulmonary, hepatic, cerebral, and spinal vasculature.

Box 38.25 Platelet count phenotypes

Platelet count ($\times 10^9$/L)	Clinical defect
>500	Haemorrhage or thrombosis
500–100	No clinical effect
100–50	Moderate haemorrhage after injury
50–20	Purpura Haemorrhage after injury
<20	Purpura common Spontaneous haemorrhage from mucous membranes Intracranial haemorrhage (rare)

Easy bruising syndrome

This is a common benign disorder often seen in otherwise healthy females. It is characterized by bruises on the arms, legs, and trunk with minor trauma, possibly because of skin vessel fragility. It may give rise to the suspicion of a serious bleeding disorder.

Senile purpura and purpura caused by steroids

These are both caused by atrophy of vascular supporting tissue.

Purpura caused by infections

This is mainly caused by damage to the vascular endothelium. The rash of meningococcal septicaemia is particularly characteristic.

Henoch–Schönlein purpura

This is mainly a disease of children. It is a type III hypersensitivity (immune complex) reaction that is often preceded by an upper respiratory tract infection. Purpura is seen mainly on the legs and buttocks. Abdominal pain, arthritis, haematuria, and glomerulonephritis also occur. Recovery is usually spontaneous, but some patients develop kidney failure.

Episodes of inexplicable bleeding or bruising

These may represent abuse, either self-inflicted or caused by others. The various forms of artificial or factitious purpura may be expressions of emotional or psychiatric disturbance.

PLATELET DISORDERS

Bleeding caused by thrombocytopenia or abnormal platelet function is characterized by purpura and bleeding from mucous membranes. Bleeding is uncommon with platelet counts of more than 50×10^9/L, and severe spontaneous bleeding is unusual with platelet counts greater than 20×10^9/L (Box 38.25).

Thrombocytopenia

This is caused by reduced platelet production in the bone marrow, excessive peripheral destruction of platelets, or sequestration in an

Box 38.26 Causes of thrombocytopenia

Impaired production
- Selective megakaryocyte depression
  - Rare congenital defects
  - Drugs, chemicals, and viruses
- As part of a bone marrow failure
  - Cytotoxic drugs and chemicals
  - Radiation
  - Megaloblastic anaemia
  - Leukaemia
  - Myelodysplastic syndromes
  - Myeloma
  - Myelofibrosis
  - Solid tumour infiltration
  - Aplastic anaemia
  - HIV infection

Excessive destruction or increased consumption
- Autoimmune: ITP
- Secondary immune: SLE, chronic lymphocytic leukaemia, viruses, drugs (e.g., heparin, bivalirudin)
- Drug induced (e.g., GP IIb/IIIa inhibitors, penicillins, thiazides)
- Alloimmune neonatal thrombocytopenia
- Post-transfusion purpura
- DIC
- TTP
- HIV

Sequestration
- Splenomegaly
- Hypersplenism

Dilutional
- Massive transfusion

DIC, Disseminated intravascular coagulation; *GP,* glycoprotein; *HIV,* human immunodeficiency virus; *ITP,* immune thrombocytopenic purpura; *SLE,* systemic lupus erythematosus; *TTP,* thrombotic thrombocytopenic purpura.

enlarged spleen (Box 38.26). The underlying cause may be revealed by history and examination, but a bone marrow examination will show whether the numbers of megakaryocytes are reduced, normal, or increased and will provide essential information on morphology. Specific laboratory tests may be useful to confirm the presence of conditions such as PNH or SLE.

Immune thrombocytopenic purpura

In ITP, thrombocytopenia is caused by immune destruction of platelets. The antibody-coated platelets are removed after binding to Fc receptors on macrophages.

Immune thrombocytopenic purpura in children

This occurs most commonly in the 2- to 6-year age group. ITP has an acute onset with mucocutaneous bleeding, and there may be a history of a recent viral infection, including varicella zoster or measles. Although bleeding may be severe, life-threatening haemorrhage is rare (~1%). Bone marrow examination is not usually performed unless there are atypical clinical or laboratory features.

Immune thrombocytopenic purpura in adults

The presentation is usually less acute than in children. ITP is characteristically seen in females and may be associated with other autoimmune disorders such as SLE, thyroid disease, and autoimmune haemolytic anaemia (Evans syndrome). It is also seen in patients with chronic lymphocytic leukaemia and solid tumours and after infections with viruses such as HIV. Platelet autoantibodies are detected in about 60% to 70% of patients. In the remaining patients, they are presumed to be present, although they are not detectable. The antibodies often have specificity for platelet membrane GPs IIb/IIIa or Ib.

Clinical features

Major haemorrhage is rare and is seen only in severe thrombocytopenia. Easy bruising, purpura, epistaxis, and menorrhagia are common. Physical examination is normal except for evidence of bleeding. Splenomegaly is rare.

Investigations

The only blood count abnormality is thrombocytopenia. The diagnosis is based on exclusion of other causes of thrombocytopenia using the history, clinical examination, blood count, and blood film. Normal or increased numbers of megakaryocytes are found in the bone marrow, but bone marrow examination is unnecessary unless there are unusual findings clinically or on blood film. The detection of platelet autoantibodies is not essential for confirmation of the diagnosis.

Management

Children

Children do not usually require treatment because the thrombocytopenia is short-lived. Treatment should be reserved for serious bleeding or urgent surgery. When treatment is necessary, corticosteroids and IVIG are effective in more than 80% of children and raise the count more rapidly than steroids alone. Chronic ITP is unusual in children and requires specialist management.

Adults

Patients with platelet counts of more than 30 × 10^9/L generally require no treatment unless they are about to undergo a surgical procedure. Those with even lower platelet counts may not require treatment unless they have spontaneous bruising or bleeding.

First-line treatment consists of oral corticosteroids (e.g., prednisolone 1 mg/kg body weight). Approximately two-thirds of patients respond to corticosteroids, but relapses are common when the dose is reduced. Only one-third of patients can expect a long-term response. Long-term remission is seen in just 10% to 20% of patients after stopping prednisolone. Patients who fail to respond to corticosteroids or require high doses to maintain a safe platelet count should be considered for alternative treatments. TPO receptor agonists such as romiplostim and eltrombopag have revolutionized management with high and durable response rates and a low level of adverse effects.

Other immune thrombocytopenias

Drugs

Many drugs have been reported to cause immune thrombocytopenia. The same drugs can be responsible for immune haemolytic anaemia, thrombocytopenia, or neutropenia in different patients.

Heparin-induced thrombocytopenia

See Chapter 40.

Neonatal alloimmune thrombocytopenia

This condition is due to fetomaternal incompatibility for platelet-specific antigens, usually for human platelet alloantigen 1a (HPA-1a). It is the platelet equivalent of HDN. The mother is HPA-1a negative and produces antibodies that destroy HPA-1a–positive

fetal platelets. Thrombocytopenia is self-limiting after delivery, but platelet transfusions may be required initially to prevent or treat bleeding associated with severe thrombocytopenia; platelets are prepared from HPA-1a–negative volunteers or the mother herself. Severe bleeding, such as intracranial haemorrhage, may occur *in utero*. Antenatal treatment of the mother, usually with IVIG, steroids, or both, has been effective in preventing haemorrhage in severely affected cases.

Post-transfusion purpura

Post-transfusion purpura (PTP) is rare, occurring 7 to 10 days after a transfusion of platelet-containing blood components, usually RBCs. PTP is associated with a platelet-specific alloantibody, usually anti–HPA-1a in an HPA-1a–negative individual. PTP always occurs in patients who have previously been immunized, either by blood transfusion or by pregnancy; hence, it is more common in females. The cause of the destruction of the patient's own platelets is not well understood, but they may be destroyed as 'bystanders' during the acute immune response to HPA-1a. PTP is self-limiting; IVIG is the treatment of choice if there is severe thrombocytopenia or bleeding.

Thrombotic thrombocytopenic purpura

TTP is a rare but very serious condition, in which platelet consumption leads to profound thrombocytopenia. There is a characteristic symptom complex of florid purpura, fever, fluctuating cerebral dysfunction, and MAHA with RBC fragmentation, often accompanied by acute kidney injury. The coagulation screen is usually normal, but LDH concentrations are markedly raised because of haemolysis. TTP stems from endothelial damage and microvascular thrombosis caused by a reduction in ADAMTS-13 (*a d*isintegrin-like *a*nd *m*etalloproteinase domain with *t*hrombospondin-type motifs), a protease that is normally responsible for regulating the size of VWF. ADAMTS-13 is needed to break down ultra-large von Willebrand factor multimers (UL VWFMs) into smaller, haemostatically active fragments that interact with platelets. Reduction in ADAMTS-13 results in the adhesion and aggregation of platelets to UL VWFMs and multiorgan microthrombi. In most sporadic cases, there is true deficiency of ADAMTS-13, associated with antibodies to ADAMTS-13. In some congenital cases, deficiency is caused by *ADAMTS-13* gene mutations. Secondary causes of acute TTP include pregnancy, oral contraceptives, SLE, infection, and drug treatment (including ticlopidine and clopidogrel). Such patients may have a variable ADAMTS-13 activity at presentation and may or may not have associated antibodies to ADAMTS-13.

Management

Plasma exchange is the mainstay of treatment. It provides a source of ADAMTS-13 and removes associated autoantibody in acute TTP. Cryoprecipitate and solvent-detergent FFP both contain ADAMTS-13. Pulsed intravenous methylprednisolone is given acutely; increasingly, rituximab is also a primary treatment of choice. Disease activity is monitored by measuring the platelet count and serum LDH. Platelet concentrates are contraindicated. The untreated condition has a mortality rate of up to 90%, but modern management has reduced this to about 10%. Recurrent and relapsing TTP occurs, often associated with a persistent lack of ADAMTS-13. In secondary TTP cases, identifiable precipitating drugs should be stopped. Caplacizumab is a single-variable-domain immunoglobulin directed to the A1 region of the VWF. It inhibits the VWF interaction with GP Ib and has recently been shown to lead to faster normalization of the platelet count and a lower rate of recurrent TTP.

Box 38.27 Causes of platelet dysfunction

Inherited

- ***Glanzmann's thrombasthenia:*** lack of platelet membrane GP IIb/IIIa complex, resulting in defective fibrinogen binding and failure of platelet aggregation
- ***Bernard–Soulier syndrome:*** lack of platelet membrane GP Ib/IX/V complex (the binding site for von Willebrand factor), causing failure of platelet adhesion and moderate thrombocytopenia
- ***Storage pool disease:*** lack of the storage pool of platelet dense bodies, causing poor platelet function

Acquired

- Myeloproliferative disorders
- Kidney and liver disease
- Paraproteinaemias
- Drug induced (e.g., NSAIDs [aspirin], platelet inhibitory drugs)

GP, Glycoprotein; *NSAID,* non-steroidal anti-inflammatory drug.

Platelet function disorders

Platelet function disorders (**Box 38.27**) are usually associated with excessive bruising and bleeding and, in some of the acquired forms, with thrombosis. The platelet count is often normal, but bleeding time is prolonged. Rare inherited defects of platelet function require more detailed investigations, such as platelet aggregation studies and factor VIII and VWF assays, because there is often a clinical overlap in symptoms with von Willebrand's disease (VWD).

If there is serious bleeding or if the patient is about to undergo surgery, drugs with antiplatelet activity should be withdrawn and any underlying condition should be treated if possible.

Platelet dysfunction can be a major component of bleeding in kidney disease. The degree of the platelet-related defect is broadly proportional to the serum urea concentration: platelet function is impaired by urea, guanidinosuccinic acid, and other phenolic metabolites that accumulate in those with CKD. Dialysis partially corrects platelet function. The Hct should be increased to over 0.30, and non-specific use of desmopressin may be helpful. Platelet transfusions may be required if these measures are unsuccessful or if the risk of bleeding is high.

Thrombocytosis

The platelet count may rise above 400 × 10^9/L as a result of:

- Splenectomy
- Malignant disease
- Inflammatory disorders (e.g., rheumatoid arthritis, inflammatory bowel disease)
- Major surgery and haemorrhage
- Myeloproliferative disorders
- Iron deficiency

Thrombocytosis is part of the acute-phase reaction. Platelet numbers are increased after splenectomy because of the loss of a major site of platelet destruction. ***Essential thrombocythaemia***, a myeloproliferative neoplasm, and other myeloproliferative conditions, such as polycythaemia vera, myelofibrosis, and chronic myeloid leukaemia, may also be associated with high platelet counts (Chapter 39).

A persistently elevated platelet count can lead to arterial or venous thrombosis. It is usual to treat the underlying cause of the

thrombocytosis but low-dose aspirin (75 mg) may also be given. In myeloproliferative diseases, the primary risk is thrombosis, and specific action is often taken to reduce the platelet count, usually with hydroxycarbamide (hydroxyurea). Paradoxically, there is also a risk of abnormal bleeding if the platelet count is very high.

Further reading

Chen M, Shortt J. Plasma cell directed therapy for immune thrombotic thrombocytopenic purpura. *Transfus Med Rev* 2022; 36:204–214.
Jiang D, Al-Samkari H, Panch SR. Changing paradigms in ITP management: newer tools for an old disease. *Transfus Med Rev* 2022; 36:188–194.
Scully M, Cataland SR, Peyvandi F et al. Caplacizumab treatment for acquired thrombotic thrombocytopenic purpura. *N Engl J Med* 2019; 380:335–346.

INHERITED COAGULATION DISORDERS

Inherited coagulation disorders are uncommon and usually involve deficiency of one factor only. Acquired coagulation disorders occur more frequently and almost always involve several coagulation factors.

Isolated deficiencies of all factors have been described in patients with inherited coagulation disorders. Those leading to abnormal bleeding are rare, apart from haemophilia A (factor VIII deficiency), haemophilia B (factor IX deficiency), and VWD.

Haemophilia A

Haemophilia A is caused by a lack of factor VIII. VWF is normal in haemophilia (Fig. 38.38). The prevalence of haemophilia A is about 1 in 5000 of the male population. It is inherited as an X-linked disorder. If a female carrier has a son, he has a 50% chance of having haemophilia; a daughter has a 50% chance of being a carrier. All female offspring of males with haemophilia are carriers; male offspring are normal.

Although many different genetic defects have been found in the factor VIII gene, a common gene inversion in intron 22 is causative in approximately 50% of families with severe disease. There is a high mutation rate, one-third of cases being apparently sporadic with no family history of haemophilia.

Clinical features and investigations

The clinical features depend on the concentration of factor VIII. The normal concentration of factor VIII is 50 to 150 IU/dL.

- ***Severe haemophilia (<1 IU/dL):*** frequent spontaneous bleeding from early life, typically into joints and muscles. Without adequate treatment, recurrent bleeding into joints leads to joint deformity.
- ***Moderate haemophilia (1–5 IU/dL):*** severe bleeding after injury and occasional spontaneous bleeds.
- ***Mild haemophilia (>5 IU/dL):*** bleeding usually only after injury or surgery. Diagnosis in this group can be delayed until late in life.

With treatment, the life expectancy is very good. The most common causes of death are cancer and heart disease comparable to the general population, rather than bleeding, although cerebral haemorrhage is more frequent. The main laboratory features are shown in Box 38.28. The abnormal findings are an isolated prolongation of the APTT and a reduced concentration of factor VIII. The PT, bleeding time, and VWF concentration are normal.

Management

Patients with bleeding are treated by administration of ***factor VIII concentrate*** by intravenous infusion to achieve normalization of concentrations. Factor VIII concentrate is available as a plasma-derived or recombinant product. Recombinant products are the treatment of choice, but economic constraints often limit availability, particularly in low-income countries.

In severe haemophilia, the aim of treatment is to prevent recurrent bleeding into joints and subsequent joint damage. Such 'prophylaxis' is usually started in early childhood using factor VIII concentrates. Prophylaxis requires two to three infusions per week.

For those not on prophylaxis, patients are treated as and when they have a bleed ('on demand'). For surgery, concentrations should be kept to normal until healing has occurred; because factor VIII has a half-life of 12 hours, it is often administered twice daily to maintain the required concentration.

Bioengineered factor VIII products are increasingly replacing standard factor VIII products. They have a half-life approximately 1.5 to 4 times that of base factor VIII, depending on the degree of bioengineering, and can be used either to reduce the frequency of infusion or to improve the concentration of haemostatic protection.

Increasingly, bispecific antibodies that bind factor IXa and factor X to mimic the effect of factor VIII are used for provision of prophylaxis in those with severe and moderate haemophilia. These can be simply administered subcutaneously every 1, 2, or 4 weeks.

Gene therapy trials, using adeno-associated virus AAV to transfer a copy of a functional VIII gene, have recently shown promise in the potential management of severe haemophilia A.

In milder haemophilia, ***synthetic vasopressin*** (desmopressin, an analogue of vasopressin)—intravenous, subcutaneous or intranasal—produces a three- to fivefold rise in factor VIII and VWF concentrations. It is very useful in patients with a baseline concentration of factor VIII above 10 IU/dL. It avoids the complications associated with blood products and is useful for treating and preventing bleeding in mild haemophilia.

People with haemophilia should be registered at a comprehensive care centre, which takes responsibility for their medical care, including social and psychological support.

Complications

Up to 30% of people with severe haemophilia will develop antibodies to factor VIII that inhibit its action. Such inhibitors usually develop after the first few treatment doses of factor VIII. Inhibitors are relatively rare in moderate and mild haemophilia and are often associated with specific molecular defects.

Management of inhibitor patients has historically been very difficult because infused factor VIII is rapidly inactivated. Acute bleeding events require treatment with agents that can bypass factor VIII, such as recombinant factor VIIa or activated prothrombin complex concentrates. Bispecific antibodies have been shown to be highly

Box 38.28 Blood changes in haemophilia A, von Willebrand's disease, and vitamin K deficiency

	Haemophilia A	von Willebrand's disease	Vitamin K deficiency
Bleeding time	Normal	↑	Normal
Prothrombin time	Normal	Normal	↑
Activated partial thromboplastin time	↑+	↑±	↑
Factor VIII:C	↓++	↓	Normal
von Willebrand factor	Normal	↓	Normal

effective at preventing bleeding in haemophilia A with inhibitors because they are independent of factor VIII. They are widely used to provide prophylaxis in inhibitor patients.

In those recently identified as having developed an inhibitor, the long-term aim is to eradicate the inhibitory antibody. This is done using immune tolerance induction strategies, sometimes with additional immunosuppression, and is successful in around 80% of cases.

Although a historical legacy of plasma-derived concentrates, the risk of ***viral transmission*** has now been virtually eliminated. Historically, many patients died as a consequence of HIV and HCV infection, and a considerable number of patients remain who have HIV or HCV infection.

Carrier detection and antenatal diagnosis

Owing to lyonization (random inactivation of one X chromosome) in early embryonic life, some female carriers may have low concentrations of factor VIII, but others have normal concentrations. Definitive carrier detection therefore requires molecular genetic testing and mutation analysis. Antenatal diagnosis may be carried out by molecular analysis of chorionic villus biopsy at 11 to 12 weeks' gestation if selective termination is being considered or by third-trimester amniocentesis if not.

Haemophilia B (Christmas disease)

Haemophilia B is caused by a deficiency of factor IX. The inheritance and clinical features are identical to those of haemophilia A, but the incidence is only about 1 in 30000 males. The half-life of factor IX is longer than that of factor VIII, at 18 hours. Haemophilia B is treated with factor IX concentrates, recombinant factor IX being generally available, and prophylactic doses are given twice a week. Bioengineered factor IX concentrates have a half-life three to five times that of regular factor IX. Desmopressin is ineffective. Gene therapy for severe haemophilia B has offered hope for some time, and recent data have demonstrated achievement of durable near-normal factor IX concentrations using a variant factor IX gene with increased activity.

von Willebrand's disease

VWD is caused by a deficiency or abnormality of VWF (Fig. 38.38). Because VWF has a critical role as an adhesive protein in the platelet–vessel wall interaction, the absence of VWF leads to impaired platelet adhesion to the subendothelium. Reduced VWF concentrations also lead to factor VIII deficiency because factor VIII is not protected from premature degradation. VWF reduction in VWD therefore leads to a dual haemostatic defect.

The *VWF* gene is located on chromosome 12 and numerous mutations have been identified. VWD has been classified into three types:

- ***Type 1*** is partial quantitative deficiency of VWF. Significant type 1 VWD usually has autosomal dominant inheritance.
- ***Type 2*** is caused by a qualitative abnormality of VWF, usually with autosomal dominant inheritance.
- ***Type 3*** is recessively inherited, and patients have virtually complete deficiency of VWF. Parents are often phenotypically normal.

Many subtypes of VWD are described, particularly type 2 variants, which reflect specific qualitative changes in the VWF protein.

Clinical features

These are very variable. Patients with types 1 and 2 usually have relatively mild clinical features. Bleeding occurs after minor trauma or surgery, and epistaxis and menorrhagia often occur. Haemarthroses are rare. Patients with type 3 have more severe bleeding but rarely experience the joint and muscle bleeds seen in those with haemophilia A.

Characteristic laboratory findings are shown in Box 38.28. Classically, VWF activity is measured as its platelet-dependent function in the presence of ristocetin.

Management

Management depends on the severity of the condition and may be similar to that of mild haemophilia, including the use of desmopressin when possible. Although recombinant VWF is available, plasma-derived factor VIII concentrates that contain intact VWF are the mainstay of replacement therapy. These specific products are used to treat patients with bleeding or to cover surgery in patients who require replacement therapy, such as those with type 3 (severe) VWD and those who do not respond adequately to desmopressin. Cryoprecipitate could be used as a source of VWF, but this is avoided if possible because it is not virus inactivated.

Further reading

Berntorp E, Fischer K, Hart DP et al. Haemophilia. *Nat Rev Dis Primers* 2021; 7:45.
Fogarty H, Doherty D, O'Donnell JS. New developments in von Willebrand disease. *Brit J Haem* 2020; 191:329–339.

ACQUIRED COAGULATION DISORDERS

Vitamin K deficiency

Vitamin K is necessary for the γ-carboxylation of glutamic acid residues on coagulation factors II, VII, IX, and X and on proteins C and S. Without it, these factors cannot bind calcium.

Deficiency of vitamin K may be caused by:

- ***Inadequate stores:*** HDN, severe malnutrition, especially when combined with antibiotic treatment
- ***Malabsorption of vitamin K:*** occurs in cholestatic jaundice owing to the lack of intraluminal bile salts preventing absorption of this fat-soluble vitamin
- ***Oral anticoagulant drugs:*** many are vitamin K antagonists.

The PT and APTT are prolonged (Box 38.28), and there may be bruising, haematuria, and GI or cerebral bleeding. Patients with minor bleeding are treated with phytomenadione (vitamin K_1) 10 mg intravenously. Some correction of the PT is usual within 6 hours, but it may not return to normal for 2 days.

Newborn babies have low concentrations of vitamin K and this may cause minor bleeding in the first week of life ***(classical haemorrhagic disease of the newborn)***. Vitamin K deficiency also causes ***late haemorrhagic disease of the newborn*** 2 to 26 weeks after birth and results in severe bleeding such as intracranial haemorrhage. Most infants with these syndromes have been exclusively breast-fed, and both conditions are prevented by administering 1 mg vitamin K intramuscularly to all neonates. Concerns about the safety of this strategy are unfounded.

Liver disease

Liver disease may result in defects in haemostasis:

- ***Vitamin K deficiency*** caused by intrahepatic or extrahepatic cholestasis
- ***Reduced synthesis*** of coagulation factors (including natural anticoagulant proteins) because of severe hepatocellular damage. The use of vitamin K does not improve the results of abnormal

coagulation tests, but it is generally given to ensure that a treatable cause of failure of haemostasis has not been missed
- ***Thrombocytopenia*** caused by hypersplenism associated with portal hypertension or from folic acid deficiency
- ***Functional abnormalities*** of platelets and fibrinogen are found in many patients with liver failure
- ***DIC*** occurs in acute hepatic failure.

Disseminated intravascular coagulation

DIC never occurs in isolation. Recognition that the patient has a clinical disorder that may result in DIC (Box 38.29) is the key to investigation and management. DIC arises because of systemic activation of coagulation either by release of procoagulant material such as tissue factor or via cytokine pathways as part of the inflammatory response. Such systemic activation leads to widespread generation of fibrin and deposition in blood vessels, leading to thrombosis and multiorgan failure. Because of widespread coagulation activation there is consumption of platelets and coagulation factors and secondary activation of fibrinolysis leading to production of FDPs and D-dimer. These further contribute to the coagulation defect by inhibiting fibrin polymerization (Fig. 38.42). The consequences of these changes are a mixture of initial thrombosis followed by a bleeding tendency due to consumption of coagulation factors and dysregulated fibrinolytic activation.

Box 38.29 Causes of disseminated intravascular coagulation

- Malignant disease
- Septicaemia (e.g., Gram-negative, meningococcal)
- Haemolytic transfusion reactions
- Obstetric causes (e.g., placental abruption, amniotic fluid embolism, pre-eclampsia)
- Trauma, burns, surgery
- Infection (e.g., *falciparum* malaria)
- Liver disease
- Snakebite

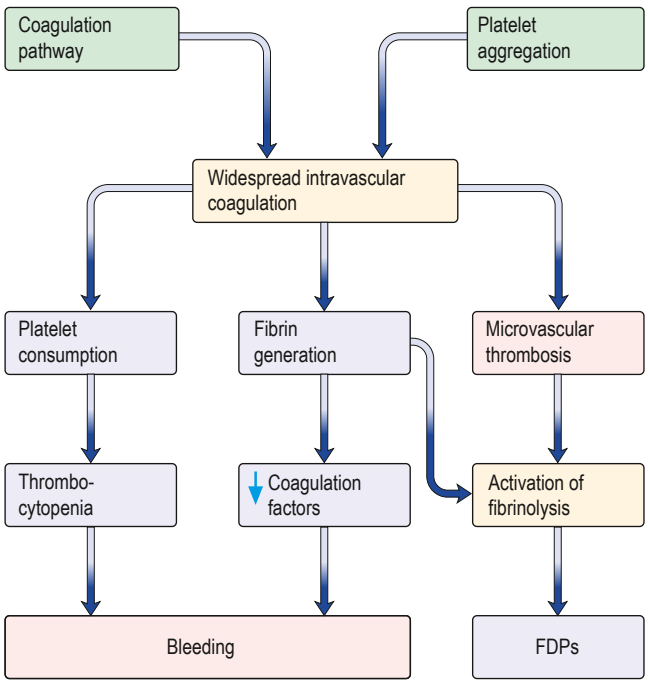


Fig. 38.42 Disseminated intravascular coagulation. *FDP,* Fibrin degradation product.

Clinical features

An underlying disorder is usually obvious. The patient is often acutely ill and shocked. The clinical presentation of DIC varies from no bleeding at all to profound haemostatic failure with widespread haemorrhage. Bleeding may occur from the mouth, nose, and venepuncture sites, and there may be widespread ecchymoses.

Thrombotic events occur as a result of vessel occlusion by fibrin and platelets. Any organ may be involved, but the skin, brain, and kidneys are most often affected.

Investigations

The diagnosis needs to encompass both clinical and laboratory aspects. It is often suggested by the underlying condition of the patient. If the patient has an underlying disorder known to be compatible with overt DIC, the International Society on Thrombosis and Haemostasis scoring system is a useful diagnostic tool and provides an objective assessment (Box 38.30).

In severe cases with haemorrhage:
- The PT, APTT, and TT are usually very prolonged, and the fibrinogen concentration is markedly reduced.
- High concentrations of FDPs, including D-dimer, are found, owing to the intense fibrinolytic activity stimulated by the presence of fibrin in the circulation.
- There is severe thrombocytopenia.
- The blood film may show fragmented RBCs.

Management

The cornerstone of management is treatment of the underlying condition and intensive support to manage hypoxia, acidosis, and organ failure. In those that are not bleeding, this is often all that is necessary. Transfusions of platelet concentrates, FFP, cryoprecipitate, and RBC concentrates are indicated in patients who are bleeding or to cover interventions. Transfusion of blood components based on coagulation tests alone is not required. Inhibitors of fibrinolysis, such as tranexamic acid, should not be used in patients with DIC because dangerous fibrin deposition may result. In cases with a dominant thrombotic component, the cautious use of unfractionated heparin should be considered. In critically ill, non-bleeding patients with DIC, thromboprophylactic doses of heparin are recommended.

Excessive fibrinolysis

Excessive fibrinolysis occurs in certain malignancies, such as acute promyelocytic leukaemia, and during surgery involving tumours of the prostate, breast, pancreas, and uterus owing to release of tissue plasminogen activators.

Primary hyperfibrinolysis is very rare. The clinical picture is like that of DIC, with widespread bleeding. Laboratory investigations are also similar, with a prolonged PT, APTT, and TT; a low fibrinogen concentration; and increased FDPs, although fragmented RBCs and thrombocytopenia are not seen because disseminated coagulation with widespread fibrin deposition is not present. If the diagnosis is certain, fibrinolytic inhibitors, such as tranexamic acid, can be given, but evidence for their efficacy is lacking.

Activation of fibrinolysis occurs in DIC as a secondary event in response to intravascular deposition of fibrin

Box 38.30 International Society on Thrombosis and Haemostasis diagnostic scoring system for disseminated intravascular coagulation

Parameter	Score[a]			
	0	1	2	3
Platelet count ($\times 10^9$/L)	>100	50 to <100	<50	
Elevated fibrin marker (e.g., D-dimer, fibrin degradation products)	No increase		Moderate increase	Strong increase
Prolonged prothrombin time (seconds)	<3	>3 to <6	>6	
Fibrinogen concentration (g/L)	>1	<1		

[a]A total score of ≥5 equals disseminated intravascular coagulation (DIC) if the score is associated with a clinical disorder known to cause DIC.

Massive transfusion

Few platelets and reduced concentrations of clotting factors are found in stored blood, although there are adequate amounts of the other coagulation factors. During massive transfusion (defined as transfusion of a volume of blood equal to the patient's own blood volume within 24 hours, e.g., >10 units in an adult), the platelet count, PT, and APTT should be checked at intervals. Transfusion of platelet concentrates, FFP, and cryoprecipitate should be given if thrombocytopenia or defective coagulation is thought to be contributing to continued blood loss. Other problems associated with transfusion are detailed in **Box 38.19**.

Inhibitors of coagulation

Factor VIII autoantibodies arise occasionally in patients without haemophilia but also with autoimmune disorders such as SLE, in older adult patients, with malignant disease, and sometimes after childbirth. There can be severe bleeding. Patients with immediate bleeding problems are managed with concentrates that bypass factor VIII activity (e.g., recombinant factor VIIa or activated prothrombin complex concentrates). Longer-term therapy aims to eliminate the autoantibody using immunosuppression (e.g., corticosteroids, cyclophosphamide, rituximab).

Lupus anticoagulant antibodies (see Chapter 40) are autoantibodies directed against phospholipids (antiphospholipid antibodies); they cause prolongation of phospholipid-dependent coagulation tests, particularly the APTT, but do not inhibit coagulation factor activity.

Further reading

Maier CL, Brohi K, Curry N et al. Contemporary management of major haemorrhage in critical care. *Intensive Care Med* 2024; 50(3):319–331.
Premkumar M, Sarin SK. Current concepts in coagulation profile in cirrhosis and acute-on-chronic liver failure. *Clin. Liver Dis* 2020; 16:158–167.

Significant websites

https://b-s-h.org.uk/guidelines/ *British Society for Haematology guidelines*
http://www.bloodline.net *General website on haematology*
http://www.hemophilia.org *US National Hemophilia Foundation*
http://www.isth.org/ *International Society on Thrombosis and Haemostasis (ISTH)*
http://www.shotuk.org *Serious Hazards of Transfusion (SHOT) scheme, covering UK and Ireland NHS and private hospitals, affiliated to the Royal College of Pathologists (based at the Manchester Blood Transfusion Centre)*
http://www.transfusionguidelines.org.uk *UK Blood Transfusion and Tissue Transplantation Services Professional Guidelines, including* Handbook of Transfusion Medicine
http://www.transfusion.org Transfusion, *the journal of the American Association of Blood Banks*
http://www.wfh.org *World Federation of Hemophilia*

Questions for this chapter can be found on Elsevier eBooks+.

39 Haematological oncology

Joost de Vos and Vanessa Foggo

Clinical skills 1328
History 1328
Assessing suitability for treatment 1328
Aims of treatment 1329
Background 1329
Aetiology 1330
Pathogenesis 1331
Investigations 1332
Leukaemia 1337
Acute leukaemias 1337
Chronic myeloid leukaemia and the myeloproliferative neoplasms 1343
Other myeloproliferative disorders 1343
Myelodysplasia 1346
Lymphomas 1348
Hodgkin lymphoma 1349
Non-Hodgkin lymphoma (NHL) 1351
B-cell lymphoma 1352
T-cell lymphoma 1356
Myeloma and other plasma cell disorders 1358

CORE SKILLS AND KNOWLEDGE

The haematological malignancies include leukaemias, the myeloproliferative neoplasms (MPNs), lymphomas, and myeloma. They are uncommon but not rare, with lymphoma being the fifth most common cancer in the UK.

Haemato-oncologists divide their time between inpatient care, outpatient clinics, chemotherapy day units, and diagnostic laboratories. In smaller centres, malignant haematology may comprise only one aspect of the work of a haematologist. In larger centres, specialist haematologists focus exclusively on the treatment of patients with malignant disease and are often engaged with clinical trials, medical education, and research.

Inpatient wards accommodate those requiring investigation and those receiving intensive chemotherapy in addition to patients requiring management of the acute complications caused by their disease or its treatment. Specialist units accommodate patients undergoing autologous or allogeneic stem cell transplantation (SCT).

Key to learning objectives for haemato-oncology include:

- Understanding acute leukaemia and the role of SCT
- Understanding lymphomas and myeloma and their different management strategies
- Having an appreciation of the key approaches to treatment with use of cytotoxics, immunotherapy, small molecules, and radiotherapy
- Managing emergencies in haemato-oncology, including leucostasis, hyperviscosity, tumour lysis syndrome (TLS), and spinal cord compression.

Specific opportunities for learning haemato-oncology include attending ward rounds on inpatient haemato-oncology wards and transplant units, spending time at haemato-oncology day units and becoming familiar with how chemotherapy is delivered, meeting patients presenting with haematological malignancy and following their diagnostic work-up, and attending haemato-oncology multidisciplinary team (MDT) meetings to review diagnostic material and discuss management options.

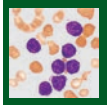
CLINICAL SKILLS

History

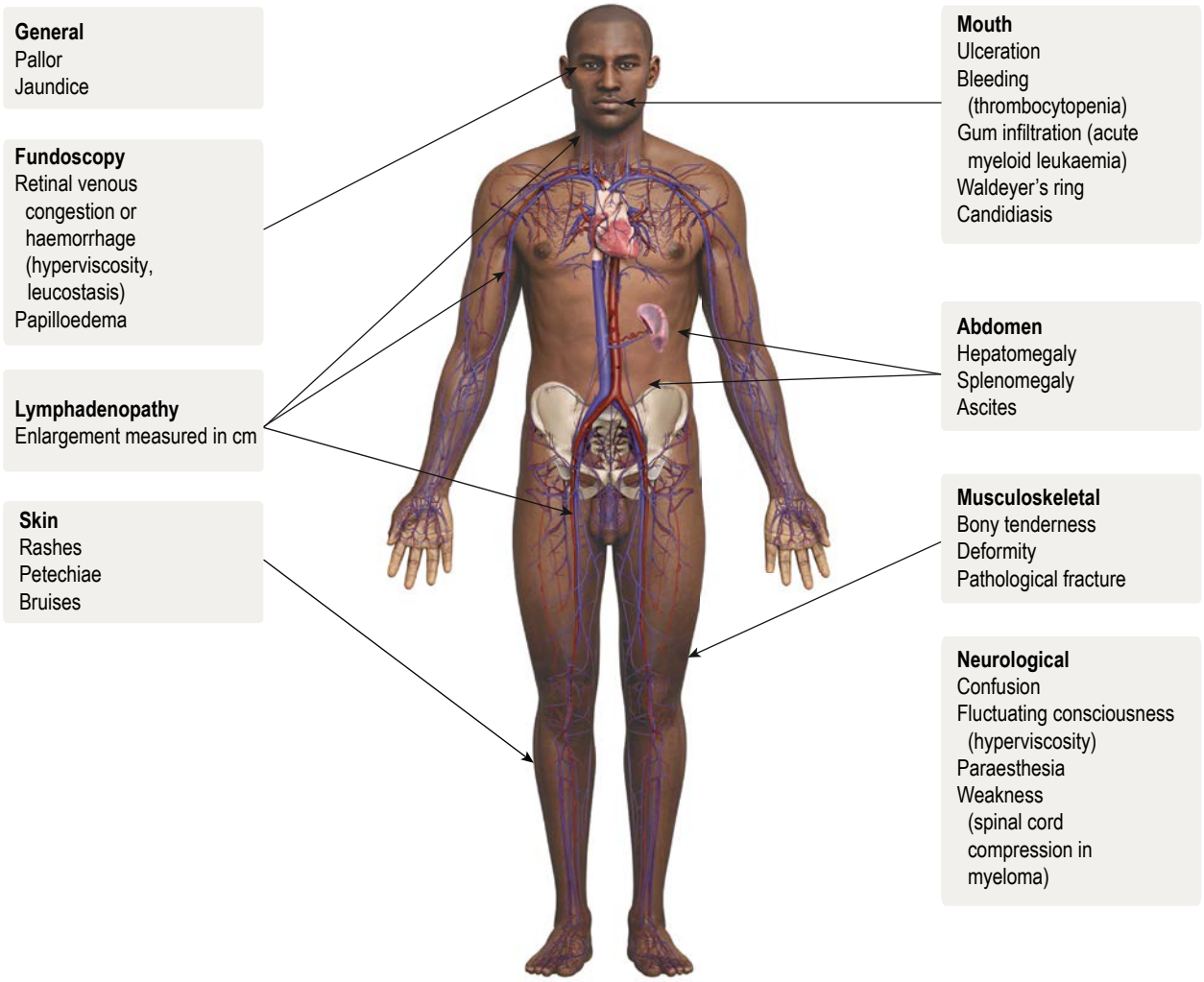
Symptoms of haematological malignancy can be acute or long-standing (Box 39.1). Patients can present with anything from an isolated lump to generalized symptoms. Aim to elicit specific symptoms, their duration, and their impact. Be alert to symptoms that may indicate an emergency presentation such as leukostasis or spinal cord compression. Always be sensitive to the challenges patients face in undergoing investigation and treatment for cancer.

Assessing suitability for treatment

A patient's ***performance status*** is a key determinant of how well they are likely to cope with intensive anticancer treatment, such as chemotherapy or SCT (see Box 41.13). Asking questions about various activities of daily living, including personal care (e.g., washing and dressing) and domestic tasks (e.g., shopping or preparing food), allows patients' performance status to be assessed and monitored over time.



Examination



Treatments

Treatment plans (Box 39.2) are guided by the large volume of clinical trial evidence that exists for all haematological malignancies, and patients can often be offered enrolment into clinical trials of novel agents.

Aims of treatment

Not all treatment intent is curative. In indolent disease, treatments are focused on slowing progression or managing symptoms and complications. Factors affecting treatment include:

- ***Patient factors.*** Performance status or frailty scores are helpful in estimating how well patients may tolerate a particular treatment. Open discussion with the patient is vital. Potential side-effects and complications of treatment must be balanced against the likelihood of a favourable outcome.
- ***Disease factors.*** Better understanding of pathophysiology has given us prognostic markers (e.g., FLT3 [FMS-like tyrosine kinase 3] in acute myeloid [or myeloblastic] leukaemia [AML]), and data gathered from large-cohort analyses inform scoring systems.
- ***Treatment factors.*** The expense of newer therapies results in restricted prescribing in most healthcare systems. Enrolment in clinical trials gives access before efficacy and cost-effectiveness have been completely evaluated.

BACKGROUND

Haematological malignancies cover a wide range of blood cancers, which can present in many ways. They can present with either ***circulating*** ('blood-based', 'leukaemic', or 'liquid') or ***solid*** disease, though often the clinical presentation is non-specific. Their diversity and nomenclature can be confusing. The updated 2022 World Health Organization (WHO) classification contains an increasing number of diagnostic entities, now well over 150. There is an increasing emphasis on underlying genetic and molecular abnormalities, as well as histology and immunophenotype. It is beyond the needs of generalists (and most haematologists) to be aware of all these conditions; however, some principles of clinical behaviour, diagnosis, as well as prognosis and treatment are helpful for all clinicians. They are discussed here.

Box 39.1 Taking a history in haemato-oncology

Symptoms of bone marrow failure
- Anaemia: fatigue, shortness of breath, reduced exercise capacity
- Thrombocytopenia: easy bruising or bleeding
- Leucopenia: recurrent infection

Symptoms caused by mass effect of malignant tissue
- Abdominal discomfort from splenic enlargement
- Neck, jaw, axillary, or groin swelling from enlarged lymph nodes

Systemic symptoms
- Malaise
- Fever: >38°C
- Sweats: particularly at night
- Weight loss: >10% unintentional weight loss over 6 months is considered significant

Family history
- Haematological disease or malignancy
- If allogeneic transplantation may be considered: health of potential donors

Risk factors
- Previous exposure to cytotoxins
- Previous exposure to ionizing radiation
- Viral infection

Fitness for treatment
- Medical co-morbidity that may influence treatment approach
- Performance status (see later)
- Impact on the individual
- Detailed social history to assess the implications of diagnosis and treatment

Box 39.2 Treatments in haematological malignancy

Cytotoxics
- Target rapidly growing cells
- Historically formed the basis of conventional anticancer chemotherapy and remain the backbone of most treatment regimens
- Classes: alkylating agents (e.g., cyclophosphamide), anthracyclines (e.g., doxorubicin), vinca alkaloids (e.g., vincristine), purine analogues (e.g., mercaptopurine, 6-MP)

Immunotherapy
- Monoclonal antibodies target a protein on the surface of the malignant cell (e.g., rituximab targets the CD20 molecule on B lymphocytes and triggers cell death). It is one of the most common monoclonal antibodies in haematological malignancy treatment.
- Drug–antibody conjugate: Antibodies target the tumour cell and act as a vector for toxins.
- Bispecific antibody therapies bind to an antigen on the tumour cell surface and also to T cells, linking the two cells and triggering cytotoxic activity.
- Chimeric antigen receptor T-cell therapy: modification of the patient's own T cells to express a receptor that will bind a particular tumour protein. The cells are expanded and re-infused to attack and destroy malignant cells.

Corticosteroid
- Synthetic high-dose corticosteroids (e.g., dexamethasone) induce cell death in malignant lymphoid cells.
- Treatment role alongside other anticancer treatments in all B-cell malignancies, including myeloma.

Molecular therapy
- Treatments targeting specific pathways:
  - FLT3 inhibitors (e.g., midostaurin) block abnormal cell signalling in acute myeloid leukaemia
  - Venetoclax blocks the mutated anti-apoptotic protein Bcl-2 in chronic lymphocytic leukaemia

Stem cell transplantation (SCT)
- Transfer of pluripotent stem cells from either the patient (autologous) or a donor (allogeneic) after myeloablative chemo- or radiotherapy.
- Generally performed after remission has been achieved with preceding chemotherapy regimens. An SCT will not treat the disease but is done to prevent disease recurrence.
- Transplanted cells permit bone marrow recovery.
- In allogeneic SCT, the patient's abnormal bone marrow is replaced with an alternative, normally functioning, haemopoietic system.

Radiotherapy
- Targets solid tumour masses (e.g., enlarged lymph nodes in early-stage lymphomas, plasmacytomas, and residual disease after other treatments).
- Surgery
- Although of limited use in treatment of patients with these conditions, surgery has a role, though rare, in obtaining diagnostic tissue through biopsy when image-guided biopsy is not feasible.
- Neurosurgery may relieve spinal cord compression, though steroids and radiotherapy are more frequently indicated.
- Orthopaedic fixation of pathological fractures may be required in patients with myeloma.

Aetiology

Like most cancers, haematological malignancies are mostly caused by the ***acquisition of genetic changes*** within a normal cell, leading to development of an abnormal population of cells. Most of these events are random, without a clear cause, although incidence generally ***increases with age***.

These changes may be large-scale chromosomal abnormalities (translocations, deletions, or duplications), which can be detected by cytogenetic analysis, or point mutations within a specific gene, which require DNA sequencing technology for analysis (Fig. 39.1). More and more cytogenetic and molecular changes are being identified. As highlighted in the 2022 WHO and European LeukemiaNet (ELN) 2022 guidelines, these changes can be diagnosis-defining and are continuously re-shaping the diagnosis, treatment, and prognosis of these malignancies.

Not all mutations necessarily lead to a malignancy. There is an increasing awareness of the accumulation of mutations as part of the normal ageing process. It can be a certain combination of changes, or *gene profile*, that leads to a haematological malignancy. It is important to know that a single finding is often not diagnostic of a type of malignancy because it can be found in people without cancer; interpretation of results should be done with caution, experience, and a knowledge of the clinical picture. Some changes are not diagnostic but carry a predisposition or risk of the patient's developing a certain haematological malignancy. Careful monitoring or intervention where appropriate might be indicated. For example, the t(14;18) translocation associated with follicular lymphoma (FL) can be found in up to 50% of healthy individuals' circulating lymphocytes.

Previous ***radiation exposure*** (usually as part of radiotherapy treatment for a previous cancer) and ***cytotoxic chemotherapy*** can cause genetic abnormalities and subsequent malignancy. This is seen most in myeloid malignancy, such as myelodysplastic syndrome (MDS) and AML, with the development of mutations that have a possible causal link with previous radiation or chemotherapy. However, the term 'therapy-related' MDS or AML is no longer being used and has been removed from classifications because these mutations can also be seen in patients without prior exposure.

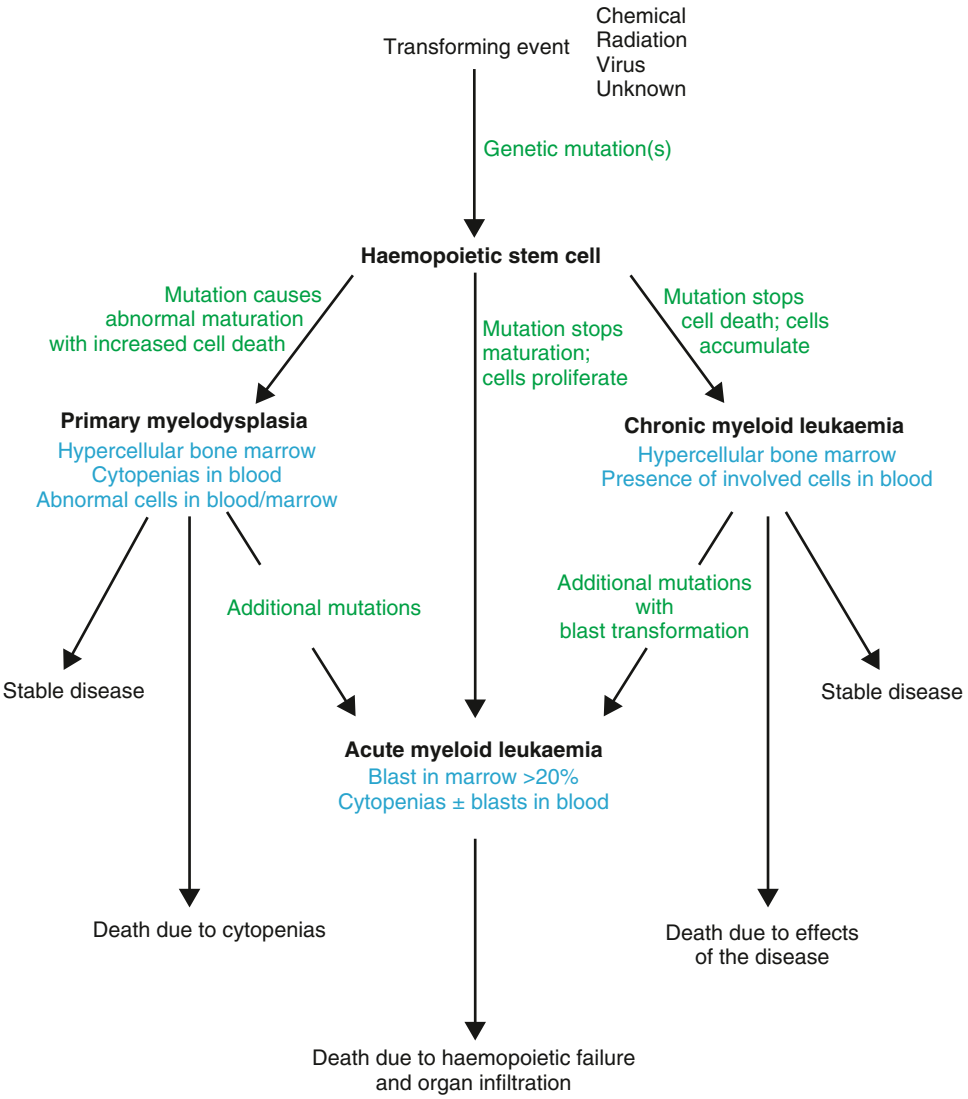


Fig. 39.1 Acquisition of clonal changes. *(Adapted from http://eclinpath.com/hematology/leukemia/transforming-aligorithm.)*

Viral infections are implicated as a possible cause of or contributor to the development of several haematological cancers. For example:

- Human T-cell leukaemia (or lymphotropic) virus type 1 (HTLV-1) in adult T-cell leukaemia/lymphoma (ATLL): Integration of viral DNA into regulatory T cells leads to malignant transformation.
- Epstein–Barr virus (EBV) plays a role in the aetiology of some lymphomas, including Hodgkin lymphoma (HL).
- Human immunodeficiency virus (HIV) infection predisposes to non-Hodgkin lymphoma (NHL) and HL development if not treated. This occurs primarily through immunosuppression, taking away essential normal immune-system control mechanisms in the development of abnormal haematopoietic cells, as well as risk of co-infection with other viruses such as EBV and human herpesvirus-8.

Immunosuppression caused by long-term use of systemic therapies can increase the risk of lymphoma, like that described in patients with HIV. This is most strikingly seen in post-transplant lymphoproliferative disorder, which can occur in patients receiving long-term immune-suppressive therapy after kidney or other transplants. In most of these cases, EBV infection or reactivation plays a significant role.

Congenital conditions that predispose patients to the development of haematological malignancy are rare, though possibly not as rare as previously thought. A small number of conditions do exist in which a germline genetic disorder increases the risk of leukaemias developing. These include Down's syndrome (trisomy 21), which can be associated with a transient myeloproliferative disorder in the neonatal period as well as an increased risk of both AML and acute lymphoblastic leukaemia (ALL) as well as and bone marrow failure syndromes such as Fanconi anaemia (see Chapter 38, 'Anaemia caused by marrow failure [aplastic anaemia]' section). In the myeloid malignancies MDS and AML, constitutional or germline mutations are increasingly identified and should be considered especially in patients in which a family history exists. If such a mutation is found, it can be challenging to determine who in a family, if any, might benefit from screening. It also has an understandable impact on donor selection when SCT is considered. These mutations remain very rare.

Pathogenesis

Haematological malignancies can be characterized based on the ***cell of origin***, the normal counterpart cell within the haemopoietic

Box 39.3 Classification of leukaemia and lymphoma

Timeframe	Myeloid (bone marrow)	Lymphoid (bone marrow)	Lymphoid (node)
Acute			
Immature cells predominate	Acute myeloid leukaemia	Acute lymphoblastic leukaemia	Lymphoma: 'high-grade', 'aggressive' (e.g., Burkitt lymphoma)
	Myelodysplastic syndrome	Myeloma	
Chronic			
Mature cells predominate	Chronic myeloid leukaemia Other myeloproliferative disorders	Chronic lymphocytic leukaemia	Lymphoma: 'low-grade', 'indolent' (e.g., follicular lymphoma)

system. For example, AML is derived from a haemopoietic stem cell (HSC) that has acquired mutations, leading to rapid proliferation of an abnormal, or ***malignant***, cell population. These are immature cells ***(blasts)*** that share some morphological and immunophenotypic appearances with normal myeloid precursor cells in the bone marrow, but their unchecked proliferation and inability to undergo differentiation lead to a neoplastic population. In contrast, in myeloma, the cell of origin is the plasma cell, a mature cell derived from a post-germinal centre B cell, whose normal function in producing antibodies is reflected in their malignant counterpart, in which an abnormal monoclonal paraprotein or light chain production is seen.

Haematological cancers can be broadly divided between ***myeloid*** and ***lymphoid*** lineages. Myeloid disorders arise from lineages within the marrow that produce granulocytes, red blood cells (RBCs), or platelets. Lymphoid disorders can arise from either B- or T-cell lineages. This classification is summarized in **Box 39.3**, and the specific subdivision of the lymphoid malignancies is illustrated in **Fig. 39.2**. This division may not, however, accurately represent or predict the cell of origin: for example, chronic myeloid leukaemia (CML) can undergo a myeloid transformation ***(blast crisis)*** into AML but can also transform into an ALL-like disease, reflecting the fact that the CML clone is derived from an HSC with both myeloid and lymphoid differentiation potential.

Early genetic changes

The number, type and sequence, and combination of acquired genetic mutations determines the phenotype of the condition. CML, for example, has a clear initiating event (the ***Philadelphia chromosome*** [Ph+] translocation) that defines the condition and has led to effective targeted therapy. Other malignancies have a wider variety of driver mutations, producing a more heterogeneous group of conditions with variable responses to different treatments and variable prognostic significance.

Cellular phenotypes

The pattern of mutations can lead to a ***dysplastic*** (abnormal appearance and function) or ***proliferative*** phenotype. The stem cell defect in MDS, for example, leads to dysplastic cells with a consequent reduction in normal haemopoiesis. AML, in contrast, is characterized by the rapid proliferation of large numbers of immature cells, which can move out of the bone marrow into the peripheral blood, causing a leukocytosis as well as failure of normal haemopoiesis often caused by extensive bone marrow infiltration.

Haematological malignancies can also be divided, based on the speed of disease evolution, into ***acute*** or ***chronic*** conditions. Whereas the most acute disorders (e.g., AML, Burkitt lymphoma) are rapidly progressive and invariably fatal within days to weeks if not treated, chronic disorders (e.g., CLL, FL) are typically indolent and slowly progressive with patients living with their conditions for long periods of time, often not even requiring treatment at diagnosis. In contrast to acute conditions, however, chronic malignancies are counter-intuitively less likely to be cured with standard treatments.

New mutational events

Additional mutational events can occur within the malignant populations, causing changes to the disease's behaviour ***(clonal evolution)***, with a subset of cells becoming dominant and leading to disease progression (**Fig. 39.3**). This can occur because of the 'selection pressure' of drug treatment, with resistant clones emerging over time, and this explains, for example, the loss of response to a drug after a period of remission. A sudden change in behaviour from a chronic to acute condition is often termed a ***high-grade transformation***, with examples including Richter's transformation of CLL to a high-grade lymphoma or blast crisis of CML to AML or ALL. High-grade transformation carries a poorer prognosis.

Leukaemic and lymphomatous presentations

A final division can be made between ***circulating*** or ***solid*** presentations, depending on whether there is primarily a marrow-based leukaemic presentation (with or without obvious circulating disease in the peripheral blood) or a nodal or extranodal lymphomatous presentation (in which soft tissue masses predominate). Often, though, these presentations are variations of the same disease. This can lead to confusing nomenclature for patients and clinicians alike.

Investigations

Full blood count

The full blood count (FBC) is the most fundamental test used to pick up most haematological cancers. This is an automated test available in all hospitals.

The basic results produced by an automated FBC are:

- Haemoglobin (Hb) concentration with RBC markers, mean corpuscular volume, and mean corpuscular haemoglobin
- White blood cell count (WCC) and proportions of cells in the different white blood cell (WBC) populations ***('differential')***
- Platelet count.

When an unexpected significant abnormal finding is detected, in either cell count or differential, a flag should be raised to laboratory staff. Often further investigation with a manual repeat with blood film (see later) is recommended. Increasingly, FBC interpretation is done automatically based on pattern recognition, though this should be interpreted with some caution. As with all laboratory tests, the FBC

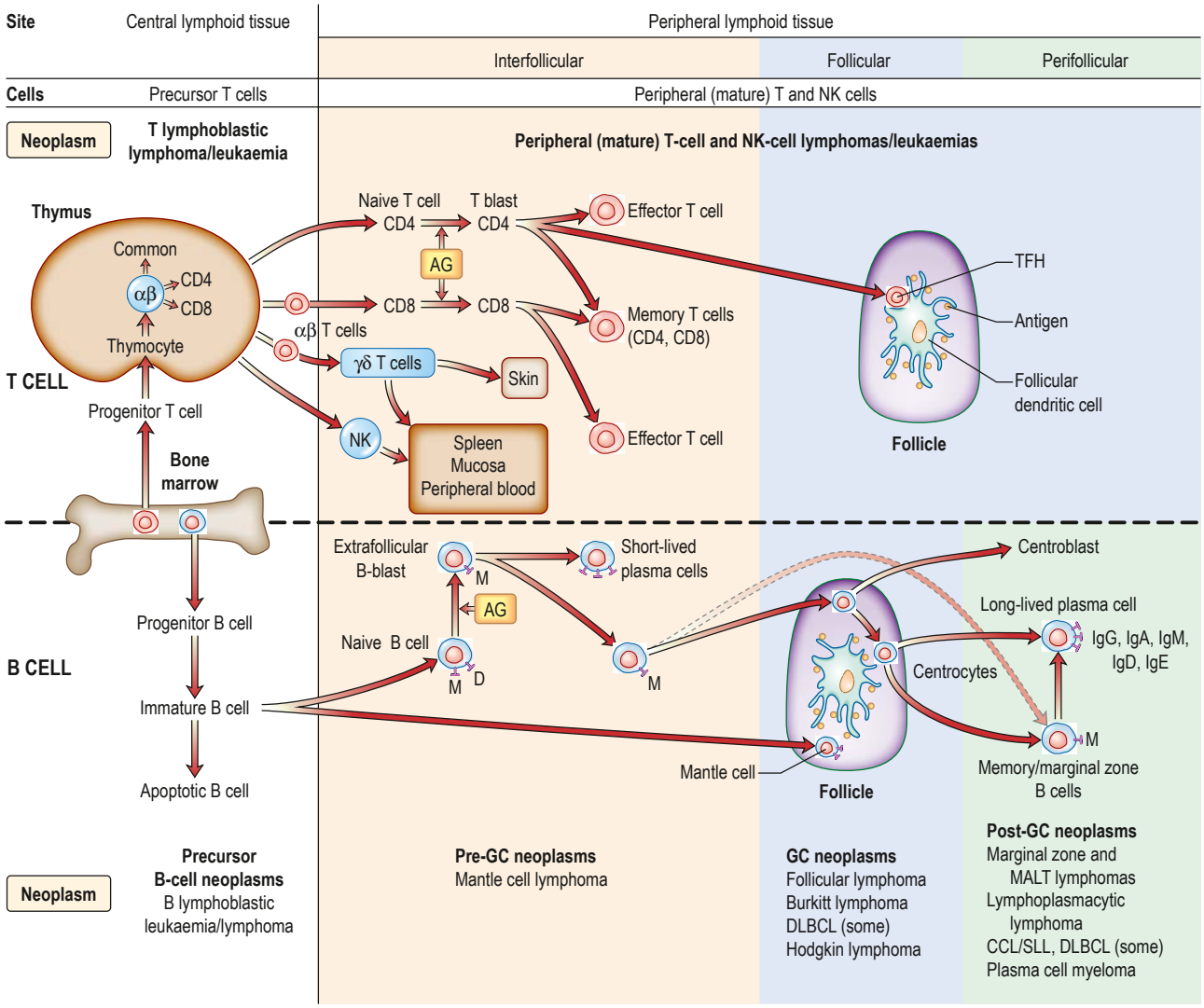


Fig. 39.2 Differentiation of T and B lymphocytes and their relationship to neoplasms. ***T-cell differentiation (above dotted line):*** Progenitor T cells from the bone marrow enter the thymus and develop into different naive T cells. αβ T cells leave the thymus, are exposed to antigen (AG), and undergo blast transformation. They then develop into CD4+ and CD8+ effector and memory T cells. CD4+ effectors cells include T regulatory cells and follicular helper T cells (TFHs) in the germinal centres (GCs). On antigenic stimulation, T-cell responses to antigenic stimulation pathways of natural killer cells (NK) and γδ T cells are unknown. ***B-cell differentiation (below dotted line):*** Precursor B cells mature in the bone marrow and undergo apoptosis or mature to naive B cells. After exposure to antigen and blast transformation, they develop into short-lived plasma cells or enter the GC. Somatic hypermutation and heavy chain class switching occur here (not shown). The transformed cells of the GC (centroblasts) undergo apoptosis or develop into centrocytes. Post-GC cells include long-lived plasma cells and memory or marginal cells. *CLL/SLL,* Chronic lymphocytic leukaemia/small lymphocytic lymphoma; *DLBCL,* diffuse large B-cell lymphoma; *Ig,* immunoglobulin; *MALT,* mucosa-associated lymphoid tissue. (*Redrawn from information in Swerdlow SH, Campo E, Harris NL et al. [eds].* WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues, *4th edn. Geneva: WHO; 2008.*)

is subject to both internal and external quality assurance testing to ensure that all laboratories report results accurately and precisely.

Morphology

Blood film

Blood cells can be easily viewed by spreading a droplet of blood into a film on a microscope slide and staining it to reveal the appearance and number of the cells (Fig. 39.4). Certain conditions have morphological features that are characteristic (e.g., Auer rods in AML cells), and although definitive diagnosis usually requires additional tests (see later), expertise at blood film examination remains a key skill for all haematologists involved in diagnosis. The initial diagnosis of acute leukaemia, of whichever lineage, mainly in patients with higher WCCs, is made on blood film. This is a medical emergency even though patients can be clinically well.

Bone marrow aspirate

Although many leukaemias are visible in the peripheral blood, other blood cancers require examination of the cells within the bone marrow to make a formal diagnosis. Slide preparation and staining techniques

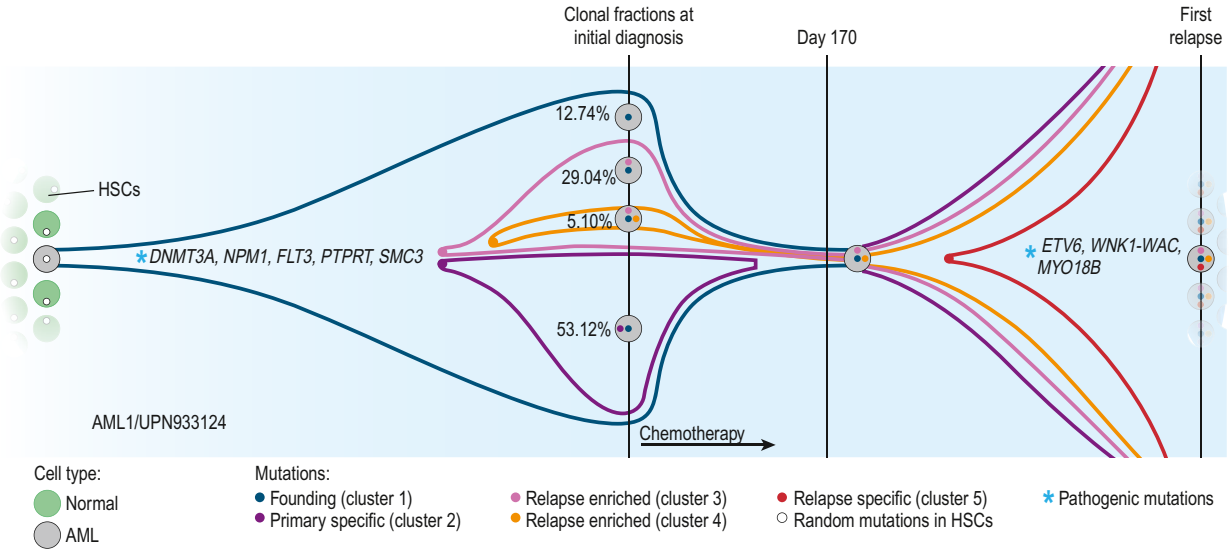


Fig. 39.3 Clonal evolution in a population of leukaemic cells from a patient with acute myeloid leukaemia (AML). The founding clone in the primary tumour contained somatic mutations in the genes *DNMT3A, NPM1, PTPRT, SMC3,* and *FLT3* that are all recurrent in AML and probably relevant for pathogenesis. One subclone within the founding clone evolved to become the dominant clone at relapse by acquiring additional mutations, including recurrent mutations in *ETV6* and *MYO18B* and a *WNK1-WAC* fusion gene. *HSC,* Haemopoietic stem cell. (*From Ding L, Ley TJ et al. Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing*. Nature *2012;481:506–510.*)

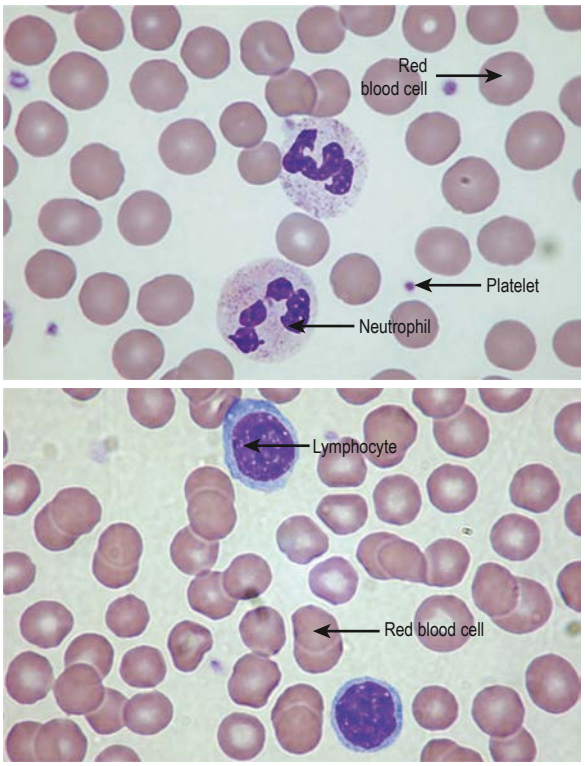


Fig. 39.4 Normal blood films.

Fig. 39.5 Normal bone marrow trephine.

are similar. The normal marrow aspirate shows a wide range of precursor cells at various stages of development.

Bone marrow trephine (trephine)

Histological examination can be performed on a bone marrow biopsy or trephine (from the Greek 'trypanon', meaning 'an instrument for boring'), which can provide an assessment of cellularity and fibrosis, tissue structure, and the presence and character of any malignant population (Fig. 39.5). Immunohistochemical staining can provide a specific diagnosis. The same staining techniques are also used on lymph node and other biopsies (e.g., in lymphomas). Time to results can vary because initial stains focus on the more likely or common diagnosis first, with stains added if no diagnosis is made. It is not that rare that repeat biopsy is required. This is essential if an accurate diagnosis cannot be made on the initial available material. This is especially seen with core biopsies, including endoscopic biopsies, in which the tissue samples obtained are small. It is for this reason that excision biopsy is recommended when possible.

Biochemistry

Routine biochemical testing is used to assess organ damage and tumour lysis. Specific tests (e.g., lactate dehydrogenase [LDH], beta-2-microglobulin [B2M]) are used in staging systems in myeloma and lymphoma.

Immunological assessment of paraproteins and serum free light chains is an essential part of the diagnosis and response assessment of myeloma and some other blood cancers (see later).

Immunophenotyping

Cell surface proteins act as markers that can be used to characterize and quantify malignant cell populations. ***Immunohistochemistry*** of solid samples (e.g., bone marrow trephines, lymph node biopsies) uses the same markers

Flow cytometry is performed on liquid samples (e.g., peripheral blood, bone marrow aspirate, cerebrospinal fluid [CSF]. In this process, cells are stained with fluorescently labelled antibodies and analyzed with specially designed analysers. The combination of cell size and appearance, together with the pattern of surface markers, identifies a profile or ***immunophenotype*** that is characteristic of a particular diagnosis (Fig. 39.6).

Immunophenotypic identification of malignant cell populations is an essential part of the diagnosis of almost all haematological malignancies. As an example, acute leukaemia cells may appear identical by morphology but have very different and specific immunophenotype profiles. It is the immunophenotype report, or ***flow***, that confirms or excludes a diagnosis; it is required before starting any treatment, except in extreme emergencies (Fig. 39.7). These results have a direct impact on treatment, different chemotherapy combinations being used for the different conditions. The markers identified can also in some cases be used to assess response to treatment or for monitoring.

Cytogenetics

Cancer is characterized by genetic abnormalities within the malignant cell populations, and these abnormalities can increasingly be detected in many haematological cancers, either to confirm a diagnosis or to act as biomarkers that can provide information on the likely timescale and behaviour of the disease ***(prognostic)*** or the likely response to a specific drug or treatment ***(predictive)***. Chromosomal abnormalities (translocations, insertions, deletions, inversions) may be detected by two major cytogenetic techniques:

- ***G-band karyotyping*** detects large structural abnormalities in chromosomes when they are condensed during metaphase. This technique detects the t(9;22) Ph+ translocation in CML (see Fig. 5.24) and many of the major structural abnormalities associated with AML prognosis (e.g., inv16, monosomy 7).
- ***Fluorescence in situ hybridization (FISH)*** assesses cells during interphase, so it is useful for slower-growing cancers; however, it only detects specific abnormalities using pre-designed fluorescently labelled probes. This technique is commonly used for detecting prognostically significant abnormalities in myeloma (e.g., t(4;14), 17p deletion) (Fig. 39.8).

Molecular genetics

Abnormalities in specific genes can be detected by DNA sequencing technologies, again providing important prognostic and predictive information. Gene mutations that are known to be of relevance may be specifically screened for. In CML, the T315I mutation in the *BCR-ABL* fusion gene, for example, is associated with poor response to standard therapy and may prompt a change to a specific tyrosine kinase inhibitor (TKI) with known activity in the presence of this mutation.

Use of high-throughput 'next-generation sequencing' (NGS) technologies allows for the investigation of several genes of interest in a single assay. These are increasingly being used as the number of known relevant mutations grows and advances in technology and bio-informatic analysis make this approach more cost-effective. As an example, *TET2*, *SF3B1*, and *ASXL1* are commonly mutated in MDS, and their detection can be helpful in making a diagnosis in unclear cases, as well as informing prognosis. With the development of novel drugs that have activity against specific mutations, the use of gene panels is likely to increase.

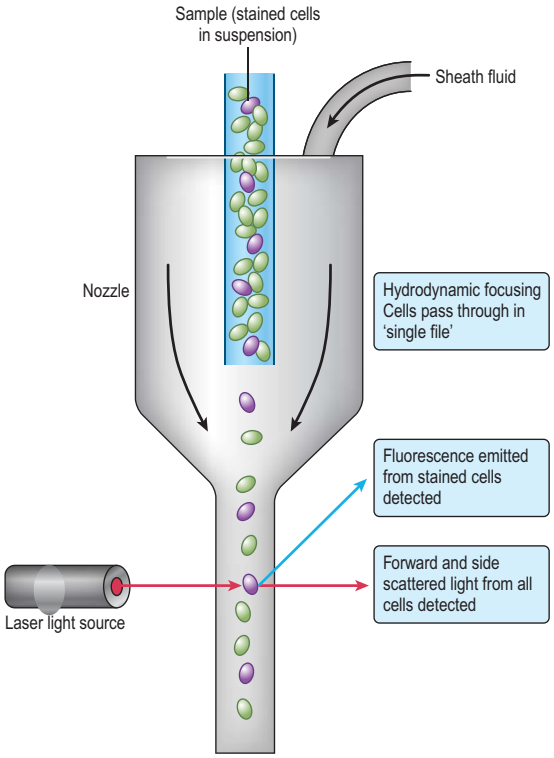


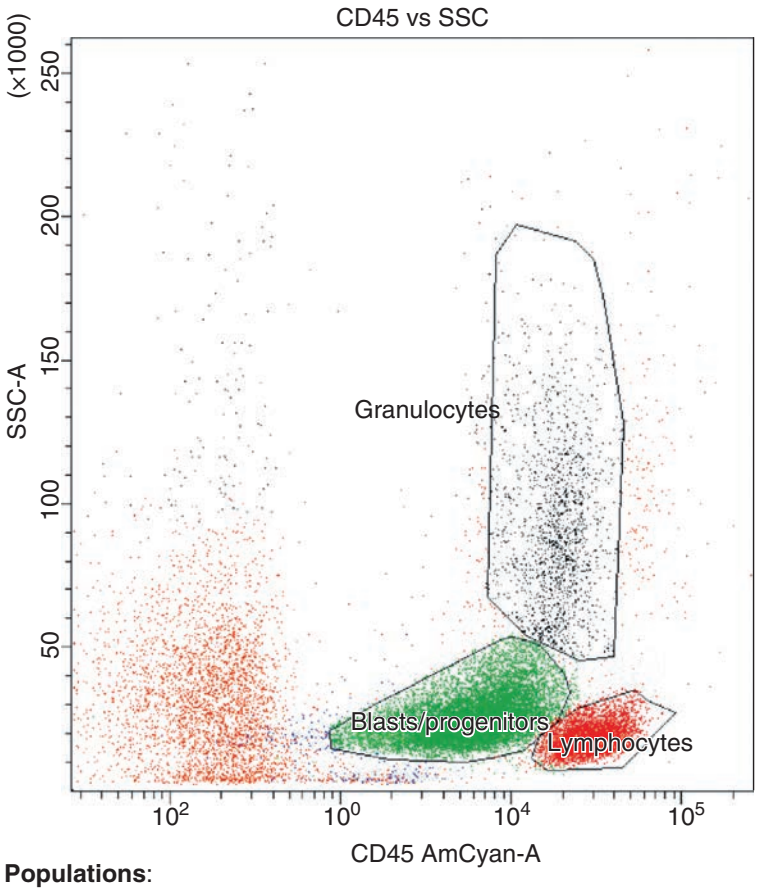
Fig. 39.6 Flow cytometer showing how cells are stained and then characterized with a laser light source and detectors. This method can produce detailed information on cell size, composition, and surface markers used to characterize both normal and malignant cell populations.

Minimal or measurable residual disease testing

Historically, the term '***minimal*** residual disease' was used. However, more appropriate is the term '***measurable*** residual disease', which is increasingly used. The presence of residual disease is far from a minimal finding because it has a significant prognostic importance and often a direct impact on treatment. It is for this reason that the authors support the use of '***measurable*** residual disease'. Flow cytometry and molecular technologies are increasingly being used to detect very low levels of malignant cells in so-called measurable residual disease (MRD) assays. An MRD-positive result can predict relapse before it is visible by morphology and may prompt an early change in treatment. Conversely, an MRD-negative result can provide confidence for de-escalating treatment or for not pursing further intensive treatments such as allogeneic transplant.

Two major methods are used for MRD testing:

- ***Flow cytometry.*** This uses the same methodology as immunophenotyping to detect low levels of malignant cells in the peripheral blood. Ideally, a sample at diagnosis is used to identify the distinctive immunophenotypic profile of the malignant cell population. This can then be used to look for persistence or early relapse of the same cell population.
- ***Quantitative polymerase chain reaction.*** This technique allows the amplification of a specific sequence of DNA or RNA,



Populations:
- Red: normal lymphocytes
- Black: granulocytes
- Green: progenitors (blasts)

Fig. 39.7 Flow cytometry plot from a patient with B lymphoblastic leukaemia. *SSC,* Side scatter. *(Courtesy of Timothy Farren, Barts Health NHS Trust, London, UK.)*

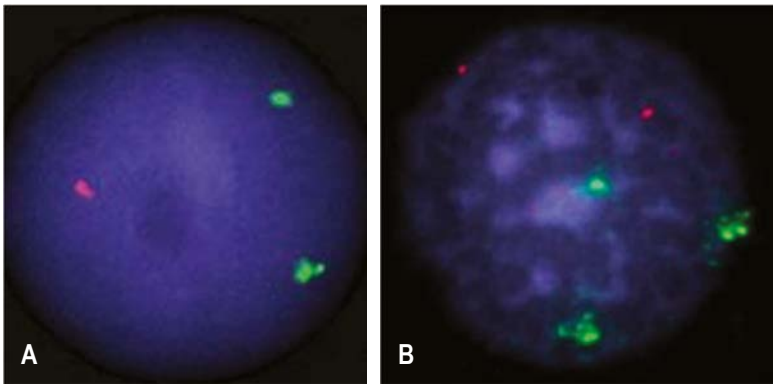


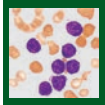
Fig. 39.8 Fluorescence *in situ* hybridization photomicrographs of a patient with chronic lymphocytic leukaemia. (A) 17p *(green probe)* shows two green signals (*TP53* deletion) and 11q *(red probe)* shows normal diploid complement of 11q. **(B)** Twelve centromere *(green probe)* shows three green signals (trisomy 12), and 13q14 *(red probe)* shows normal diploid complement of 13q. (*Courtesy of Debra Lillington, Barts and the London NHS Trust*.)

and it can be used to detect low levels of disease in several conditions:

- ***Gene fusions*** (e.g., *BCR-ABL* in CML)
- ***Immunoglobulin gene rearrangements*** in B-cell ALL
- ***T-cell receptor gene rearrangements*** in T-cell ALL

Tissue typing

For patients who may go on to allogeneic SCT, it is important to collect a blood sample for tissue typing at diagnosis so that potential donors may be identified. Testing identifies donors who match the patient's human leucocyte antigen (HLA) type and is done by



genetic testing, ideally at the HLA-A, B, C, DP, DQ, and DR loci. Testing packs can be sent out to siblings or potential haplo donors such as parents and children, to be returned for testing at the regional laboratory. If no related donors are identified, a search can be performed on national and international donor registries.

Radiological imaging

Although a tissue sample is required to make a diagnosis and commence treatment, radiological imaging plays a key role in the staging, response assessment, and investigation of haematological cancers, mainly the lymphomas and myeloma.

- Plain films are not routinely used for formal staging but may be helpful in initial investigation and assessment of complications (e.g., long bone fractures as a presenting symptom of myeloma).
- Ultrasonography assesses soft tissue masses (e.g., characterization of enlarged lymph nodes). Ultrasound-guided biopsies are a common method of making an initial lymphoma diagnosis, though excision biopsy is generally preferred.
- Computed tomography (CT) provide accurate imaging to stage tumours at diagnosis and assess response. Low-dose, non-contrast CT scans may be sufficient in myeloma, in which the main site of organ damage is the bones. Assessment of soft tissue masses (e.g., lymphoma) requires contrast. CT-guided biopsies are an effective method of obtaining tissue from deeper-lying areas. Positron emission tomography (PET) scans in combination with CT scans have become the standard in many centres.
- Nuclear medicine scans, such as fluorodeoxyglucose positron emission tomography (FGD-PET), are increasingly used in combination with CT (CT-PET) to visualize metabolically active disease. This technique is helpful for assessing response after therapy and to distinguish active residual tumour from fibrotic tissue after treatment.
- Magnetic resonance imaging (MRI) provides detailed views of organs and has benefits in a number of specific situations (e.g., bone marrow, brain, spinal cord). Diffusion-weighted MRI can provide functional information and is increasingly being used in myeloma assessment.

Specialist integrated haematological malignancy diagnostic service and multidisciplinary team

Several different techniques are used in establishing a diagnosis of a haematological cancer, as well as in providing relevant staging and prognostic information to guide treatment. Diagnostic reports can be integrated into a single document, incorporating all relevant pathological information. This is often carried out as part of a specialist integrated haematological malignancy diagnostic service, which may be run on a regional basis.

Review of all diagnostic information, choice of treatment, response assessment, and disease progression should (when possible) be done in the setting of an MDT. Emergency treatment should not be delayed and should be initiated as soon as possible with retrospective MDT discussion.

Further reading

Bain BJ. *Blood Cells: A Practical Guide*, 5th edn. Oxford: Wiley–Blackwell.

LEUKAEMIA

There are four distinct subtypes of leukaemia: AML, ALL, CML, and chronic lymphocytic leukaemia (CLL).

The leukaemias are relatively rare diseases; the different subtypes have distinct diagnostic features. Though they can occur at any age, they tend toward an age distribution: ALL is mainly seen in childhood, but the median age at presentation for AML is 65 years; CML is seen across the ages, but CLL is almost exclusively a disease of older adults.

The diagnosis of leukaemia is remarkably straightforward on a routine blood film examination, available in every hospital, though confirmation of the diagnosis and subtype requires immunophenotyping, cytogenetics, molecular genetics, and a bone marrow investigation. The diagnosis is almost invariably unexpected and will worry both the clinician and patient. It is essential the patient is informed about the results in a timely manner.

Aetiology

In the majority of patients, the cause is unknown; however, several factors have been associated:

- ***Radiation.*** This can induce genetic damage to haemopoietic precursors; ALL, AML and CML have increased incidences in survivors of Hiroshima, Nagasaki, and Chernobyl and in patients treated with ionizing radiation.
- ***Chemicals and drugs.*** Exposure to benzene may lead to bone marrow damage. AML can also occur after treatment with chemotherapy drugs such as alkylating agents (e.g., melphalan) and topoisomerase-II inhibitors (e.g., etoposide). Historically, the term 'secondary AML' was used for AML after prior chemotherapy, though this entity is no longer present in the 2022 leukaemia classification.
- ***Genetic factors.*** Leukaemia risk is elevated in several germline conditions that result in genetic instability or bone marrow failure. These include Fanconi anaemia, ataxia telangiectasia, and Li–Fraumeni syndrome. The risk is also elevated some 30 times in Down's syndrome (trisomy 21), in which there is also a specific subtype in newborns. There is a high degree of concordance among monozygotic twins. Several genes have been associated with familial AML, including *CEBPA* and *RUNX1.* This remains an area of ongoing research.
- ***Viruses.*** ATLL is a rare type of leukaemia associated with HTLV-1, found particularly in Japan and the Caribbean. No other viral links are known.
- Lifestyle factors are not implicated.

Acute leukaemias

AML has a median age at presentation of 65 years and may arise *de novo* or against a background of myelodysplasia (MDS). The different subtypes within AML are increasingly defined by their cytogenetic and molecular findings, which in many cases carry clear prognostic value, as well as being potential treatment-targets.

ALL has a substantially lower median age at presentation; it is the most common malignancy in childhood. The genetic findings in ALL vary greatly between age groups, which explains the improved prognosis in children compared with the ongoing poor prognosis in adults.

A small number of acute leukaemias have an indeterminate immunophenotype between myeloid and lymphoblastic, labelled 'mixed phenotype leukaemia'. Genetic, molecular, histological, and clinical features determine the best treatment strategy.

Clinical features

Most patients with acute leukaemia, regardless of subtype, present with one or more of the following:

- Cytopaenia-related symptoms (Box 39.4) caused by inadequate haemopoiesis secondary to infiltration of the bone marrow by leukaemia
- Symptoms caused by tissue infiltration by leukaemic cells
- Clinical consequences of a high WCC in the circulation
- Symptoms caused by substance release from tumour cells.

Investigations

- ***Blood count.*** Hb and platelets are frequently low, though they can be deceivingly normal. The WCC is often raised, reflecting the proliferating population, but equally can be low if the leukaemic cells are confined to the bone marrow.
- ***Blood film.*** Blast cells are almost invariably seen (Fig. 39.9A). The lineage may be identified morphologically (e.g., the presence of Auer rods is pathognomonic for a diagnosis of AML). In emergencies, a blood film can suffice for diagnosis.
- ***Bone marrow aspirate.*** Increased cellularity (Fig. 39.9B), reduced erythropoiesis, and reduced megakaryocytes are usual. Replacement by blast cells traditionally had to be more than 20% for diagnosis. This diagnostic threshold has now been lowered to 10% in the presence of certain genetic abnormalities. Because AML is increasingly defined by its molecular and genetic basis, the presence of any change carries the most diagnostic weight, almost irrespective of blast percentage. Blast population and lineage is confirmed by immunophenotyping (see Fig. 39.7):
  - CD45: pan-leukocyte marker: defines presence of a blast population in both AML and ALL
  - AML: CD33 or CD13
  - B lineage ALL: CD10 and CD19
  - T lineage ALL: CD3

  Cytogenetic and FISH analysis, as well as molecular genetics, may be used for prognostication. NGS is increasingly used, though it is not available in all centres.
- ***Chest X-ray.*** Mediastinal widening is often present in T-lymphoblastic leukaemia.
- ***CSF examination.*** This should be done in all patients with ALL because the risk of central nervous system (CNS) involvement is high, though it is not indicated at diagnosis unless clear clinical concern is present. Intrathecal chemotherapy is part of standard ALL treatment regimens with its first administration usually around day 14 of treatment. CNS investigation is not routinely performed in AML unless there is a clinical concern.
- ***Coagulation profile.*** This is performed to exclude the presence of disseminated intravascular coagulation (DIC), particularly associated with acute promyelocytic leukaemia (APML), a specific AML subtype. Patients with DIC patients typically have

Box 39.4 Symptoms and signs of leukaemia

	Symptoms	Signs	Type
Marrow failure			
Anaemia	Breathlessness Fatigue Angina Claudication	Pallor Cardiac flow murmur	ALL or AML
Neutropenia	Infections	Fever Mouth ulcers Septic focus	
Thrombocytopenia	Bleeding and bruising	Petechiae Gum bleeding Fundal haemorrhage	
High WCC			
Leucostasis	Breathlessness	Hypoxia, pulmonary infiltrates	ALL or AML
	Headache	Reduced GCS	
	Confusion		
	Visual problems	Retinal vein dilation, papilloedema, fundal haemorrhage	
Tissue infiltration			
Marrow	Bone pain		ALL or AML
Gums		Gum hypertrophy	AML
Skin		Violaceous skin deposits	AML
Liver or spleen		Hepatosplenomegaly	ALL or AML
Lymph nodes		Lymphadenopathy	ALL or AML
CNS	Headache	Cranial nerve palsies	ALL (or AML)
Testes		Testicular enlargement	ALL
Mediastinum		Mediastinal mass, SVCO	ALL
Substance release			
DIC	Bleeding and bruising	Ecchymoses Bleeding i.v. sites	AML
Hyperuricaemia	Acute gout	Kidney stones, tophi	ALL or AML
	Tumour lysis	Acute kidney failure	

ALL, Acute lymphoblastic leukaemia; *AML,* acute myeloid leukaemia; *CNS,* central nervous system; *DIC,* disseminated intravascular coagulation; *GCS,* Glasgow Coma Scale *SVCO,* superior vena cava obstruction; *WCC,* white blood cell count.

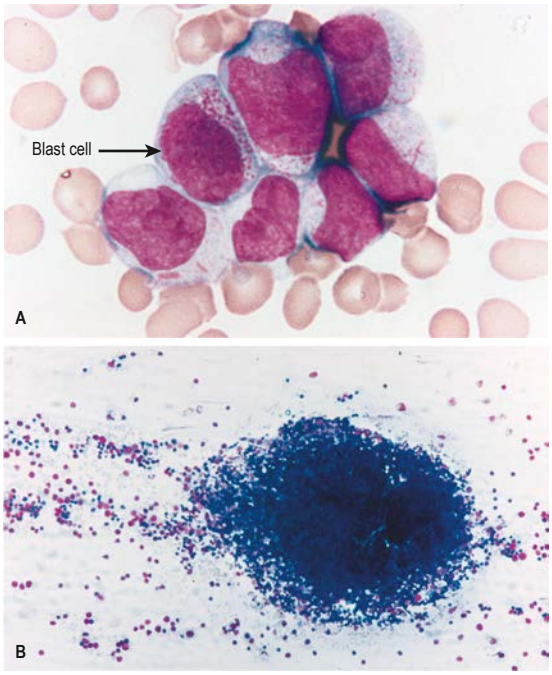


Fig. 39.9 Acute leukaemia. **(A)** Peripheral blood film showing characteristic blast cells. The *arrow* points to the abnormal blast cell. **(B)** Bone marrow aspirate showing a particle with increased cellularity. (*Courtesy of Dr Manzoor Mangi*.)

a raised prothrombin time and activated partial thromboplastin time (APTT), reduced fibrinogen, and increased fibrinogen degradation products (D-dimers). Correction of clotting and platelet count during the initial few weeks of APML treatment is lifesaving.

- ***Biochemistry.*** This should include serum urate, kidney, and liver function tests.
- ***Cardiac function.*** Electrocardiography and echocardiography are used to assess cardiac function before intensive chemotherapy, which may include key cardiotoxic anthracycline drugs. In emergencies, treatment can start without an echo being available.
- ***HLA type and tissue type.*** These are performed to identify potential stem cell donors for allogeneic transplantation. They are recommended on appropriate family members in the first cycle of treatment for patients with AML and ALL. An allograft might not be indicated, but work-up for transplant can take a significant amount of time.
- ***HIV and hepatitis B screening.***

Management

If untreated, acute leukaemia is invariably fatal with a prognosis of weeks to months, though with best supportive care and early involvement of palliative care, this can be longer. If intensive chemotherapy is not appropriate, this should be discussed with the patient and their family at the earliest possibility.

Treatment with curative intent is appropriate in younger and fitter patients, generally those without significant co-morbidity or frailty. Cytogenetic results often define the patient's risk at diagnosis, and it is increasingly recommended, especially for those with AML, to wait for results before commencing treatment provided that a wait is clinically safe because results may determine the best treatment approach, as well as clinical trial entry. Leukaemia that does not respond to frontline treatment is termed 'refractory' leukaemia. Leukaemia that returns after a period of remission is termed 'relapsed' leukaemia. Both scenarios carry a poor prognosis.

At initial presentation, acute leukaemias range from being probably curable (e.g., childhood ALL) through to probably incurable (e.g., AML with adverse cytogenetic features such as p53 mutation). Because curative treatment carries considerable morbidity and potential mortality, it is essential for the risk-to-benefit ratio to be understood by both the physician and patient. In many cases with poor prognostic genetic features, cure requires allogeneic SCT.

The decision to treat with curative intent severely disrupts normal life for the patient and family for at least 6 months and often up to 1 year, if not beyond. In the short term, it may require transfer to another hospital, farther from the patient's home, because patients with acute leukaemias are treated in specialist centres. It is standard to admit the patients to hospital for the first cycle (AML) or phase (ALL) of treatment meaning at least 3 to 4 weeks in hospital. If response is favourable, this will be followed by further, partly predictable, lengthy admissions for further cycles or phases of treatment. It is a commitment that is daunting and hard for patients and relatives; though they should be aware of the longer-term treatment plan, it is advisable that the focus should be on one cycle or phase of treatment at a time because treatment is response dependent.

The decision to treat with curative intent implies that the chance of cure justifies the risks of the therapy. It does not imply that cure is guaranteed or expected.

Supportive care

This is integral, irrespective of treatment approach.

- ***Anaemia.*** Repeated transfusion of packed RBCs are given to keep Hb above 70 g/L. Irradiation of cells may be required (e.g., in patients who have received purine analogue drugs such as fludarabine).
- ***Thrombocytopaenia.*** Prophylactic platelet transfusion has been evaluated and remains superior to any alternative, certainly for patients with AML. Platelet count is maintained above 10×10^9/L in a patient that is stable or above 20×10^9/L if septic.
- ***Coagulation abnormalities.*** Fresh-frozen plasma is used to keep the APTT ratio and international normalized ratio below 1.5 times normal, with cryoprecipitate to keep the fibrinogen concentration above 1.5 g/dL. This is a particular risk in APML, which is associated with activation of the coagulation system (DIC).
- Norethisterone can be given to those who are menstruating to avoid menorrhagia during a thrombocytopenic phase.
- ***Infection*** (see **Box 11.98**)
  - ***Prophylaxis.*** Education of patients, relatives, and staff about diet, hand-washing, and isolation facilities should be done. Use of prophylactic antibiotics is controversial and varies greatly; local guidelines should be followed. Antiviral prophylaxis is recommended throughout. *Pneumocystis jirovecii* pneumonia prophylaxis should be considered, though the treatment itself can influence blood counts. Antifungal prophylaxis should be given to all patients during neutropaenic phases.
  - ***Therapeutic.*** All fevers should be managed following local 'neutropaenic sepsis' guidelines.
  - ***SARS-CoV-2.*** Vaccination is recommended, and all available treatments can be given.
- ***Control of hyperuricaemia (tumour lysis).*** This is achieved with hydration and prophylactic allopurinol. Rasburicase is used

for prophylaxis and treatment of patients with proliferative or high-count acute leukaemia.
- Indwelling venous devices, such as a peripherally inserted central catheter or Hickman line, are required to allow easy access to the blood for tests and administration of therapy.
- In patients of reproductive age, fertility preservation should be considered and offered, time and clinical circumstances allowing.
- Psychological support should be offered and available to all patients.

Intensive chemotherapy

For all acute leukaemias, the initial aim of therapy is to achieve remission, meaning the peripheral blood and bone marrow return to normal (complete remission [CR]). This 'induction chemotherapy' is tailored to the particular leukaemia and the individual patient's risk factors. Because treatments are not targeted, they also impair bone marrow function, which leads to a major risk of life-threatening infection, requiring a prolonged hospital admission.

Successful remission induction should always be followed by further treatment (consolidation) when possible. This consists of further cycles or phases of chemotherapy, which in some cases are subsequently followed by allogeneic SCT, when an appropriate donor is identified. Without consolidation treatment, recurrence is almost inevitable. Cytogenetics and molecular genetic techniques are now routinely used to identify residual leukaemia, not detected morphologically. This is termed 'detection of MRD'. The absence or presence of MRD is highly predictive of recurrence. In ALL, the importance of MRD and when to test are clearly understood and standardized because they have been an integral part of international trials for decades. In AML, this is less clear and standardized, though the role of MRD is clearly important.

Refractory and relapsed disease

Failure to achieve morphological CR with two cycles or phases of therapy ('refractory leukaemia') carries a poor prognosis. Salvage regimens exist in both acute leukaemias, though overall survival is significantly reduced. Salvage treatment only has a curative endpoint if remission is achieved and it is followed by an allogeneic stem cell transplant.

The treatment of patients with relapsed leukaemia, which also carries a poor prognosis, is identical.

Allogeneic stem cell transplant

Colloquially, the term 'bone marrow transplant' is used, though 'stem cell transplant' is correct. Whereas now stem cells are almost invariably collected ('harvested') from the peripheral blood, in the past, they were obtained via bone marrow aspirate. The terminology can be confusing for patients and doctors.

The decision on whether to proceed to allogeneic SCT is based on several factors:
- ***Relapse risk.*** This depends on cytogenetic and molecular risk factors at diagnosis. These have become increasingly defined and have even become part of the official diagnostic nomenclature in the most recent ELN classification (2022).
- ***Response to induction therapy.*** Allogeneic transplant can only proceed if the patient is in morphological remission. The prognosis of patients who have MRD going into allograft is significantly better than the prognosis of patients who have MRD. Further treatment to try to achieve an MRD result before transplant should be undertaken whenever possible.
- ***Transplant-related mortality.*** This varies, depending on frailty and co-morbidity, as well of the type of transplant and quality of the donor match.

Patients should proceed to transplantation only if the reduction in relapse risk outweighs the likely treatment-related mortality, leading to an improved overall survival. The Haematopoietic Stem Cell Transplant Comorbidity Index (HSCT-CI) is a useful tool that incorporates age, co-morbidities, and additional patient-related risk factors (e.g., obesity). All decisions for transplant should be made via an MDT.

Acute myeloid leukaemia

Prognosis

The prognosis depends on a range of key variables, mainly age, cytogenetics, specific gene mutations (Fig. 39.10), and the presence of MRD after initial therapy (Box 39.5).

Management

Treatment with intensive chemotherapy with curative intent is undertaken in most adults younger than the age of 70 years, provided there is no significant co-morbidity. Improved supportive care and reduced-intensity condition transplants have led to a steady increase in the upper age limit of patients considered for allograft. Treatment success reflects the cytogenetic pattern. A CR is achieved in about 80% of patients younger than the age of 70 years. Failure of treatment is caused by either resistant leukaemia (10%) or treatment-related mortality (10%), most commonly infection, sepsis, or haemorrhage. Approximately 50% of those entering CR will be cured (i.e., ~40% overall), although this varies from 60% to 70% in a favourable cytogenetic group to 10% to 20% in an adverse cytogenetic group.

Patients with 'favourable'-risk AML, depending on cytogenetic risk stratification, are treated with intensive combination chemotherapy. The usual approach is two cycles of induction chemotherapy (including an anthracycline) followed by two cycles of consolidation chemotherapy. Each cycle averages 3 to 4 weeks in hospital with a 1- to 2-week period at home in between. A large international AML trial more than a decade ago, which randomized patients to a total of either three or four cycles of chemotherapy, has yet to publish its results. Patients with favourable-risk disease do not benefit from allogeneic SCT during their first CR because the relapse risk is low. This group of patients have been termed as having 'good-risk' AML in the past, but using 'good' in the description of any AML subgroup seems a misnomer and should be avoided.

Patients with 'intermediate'-risk or 'standard'-risk AML are a heterogeneous and increasingly complex group. The initial treatment approach is the same as for those with favourable-risk disease when possible. In the past, most patients in this group were recommended to proceed to allogeneic transplant; however, in the past decade, MRD status has become a vital part of risk stratification. The confirmation of deep remission by MRD negativity is considered sufficient to remove the need for allogeneic transplant for most patients, though this should be assessed on a case-by-case basis. The shift of using MRD status to determine patients requiring allograft in first remission has led to overall 'poorer'-risk patients proceeding to allograft compared with historical approaches. This should be kept in mind when interpreting more recent allograft outcomes and comparing them with historical data.

Patients with 'adverse'-risk cytogenetics respond poorly to conventional induction and consolidation chemotherapy alone and have a high risk of relapse. If suitable, allogeneic SCT is

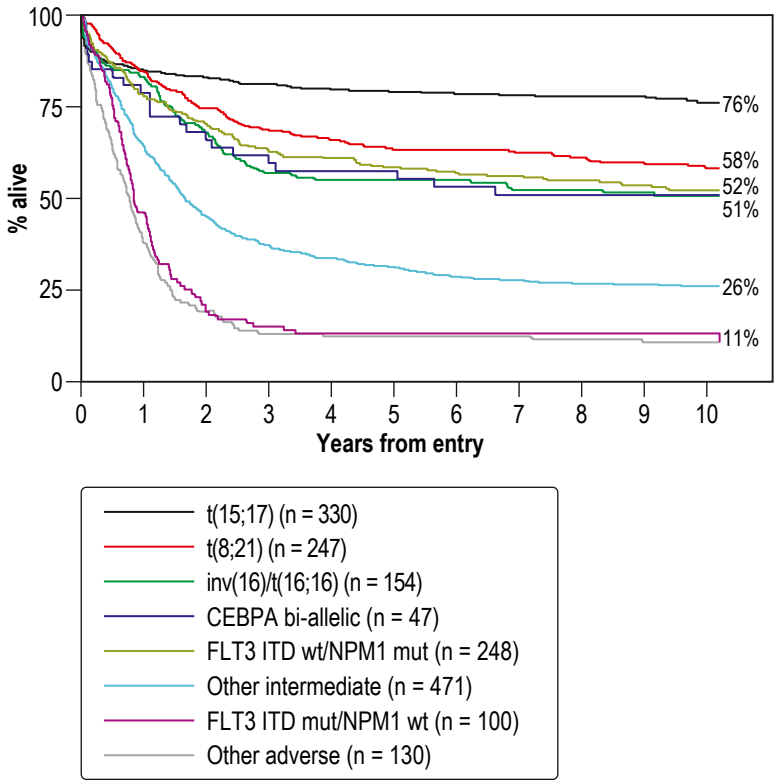


Fig. 39.10 Prognosis related to cytogenetics and molecular data in acute myeloid leukaemia (AML). Outcome of younger adults with AML treated in the Medical Research Council AML10 and AML12 trials, stratified according to cytogenetic and molecular abnormalities. (*Reproduced with permission from Smith ML, Hills RK, Grimwade D. Independent prognostic variables in acute myeloid leukaemia.* Blood Rev *2011; 25:39–51.*)

Box 39.5 Risk factors in acute myeloid leukaemia

Good risk
- *De novo* disease
- Favourable cytogenetics: t(15; 17) t(8; 21) or inv(16) or its variant t(16; 16)
- *CEBPA* bi-allelic mutation
- *NPM1* mutation with FLT3 wild type

Poor risk
- Age older than >60 years
- Male gender
- Secondary disease (e.g., prior MDS or MPN)
- High WCC
- Adverse cytogenetics: –5, del(5q), –7, abnormal 3q26 or a complex karyotype
- FLT3 internal tandem duplication mutation
- MRD positivity post induction chemotherapy

FLT3, FMS-related tyrosine kinase-3; *MRD,* measurable residual disease; *MDS,* myelodysplastic syndrome; *MPN,* myeloproliferative neoplasm; *WCC,* white blood cell count.

performed after they achieve a CR. In this strategy, allograft will often replace the fourth cycle of chemotherapy and occasionally the third.

Since 2020, the combination therapy of venetoclax and azacitidine has become another standard frontline therapy option. It is the first new option in frontline therapy for AML patients in more than 20 years. Though initially used in older and frailer patients not fit for intensive chemotherapy, there is emerging use in other patients as well. It allows more outpatient-based therapy. The role in refractory or relapsed patients is not clearly defined but may be significant.

Allografts, with many different conditioning regimens and protocols, as well as the vastly increased donor potential because of haplo-identical allografts becoming standard in recent years, fall outside the scope of this text.

The management of recurrence is undertaken on an individual basis because the overall prognosis is very poor, even though second remissions can be achieved. Long-term survival after recurrence is rarely achieved without allogeneic transplantation. The use of MRD monitoring may identify patients in CR who are in the early stages of relapse and need pre-emptive therapy before frank marrow relapse occurs.

Acute promyelocytic leukaemia

APML is a variant of AML, occurring in 10% to 15% of cases. It is characterized by the translocation t(15; 17), which produces a *PML-RARA* fusion gene with particular morphological features. There is an almost invariable coagulopathy, which remains a major cause of early death. The empirical discovery that all-*trans*-retinoic acid (ATRA) causes differentiation of promyelocytes and rapid reversal of the bleeding tendency was a major breakthrough. ATRA alone can achieve remission. If APML is suspected, ATRA should be commenced as soon as possible. Any coagulopathy and thrombocytopaenia should also be aggressively corrected for the initial few weeks. Intracranial haemorrhages occur frequently as a devastating and terminal event in the first few weeks of treatment. After remission is achieved,

prognosis is excellent, with relapses increasingly rare. There is therefore no need for allogeneic transplant in first remission.

APML has two curative treatment options in the front-line setting. Patients with high-risk APML patients (high WCC at diagnosis) are treated with ATRA and combination chemotherapy similar to chemotherapy outlined earlier. Those with low-risk APML (low WCC at diagnosis) are treated with ATRA and arsenic trioxide, a chemotherapy-free, mainly outpatient-based treatment. Both treatments have comparable outcomes.

Acute lymphoblastic leukaemia

This condition may present in leukaemic phase with significant marrow involvement (ALL) or as localized bulky disease, typically a mediastinal mass (lymphoblastic lymphoma). The tumour cells in each condition are indistinguishable, and similar therapies are therefore used.

Prognosis

Several clinical and laboratory features predict treatment response and survival in patients with ALL (**Box 39.6**). Increasingly, therapeutic strategies based on prognostic risk are used in management. MRD stratification is increasingly used to select patients who are at high risk (for treatment intensification) or low risk (for treatment de-escalation).

The prognosis of ALL in childhood is excellent: CR is achieved in almost all, with up to 80% being alive without recurrence at 5 years. Failure occurs most frequently in those with high presentation blast counts or patient with clearly defined adverse cytogenetics. Current treatment strategies de-escalate therapy for favourable-risk children to avoid some of the long-term consequences of therapy, such as avascular necrosis of bone, infertility, neurotoxicity, and cardiotoxicity.

The prognosis of ALL in adults remains very poor. Adverse cytogenetics are more commonly seen with advancing age. Co-morbidities further complicate the intense treatment approach. Long-term remission is only achieved in 25% to 30% of patients. Allogeneic SCT is recommended in almost all adults achieving a remission, though there are exceptions. Novel targeted immuno- and antibody therapies have led to improvement in patients achieving MRD status, allowing more patients to proceed to allograft. Their role in the relapse setting is less well defined. As with AML, most relapses occur within the first 2 years of initial treatment. Outcomes in those with relapsed ALL are extremely poor.

Management

As with AML, the overall strategy is to achieve an initial remission with induction chemotherapy followed by consolidation phases, an allogeneic transplant, or both. The drug combinations in ALL differ from those used in AML, with initial induction typically consisting of vincristine, a glucocorticoid, an anthracycline and Peg-asparaginase; different agents are used in the consolidation cycles. Every phase of ALL treatment is different, which makes the regimens appear very complex because they are not repeated. Different phases are also skipped or added depending on the patient route to transplant and the type of transplant being planned. To complicate things further, in patients not proceeding to allograft, the induction and consolidation phases are followed by lengthy maintenance phases to reduce the risk of disease recurrence, typically continuing for another 2 years.

A major difference between treatment for ALL and AML is the need for CNS-directed therapy in ALL because of the higher risk of CNS infiltration or relapse. Prophylaxis should be given with intrathecal chemotherapy (under platelet cover if necessary) as soon as blasts are cleared from the blood. The first planned intrathecal administration is usually around day 14 of the first induction phase. Cranial irradiation was used routinely in the past to reduce the risk of relapse within the CNS; however, this has been replaced by an intensification phase with high-dose methotrexate (which crosses the blood–brain barrier), usually after two induction phases. This phase was added to standard treatment, which historically included a number of intrathecal chemotherapy administrations; no randomized data exist comparing the two or a combination approach.

The presence of the t(9;22) Ph+ (see **Fig. 5.24**) has traditionally been associated with increased risk; however, the advent of imatinib and the other TKIs, originally developed for CML, have overcome this risk to some extent. These agents can be used in combination with intensive or non-intensive chemotherapy.

Chimeric antigen receptor T-cell (CAR T) therapy has rapidly become a treatment option for patients with relapsed ALL and for patients with no available donor. CAR T therapy comes with its own unique and potentially dangerous set of side-effects. An important difference between CAR T therapy and allogeneic SCT is that the patient does not need to be in remission to proceed to CAR T therapy. Indications for CAR T therapy will no doubt increase significantly in the coming years.

Further reading

Döhner H, Wei AH, Appelbaum FR et al. Diagnosis and management of AML in adults: 2022 recommendations from an international expert panel on behalf of the ELN. *Blood.* 2022;140(12):1345–1377.

Box 39.6 Risk factors in acute lymphoblastic leukaemia

Risk factor	Good	Poor
Age	Younger	Older
WCC	<30 × 10^9/L for B-cell lineage <100 × 10^9/L for T-cell lineage	>30 × 10^9/L for B-cell lineage >100 × 10^9/L for T-cell lineage
Immunophenotype	CD10+ common ALL	Pro-B ALL
Cytogenetic aberrations	t(12; 21) hyperdiploidy	t(9; 22) or t(4; 11) hypodiploidy
Time to response	Early clearance of blasts	Failure to achieve CR within 3–4 weeks
Minimal residual disease	MRD negative	MRD positive
Extramedullary disease	CSF clear	CSF positive

ALL, Acute lymphoblastic leukaemia; *CR*, complete remission; *CSF*, cerebrospinal fluid; *MDS*, myelodysplastic syndrome; *MPN*, myeloproliferative neoplasm; *MRD*, measurable residual disease; *WCC*, white blood cell count.

Chronic myeloid leukaemia and the myeloproliferative neoplasms

The MPNs are clonal stem cell disorders characterized by uncontrolled proliferation of one or more of the cell lines in the bone marrow, usually erythroid, myeloid, and/or megakaryocyte lines. Myeloproliferative disorders include:

- CML
- Polycythaemia vera (PV)
- Essential thrombocythaemia (ET)
- Primary myelofibrosis
- Chronic neutrophilic leukaemia
- Chronic eosinophilic leukaemia
- Juvenile myelomonocytic leukaemia
- MPN, not otherwise specified.

Chronic myeloid leukaemia

CML, which accounts for about 14% of all leukaemias, is a member of the family of MPNs. It is almost exclusively a disease of adults, with presentation between 40 and 60 years of age. It is defined by the presence of the Ph+ (see Fig. 5.24), which is demonstrated either by cytogenetic analysis alone (95%) or by FISH or reverse transcription polymerase chain reaction (RT-PCR) (5%). Unlike the acute leukaemias, which are rapidly fatal without urgent treatment, CML has a more slowly progressive course. If left untreated, the disease will evolve from the chronic phase to the blast phase. When it evolves, it can transform to an AML (75%) or an ALL (25%) with death in a median of 3 to 4 years. Fortunately, targeted therapy to arrest disease progression is highly effective, and the survival rate is excellent.

Clinical features

CML usually presents in the chronic phase, and some patients have no symptoms.

Symptoms

When present, symptoms include:

- Symptomatic anaemia (e.g., shortness of breath)
- Abdominal discomfort caused by splenomegaly
- Weight loss
- Fever and sweats in the absence of infection
- Headache (occasionally) or priapism caused by hyperleucocytosis
- Bruising and bleeding (uncommon)

Signs

- Pallor
- Splenomegaly, often massive
- Lymphadenopathy (uncommon; suggests blast crisis)
- Extramedullary soft tissue leukaemic deposit: 'chloroma' (indicates blast crisis)
- Retinal haemorrhage caused by leucostasis

Investigations

- ***Blood count.*** Hb is low (normochromic and normocytic) or normal; the WCC is high (usually >100 × 10^9/L); the platelet count can be low, normal, or high.
- ***Blood film.*** There is neutrophilia with the whole spectrum of mature myeloid precursors. Basophils and eosinophils are elevated. Increased numbers of blasts indicate evolution to blast crisis (Fig. 39.11).
- ***Bone marrow aspirate.*** Increased cellularity is seen, with greater numbers of myeloid precursors.

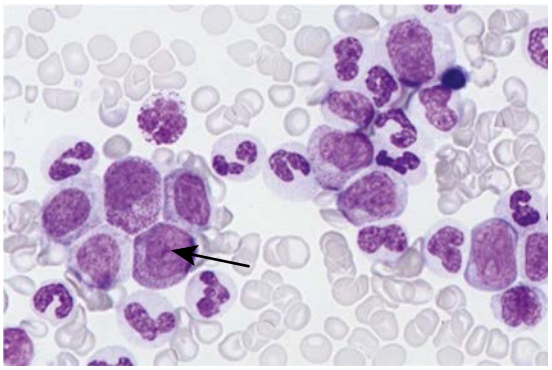


Fig. 39.11 Blood film showing blast cells *(arrow)* in chronic myeloid leukaemia.

- ***Cytogenetic analysis (G-banding).*** This reveals a t(9; 22) translocation (Ph+; Fig. 5.24).
- ***FISH or RT-PCR.*** These demonstrate the BCR:ABL fusion. These techniques are also used to quantify response to therapy.

Management

Most patients diagnosed with chronic-phase CML have a normal life expectancy. Treatment is with TKIs, which specifically block the enzymatic action of the BCR-ABL fusion protein. Imatinib is first-line treatment for patients in the chronic phase, although second-generation TKIs may be used. Imatinib produces a complete cytogenetic remission (no detectable Ph+ cells in the marrow) in 70% of patients. A significant proportion lose molecularly detectable BCR-ABL transcripts from the blood, achieving a molecular response (MR) of varying depths: MR3 (a log reduction of 3; <0.1%), MR4 (<0.01%), or MR5 (<0.001%). Side-effects of imatinib include nausea, headache, rash, and cytopenias. A goal of therapy is to achieve deep molecular response (DMR) to allow reduction and then stop treatment for a period of treatment-free remission (TFR). Discontinuation can be considered in those who respond optimally and achieve a sustained DMR (MR4). Patients need to be closely monitored, with around two-thirds able to remain off treatment without disease recurrence.

Resistance to imatinib as a single agent may develop as a result of secondary BCR-ABL kinase mutations beyond t(9; 22). If patients do not tolerate imatinib or have a suboptimal response, mutation analysis should be performed, and a second-generation TKI (e.g., dasatinib, nilotinib, bosutinib) should be commenced. Those who have developed a T315I mutation should be offered ponatinib.

In the blast phase, the approach is to combine a second-line TKI with intensive chemotherapy, as for acute leukaemia. SCT should then be considered with the aim of achieving a durable remission.

Other myeloproliferative disorders

The other myeloproliferative disorders are often characterized by a mutation in *JAK2*, calreticulin *(CALR)*, or thrombopoietin receptor *(MPL)*. These driver mutations cause proliferation of the myeloid cell lines. This is typically gradual, but all can transition to more aggressive forms, either myelofibrosis or AML, with the accumulation of secondary cytogenetic or molecular abnormalities.

Polycythaemia vera

The investigation of polycythaemia is covered in Chapter 38 (see 'Polycythaemia' section). Although it is important to exclude secondary causes of polycythaemia, the presence of a mutation of the

JAK2 gene is strongly suggestive of primary polycythaemia (PV). This is a clonal stem cell disorder in which there is an excessive proliferation of erythroid, myeloid, and/or megakaryocytic progenitor cells. Almost all patients with PV have acquired mutations of the *JAK2* gene. In 96% to 97% of patients, this is a point mutation that causes the substitution of phenylalanine for valine at position 617 (JAK2V617F). Most of the remainder have rarer *JAK2* exon 12 mutations. *JAK2* is a cytoplasmic tyrosine kinase that transduces signals, especially those triggered by haemopoietic growth factors such as erythropoietin (EPO), in normal and neoplastic cells. The discovery of the *JAK2* mutation in MPNs is significant for the diagnosis of PV and as a target for therapies.

Clinical features

The onset is insidious. PV is more common in patients older than 60 years of age and may present with fatigue, itching, vertigo, headache, and visual disturbance. These symptoms are common in the older population, so PV is easily missed. Increasingly, PV is detected on routine blood tests conducted for other reasons. Severe itching may occur after a hot bath or when the patient is warm. Gout caused by increased cell turnover may be a feature, and peptic ulceration occurs in a minority of patients. Thrombosis and haemorrhage are the major complications of PV.

The patient may be plethoric and have a dusky cyanosis and injection of the conjunctivae. The spleen is palpable in 70% and is useful in distinguishing PV from secondary polycythaemia. The liver is enlarged in 50% of patients.

Diagnosis

JAK2 mutation screening is routine in the investigation of polycythaemia and incorporated into diagnostic criteria. Although WHO criteria require a bone marrow biopsy for diagnosis in adults, current British guidelines do not mandate this in the presence of a confirmed *JAK2* mutation (**Box 39.7**).

The measurement of RBC mass and plasma volume is not routinely required or available. Measurement of serum ferritin can be useful to detect cases of iron-deficient PV. There may be a raised serum uric acid, leucocyte alkaline phosphatase, and raised serum vitamin B_{12} and vitamin B_{12} binding protein (transcobalamin 1).

Box 39.7 Recommended diagnostic criteria for polycythaemia vera (PV)

***JAK2*-positive PV (requires both criteria)**

- Haematocrit >0.52 in males, >0.48 in females, *or* raised RBC mass >25% above predicted
- Mutation in *JAK2*

***JAK2*-negative PV (requires A1–A4 plus another A or two B criteria)**

- A1: haematocrit ≥0.60 in males, ≥0.56 in females, *or* raised RBC mass >25% above predicted
- A2: absence of mutation in *JAK2*
- A3: no secondary cause of erythrocytosis
- A4: bone marrow histology consistent with PV
- A5: palpable splenomegaly
- A6: acquired genetic abnormality (excluding *BCR-ABL*) in haemopoietic cells
- B1: thrombocytosis (platelet count >450 × 10^9/L)
- B2: neutrophil count >10 × 10^9/L in non-smokers, ≥12.5 × 10^9/L in smokers
- B3: X-ray evidence of splenomegaly
- B4: low serum erythropoietin

Adapted from BSH Guidelines: https://onlinelibrary.wiley.com/doi/full/10.1111/bjh.15648.

Management

Course and treatment

Patients should be risk assessed for the likelihood of vascular events. Treatment is designed to prevent thrombotic complications of the disease and to reduce the long-term risk of progression and transformation. Additional modifiable cardiovascular risk factors should be actively addressed.

There are three types of specific treatment:

- ***Low-dose aspirin.*** Aspirin 75 mg/day is routinely used for patients with PV in the absence of contraindications to reduce the risk of cardiovascular complications.
- ***Venesection.*** During initial phases of treatment, the removal of 400 to 500 mL of blood weekly successfully relieves many of the symptoms of PV. Iron deficiency eventually limits erythropoiesis, and only infrequently is venesection required. Venesection is often used as the sole treatment, with the aim of maintaining a haematocrit below 0.45.
- ***Chemotherapy.*** Continuous or intermittent treatment with hydroxycarbamide (hydroxyurea) is used in patients who do not tolerate venesection or are at higher risk of thrombosis. It also helps those who have other poorly controlled disease features such as thrombocytosis, symptomatic splenomegaly, or thrombosis. Interferon-alfa is also effective; it is administered by subcutaneous injection but can cause unacceptable toxicity. Low-dose intermittent busulfan may be more convenient, but use must be weighed against potential long-term complications, such as the increased risk of leukaemia development.
- ***Ruxolitinib.*** This *JAK2* inhibitor can be used in patients who are intolerant or unresponsive to hydroxycarbamide.
- Iron replacement should be avoided in patients with iron-deficient PV because it can result in a dangerous rebound polycythaemia.

General measures

- ***Allopurinol.*** This can be given to block uric acid production.
- ***Avoiding very hot baths.*** This lessens pruritus.
- ***Surgery.*** Polycythaemia should be controlled before surgery. Patients with uncontrolled PV have a high operative risk. In an emergency, the haematocrit must be reduced by venesection and appropriate fluid replacement.

Prognosis

PV develops into myelofibrosis in 10% to 15% of cases by 10 years and into AML in 2% as part of the natural history of the disease.

Essential thrombocythaemia (ET)

ET is an MPN, closely related to PV, which is characterized by a persistently raised platelet count. Patients usually have normal Hb concentrations, although mild anaemia may occur. The WCC is high in some patients. At diagnosis, the platelet count is usually above 600 × 10^9/L and may be as high as 2000 × 10^9/L or rarely even higher.

Clinical features

ET presents either symptomatically with thromboembolic or bleeding (less common) or incidentally (e.g., at a routine medical check). Erythromelalgia (severe burning pain), erythema, and warmth of the extremities (primarily the feet and, to a lesser extent, the hands) may occur.

Investigations

The diagnosis of ET is based on a sustained platelet count of 450 × 10^9/L or more, in the absence of a possible reactive cause of thrombocytosis (Box 39.8) and of other haematological malignancy associated with thrombocytosis (e.g., PV, myelofibrosis). Mutations in *JAK2, CALR,* or *MPL* genes are detected in 50% to 60%, 25% to 30%, and 3% to 10% of cases, respectively. As a generalization, a person with a very high platelet count (>1000 × 10^9/L) who is clinically well with good health will most likely prove to have ET. In a patient who has a lower platelet count (e.g., 600 × 10^9/L) or infective or inflammatory conditions, the diagnosis can be more difficult.

Management

Because of the thrombotic risk, aspirin should be considered in all patients without contraindications, especially those with *JAK2-V617F* mutations or cardiovascular risk factors. Treatment to reduce the platelet count ('cytoreduction') is recommended for patients with any high-risk feature:

- Age older than 60 years
- Platelet count greater than 1500 × 10^9/L
- History of thrombosis or haemorrhage.

Treatment is usually with hydroxycarbamide (hydroxyurea) to control the platelet count to below 400 × 10^9/L. Alternative cytoreductive agents include anagrelide or busulfan. Interferon-alfa is also effective; it is administered by subcutaneous injection but can be difficult to tolerate because of side-effects. ET may eventually transform into PV, myelofibrosis, or acute leukaemia, but the disease may not progress for many years.

Secondary thrombocytosis

Other disorders that may give rise to reactive high platelet counts include iron deficiency, infection, inflammation, and malignancy. Individuals who have had a splenectomy often have high platelet counts.

? Box 39.8 Causes of thrombocythaemia

Primary ('haematological')

- Essential thrombocythaemia
- Polycythaemia vera
- Myelofibrosis
- CML
- Myelodysplasia with del(5q)
- MDS/MPN overlap syndromes (e.g., CMML, atypical CML)

Secondary ('reactive')

- Infection
- Inflammation
- Haemorrhage
- Postoperative
- Inflammatory conditions (e.g., IBD, RA)
- Haemolysis
- Hyposplenism: functional or surgical
- Iron deficiency
- Drug therapy (e.g., corticosteroids, thrombopoietin receptor agonists)
- Rebound after chemotherapy
- Solid malignancy

CML, Chronic myeloid leukaemia; *CMML,* chronic myelomonocytic leukaemia; *IBD,* inflammatory bowel disease; *MDS,* myelodysplastic syndrome; *MPN,* myeloproliferative neoplasm; *RA,* rheumatoid arthritis.

Adapted from British Society for Haematology Thrombocythemia guidelines https://onlinelibrary.wiley.com/doi/full/10.1111/j.1365-2141.2010.08122.x.

Myelofibrosis

Myelofibrosis is an MPN characterized by clonal proliferation of megakaryocytes and abnormal myeloid cells in the bone marrow. It may be primary or develop late in the course of ET or PV. Abnormal megakaryocytes release fibroblast-stimulating factors such as platelet-derived growth factor, which cause reactive fibrosis in the bone marrow.

Clinical features

The disease presents with lethargy, weakness, and weight loss. Patients often experience a 'fullness' in the upper abdomen because of splenomegaly, which may be massive. Severe pain related to respiration may indicate perisplenitis secondary to splenic infarction, and bone pain and attacks of gout can complicate the illness. Extramedullary haemopoiesis can develop, typically in the spleen but also affecting other organs. Bruising and bleeding occur because of thrombocytopenia or abnormal platelet function.

Investigations

- ***FBC*** shows anaemia with leucoerythroblastic features. Poikilocytes and RBCs with characteristic teardrop forms are seen (Fig. 39.12). The WCC may be greater than 100 × 10^9/L, and the differential WCC may be very similar to that seen in CML; leucopenia may develop later. The platelet count may be very high but in later stages thrombocytopenia occurs.
- ***Bone marrow aspiration*** is often unsuccessful (a 'dry tap') owing to the fibrotic state of the bone marrow, which may give a clue to the presence of the condition. A bone marrow trephine is necessary to show the markedly increased fibrosis. Increased numbers of megakaryocytes may be seen.
- ***Cytogenetic and molecular analysis*** shows the absence of the Ph+ (*BCR-ABL* fusion gene); this helps to distinguish myelofibrosis from CML. A *JAK2* mutation is present in approximately 65% of cases, a *CALR* mutation in 25%, and *MPL* mutation in 8%.

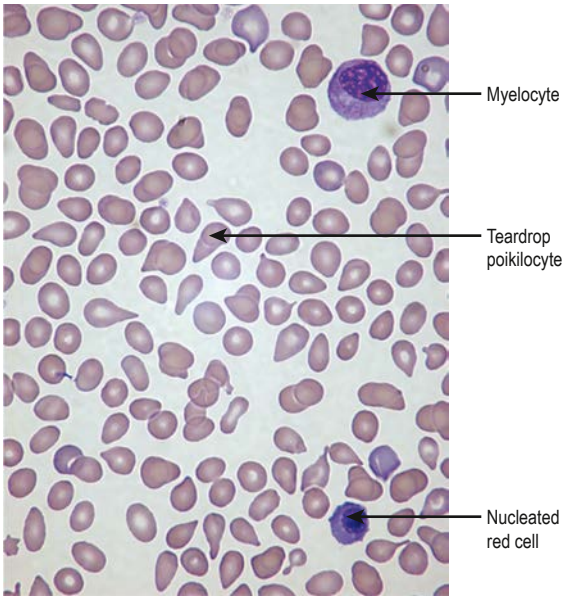


Fig. 39.12 Leucoerythroblastic blood film with teardrops in myelofibrosis.

Management

Myelofibrosis is incurable in all but the minority of patients who are fit enough to undergo allogeneic SCT. For most patients, treatment is aimed at managing symptoms of anaemia, splenomegaly, and constitutional symptoms. Prognostic scoring systems (e.g., international prognostic scoring system [IPSS] and dynamic international prognostic scoring system [DIPSS]) help estimate median survival.

JAK inhibitor treatments are now available to treat patients with myelofibrosis. Ruxolitinib results in substantial spleen reduction and improvement in symptoms in many patients with higher-risk myelofibrosis. Fedratinib improves spleen and symptom burden. Newer JAK inhibitor treatments are in the pipeline.

Anaemia is managed with transfusions, danazol, and EPO.

In the proliferative stages, hydroxycarbamide, interferon, and anagrelide can be used to control counts. Thalidomide and lenalidomide can improve cytopenias and reduce spleen size. Patients with massive splenomegaly can be treated with radiotherapy, but this has limited benefit in those with myelofibrosis.

Prognosis

Patients may survive for 10 years or more, though the median survival period is 4 to 5 years. Poor prognostic factors include cytopenias, the presence of circulating blast cells, constitutional symptoms, and the presence of poor-risk cytogenetics (e.g., monosomy 7) or molecular genetic mutations (e.g., *ASXL1* or *EZH2*). Death occurs in 10% to 20% of patients by transformation to AML, which is often resistant to intensive chemotherapy. The other most common causes of death are progressive bone marrow failure and infection.

Other myeloproliferative neoplasms

Rare MPNs include the following:

- ***Chronic eosinophilic leukaemia:*** associated with gene fusions that activate tyrosine kinase activity and may be treated with imatinib, as for CML
- ***Chronic neutrophilic leukaemia***
- ***'PMH, not otherwise specified'***

Further reading

Etienne G, Guilhot J, Rea D et al. Long-term follow-up of the French Stop Imatinib (STIM1) study in patients with chronic myeloid leukemia. *J Clin Oncol* 2017; 35:298–305.
Hehlmann R. The new ELN recommendations for treating CML. *J Clin Med.* 2020;9(11):3671.

MYELODYSPLASIA

'Myelodysplastic neoplasms', or MDS, is the collective term for a group of acquired bone marrow failure disorders caused by a defect in haematopoietic stem cells. They are characterized by progressive bone marrow failure with quantitative and qualitative abnormalities in at least one of the three myeloid cell lines (RBCs, granulocyte/monocytes, and platelets). MDS covers a heterogeneous group of conditions with varying presentations and prognosis, for which there is a wide range of treatment approaches ranging from observation to allogeneic SCT. Although the natural history of MDS is variable, there is high morbidity and mortality owing to bone marrow failure and transformation to AML, which occurs in about 30% of cases. MDS and all its subtypes are considered malignancies.

The molecular and genetic discoveries during the past decade have had a huge impact on the understanding and classification of MDS and led to significant changes in treatment approaches. The classification of MDS has changed dramatically over the decades; often old terminology is used, which should be avoided. The latest classification by the WHO in 2022 identifies subtypes defined by their genetic abnormality (a trend that will likely continue in future classifications); it also lowers the blast percentage that defines AML from 20% to 10% in some subtypes. This is in keeping with the current clinical approach in which 'high-risk' MDS in younger or fitter patients is managed similar to AML as it is considered to be the same disease by most. The current WHO classification of MDS is shown in **Box 39.9**.

Clinical features

MDS occurs mainly in older adults and presents with symptoms of anaemia, infection, or bleeding because of thrombocytopaenia. MDS should be suspected in patients with otherwise unexplained cytopaenias or macrocytosis. Labelling a patient with MDS without absolute certainty should be avoided. If the diagnosis is in doubt, investigations, primarily of the bone marrow, should be repeated 3 to 4 months after the initial one.

Investigations

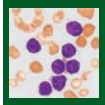
Serial blood counts show evidence of increasing bone marrow failure with anaemia, neutropenia, and thrombocytopenia, either alone or in combination.

A bone marrow investigation, both an aspirate and trephine, are required for diagnosis. The bone marrow cellularity can be increased, normal, or reduced, irrespective of the pancytopaenia. Dysplasia in one or more of the erythroid, granulocytic, and megakaryocytic lineages is usually present (**Fig. 39.13**). The description comes with an element of subjectivity, and an understanding of the clinical context is required. Patients who are otherwise unwell often have minor reactive dysplastic features. The percentage of blast cells in the bone marrow is important for classification and prognostic risk stratification in MDS. If the percentage rises above 20%, the diagnosis changes from MDS to AML. In some subtypes, this percentage has dropped to 10% in the presence of certain cytogenetic abnormalities, blurring the lines further between MDS and AML.

Cytogenetic abnormalities are common and increasingly important for prognostic stratification. The discovery of new cytogenetic abnormalities has greatly changed the classification and approach to treatment of patients with MDS. NGS has and will continue to identify further mutations with prognostic relevance. Any cytogenetic change can confirm a diagnosis of MDS if morphological changes are not sufficient. If there is diagnostic uncertainty, a repeat of investigations some months down the line can clarify whether a patient has MDS.

Management

MDS is a heterogeneous disease, and management is dependent on clinical and patient-dependent features, as well as the predicted prognosis estimated using international prognostic scoring systems incorporating blood counts, bone marrow blast percentage, and cytogenetics. Treatment strategy is based on the behaviour and clinical burden of the disease, the likely prognosis, and the frailty and co-morbidity of the patient. For example, an older patient may be managed with transfusion support alone, but a younger, fitter patient with high-risk disease may be considered for the higher-risk but potentially curative strategy of allogeneic SCT.



Box 39.9 Classification and defining features of myelodysplastic syndrome (MDS)

	Blasts	Cytogenetics	Mutations
MDS with defining genetic abnormalities			
MDS with low blasts and isolated 5q deletion (MDS-5q)	<5% BM and <2% PB	5q deletion alone, or with 1 other abnormality other than monosomy 7 or 7q deletion	
MDS with low blasts and *SF3B1* mutation[a] (MDS-SF3BI)		Absence of 5q deletion, monosomy 7, or complex karyotype	*SF3B1*
MDS with bi-allelic *TP53* inactivation (MDS-biTP53)	<20% BM and PB	Usually complex	Two or more *TP53* mutations or one mutation with evidence of *TP53* copy number loss or cnLOH
MDS, morphologically defined			
MDS with low blasts (MDS-LB)	<5% BM and <2% PB		
MDS, hypoplastic[b] (MDS-h)			
MDS with increased blasts (MDS-IB)			
MDS-IB1	5%–9% BM or 2%–4% PB		
MDS-IB2	10%–19% BM or 5%–19% PB or Auer rods		
MDS with fibrosis (MDS-f)	5%–19% BM; 2%–19% PB		

[a]Detection of >15% ring sideroblasts may substitute for *SF3B1* mutation. Acceptable related terminology: MDS with low blasts and ring sideroblasts.
[b]By definition, <25% bone marrow cellularity, age adjusted.
BM, Bone marrow; *cnLOH,* copy neutral loss of heterozygosity; *PB,* peripheral blood.

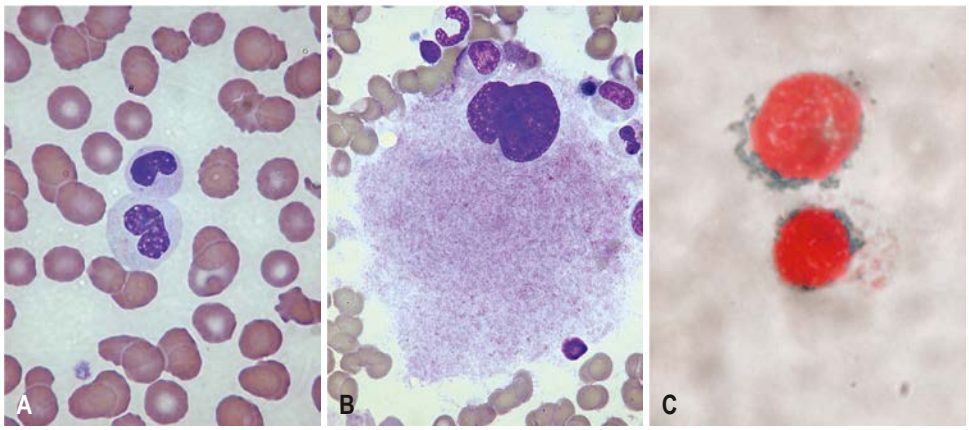


Fig. 39.13 Myelodysplastic syndrome. **(A)** Pseudo-Pelger neutrophil with bilobed nucleus. **(B)** Dysplastic megakaryocyte in the bone marrow. **(C)** Iron stain of the bone marrow showing 'ring sideroblasts'.

Treatment options include the following:

- ***Best supportive care*** is the mainstay for most patients with MDS and symptomatic low blood counts. It includes RBC and platelet transfusion and the prevention and treatment of MDS-associated bleeding and infection. Many patients with low-risk MDS can maintain a stable clinical picture for years.
- ***EPO*** can be used in selected patients with low-risk MDS, symptomatic anaemia, and inappropriately low endogenous EPO levels.
- ***Iron chelation*** should be considered for patients with good prognostic MDS who develop secondary iron overload (transfused >20 units of RBCs or serum ferritin >1000 μg/L).
- ***Hypomethylating agents*** (e.g., azacitidine, decitabine). Azacitidine, licensed in 2011, was the first MDS-specific treatment available. Significant clinical benefits are seen in half of patients with an improvement in quality of life mainly due to reduced transfusion requirements and a reduced risk of leukaemic transformation. It is a disease-stabilizing treatment, not a curative strategy.
- ***Immunosuppressive agents***, such as ciclosporin and antithymocyte globulin, may be used in selected patients with a hypocellular bone marrow. They are hypothesized to modify an immune driver in hypoplastic MDS, similar to aplastic anaemia.
- ***Intensive chemotherapy*** schedules used in AML should be considered in high-risk MDS. It is in many ways considered the same disease. Remission rates, however, are lower in patients with MDS, and more prolonged periods of pancytopaenia can be seen after treatment in this patient group.

- ***Lenalidomide*** (a thalidomide analogue) is successful in the treatment of early-stage MDS with a chromosome 5q deletion (the 5q– syndrome). This is a now clearly defined subtype of MDS.
- ***Allogeneic SCT*** carries the only possibility of cure in appropriate patients with MDS, similar to the case in AML. Patients (and donors) should be carefully selected. This treatment is restricted to patients with higher-risk MDS without major co-morbidities and frailty.

Myelodysplastic syndrome–myeloproliferative neoplasm overlap syndromes

This is a mixed group of conditions that have both dysplastic and proliferative features. It is often tricky to get to an accurate diagnosis. Repeat bone marrow biopsies may help elucidate the problem. It includes:

- Chronic myelomonocytic leukaemia
- Atypical chronic myeloid leukaemia (*BCR-ABL* negative)
- Juvenile myelomonocytic leukaemia
- MDS/MPN, unclassifiable.

Because of the rarity of these conditions, protocols are less well established. Treatment principles, however, are similar to those for MDS (e.g., supportive care for cytopaenias) and MPNs (e.g., cytoreduction for elevated blood counts). Similarly, more intensive approaches, including AML-type treatment and SCT, should be considered for selected patients. Patients with these conditions should be cared for in specialist centres.

Further reading

Cross CP, Godfrey AL, Cargo C et al. The use of genetic tests to diagnose and manage patients with myeloproliferative and myeloproliferative/myelodysplastic neoplasms, and related disorders: a British Society for Haematology guideline. *Br J Haematol.* 2021; 195:338–351.

LYMPHOMAS

The lymphomas are malignancies of the lymphoid ('lymphatic') system. They are a heterogenous group of malignancies that can arise anywhere in the body where lymphoid tissue is present. The overall incidence of lymphoma ranges 15 to 20 per 100 000 population; it is the fifth most common malignancy in the Western world. Clinical presenting features and symptoms can range wildly between the different types of lymphoma, from entirely asymptomatic to gravely ill patients. Most commonly, patients present with peripheral lymphadenopathy or symptoms caused by occult lymph nodes; however, approximately 20% of patients present with disease in extranodal sites. Only a small proportion of patients present with lymphoma-associated B symptoms: weight loss, fever, and night sweats (despite prominence in medical textbooks!). Significant disease can be present in the absence of these symptoms and vice versa. Although 'B symptoms' are important, they are non-specific and can be present in many other medical conditions. The natural history, clinical course, and treatment vary widely between lymphoma subtypes.

A significant proportion of patients are cured. However, in many types, the aim of treatment is control of the disease, which can be for many years.

The increasingly complex WHO classification of tumours of haemopoietic and lymphoid tissues primarily distinguishes HL from NHL, an umbrella term covering a multiply subclassified spectrum of B- and T-cell malignancies. It was last updated in 2022.

Overall management strategy common to all lymphomas

A suspected diagnosis of lymphoma should always be confirmed by an excision biopsy of the affected area to allow histological, immunological, and molecular analysis. Core biopsy is an acceptable substitute where excision is not possible. Fine-needle aspiration is not adequate and often leads to confusing reports unable to exclude lymphoma. Every histological diagnosis should be confirmed by an expert haemato-pathologist and be reviewed by an MDT. Use of steroids before biopsy should be avoided when clinically possible.

Investigations

After the diagnosis is established, the treatment strategy depends on the outcome of staging investigations that are common to all lymphomas (**Box 39.10**). These investigations allow the clinician and patient to better understand the extent of disease; staging can also determine treatments and inform prognosis, though not in all cases. The Ann Arbor classification is generally used, though historically, this was developed HL. When staging investigations would delay treatment to the detriment of the patient, common sense should be applied.

The tests listed in **Box 39.10** are essential for planning specific therapy. LDH and urate concentrations are helpful, particularly in lymphomas that carry a risk of TLS. An echocardiogram is recommended if anthracyclines are planned because these agents can be cardiotoxic. All patients should be screened for HIV and HBV, without exception. Staging PET/CT scans are recommended for all patients. Bone marrow investigations are recommended for staging in all NHLs but are performed less frequently in HL.

Treatment response, aside from obvious clinic change in symptoms, is assessed at standardized time points during treatment. In most cases, an end of treatment scan is recommended to assess treatment response but also to function as a new baseline scan for any possible future concerns. After treatment, patients with lymphoma should be followed up at regular intervals, which can decrease in frequency as time goes on. Most centres follow up with

Box 39.10 Investigation of the patient with lymphoma

Diagnosis

- Clinical history and examination
- Performance status
- Comprehensive geriatric assessment
- Chest radiograph for mediastinal widening (see **Fig. 39.15**)
- CT/PET scan (see **Fig. 39.16**)
- Bone marrow biopsy (if PET not used to assess bone marrow involvement)
- Blood count, differential, film

For planning specific therapy

- Electrolytes, kidney and liver function
- Serum uric acid
- Virology: HIV, hepatitis B and C
- Cardiac function: e.g., echocardiogram
- Respiratory function: e.g., lung function testing
- Referral for fertility preservation

CT, Computed tomography; *HIV,* human immunodeficiency virus; *PET,* positron emission tomography.

patients for 5 years (i.e., every 3 months for the first year, then every 4 months during the second year, 6 monthly in the third year, and then annually). Patients should be advised to contact the team if they have any concerns between appointments. ***Patient-initiated follow-up*** might be appropriate for certain patient groups.

Hodgkin lymphoma

HL has an incidence of 3 per 100 000 in the Western world. There is a slight male predominance of approximately 1.3 to 1.0. The majority of cases are diagnosed between the ages of 16 and 65 years, with a peak in the third decade of life. It is one of the most common cancers in teenagers. It has led to many countries providing more focused care and designated staff for teenagers and young adults with cancer.

Aetiology

The aetiology of HL is unknown. Positive antibody titres against EBV are found in 40%, providing evidence for an epidemiological link between the two in some cases. Other viruses have not been linked in the same way. Smoking and other lifestyle factors have not been linked, although environmental and occupational exposures to pathogens have been postulated. Large genetic studies have not found any predisposing or causal genetic changes linked to HL to date. It is not an inherited disease.

Diagnosis

HL is divided into subtypes, according to the WHO classification (Box 39.11):

- ***Classic Hodgkin lymphoma (cHL)***, the hallmark of which is the Reed–Sternberg cell (Fig. 39.14). cHL accounts for 90% to 95% of cases and is further subdivided into four distinct categories, with similar treatment strategies.
- ***Nodular lymphocyte-predominant HL (NLPHL)***, characterized by the lymphocyte-predominant or 'popcorn cell'. Reed-Sternberg cells are not seen in NLPHL.

Clinical features

Most patients are asymptomatic and well at diagnosis. The most common presentation of HL is painless lymphadenopathy, most often cervical. The lymphadenopathy has often been present for some time. Lymphadenopathy can be multiple or in other areas. Most disease has limited stage at diagnosis. More unusual presentations are caused by mediastinal lymph nodes and include cough, a feeling of 'pressure' in the chest, and shortness of breath (Fig. 39.15). Occasionally patients present with non-specific symptoms, including constitutional B symptoms or symptoms linked to hepatosplenomegaly. Rarer symptoms directly related to HL, but often not initially recognized as such, are pruritus and alcohol-related pain at the site of lymphadenopathy.

Box 39.11 World Health Organization pathological classification of Hodgkin lymphoma (HL)

Type	Histology classification	Proportion of HL (%)
Nodular lymphocyte-predominant HL		5
Classical HL	Nodular sclerosing	70
	Mixed cellularity	20
	Lymphocyte rich	5
	Lymphocyte depleted	Rare

Courtesy of WHO.

Investigations

These are summarized in Box 39.10. Baseline PET/CT is the gold standard to establish the extent and distribution of metabolically active sites of disease (Fig. 39.16). This enables staging according to the Cotswolds modification of the Ann Arbor classification (Box 39.12). Patients with early-stage disease are then categorized as having favourable or unfavourable disease depending on defined criteria (Box 39.13). The Hasenclever International Prognostic Score has been in use for decades for prognostication (Box 39.14). Stage and other prognostic factors divide patients with HL into three distinct groups (see Box 39.13). These groups determine clearly defined treatment paths for patients with HL.

Management

Principles of management

Patients with HL are invariably treated with a curative intent with a high success rate. In older or frailer patients, treatment modifications or alternative regimens should be reviewed on an individual basis by an experienced haematologist. Patients should be discussed by

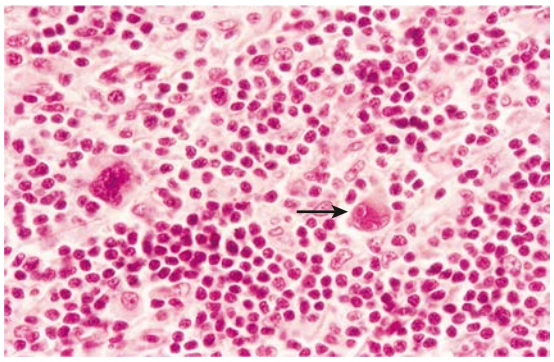


Fig. 39.14 Histological appearance of Hodgkin lymphoma. Scattered mononuclear Hodgkin cells, including a morphologically classical malignant binucleate Reed–Sternberg cell *(arrow)* on a background of benign, normal lymphoid tissue. (*Courtesy of Dr AJ Norton*.)

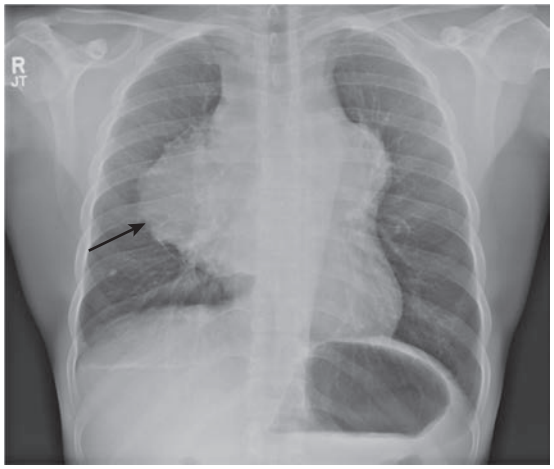


Fig. 39.15 Chest X-ray showing a mediastinal mass *(arrow)* caused by Hodgkin lymphoma.

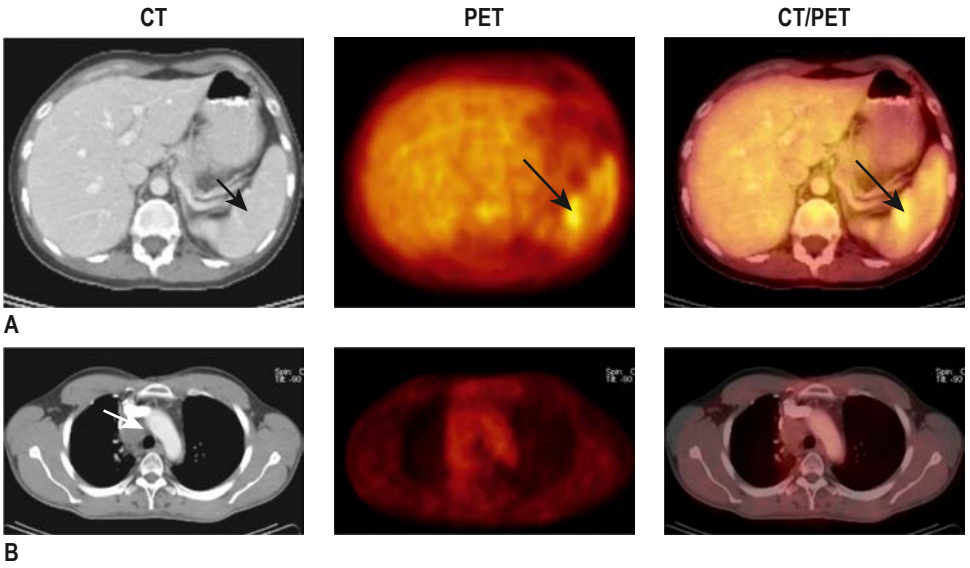


Fig. 39.16 Positron emission tomography (PET) and computed tomography (CT) in the investigation of Hodgkin lymphoma. **(A)** Lymphoma in the spleen *(arrow)* not seen on CT but detected with PET *(centre)* and CT/PET. **(B)** Lymphoma: mediastinal mass on CT scan shown to be metabolically inactive on PET *(centre)* and PET/CT *(right)*. (*Courtesy of Dr N Avril.*)

Box 39.12 Cotswolds modification of Ann Arbor staging of Hodgkin lymphoma

Stage	Description
I	Involvement of a single lymph node region or lymphoid structure (e.g., spleen, thymus, Waldeyer's ring) or involvement of a single extralymphatic site
II	Involvement of two or more lymph node regions on the same side of the diaphragm; localized contiguous involvement of one extranodal organ or site and lymph node region(s) on the same side of the diaphragm (IIE). The number of anatomical regions involved is indicated by a subscript (e.g., II_3)
III	Involvement of lymph node regions on both sides of the diaphragm, which may be accompanied by involvement of the spleen (IIIS), localized involvement of only one extranodal organ site (IIIE), or both (IIISE)
IV	Diffuse or disseminated involvement of one or more extranodal organs or tissues, with or without associated lymph node involvement
Designations applicable to any disease state	
A	No symptoms
B	Fever (temperature >38°C), drenching night sweats, unexplained loss of >10% of body weight within previous 6 months
X	Bulky disease (a widening of the mediastinum by more than one-third or the presence of a nodal mass with a maximal dimension >10 cm)
E	Involvement of a single extranodal site that is contiguous or proximal to the known nodal site

Adapted from Diehl V, Thomas RK, Re D et al. Hodgkin's lymphoma—diagnosis and treatment. Lancet Oncol *2004; 5:19–26, with permission from Elsevier.*

Box 39.13 Hodgkin lymphoma: prognostic groups

Prognostic group	Description
Early favourable	Stage I + II without unfavourable prognostic factors
Early unfavourable	Stage I + II with unfavourable factors
Advanced	The remainder

Box 39.14 Advanced-stage Hodgkin lymphoma (Hasenclever score)

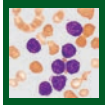
Clinical prognostic factors	Score	
	0	1
Serum albumin	Normal	<40 g/L
Haemoglobin (g/L)	>105	<105
Age	Younger than 45 years	Older than 45 years
Sex	Female	Male
Stage	<IV	IV
Leucocytosis	$<15 \times 10^9$/L	$>15 \times 10^9$/L
Lymphopenia	$>0.6 \times 10^9$/L	$<0.6 \times 10^9$/L
Cumulative score		
Score 0: 84% freedom from progression at 5 years; 89% overall survival Score 1: 77%; 90% Score 2: 67%; 81% Score 3: 60%; 78% Score 4: 51%; 61% Score ≥5: 42%; 56%		

an expert MDT before treatment, clinical picture allowing. There is no change in treatment for patients living with HIV. NLPHL is often CD20+ on histology, hence rituximab should be added to standard chemotherapy. Patients with early-stage favourable disease are often treated with involved-field radiotherapy alone.

Treatment

Frontline treatment

Standard frontline treatment for all patients with cHL, irrespective of prognostic group and staging is with ABVD chemotherapy (doxorubicin, bleomycin, vinblastine, and dacarbazine) if patients



are fit enough to receive this. It is highly emetogenic, but the risks of infertility and secondary malignancy are low. The number of cycles of treatment required depends on staging and prognostic group at diagnosis. The biggest recent change is that HL treatment is now recommended to be response adapted: an early PET/CT scan is done after two cycles of treatment, and depending on the response, the treatment is appropriately adapted. Treatment is escalated where necessary and clinically appropriate, as well as de-escalated when a complete metabolic remission (CMR) is achieved. (Most guidelines recommend dropping the bleomycin from the regimen at CMR.) Depending on treatment response, two to six cycles of ABVD are recommended. Radiotherapy remains in use for localized stage 1 disease, as well as for bulky disease; however, the risks and benefits of radiotherapy should be individually assessed. Radiotherapy is used less now than in the past, mainly because of the superiority of PET/CT to assess response compared with CT. The cure rate is 90% for frontline therapy. For patients who are not fit enough for ABVD, alternative milder treatments are available, often also with significant responses, though cure is less likely.

In NLPHL, rituximab is added to ABVD to target CD20 positive B-cells. It is a different disease entity entirely. Many patients with NLPHL are asymptomatic, and their disease progression is often very slow. Expectant management is often recommended in early-stage, asymptomatic disease, though practices vary.

Salvage treatment

Although most cases patients are cured with frontline therapy, a number of patients either do not respond ('refractory') or relapse. Relapses are most common in the first few years after treatment, though late relapses also happen. Treatment for patients with refractory or relapsed HL remains treatment with a curative intent, though success rates are lower. Often salvage regimens require multiple elective hospital admissions to allow administration. Many different salvage regimens exist; the discussion of these falls outside of the scope of this text. No salvage regimens is clearly superior. Salvage treatment has greater toxicity and risk of late effects. When remission is achieved, strong consideration for consolidation with an autologous SCT (autograft) to achieve a lasting remission should be given to patients who are fit enough to undergo it. Peripheral blood stem cells are taken ('harvested') and stored. The removal of stem cells allows high dose therapy to be given, a treatment most patients would not naturally recover from if their stem cells were not re-infused at a clearly defined time point after high-dose therapy. Different high-dose therapy combinations are used in autografts, most commonly BEAM (carmustine, etoposide, cytarabine, and melphalan) or LEAM (Lomustine, etoposide, cytarabine, and melphalan). Outcomes of autologous SCTs are vastly improved in patients who are in a CR before transplant.

Novel and developing approaches

With such excellent first- and second-line therapies, the number of patients who experience treatment failure is small. However, this subset of patients has a very poor outcome and has been the topic of research, trials, and changing approaches. Brentuximab, an anti-CD30 immunoconjugate, can be used as a bridge to transplant when conventional chemotherapy regimens have failed. Likewise, the use of checkpoint inhibitors pembrolizumab and nivolumab should be considered. Allogeneic SCT (allograft) is only rarely required or considered in patients with HL. It is usually only considered where an autograft has failed. Remission is required before an allograft, which at this stage of disease can be extremely challenging.

Long-term follow-up

The risks of late effects of therapy, particularly second malignancy and cardiac and endocrine problems, require appropriate awareness and surveillance; this may be more significant for patients with HL because they are often at a younger age when diagnosed and treated.

Non-Hodgkin lymphoma (NHL)

The NHLs are a heterogenous, reflected in the WHO classification, last updated in 2022 (**Box E39.1**). A distinction is made between B-cell and T-cell lymphomas, depending on the cell of origin. Approximately 80% of NHL cases are of B-cell origin, with the remaining 20% is T-cell origin. Generally, B-cell lymphomas have a superior prognosis compared with T-cell lymphomas. It is essential a full histological diagnosis is made to inform patient information and treatment.

Aetiology

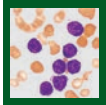
No clear aetiology is known for most NHL, including lifestyle factors. EBV is not clearly linked to NHL with the exception of endemic Burkitt lymphoma, a specific high-grade NHL subtype; sporadic Burkitt lymphoma has a less strong causal link with EBV. Another viral precipitant is HTLV-1 infection, which is casually linked to ATLL. Overall, NHL is seen more often in patients who are immunocompromised for a prolonged period, including patients living with HIV, patients on long-term immunosuppressive agents, and patients with inherited immune deficiency syndromes. Occasionally, bacterial infection may be linked (e.g., *Helicobacter pylori* in gastric mucosa-associated lymphoid tissue [MALT] lymphoma).

Pathogenesis

Lymphocytes develop in the bone marrow, the haematopoietic system, from the haematopoietic stem cells. They go through different stages of development in the bone marrow, and then they are released into the circulation, ultimately residing in the lymphoid tissue, where further maturation happens. Errors in development leading to malignant clonal expansion of lymphocytes can occur at any stage or in any site of a lymphocyte's development. Depending on where the error occurs, this will lead to the different subtypes of lymphoma (see **Fig. 39.2**). In general, neoplasms arising from more mature lymphocytes are more indolent, and those arising in earlier stages of cell development (e.g., lymphoblasts in the bone marrow) are more aggressive. The cell of origin (T or B lineage) and the stage of lymphocytic maturation at which the error occurs lead to the subclassification of NHL.

Genetic features

Unlike HL, multiple cytogenetic changes have been identified in certain NHL subtypes. Burkitt lymphoma was the first haematological malignancy in which a cytogenetic change was shown to involve the translocation of a specific gene: the *MYC* oncogene is translocated from chromosome 8 to near the immunoglobulin heavy chain gene (IGH) on chromosome 14. In most known cytogenetic changes in NHL, there is up- or downregulation of a specific regulator, inhibitor, or protein or pathway promotor in the cell cycle. Other cytogenetic abnormalities associated with NHL are t(14;18) in FL involving upregulation of *BCL2* and t(11; 14) in mantle cell lymphoma (MCL) involving upregulation of the cell cycle



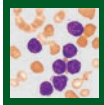
Box E39.1 World Health Organization (WHO) classification of lymphoid neoplasms (2022)

WHO Classification, 5th edition	**Tumour-like lesions with T-cell predominance**
Tumour-like lesions with B-cell predominance	Kikuchi-Fujimoto disease
Reactive B cell–rich lymphoid proliferations that can mimic lymphoma	Indolent T-lymphoblastic proliferation
IgG4-related disease	Autoimmune lymphoproliferative syndrome
Unicentric Castleman disease	**Precursor T-cell neoplasms**
Idiopathic multicentric Castleman disease	*T-lymphoblastic leukaemia/lymphoma*
KSHV/HHV8-associated multicentric Castleman disease	T-lymphoblastic leukaemia/lymphoma, NOS
Precursor B-cell neoplasms	Early T-precursor lymphoblastic leukaemia/lymphoma
B-cell lymphoblastic leukaemias/lymphomas	(Entity deleted)
B-lymphoblastic leukaemia/lymphoma, NOS	**Mature T-cell and NK-cell neoplasms**
B-lymphoblastic leukaemia/lymphoma with high hyperdiploidy	*Mature T-cell and NK-cell leukaemias*
B-lymphoblastic leukaemia/lymphoma with hypodiploidy	T-prolymphocytic leukaemia
B-lymphoblastic leukaemia/lymphoma with iAMP21	T-large granular lymphocytic leukaemia
B-lymphoblastic leukaemia/lymphoma with *BCR*::*ABL1* fusion	NK-large granular lymphocytic leukaemia
B-lymphoblastic leukaemia/lymphoma with *BCR*::*ABL1*-like features	Adult T-cell leukaemia/lymphoma
B-lymphoblastic leukaemia/lymphoma with *KMT2A* rearrangement	Sezary syndrome
B-lymphoblastic leukaemia/lymphoma with *ETV6*::*RUNX1* fusion	Aggressive NK-cell leukaemia
B-lymphoblastic leukaemia/lymphoma with *ETV6*::*RUNX1*-like features	*Primary cutaneous T-cell lymphomas*
B-lymphoblastic leukaemia/lymphoma with *TCF3*::*PBX1* fusion	Primary cutaneous CD4-positive small or medium T-cell lymphoproliferative disorder
B-lymphoblastic leukaemia/lymphoma with *IGH*::*IL3* fusion	Primary cutaneous acral CD8-positive lymphoproliferative disorder
B-lymphoblastic leukaemia/lymphoma with *TCF3*::*HLF* fusion	Mycosis fungoides
B-lymphoblastic leukaemia/lymphoma with other defined genetic abnormalities	Primary cutaneous CD30-positive T-cell lymphoproliferative disorder: Lymphomatoid papulosis
Mature B-cell neoplasms	Primary cutaneous CD30-positive T-cell lymphoproliferative disorder: Primary cutaneous anaplastic large cell lymphoma
Pre-neoplastic and neoplastic small lymphocytic proliferations	Subcutaneous panniculitis-like T-cell lymphoma
Monoclonal B-cell lymphocytosis	Primary cutaneous gamma/delta T-cell lymphoma
Chronic lymphocytic leukaemia/small lymphocytic lymphoma	Primary cutaneous CD8-positive aggressive epidermotropic cytotoxic T-cell lymphoma
(Entity deleted)	Primary cutaneous peripheral T-cell lymphoma, NOS
Splenic B-cell lymphomas and leukaemias	*Intestinal T-cell and NK-cell lymphoid proliferations and lymphomas*
Hairy cell leukaemia	Indolent T-cell lymphoma of the gastrointestinal tract
Splenic marginal zone lymphoma	Indolent NK-cell lymphoproliferative disorder of the gastrointestinal tract
Splenic diffuse red pulp small B-cell lymphoma	Enteropathy-associated T-cell lymphoma
Splenic B-cell lymphoma/leukaemia with prominent nucleoli	Monomorphic epitheliotropic intestinal T-cell lymphoma
Lymphoplasmacytic lymphoma	Intestinal T-cell lymphoma, NOS
Lymphoplasmacytic lymphoma	*Hepatosplenic T-cell lymphoma*
Marginal zone lymphoma	Hepatosplenic T-cell lymphoma
Extranodal marginal zone lymphoma of mucosa-associated lymphoid tissue	*Anaplastic large cell lymphoma*
Primary cutaneous marginal zone lymphoma	ALK-positive anaplastic large cell lymphoma
Nodal marginal zone lymphoma	ALK-negative anaplastic large cell lymphoma
Paediatric marginal zone lymphoma	Breast implant-associated anaplastic large cell lymphoma
Follicular lymphoma	**Nodal T-follicular helper (TFH) cell lymphoma**
In situ follicular B-cell neoplasm	Nodal TFH cell lymphoma, angioimmunoblastic-type
Follicular lymphoma	Nodal TFH cell lymphoma, follicular-type
Paediatric-type follicular lymphoma	Nodal TFH cell lymphoma, NOS
Duodenal-type follicular lymphoma	*Other peripheral T-cell lymphomas*
Cutaneous follicle centre lymphoma	Peripheral T-cell lymphoma, not otherwise specified
Primary cutaneous follicle centre lymphoma	*EBV-positive NK/T-cell lymphomas*
Mantle cell lymphoma	EBV-positive nodal T- and NK-cell lymphoma

Continued

Box E39.1 World Health Organization (WHO) classification of lymphoid neoplasms (2022)—cont'd

In situ mantle cell neoplasm	Extranodal NK/T-cell lymphoma
Mantle cell lymphoma	*EBV-positive T- and NK-cell lymphoid proliferations and lymphomas of childhood*
Leukaemic non-nodal mantle cell lymphoma	Severe mosquito bite allergy
Transformations of indolent B-cell lymphomas	Hydroa vacciniforme lymphoproliferative disorder
Transformations of indolent B-cell lymphomas	Systemic chronic active EBV disease
Large B-cell lymphomas	Systemic EBV-positive T-cell lymphoma of childhood
Diffuse large B-cell lymphoma, NOS	
T-cell/histiocyte-rich large B-cell lymphoma	
Diffuse large B-cell lymphoma/high grade B-cell lymphoma with *MYC* and *BCL2* rearrangements	
ALK-positive large B-cell lymphoma	
Large B-cell lymphoma with *IRF4* rearrangement	
High-grade B-cell lymphoma with 11q aberrations	
Lymphomatoid granulomatosis	
EBV-positive diffuse large B-cell lymphoma	
Diffuse large B-cell lymphoma associated with chronic inflammation	
Fibrin-associated large B-cell lymphoma	
Fluid overload-associated large B-cell lymphoma	
Plasmablastic lymphoma	
Primary large B-cell lymphoma of immune-privileged sites	
Primary cutaneous diffuse large B-cell lymphoma, leg type	
Intravascular large B-cell lymphoma	
Primary mediastinal large B-cell lymphoma	
Mediastinal grey zone lymphoma	
High-grade B-cell lymphoma, NOS	
Burkitt lymphoma	
Burkitt lymphoma	
KSHV/HHV8-associated B-cell lymphoid proliferations and lymphomas	
Primary effusion lymphoma	
KSHV/HHV8-positive diffuse large B-cell lymphoma	
KSHV/HHV8-positive germinotropic lymphoproliferative disorder	
Lymphoid proliferations and lymphomas associated with immune deficiency and dysregulation	
Hyperplasias arising in immune deficiency/dysregulation	
Polymorphic lymphoproliferative disorders arising in immune deficiency/ dysregulation	
EBV-positive mucocutaneous ulcer	
Lymphomas arising in immune deficiency/dysregulation	
Inborn error of immunity-associated lymphoid proliferations and lymphomas	
Hodgkin lymphoma	
Classic Hodgkin lymphoma	
Nodular lymphocyte predominant Hodgkin lymphoma	
Plasma cell neoplasms and other diseases with paraproteins	
Monoclonal gammopathies	
Cold agglutinin disease	
IgM monoclonal gammopathy of undetermined significance	
Non-IgM monoclonal gammopathy of undetermined significance	
Monoclonal gammopathy of renal significance	
Diseases with monoclonal immunoglobulin deposition	
Immunoglobulin-related (AL) amyloidosis	



Box E39.1 World Health Organization (WHO) classification of lymphoid neoplasms (2022)—cont'd

Monoclonal immunoglobulin deposition disease	
Heavy chain diseases	
Mu heavy chain disease	
Gamma heavy chain disease	
Alpha heavy chain disease	
Plasma cell neoplasms	
Plasmacytoma	
Plasma cell myeloma	
Plasma cell neoplasms with associated paraneoplastic syndrome	
POEMS syndrome	
TEMPI syndrome	
AESOP syndrome	

AL, Anaplastic lymphoma kinase; *EBV*, Epstein–Barr virus; *HHV8*, human herpesvirus-8; *Ig*, immunoglobulin; *KSH*, Kaposi sarcoma–associated herpesvirus; *NK*, natural killer; *NOS*, not otherwise specified.

Box 39.15 Chromosome translocations in non-Hodgkin lymphoma

Type	Translocation	Genes	Function
Follicular	t(14; 18)	*IGH/BCL2*	Suppressor of apoptosis
Mantle cell	t(11; 14)	*CCND1/IGH*	Cell cycle regulator
Diffuse large B cell	t(3; 4)	*BCL6*	Cell cycle regulator
Burkitt	t(8; 14)	*MYC/IGH*	Transcription factor
Anaplastic	t(2; 5)	*NPM1/ALK*	Tyrosine kinase
MALT	t(11; 18)	*BIRC3/MALT1*	Suppressor of apoptosis

MALT, Mucosa-associated lymphoid tissue.

regulator cyclin D1. These changes become diagnostic for certain subtypes of NHL and are included in the WHO classification. They very often carry prognostic significance as well. Ongoing developments in gene expression profiling and NGS will continue to identify new molecular subclasses of lymphoma with prognostic significance, and any genetic change identified could also be a potential therapeutic target (Box 39.15).

Clinical features

The presentation can range from patients being completely well to patients with marked constitutional symptoms. Painless lymphadenopathy remains a common presentation (and cannot clinically distinguish between HL and NHL). Lymphoma identified in the CNS is almost invariably NHL and is likely to be a more clinically aggressive type. CNS lymphoma is exceedingly rare in HL.

The most common and clearly distinguished subtypes of NHL are described next.

B-cell lymphoma

See Fig. 39.17.

Diffuse large B-cell lymphoma

Diffuse large B-cell lymphoma (DLBCL) is the most common adult lymphoma worldwide (with increasing incidence with age) and the second most common lymphoma in childhood. It accounts for approximately 30% of all cases. Our understanding of DLBCL's molecular and genetic features has expanded exponentially in the past decade, leading to the description of different subgroups. Gene expression profiling has recognized germinal centre B cell–like (GCB) and activated B cell–like (ABC) subgroups, with the latter having an inferior clinical outcome. There is a focus of research into novel treatments to try to overcome this distinction (Fig. 39.18). Patients with a *MYC*, *BCL2*, and/or *BCL6* rearrangement represent a subtype of DLBCL that does considerably less well with conventional therapy. The increased understanding of variability within the disease may help to define groups of patients who are likely to do less well with conventional treatment and toward whom future targeted therapies may be directed. The development of new agents has happened alongside an understanding of molecular and genetic features that was not available for patients in the past, so historical trial data need to be interpreted with caution.

Clinical features

DLBCL is classed as one of the more aggressive subtypes, or 'high-grade' NHL. More patients have B symptoms, though asymptomatic patients remain common. Intra-abdominal disease presents with bowel symptoms caused by compression or infiltration of the gastrointestinal (GI) tract. In a small proportion, there

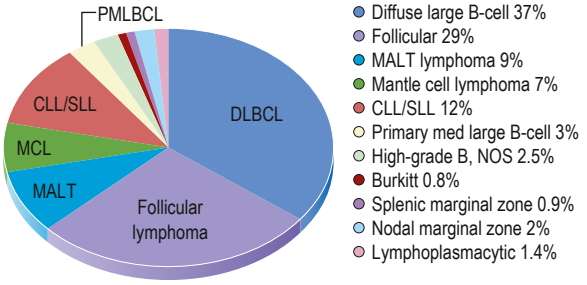


Fig. 39.17 Relative frequencies of B-cell lymphoma subtypes in adults. *CLL/SLL,* Chronic lymphocytic leukaemia/small lymphocytic lymphoma; *MALT,* mucosa-associated lymphoid tissue; *MCL,* mantle cell lymphoma; *NOS,* not otherwise specified; *PMLBCL,* primary mediastinal b-cell lymphoma. (*From Swerdlow SH, Campo E, Harris NL et al [eds].* WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues, *4th edn. Geneva: WHO; 2008, with permission.*)

is a primary mediastinal presentation, with symptoms and signs akin to those of HL. DLBCL is often of higher stage at diagnosis than HL. DLBCL has a rapidly progressive disease course, and treatment when appropriate should be recommended as soon as investigations are complete. In rare cases patients are gravely ill. Patients presenting with life- or organ-threatening disease and initial results confirming NHL but no conclusive subtype should be treated as having DLBCL until proven otherwise, after clear explanation and discussion with the patient. LDH concentrations are often raised.

Management

First-line treatment

Treatment is with a curative intent in patients without significant co-morbidity. The revised International Prognostic Index (R-IPI) is used to inform prognosis and guide treatment (Box 39.16 and Fig. 39.19).

The age of the patient and co-morbidity are critical for both selection of treatment and prognosis.

Six cycles of R-CHOP (**r**ituximab, **c**yclophosphamide, **h**ydroxydaunorubicin, **O**ncovin [vincristine], and **p**rednisolone) have been standard treatment since the early 2000s when rituximab was added to CHOP chemotherapy. However, recently, substituting vincristine with polatuzumab (a CD79b-directed antibody–drug conjugate) has demonstrated superior outcomes for patients with R-IPIs greater than 2. This combination, R-PolaCHP, is now widely used.

In stage IA disease, three cycles of R-CHOP or R-PolaCHP followed by involved-field radiotherapy may be considered, though many centres give six cycles of R-CHOP.

The dose of the anthracycline and other cytostatic agents can be reduced in frail patients ('mini'-R-CHOP). In those with cardiac

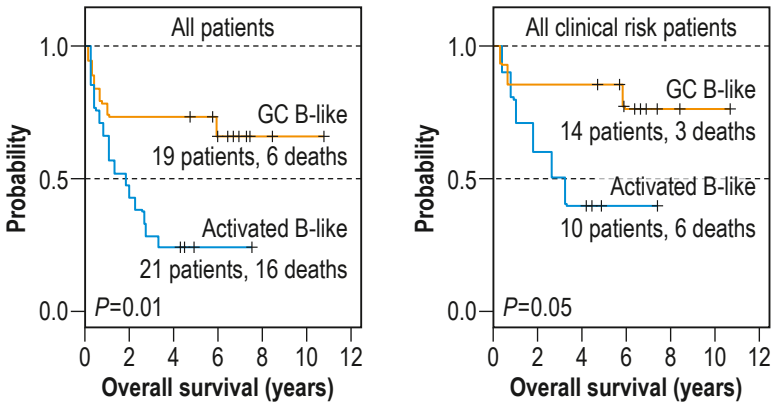


Fig. 39.18 Kaplan–Meier plot of overall survival of diffuse large B-cell lymphoma patients. Patients are grouped on the basis of gene expression profiling. *GC,* Germinal centre cell. (*Reproduced with permission from Alizadeh AA. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling*. Nature *2000; 403:503–511*.)

Box 39.16 Adverse prognostic factors in diffuse large B-cell lymphoma

- Age older than 60 years
- Stage III or IV
- High serum lactate dehydrogenase
- More than one extranodal site involved
- ECOG performance status 2 or more (see Box 41.13)

ECOG, Eastern Cooperative Oncology Group.

impairment, R-GCVP can be considered, in which gemcitabine replaces the anthracycline, doxorubicin. R-maxi-CHOP, in which the cyclophosphamide and daunorubicin doses are increased, is used in patients with mantle cell lymphoma, a type of high-grade NHL. Radiotherapy to bulky disease should be considered on a case-by-case basis at the end of chemotherapy depending on the site and size of the radiotherapy field.

Involvement of the CNS, most often meningeal, is an uncommon but devastating complication with a very poor prognosis. Distinction should be made between primary CNS lymphoma and CNS relapse or involvement of systemic DLBCL. Patients with a high LDH and more than two extranodal sites of disease have a higher risk of CNS involvement and relapse and should be considered for CNS prophylaxis. Historically, three or four injections of intrathecal methotrexate, in addition to R-CHOP, were administered in the earlier treatment cycles. More recently, two or three cycles of high-dose methotrexate (which crosses the blood–brain barrier), alternating with R-CHOP, are preferentially given. Randomized controlled trials are lacking in this area.

Second (and subsequent) therapy

Although the patient may respond to alternative chemotherapy, the prognosis is poor after failure of initial therapy or subsequent progression. The major issues is whether treatment is to be undertaken with palliative or curative intent, and if the latter, what the chances of success are. The ***palliative approach*** involves both chemotherapy and irradiation. The ***curative approach*** involves re-evaluation followed by second-line (salvage) chemotherapy with an approach and regimens identical to salvage treatment for HL (see earlier). If response is achieved with salvage, it should be followed by high-dose therapy and autograft. The best prognosis is for those for whom a second CR can be achieved. When disease does not respond to salvage chemotherapy, clinical trials, including novel agents, should be considered. The prognosis remains poor. CAR T therapy is a rapidly evolving and promising new treatment.

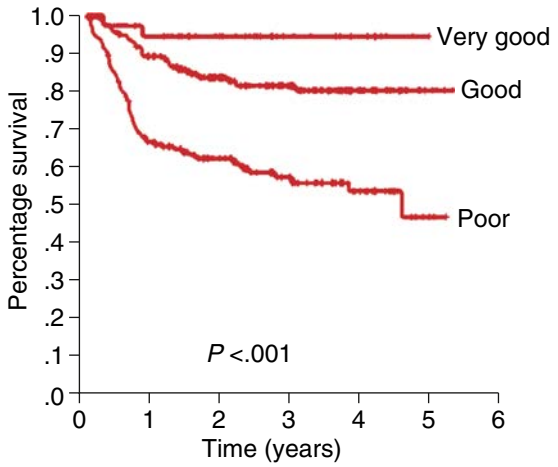


Fig. 39.19 Survival curves for the revised International Prognostic Index (R-IPI) in diffuse large B-cell lymphoma. (*Adapted from Sehn LH, Berry B, Chhanabhai M et al. The revised IPI is a better predictor of outcome than the standard IPI for patients with DLBCL treated with RCHOP*. Blood *109; 1857–1861*.)

Follicular lymphoma

FL is the second most common lymphoma and accounts for approximately 20% of lymphomas worldwide. It is classed as a 'low grade' and often 'indolent' NHL, though within FL, there are further subtypes, which can vary greatly in clinical phenotype. FL is further divided into different grades as defined by morphology (grade 1, grade 2, grade 3a, grade 3b). The grading often correlates with the clinical speed of expected progression. Grade 3b is the highest grade and is now considered equivalent in nature to DLBCL, with the same treatment. In rare cases, FL can transform to high-grade lymphoma; this is prognostically poor.

Clinical features and course

FL occurs in middle to late life; it is very rare in childhood. Most patients present with painless lymphadenopathy, which is often present at more than one site and may have been present for a long time. Results of bloods tests in this group are often normal,

including LDH. In a small proportion, clinical features might be more in keeping with high-grade NHL with the presence of B symptoms and raised LDH. Excision biopsy as with all lymphomas is essential to confirm the diagnosis and establish the grading (see later), and it may reveal transformation to DLBCL. Ideally, the largest lymph node or mass or the most ^{18}F-fluorodeoxyglucose avid on PET/CT should be biopsied if feasible. If a peripheral node is biopsied and shows FL, in the presence of a larger or more avid mass, transformation cannot fully be excluded. Clinical acumen and a repeat biopsy might be indicated in these cases. Bone marrow investigations are done less frequently but remain essential in FL and all low-grade NHL because potential upstaging has significant prognostic and treatment implications.

Most FLs progress very slowly, and in asymptomatic patients, expectant management should be adopted. This can be tricky to explain to the patient. It has been confirmed in large international trials that early treatment in this group of patients is equivalent to a 'watch-and-wait' approach. Treatment, if indicated, is with chemo-immunotherapy ('R-chemo'). FL is not considered a curable condition in most patients, often with a remitting–recurring after diagnosis. However, remissions can be for many years, if not decades. The advent of new combination therapies with favourable side-effect profiles have improved median survival, which is now greater than 10 to 15 years. In older or frailer patients, rituximab can be used as a single agent. The Follicular Lymphoma International Prognostic Index (FLIPI) continues to be used for prognosis and treatment, though it precedes advent of novel therapies (**Box 39.17**).

Management

General management

Asymptomatic patients without organ impairment, 'bulky disease', or evidence of rapid progression or transformation should be managed expectantly. This accounts for around 50% of patients at diagnosis. Up to 15% of patients require no treatment for more than 10 years from diagnosis. Indications for treatment of patients with low-grade NHL are shown in **Box 39.18**. It is vital that an experienced clinician and MDT are involved in the treatment decision. Careful thought should be given to alternative causes for the symptoms, as well as the anxiety of a 'watch-and-wait' approach, which can heighten symptom experience.

Box 39.17 Adverse prognostic factors in non-follicular lymphoma

- Haemoglobin <120 g/L
- More than four nodal sites involved
- Age older than 60 years
- Stage III or IV disease
- Elevated serum lactate dehydrogenase

Box 39.18 Indications for treatment of low-grade non-Hodgkin lymphoma at presentation or progression

- Stage I disease; possibly limited stage II disease
- Rapidly progressive disease
- Organ impairment bone marrow failure, urinary tract obstruction
- Bulky disease (i.e., lymph node mass >7 cm)
- Histological transformation
- Ascites or pleural effusion
- Shared decision-making between patient and healthcare professional

Frontline treatment: early disease

Patients with stage IA or IIA non-bulky disease are treated with involved field radiotherapy, with a possible curative intent. It has a high CR rate (50% of patients will be disease free after 10–15 years).

Frontline treatment: advanced disease (stages II–IV)

Chemo-immunotherapy including an anti-CD20 monoclonal antibody (rituximab or less commonly obinutuzumab) is the treatment of choice, often referred to 'R-chemo'. The combination regimen added to rituximab various widely across the world. R-CHOP and R-bendamustine are most often used. Consideration should be given to O-CHOP (obinutuzumab substituted for rituximab) and O-bendamustine if the FLIPI score is high. Rituximab is widely used with no evidence for a survival advantage between rituximab and obinutuzumab even though progression-free survival at 3 years appears improved with obinutuzumab. Continuing rituximab or obinutuzumab 'maintenance' every 2 months for 2 years improves the progression-free survival, although an overall survival benefit is also not seen. It should be considered that maintenance trials were done in the pre-PET/CT era; it has been postulated that the benefit of maintenance therapy may be greater with PET positivity after R-chemo. Single-agent rituximab is an option for frailer patients.

Second-line therapy and beyond

Patients are managed expectantly in the first instance as long as full re-evaluation, including repeat biopsy, reveals no evidence of transformation. Several options are available, determined by individual patient and the first-line therapy considerations. The same first-line therapy is not usually revisited. Further maintenance treatment (every 3 months) for another 2 years can be given. In second remission, high-dose treatment and autologous SCT should be considered for appropriate patients. Allogeneic HSCT is rarely considered but should not be forgotten as an option in unusually young and fit patients.

Prognosis

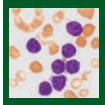
There has been a dramatic improvement in the overall survival of FL since anti-CD20 therapy addition to chemotherapy. The median survival time has been extended well beyond 10 years. Improvements in disease-free survival, after both initial and second-line therapy, are continuing to improve with novel targeted therapies and combinations.

Burkitt lymphoma

Burkitt lymphoma is the most rapidly proliferating lymphoma, with almost invariable, rapid clinical deterioration. It is the most common childhood malignancy worldwide, but it can occur at all ages, with a male-to-female ratio of approximately 3 to 1. There are three subtypes:

- Endemic Burkitt lymphoma is always EBV related and has a similar distribution to malaria. It occurs mainly in children and most frequently presents with a large jaw tumour.
- Sporadic Burkitt lymphoma is linked to EBV in 30% of cases and is the most common type in the Western world.
- Immunodeficiency-related Burkitt lymphoma is seen in patients living with HIV.

Extranodal disease is common at presentation, with renal and testicular disease frequently seen. Both the bone marrow and CNS are often involved; a staging bone marrow biopsy and lumbar puncture are required.



Management

Patients with Burkitt lymphoma are treated with curative intent when possible. The treatment approach is the same across the subtypes. Patients might not be well at diagnosis, though a treatment delay should be avoided because of the aggressive nature of the disease. Patients often have high LDH, with features of TLS before treatment has even started. Rasburicase prophylaxis (3 mg intravenously [i.v.] once a day for up to 7 days) should be given before treatment. The rasburicase treatment dosage (200 mcg/kg i.v. once a day for up to 7 days) should be given if any TLS features are present before starting treatment. Twice-daily monitoring of kidney function and electrolytes is recommended for the first few days of treatment.

The mainstay of treatment has been with alternating cycles of R-CODOX-M (**r**ituximab, **c**yclophosphamide, **O**ncovin [vincristine], **dox**orubicin, and **m**ethotrexate) and R-IVAC (**r**ituximab, **i**fosfamide, etoposide [**V**ePesid], and **C**ytarabine) for more than 2 decades. It is an intense, alternating, non–cross-resistant regimen. The number of cycles required, and whether both regimens are used and alternated depends on risk stratification at diagnosis. All patients should have prophylactic CNS therapy, either with intrathecal methotrexate or cytarabine or by adding high-dose methotrexate. (The i.v. dose of methotrexate in R-CODOX-M is not high enough for CNS penetration.)

Prognosis

In the Western world, the prognosis of Burkitt lymphoma has improved markedly in recent years, most likely because of better support measures and improved environmental factors. Cure rates are 90% for 'low-risk' patients and exceed 60% for 'poor-risk' patients, including those living with HIV, provided all treatment can be administered. Autologous or allogeneic SCT has no role in first remission. The prognosis is poor if remission is not achieved or relapse occurs. In relapsed or refractory Burkitt lymphoma, it is exceptionally difficult to achieve remission. Even if remission is achieved and autologous and/or allogeneic transplant is performed, the likelihood of success remains low.

Mantle cell lymphoma

MCL is one of the less common B-cell lymphomas. It is discussed here because of its specific diagnostic and clinical features (with aggressive and indolent clinical subtypes), as well as its unique treatment algorithm. MCL usually presents later in life and has a male-to-female ratio of 3 to 1. The most frequent presentation is with painless lymphadenopathy, often generalized. There may be non-specific symptoms of tiredness or GI symptoms. B symptoms occur in fewer than 50% of patients. Examination and standard investigation usually confirm generalized lymphadenopathy with or without hepatosplenomegaly. In patients with bowel symptoms, multiple lesions are frequently found on endoscopy. The bone marrow is usually involved, and analysis is recommended when MCL is suspected. Many patients with MCL, though not all, have circulating disease; a lymphocytosis can be the first and often an incidental finding that triggers further investigations.

Cytogenetically, MCL is characterized by a t(11; 14), with dysregulation of cyclin D1. *IGHV* gene mutations and expression of *SOX11* can identify a subtype of indolent disease that can be monitored without the need for immediate treatment. Given that most patients with MCL have an aggressive clinical course and rapid need for treatment, it is surprising there are patients with MCL with distinct low-grade features that might not require treatment for many years.

Management

In most patients, therapy is started after investigation has been completed. A prognostic index (Mantle Cell Lymphoma International Prognostic Index) is used to help determine the best treatment option for each patient. Older, less fit patients are treated with less intense chemo-immunotherapy combinations. In younger and fitter patients, the 'Nordic' regimen is used, an intensive chemo-immunotherapy protocol including alternating cycles of high-dose cytarabine and R-Maxi-CHOP with rituximab, followed by high-dose therapy and autologous SCT (autograft). Maintenance rituximab is commonly given for 2 years after autograft.

Second-line treatment is with a (Bruton's tyrosine kinase [BTK]) inhibitor. It is an oral agent that continues indefinitely or until progression; it is generally very well tolerated.

Historically, if patients progressed after autograft, allogeneic SCT (allograft) was considered, but since the availability of BTK inhibitors as a possible relapse treatment option in these patients, this is done less and less, and the role of allograft is less clear.

Prognosis

Untreated MCL has a natural history that lies between that of DLBCL and FL with features of high- and low-grade lymphoma. It is generally not considered a curable NHL, though if allogeneic SCT is considered, it is done so with curative intent. The prognosis used to be 8 to 12 years in patients well enough to tolerate intensive therapy. However, the advent of the first licensed BTK inhibitor for MCL, ibrutinib, in 2013 has improved the median prognosis, providing a valid treatment option in the relapse setting, also suitable for frailer patients.

Lymphoplasmacytic lymphoma

Lymphoplasmacytic (LPC) lymphoma is a relatively rare type of B-cell lymphoma, also known as ***Waldenström's macroglobulinaemia*** when associated with an immunoglobulin M (IgM) paraprotein. It usually occurs in later life. The presentation is often with cytopaenias related to bone marrow infiltration and the typical presence of a (mostly) IgM paraprotein. High levels of the IgM paraprotein can cause serum hyperviscosity symptoms, such as headaches, visual disturbance, dizziness, confusion, or seizures; they can be the presenting feature. Lymphadenopathy is typically minimal; splenomegaly is more common.

Management

Treatment depends on symptoms and cytopaenias. Many patients are asymptomatic and can be managed expectantly. In the presence of symptoms, rituximab-containing regimens are recommended. R-Bendamustine and DRC (dexamethasone, rituximab, and cyclophosphamide) are most commonly used. Caution should be given to using rituximab in the first cycle in patients with high tumour burden or high IgM paraprotein because there can be transient increase in the paraproteinaemia after treatment has been given. It is common to omit rituximab from cycle 1, adding it back in for subsequent cycles.

Although hyperviscosity is rare, it is an emergency, and patients should undergo urgent plasma exchange at the first instance.

Ibrutinib is used in patients with relapsed or refractory disease and continues indefinitely. The role of autograft should continue to be considered in second remission in appropriate patients, as should allogeneic transplant further down the line.

Prognosis

LPC lymphoma may also follow an indolent course for many years without therapy. It is usually chemotherapy and immunotherapy

sensitive. The 5-year survival rate is around 78%. The advent of BTK inhibitors will no doubt continue to improve prognosis.

Primary central nervous system lymphoma

Primary CNS lymphoma is histologically a diffuse large B-cell lymphoma. It occurs in patients that are immunocompetent (predominantly aged 55–70 years) and in patients that are immunosuppressed (mainly in the context of HIV infection or after solid organ transplantation). Its incidence has dramatically decreased since the introduction of highly active anti-retroviral therapy for patients with HIV. It presents with symptoms caused by single or multiple parenchymal mass lesions, often mimicking a stroke. The diagnosis is confirmed on a biopsy, particularly in those who are immunocompromised in whom an infectious aetiology, such as toxoplasmosis, is possible. A contrast-enhanced MRI scan is the first choice of imaging. CSF, if an LP is safe to perform, is usually normal. Further investigation is necessary to exclude generalized disease, hence staging PET/CT is recommended.

Management

Intensive inpatient chemotherapy is recommended for patients who are fit enough. It is given with curative intent. Different regimens are in use. Methotrexate is the key element of all regimens and needs to be given at a high enough dose to cross the blood–brain barrier. It is only since high-dose methotrexate has been used that cure rates are seen. Methotrexate is often given in combination with cytarabine and thiotepa, with rituximab routinely added. Consolidation with autologous SCT is recommended in those fit enough with adequate response to treatment. Whole-brain radiotherapy is used in less fit or relapsed patients.

Marginal zone lymphoma

This is an uncommon B-cell lymphoma associated with chronic inflammation or infection. (MALT is the commonest form of marginal zone lymphoma.) MALT of the salivary glands is linked with Sjögren's syndrome. Gastric MALT lymphoma is closely associated with *H. pylori* infection and presents with symptoms of gastric ulceration or a mass, indigestion, or bleeding; the diagnosis is made by endoscopic biopsy to include confirmation of both lymphoma and *H. pylori* status. Careful assessment of extranodal sites is required.

Management

Management depends on disease localization and symptoms. *Helicobacter* eradication therapy (see Chapter 22, 'Eradication therapy' section) is the treatment of choice for patients with 'low-grade' gastric MALT and almost invariably alleviates symptoms. Re-evaluation is carried out after 6, 12, and 24 months with endoscopy and repeat biopsies. In most patients, a conservative approach is followed because responses may take many months, and rapid progression is very unlikely. Failure is uncommon, but if it does occur, biopsy should be repeated. If the histology is unchanged after further *Helicobacter* eradication therapy, rituximab in combination with chemotherapy, single-agent rituximab, or low-dose involved-field radiotherapy is likely to be effective and possibly curative. Overall, the prognosis is very good, with a large proportion of patients being alive 10 years after diagnosis.

T-cell lymphoma

T-cell lymphomas can present in nodal, extranodal, cutaneous, or predominantly leukaemic forms (Fig. 39.20). They are a large group of rare diseases, less common than their B-cell counterparts, and varying in behaviour from indolent to aggressive. Many subtypes carry poor prognoses because of disappointing responses to treatment.

The two most common subtypes of nodal T-cell lymphoma are '*peripheral T-cell lymphoma, not otherwise specified (NOS)*' and '*angio-immunoblastic T-cell lymphoma*', which together account for about 50% of T-cell lymphomas. Both are rare at a young age. Most patients present with lymphadenopathy. In contrast to B-cell lymphomas, B symptoms are very common, and most patients have constitutional symptoms. Patients with angio-immunoblastic T-cell lymphoma can also present with inflammatory symptoms of fever, rash, and arthritis. *Anaplastic T-cell lymphoma* (ALCL), particularly when anaplastic lymphoma kinase (ALK) positive, has a better prognosis than other subtypes.

Management

After standard investigation, which usually reveals widespread disease, patients are treated with cyclical combination chemotherapy as for DLBCL. CD20 is not expressed on T cells, so rituximab is not used; as yet, there is no equivalent drug used routinely in T-cell lymphoma. The exception is the use of brentuximab vedotin, a CD30 antibody–drug conjugate, in combination with chemotherapy, in the treatment of patients with systemic ALCL, an aggressive subtype of T-cell lymphoma characterized by the uniform expression of CD30. Symptoms almost invariably resolve, although they may recur between cycles. Overall, the outcome of treatment is worse than for DLBCL in terms of quality of response, duration of response, and overall survival. Other than in ALK-positive ALCL, autologous SCT or indeed allograft should be considered in first remission. Second-line therapy is rarely satisfactory; patients should be considered for allogeneic SCT if a second remission is achieved, though this is rare.

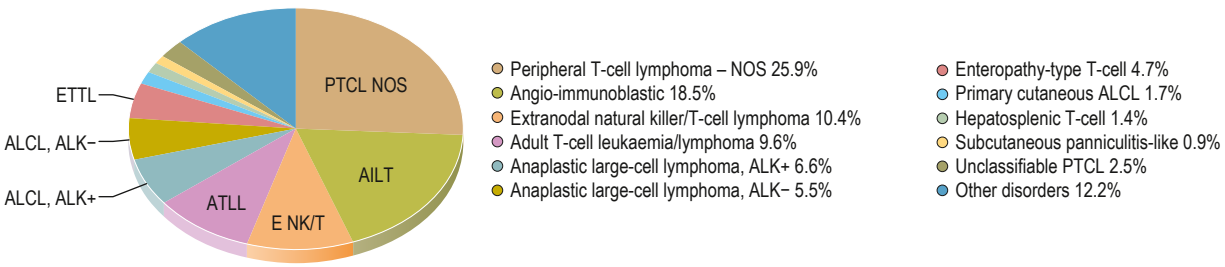
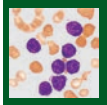


Fig. 39.20 Relative frequencies of T-cell lymphoma subtypes in adults. *ALCL,* Anaplastic large-cell lymphoma; *ALK,* anaplastic lymphoma kinase; *ETTL,* enteropathy-type T-cell lymphoma; *NOS,* not otherwise specified; *PTCL,* peripheral T-cell lymphoma. (*Reproduced with permission from WHO. Swerdlow SH, Campo E, Harris NL et al [eds].* WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues, *4th edn. Geneva: WHO; 2008;Fig 8.07.*)



Mycosis fungoides and Sézary's syndrome

Mycosis fungoides is the most common cutaneous lymphoma. It predominantly arises in and is confined to the skin, spreading to other organs only later in the disease. The rare leukaemic variant is known as Sézary's syndrome. It has a long natural history, presenting with multiple erythematous lesions, plaques, and tumours. Often multiple skin biopsies are required to establish a diagnosis. A wide range treatments from topical to systemic treatments are used depending on the extent of disease. It is rare for patients to achieve a durable response, and the prognosis is poor. Patients with mycosis fungoides and Sézary's syndrome should be looked after in a specialist centre.

Chronic lymphocytic leukaemia

CLL is the most common type of leukaemia, increasing in incidence with age; the median age of presentation is around 70 years. CLL results from the clonal expansion of small, mature B lymphocytes with a characteristic immunophenotype. Most patients are asymptomatic and diagnosed after an incidental finding of a raised WCC on a blood test done for another indication. The WCC can be significantly raised. Some patients have symptoms of marrow failure or immunosuppression.

CLL most commonly presents in 'leukaemic phase', with bone marrow or blood involvement. It can also present in non-leukaemic form with isolated localized lymphadenopathy only, for which the terminology is small lymphocytic lymphoma (SLL), blurring the line between leukaemia and lymphoma. This terminology can be confusing. It is important to note that CLL and SLL are different clinical presentations of the same disease and are managed the same way. Most clinicians use the terminology CLL/SLL to reflect this and avoid confusion.

Nearly all cases of CLL are preceded by a monoclonal B-cell lymphocytosis (MBL) when there are fewer than 5×10^9/L circulating clonal B cells. However, this condition is also present in approximately 12% of the healthy population. The relationship between MBL and CLL is analogous to that between monoclonal gammopathy of uncertain significance (MGUS) and myeloma. MBL is not considered a neoplasm.

Although CLL is not considered a curable condition, the prognosis is good for most patients. Before the development and availability of many novel agents in the past decade, the median survival time was 10 years. Novel treatments and treatment approaches have no doubt improved the prognosis further (and will continue to do so), though mature data are not available yet.

Clinical features

Most patients are asymptomatic at presentation. If present, symptoms are non-specific. Common symptoms are:

- Recurrent infection caused by leukopenia and immune suppression: reduced immunoglobulins)
- Anaemia caused by marrow infiltration or haemolysis: often clinically obvious
- Painless small-volume lymphadenopathy: easily palpable and often present for long periods of time before diagnosis
- Splenomegaly related symptoms: left upper quadrant discomfort or early satiety.
- Constitutional symptoms: similar to B symptoms

Investigations

- ***FBC*** reveals a normal or low Hb, a raised WCC, (almost exclusively a lymphocytosis), and normal or low platelets.
- ***Blood film*** demonstrates small- or medium-sized mature and normal-appearing lymphocytes. Smudge cells may be seen *in vitro.* No immature forms (blasts) are evident, distinguishing CLL from ALL.
- ***Bone marrow*** reflects the peripheral blood findings. Often the marrow is heavily infiltrated with mature lymphocytes, with remarkable preservation of Hb and platelet count.
- ***Immunophenotyping*** shows a characteristic immunophenotype. The presence of specific markers give a score (the CLL score) that indicates the likelihood of CLL.
- ***Cytogenetics and FISH analyses*** can identify abnormalities, with known prognostic value. Abnormalities are seen in up to 90% of patients.
- ***DAT screen, direct Coombs test, and other haemolysis markers*** may be positive if there is haemolysis. CLL has a known association with (secondary) Evans syndrome, a rare disease condition presenting with a combination of autoimmune haemolytic anaemia and immune thrombocytopaenia.
- ***Immunoglobulins*** are low or normal.

Prognosis

The prognosis of CLL correlates with clinical stage at presentation (**Box 39.19**). Although different staging systems are used

Box 39.19 The Rai and Binet staging systems for chronic lymphocytic leukaemia[a]

Stage	Risk	Manifestations	Recommended treatment
Rai staging system			
0	Low	Lymphocytosis	Watch and wait
I	Intermediate	Lymphadenopathy	Treat only with progression[b]
II	Intermediate	Splenomegaly, lymphadenopathy or both	Treat only with progression[b]
III	High	Anaemia, organomegaly	Treat in most cases
IV	High	One or more of the following: anaemia, thrombocytopenia and organomegaly	Treat in most cases
Binet staging system			
A	Low	Lymphocytosis, <3 lymphoid areas enlarged[c]	Watch and wait
B	Intermediate	≥3 lymphoid areas enlarged[c]	Treat only with progression
C	High	Anaemia, thrombocytopenia, or both	Treat in most cases

[a]Lymphocytosis is present in all stages of the disease.
[b]Progression: weight loss, fatigue, fever, massive organomegaly, rapidly increasing lymphocyte count.
[c]Lymphoid areas include the cervical, axillary, and inguinal lymph nodes; spleen; and liver.

worldwide, key prognostic elements are not the presenting WCC but the speed at which it changes ('doubling time'), as well as the presence (or absence) of an associated anaemia or thrombocytopaenia. Several cytogenetic and molecular abnormalities also have prognostic significance; for example, the significant finding of a 17p deletion (which often coincides with a TP53 mutation) influences choice of therapy because it has a known association with rapid progression and poor response to certain treatments. NGS has recently identified additional gene mutations that influence prognosis.

Management

Asymptomatic patients with early-stage CLL should be managed expectantly. Up to 30% of CLL patients never require treatment. The most important decision is to know if or when to treat. Those with advanced-stage disease, either at diagnosis or if this develops during monitoring of the patient, should be treated. Although patient factors, such as age and fitness, play an important role, general indications for treatment are:

- Bone marrow failure with worsening anaemia or thrombocytopenia that is not autoimmune
- Massive or progressive splenomegaly or lymphadenopathy
- Progressive disease demonstrated by doubling of the lymphocyte count in less than 6 months
- Systemic symptoms: fever, night sweats, or weight loss
- Presence of haemolysis or other immune-mediated cytopaenias that are poorly responsive to steroids.

General and supportive measures

Patients with CLL with an associated autoimmune haemolytic anaemia should be treated with steroids in the first instance. Those with anaemia and thrombocytopenia caused by marrow infiltration should be treated with CLL-specific treatments.

Patients with infections should be treated as indicated; prophylactic antibiotic, antiviral, anti-*Pneumocystis*, and antifungal therapy should be considered for higher-risk patients during periods of treatment. Immunoglobulin replacement can be considered in patients with CLL with hypogammaglobulinaemia and recurrent infections, though availability of immunoglobulin products has dwindled in recent years. Pneumococcal, *Haemophilus influenzae* type B (Hib), influenza, and SARS-CoV-2 vaccinations are recommended. Live vaccines should be avoided. Allopurinol is given to prevent hyperuricaemia.

Treatments

The choice of therapy depends on patient-specific factors, including adverse prognostic features, anticipated response, and toxicities of therapy. The range of therapies has increased exponentially in the past decade. The availability of treatments and combinations is variable. Treatments are tailored to the patient's age, fitness, prior therapy, and co-morbidities.

- The ***BTK inhibitors*** ibrutinib and acalabrutinib, first approved in 2014, have greatly improved and forever changed the treatment approach for many patients with CLL. They are well-tolerated oral treatments that can often overcome poor prognostic cytogenetic findings. They are taken indefinitely.
- ***Fludarabine***, a purine analogue, in combination with cyclophosphamide and rituximab, was the standard first-line treatment for many years. Although novel therapies have superseded this for many patients, it may still be used for younger and fitter patients and remains a valid (and cheap) treatment option.
- ***Venetoclax***, a BCL-2 inhibitor, has incredible response rates and is often used as a second-line treatment option.
- ***'Old-fashioned'*** drugs, such as chlorambucil and bendamustine, continue to play a role, especially in refractory patients and in those intolerant of more novel agents.
- The role of allogeneic transplant in the era of novel therapies is slowly disappearing.

Richter's transformation

In rare cases, CLL can undergo transformation to high-grade lymphoma, usually DLBCL. Treatment is similar to that of DLBCL. This is a feared complication with a poor prognosis.

Hairy cell leukaemia

Hairy cell leukaemia (HCL) is a clonal proliferation of abnormal B (or very rarely T) cells, which accumulate in the bone marrow and spleen. It is a rare disease with a median age at presentation of 52 years. There is a male-to-female ratio of 4 to 1.

The name relates to the typical appearance of the WBCs down the microscope on both blood film and bone marrow.

HCL is almost always an indolent disease with a very slow course of disease progression. It can have a delayed diagnosis because findings on routine blood tests are minimal, with often only a minor neutropaenia in the first instance. Despite the striking features of the 'hairy cell', they are only present in low numbers on blood film and hence can easily be missed. Most patients are asymptomatic.

Clinical features

Clinical features can include anaemia, recurrent infections secondary to neutropaenia, and abdominal discomfort because of splenomegaly. Splenomegaly occurs in 80% of patients and can be massive. Constitutional symptoms and lymphadenopathy are rare. An unusual but typical finding in HCL is a peripheral blood monocytopaenia.

Management

The purine analogues cladribine and pentostatin have specific activity in this condition and are ideally given in combination with rituximab. A rituximab–cladribine combination is most used. It is a 5-day treatment, administered subcutaneously, leading to a CR in 95% of patients with just one cycle of treatment. Remissions can last for many years, and patients can be successfully retreated with the same treatment. Splenectomy is no longer routinely recommended.

Further reading

Walewska R, Parry-Jones N, et al. Guideline for the treatment of chronic lymphocytic leukaemia. *Br J Haematol*. 2022;197:544–597.

MYELOMA AND OTHER PLASMA CELL DISORDERS

The classification of this group of related conditions is presented in **Box 39.20**; the most significant is myeloma.

Myeloma

Myeloma (or multiple myeloma) is a cancer of plasma cells. These cells usually produce immunoglobulin, but with malignant plasma cells, this is a non-functioning monoclonal paraprotein (or 'M band'). The paraprotein is usually an intact immunoglobulin, most commonly IgG (55%) or IgA (20%) and more rarely IgM or IgD. There is normally also an imbalance in free light chains (kappa or lambda),

Box 39.20 World Health Organization classification of plasma cell neoplasms (2017)

- Monoclonal gammopathy of undetermined significance
- Plasma cell myeloma
- Plasma cell myeloma variants:
  - Smouldering (asymptomatic) plasma cell myeloma
  - Non-secretory myeloma
  - Plasma cell leukaemia
- Plasmacytoma:
  - Solitary plasmacytoma of bone
  - Extraosseous plasmacytoma
- Monoclonal immunoglobulin deposition diseases
  - Primary amyloidosis (see Chapter 26)
  - Light chain diseases
  - Heavy chain diseases
    - Gamma (γ) heavy chain disease
    - Mu (μ) heavy chain disease
    - Alpha (α) heavy chain disease

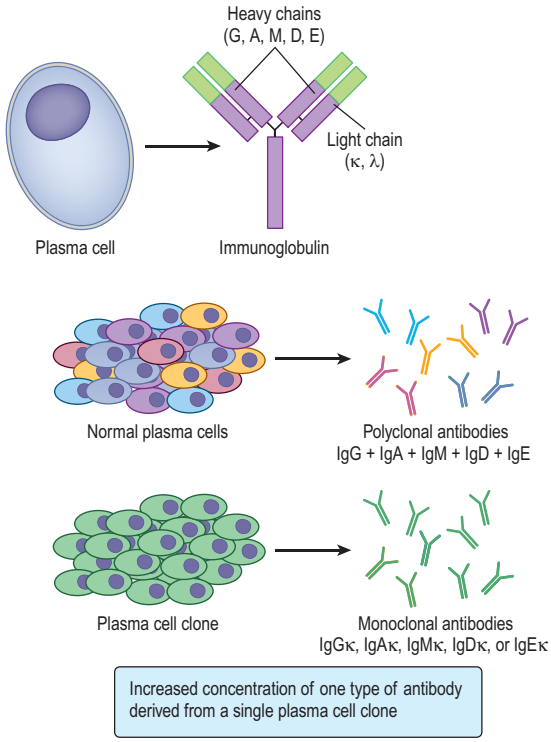


Fig. 39.21 Plasma cells produce immunoglobulins (Igs; antibodies) with different heavy- and light-chain combinations. A mixed population of plasma cells produces a polyclonal mixture of different immunoglobulins. In myeloma, the malignant plasma cell population is derived from a single clone, which produces a single type of Ig that can be detected by blood or urine assays.

which can be detected in the blood (serum free light chains [SFLCs]) and urine (Bence Jones protein) (Fig. 39.21). Around 15% of patients have no intact immunoglobulin and secrete an excess of light chain only, and 5% of cases are non-secretory.

Other conditions also produce paraproteins. B-cell lymphoproliferative disorders, such as low-grade B-cell NHL and CLL, can produce a paraprotein, and Waldenström's macroglobulinaemia is an IgM-producing lymphoplasmacytic lymphoma. Non-malignant disorders producing a paraprotein include poorly controlled HIV infection and autoimmune conditions, although the proteins are likely

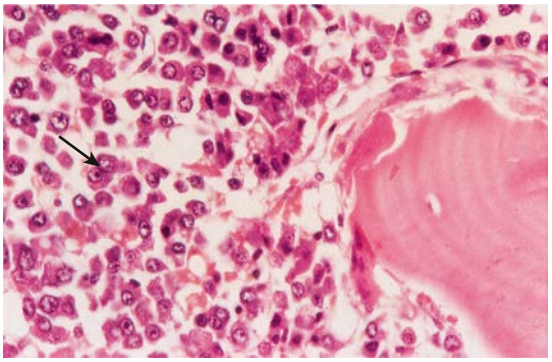


Fig. 39.22 Multiple myeloma. Histology shows replacement of the medullary cavity by abnormal plasma cells with some binucleate forms *(arrow)*. A residual bony trabeculum is present toward the *right*. (*Courtesy of Dr AJ Norton*.)

Box 39.21 Life-threatening complications of myeloma

- ***Kidney injury pr failure:*** requires urgent attention; renal replacement therapy may be indicated; haemodialysis
- ***Hypercalcaemia:*** treat with rehydration and bisphosphonates (e.g., pamidronate)
- ***Spinal cord compression:*** treat with dexamethasone followed by radiotherapy
- ***Hyperviscosity:*** caused by high circulating concentrations of paraprotein; can be managed with plasmapheresis

to be transient and should resolve with treatment of the underlying condition.

The abnormal plasma cell population can be identified on bone marrow aspirate or trephine biopsy, with the imbalance of kappa and lambda restricted plasma cells demonstrated by flow cytometry (aspirate) or immunohistochemistry (biopsy) (Fig. 39.22).

Myeloma is predominantly a disease of older adults, with the median age at presentation being older than 60 years. It is rare before 40 years of age. The annual incidence is 4 per 100 000, and disease is more common in males and in those of Black ethnicity.

Clinical features

The clinical features include:

- Bone pain: most commonly backache, owing to vertebral involvement (60%)
- Symptoms of anaemia
- Recurrent infections
- Symptoms of kidney failure (20%–30%)
- Symptoms of hypercalcaemia
- Rarely, symptoms of hyperviscosity or bleeding because of thrombocytopenia.

Patients can be asymptomatic, with abnormalities picked up on routine blood tests. Life-threatening complications are shown in Box 39.21.

The clinical features of myeloma are related to the underlying pathology:

- Bone marrow infiltration with plasma cells: anaemia, neutropenia, thrombocytopenia
- Paraprotein secretion. A reduction normal immunoglobulins (immune paresis) contributes to recurrent infections (Fig. 39.23). Very high paraprotein concentrations can cause symptoms of serum hyperviscosity, such as headaches, visual disturbance, dizziness, confusion, or seizures.

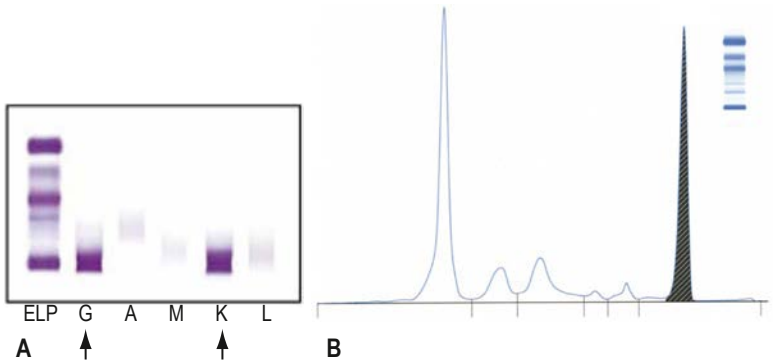


Fig. 39.23 Serum protein electrophoresis and immunofixation. **(A)** Immunofixation of immunoglobulin G (IgG) kappa monoclonal paraprotein in a patient with an IgG K paraprotein. The *column* on the *far left* is the total serum electrophoresis stained with anti-protein reagents. Albumin is at the *top* and the increased immunoglobulins at the *bottom*. These are shown to stain with anti-IgG and anti-kappa light chain reagents *(arrows)*. **(B)** Quantitation and serum electrophoresis of IgG paraprotein. The paraprotein is quantified *(right)* by calculating the area under the paraprotein spike and comparing it with the albumin concentration. *A,* Immunoglobulin A; *ELP,* electrophoresis; *G,* immunoglobulin G; *K,* kappa; *L,* lambda; *M,* immunoglobulin M.

- Free light chain secretion. This is often deposited in the renal tubules, causing kidney impairment by cast nephropathy. Other factors such as hypercalcaemia and (rarely) the deposition of AL amyloid can also contribute to kidney injury (see Chapter 26).
- Bone disease. Dysregulation of bone remodelling leads to typical lytic lesions, usually in the spine, skull, long bones, and ribs (Fig. 39.24). There is increased osteoclastic activity without increased osteoblast formation of bone, causing long bone fractures, vertebral collapse, and hypercalcaemia. Soft tissue plasmacytomas also occur and are the usual cause of spinal cord compression.

Diagnosis

The spectrum of plasma cell disorders is classified based on the degree of plasma cell infiltration and the presence of end-organ damage (Box 39.22). The presence of a paraprotein does not in itself make a diagnosis of myeloma. A diagnosis of myeloma requires more than 10% plasma cell infiltrate or a biopsy-proven bony or extramedullary plasmacytoma, together with at least one myeloma-defining event, which include 'CRAB' and biomarker criteria:

- Hyper***c***alcaemia: calcium greater than 2.75 mmol/L
- ***R***enal impairment: creatinine clearance less than 40 mL/min or serum creatinine greater than 177 μmol/L
- ***A***naemia: Hb less than 100 g/L or greater than 20 g/L below normal
- ***B***one involvement: one or more osteolytic lesions of 5 mm or larger on X-ray, CT, or PET scan

Biomarker criteria include:

- Involved-to-uninvolved (e.g., kappa–lambda) SFLC ratio of 100 or greater
- Bone marrow plasma cell infiltration of 60% greater than
- Two or more focal lesions on MRI scan.

Other related manifestations of end-organ damage (hyperviscosity, recurrent bacterial infections, AL (amyloid light chain) amyloidosis, peripheral neuropathy) are non-specific and not diagnostic of myeloma.

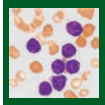
Investigations

General

- ***FBC:*** Hb, WCC, and platelet count are normal or low.
- ***Erythrocyte sedimentation rate*** is often high.

Fig. 39.24 Pathological fracture of long bone. A plain X-ray showing a spiral fracture through the right humerus.

- ***Blood film:*** rouleaux formation because of the paraprotein. Circulating plasma cells are present in the aggressive plasma cell leukaemia variant of myeloma.
- ***Serum creatinine and electrolytes:*** kidney injury (see earlier).
- ***Serum calcium:*** normal or raised. Serum alkaline phosphatase is usually normal.



Box 39.22 Classification and definitions of plasma cell disorders (IMWG, 2016)

	MGUS	Smouldering myeloma	Active myeloma
Clonal plasma cells in bone marrow (%)	<10	10–60	>10
Presence of myeloma-defining events	None	None	Yes
Likelihood of progression to myeloma	~1% per year	~10% per year	

IMWG, International Myeloma Working Group; *MGUS*, monoclonal gammopathy of undetermined significance.

- ***Total protein:*** normal or raised.
- B2M and LDH are used in prognosis. B2M is a marker of renal impairment and cell turnover.

Immunology

- ***Serum protein electrophoresis and immunofixation:*** characteristically shows a monoclonal band and immune paresis (see Fig. 39.23). The SFLC assay may show an abnormal ratio and an increased total amount of free kappa or lambda chains.
- ***Urine electrophoresis and immunofixation:*** assess light-chain excretion, although this is now superseded by the SFLC assay.

Radiology

- ***Skeletal imaging:*** CT, CT/PET, or whole-body MRI shows characteristic lytic lesions. Plain X-ray skeletal surveys are outdated because of poor sensitivity, although plain films may pick up a presenting pathological fracture.
- ***MRI of the spine:*** useful if there is back pain; it may show imminent compression or collapse. It can also pick up occult lesions not seen on CT (as per updated diagnostic criteria).

Histology

Bone marrow aspirate or trephine shows characteristic infiltration by clonal plasma cells (see Fig. 39.22).

Staging and prognosis

Patients can be risk stratified using the International Staging System (ISS) based on concentrations of albumin and B2M:

- ***ISS stage I:*** B2M less than 3.5 mg/L and serum albumin 35 g/L or greater
- ***ISS stage II:*** neither stage I nor stage III
- ***ISS stage III:*** B2M 5.5 mg/L or greater

Cytogenetic abnormalities can be detected via FISH in most cases of myeloma and have a bearing on prognosis. Hyperdiploidy is associated with a good prognosis, but some translocations affecting the IgH locus on chromosome 14 (e.g., t(4;14), t(14;16), t(14;20)) are associated with poor prognosis, as are other findings such as 17p–, 1q+, and 1p–. The most important of these, together with LDH concentrations, have been incorporated into a revised ISS staging system (R-ISS):

- ***R-ISS I:*** ISS stage I *and* normal LDH *and* no del(17p), t(4;14) or t(14;16) by FISH
- ***R-ISS II:*** neither stage I nor stage III
- ***R-ISS III:*** ISS stage III *and* elevated LDH *and/or* del(17p), t(4;14) or t(14;16) by FISH

This system is based on clinical trial data and clearly divides patients by overall survival: 10 years, 6 years, and 1.3 years for R-ISS groups I, II, and III respectively.

Management

Supportive therapy

- ***Anaemia:*** blood transfusion. In patients with very high paraproteins, there is a risk of precipitating or worsening symptoms of hyperviscosity.
- ***Hyperviscosity:*** plasmapheresis
- ***Hypercalcaemia and kidney injury:*** hydration, bisphosphonates, and high-dose steroids
- ***Infection:*** prompt treatment with antibiotics
- ***Bone pain:*** radiotherapy, high-dose dexamethasone, and systemic chemotherapy. NSAIDs are avoided because of the risk of acute kidney injury. Bisphosphonates such as zoledronate, which inhibit osteoclast activity, treat patients with hypercalcaemia and help reduce pathological fractures, cord compression, and bone pain.
- ***Pathological fractures:*** prevention by pinning of lytic bone lesions at critical sites such as the femoral shaft. Kyphoplasty and vertebroplasty may be useful in treating vertebral fractures.

Specific therapy

Although myeloma remains incurable for most patients, there are now a large number of drugs that can be used in combination and in sequence to prolong survival and reduce complications.

- ***Steroids.*** Dexamethasone and prednisolone are included in combination therapies and in initial management of acute complications.
- ***Cytotoxic chemotherapy drugs*** (e.g., melphalan and cyclophosphamide). Melphalan is used as high-dose therapy in autologous SCT.
- ***Proteasome inhibitors.*** Bortezomib was the first of these agents and is commonly given subcutaneously. Carfilzomib, an irreversible proteasome inhibitor, may have greater efficacy but can have cardiotoxic side-effects. Ixazomib is a newer oral agent.
- ***Immunomodulatory drugs.*** Thalidomide, lenalidomide, and pomalidomide act in conjunction with the cereblon protein to alter the transcriptional profile of myeloma cells, leading to cell death. They can be given over an extended period.
- ***Monoclonal antibodies.*** Daratumumab and isatuximab (anti-CD38), as well as elotuzumab (anti-SLAMF7), have anti-myeloma activity, particularly when given in combination with other agents.
- ***New and experimental immunotherapies.*** These include antibody–drug conjugates, bispecific T-cell engagers, and CAR T therapies. These are likely to be powerful tools in the future.

Anti-myeloma agents typically work most effectively when given in combination: daratumumab, bortezomib, thalidomide, and dexamethasone are highly effective when used together. Combinations are selected based on efficacy, side-effect profile, previous treatment history, and funding availability.

Younger patients who are fit enough for intensive chemotherapy will typically have an autologous SCT in their first and sometimes second remission to consolidate a good response to therapy. This typically increases the duration of remission. Low-level maintenance chemotherapy after high-dose therapy with stem cell rescue further extends the duration of response and overall survival. Myeloma inevitably relapses after each initial response, but there are now many effective treatments for patients with relapsed disease. Each subsequent remission tends to be less sustained with each treatment course as the cancer becomes more resistant to treatment.

Immunological therapies have the potential to cure a small subset of patients, with the use of allogeneic transplant sometimes considered for young patients with high-risk disease. Newer immunotherapies including CAR T, bispecifics, and antibody–drug conjugates are likely to transform future treatment algorithms.

Asymptomatic (smouldering) myeloma

In this condition, there is a significant paraprotein (IgG or IgA >30 g/L or urinary light chain excretion >0.5 g/day) or a marrow plasma cell infiltration of more than 10% but no end-organ damage or biomarkers of symptomatic myeloma. Most, but not all, patients eventually progress to symptomatic myeloma, with a median time to progression of about 5 years. The risk is highest for those with an elevated SFLC ratio, high bone marrow plasma cell infiltration, and paraprotein over 30 g/L. These patients need close monitoring in clinic.

Monoclonal gammopathy of undetermined significance

MGUS is an isolated finding of a monoclonal intact paraprotein or free light chain in serum or urine that does not fulfil the diagnostic criteria for asymptomatic or symptomatic myeloma. The risk of progression to myeloma is approximately 1% per year, with risk factors including and elevated SFLC ratio, paraprotein of more than 15 g/L and non-IgG subtype.

Solitary plasmacytoma

This is an isolated tumour of neoplastic plasma cells. It may be a solitary plasmacytoma within the skeleton or a soft tissue extramedullary plasmacytoma outside the marrow cavity. These can be treated curatively with radiotherapy, although a high proportion (especially bone plasmacytomas) will eventually progress to systemic myeloma.

Heavy-chain diseases

These rare conditions are characterized, as the name suggests, by the production of a heavy-chain-only paraprotein. They are usually caused by a B-cell lymphoproliferative condition.

Monoclonal gammopathies of clinical significance

There are several other conditions in which the clonal population is relatively small, but secondary clinical effects occur. These are discussed in more detail in the relevant organ-based chapters.

- ***AL amyloidosis*** involves organized fibrillary deposits in kidneys, heart, GI tract, skin, and nerves. It is usually caused by a lambda-light chain isotype. Other subtypes of amyloid (e.g., hereditary ATTR amyloid, and reactive AA amyloid) are not caused by a plasma cell disorder (see Chapter 26).
- ***Monoclonal gammopathies of renal significance*** is a group of renal pathologies caused by immunoglobulin deposition (see Chapter 26). Like AL amyloid, all of these may be treated with myeloma-directed therapies (e.g., bortezomib, cyclophosphamide, and dexamethasone).
- ***Plasma cell neoplasms with associated paraneoplastic phenomena (POEMS) syndrome*** is a rare disease characterized by ***p***olyneuropathy, ***o***rganomegaly, ***e***ndocrinopathy, ***m***onoclonal plasma cell disorder and ***s***kin changes (see Chapter 28). High concentrations of vascular endothelial growth factor are common and may be the mediating mechanism.

Further reading

Fermand JP, Bridoux F, Dispenzieri A et al. Monoclonal gammopathy of clinical significance: a novel concept with therapeutic implications. *Blood.* 2018; 132:1478–1485.
Go RS, Rajkumar SV. How I manage monoclonal gammopathy of undetermined significance. *Blood.* 2018; 131:163–173.
Moreau P et al. Treatment of relapsed and refractory multiple myeloma: recommendations from the International Myeloma Working Group. *Lancet Oncol.* 2021;22(3):e105–e118.
Rajkumar SV. Multiple myeloma: 2022 update on diagnosis, risk-stratification, and management. *Am J Hematol.* 2022; 97(8):1086–1107.
Tandon N, Rajkumar SV, LaPlant B et al. Clinical utility of the Revised International Staging System in unselected patients with newly diagnosed and relapsed multiple myeloma. *Blood Cancer J.* 2017;7:e528.

Questions for this chapter can be found on Elsevier eBooks+.

40 Venous thromboembolic disease

Peter K. MacCallum, K. John Pasi, and Suthesh Sivapalaratnam

Clinical skills 1363
Diagnosis 1363
Examination 1363
Investigation 1363
Background 1364
Pathogenesis of thrombosis 1364
Acute venous thromboembolism 1365
Long-term management of venous thromboembolism 1365
Prediction of recurrent venous thromboembolism 1365
Complications 1366
Investigation after venous thrombosis 1366
General investigations 1366
Thrombophilia testing 1367
Prevention 1368
Mechanical prophylaxis 1369
Pharmacological prophylaxis 1369
Anticoagulant agents 1369
Parenteral anticoagulants 1369
Oral anticoagulants 1371

CORE SKILLS AND KNOWLEDGE

Venous thromboembolism (VTE) is common, affecting more than 5% of people during their lifetimes. The risk increases with age. VTE presents as an acute episode of deep vein thrombosis (DVT) or pulmonary embolism (PE) and is frequently recurrent.

Patients with a suspected episode of VTE may present to their general practitioners or directly to the hospital emergency department, or VTE may occur while care is being given to patients in hospital for other medical problems. Many clinical teams may therefore be involved in the initial diagnosis and management, but most commonly, acute physicians play a leading role. Haematologists often provide advice on acute management and frequently oversee long-term treatment, and some chest physicians have a particular interest in the management and long-term complications of PE.

Key learning objectives for VTE include:

- Being familiar with the main risk factors for VTE and recognizing its presenting clinical features
- Knowing how to manage suspected VTE
- Understanding the pharmacology of the different anticoagulants used to treat patients with VTE, their use in clinical practice, and the management of anticoagulant-related bleeding.

Opportunities for learning about VTE include observing the work of a DVT diagnostic service (linked in many hospitals to the emergency department or acute medical unit) or attending an outpatient anticoagulant service, overseen in many hospitals by the haematology department.

CLINICAL SKILLS

Diagnosis

In DVT, a history should be elicited of leg pain and swelling. In PE, the patient has breathlessness (sometimes only on exertion), chest pain (often pleuritic in nature), and haemoptysis or collapse or syncopal episodes. A personal or family history of VTE should be sought along with information about any current or past anticoagulant therapy, including adherence. Risk factors for VTE include recent surgery, hospital admission, trauma, immobility, cancer and its treatment, oestrogen therapy (contraception, hormone replacement, and pregnancy), and intravenous drug use. Pregnancy history includes current or recent pregnancy and history of pregnancy complications, including miscarriage, intrauterine death, and preterm delivery because of pre-eclampsia. Around 5% of patients presenting with an unprovoked DVT or PE have an underlying cancer, so note should be made of weight loss or other system-related symptoms.

Examination

Important features include tachycardia, reduced oxygen saturation (<95% while breathing room air), leg swelling (particularly if unilateral and the calf is at least 3 cm larger on the symptomatic side), tenderness along the deep veins, and pitting oedema affecting the symptomatic leg. A pleural rub might be heard on chest auscultation along with a split second heart sound. Haemodynamic instability characterized by a systolic blood pressure below 90 mmHg not caused by arrhythmia, hypovolaemia, or sepsis, and features of end-organ hypoperfusion (altered mental status; cold, clammy skin) is particularly concerning because it may indicate a high-risk PE and the need for urgent treatment escalation.

Investigation

These are dictated by whether a diagnosis of DVT or PE is considered and the likelihood of the diagnosis being present. If a DVT or PE is suspected, blood is typically taken for a full blood count (FBC); kidney and liver function tests; and a clotting screen, including D-dimer. Imaging for DVT is typically with ultrasonography. Most

often, computed tomography pulmonary angiography (CTPA) is performed if imaging is required for a possible PE. Patients presenting with chest pain also typically have an electrocardiogram and serum troponin performed. If a DVT or PE is confirmed, routine tests, including serum calcium, urinalysis, and a chest X-ray (if a CTPA has not been performed), should be undertaken, with further screening for an underlying cancer performed only if clinical features or routine investigations suggest the need to pursue this possibility. Thrombophilia testing is not indicated at the time of diagnosis, although testing for antiphospholipid antibodies might be considered, especially in those with an unprovoked episode because it might influence choice of anticoagulant. A pregnancy test should be considered if pregnancy is possible because it may influence treatment.

BACKGROUND

PE is the third most common cause of cardiovascular death after acute myocardial infarction and stroke. As well as leading to PE, DVT frequently results in the post-thrombotic syndrome (PTS), which is a major cause of long-term disability. 'VTE' is the term describing both conditions. It is multicausal, often arising as a result of transient provoking factors or in individuals with a heritable or acquired predisposition, although no cause can be established in up to 50% of patients.

Every clinician must be familiar with VTE because:

- It is common, particularly in those admitted to hospital, after surgery or trauma, or if there is another cause of reduced mobility.
- It is life-threatening.
- It is preventable. PE is the most common avoidable cause of death in patients admitted to hospital, and the single leading direct cause of death during pregnancy and the puerperium in the UK.
- It can be difficult to diagnose. Clinical features (Fig. 40.1) and routine initial investigations, particularly for PE, are often non-specific, leading to diagnostic delay with potentially serious consequences.
- Its treatment can be hazardous. Anticoagulant therapy, which may be long-term, is effective but carries the risk of major, even fatal, bleeding, and it can be particularly challenging in patients with co-morbidities.
- It is a significant cause of morbidity. Long-term sequelae are frequent and significantly impair quality of life.

PATHOGENESIS OF THROMBOSIS

Thrombosis is the pathological process by which a localized solid mass of blood constituents (a blood clot or ***thrombus***) forms within a blood vessel, mostly because of fibrin formation, with a variable contribution from platelets and other cells. This differentiates it from physiological haemostasis in which a fibrin-rich blood clot occurs outside the vessel-wall lining (endothelium) as a result of injury. Thrombi form on and are attached to the vessel wall, but fragments ***(emboli)*** may break off and occlude downstream vessels.

Arterial and venous thrombosis

Thrombosis can occur in both arteries and veins. For example, the usual cause of myocardial infarction is thrombosis within a coronary artery. The pathogenesis of thrombosis in these two sites is different, reflecting the different shear stresses in arteries and veins and the contribution that rupture of atheromatous plaques makes

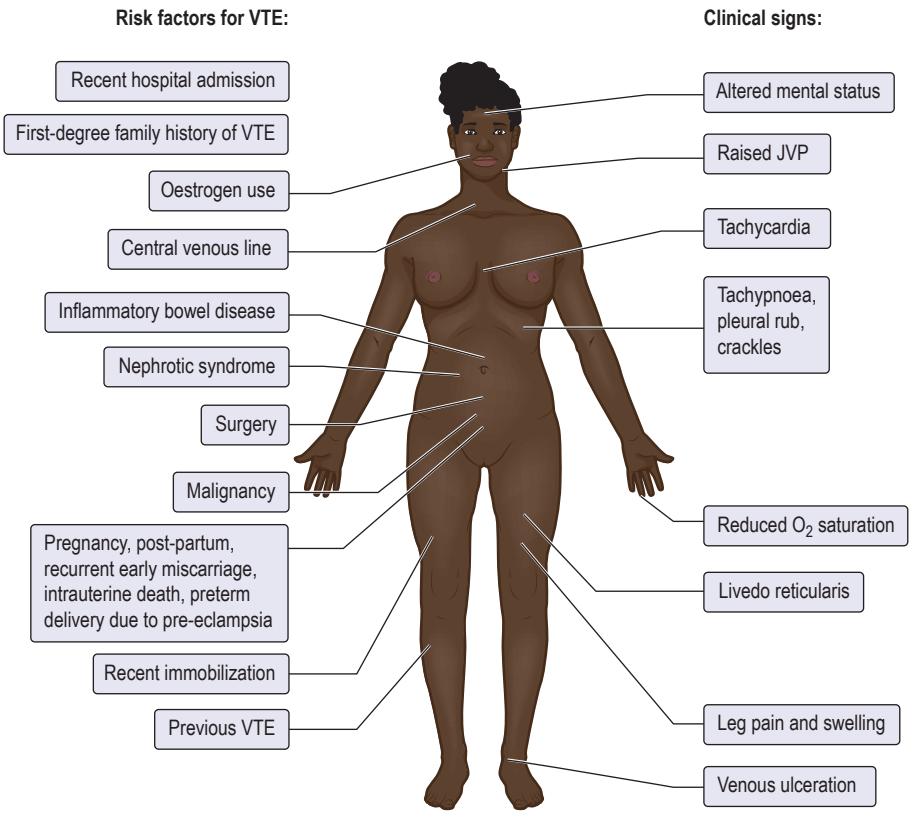


Fig. 40.1 Diagnosis of venous thromboembolism. *JVP,* Jugular venous pressure; *VTE,* venous thromboembolism.

to initiation of arterial thrombosis. Arterial clots are described as white thrombi, reflecting the contribution that platelets as well as fibrin make, and venous clots as red thrombi because of red cells and fibrin. Thrombosis is thought to be the cause of about 25% of deaths worldwide.

Factors influencing thrombosis

Thrombosis is considered to arise from the interplay between the three factors that make up Virchow's triad:

- Changes in blood flow (stasis or turbulence)
- Vessel wall dysfunction
- Changes in blood components, leading to hypercoagulability

The importance of the individual components of Virchow's triad varies between arterial and venous thrombosis:

- ***Turbulence*** and ***vessel wall dysfunction*** caused by atheromatous plaques are factors in arterial thrombosis.
- ***Stasis*** and ***hypercoagulability*** are more relevant in the pathogenesis of venous thrombosis.

Further reading

Furie B, Furie BC. Mechanism of thrombus formation. *N Engl J Med* 2008; 359:938–949.
Wolberg AS, Rosendaal FR, Weitz JI et al. Venous thrombosis. *Nat Rev Dis Primers* 2015; 1:1–15.

ACUTE VENOUS THROMBOEMBOLISM

Definitions, epidemiology, clinical features, and the diagnosis and primary management of VTE are detailed in Chapter 11.

LONG-TERM MANAGEMENT OF VENOUS THROMBOEMBOLISM

Anticoagulation needs to continue for a minimum of 3 months in all patients with proximal DVT or PE because earlier discontinuation is associated with a higher risk of recurrence. After 3 months, the decision to stop treatment or continue long-term anticoagulation for secondary prevention should be made on an individual patient basis, taking account of the risks of recurrent thrombosis, bleeding, and the person's own views.

Anticoagulants reduce the relative risk of recurrent thrombosis by 80% to 90% but carry a 1% to 3% absolute risk per year of major bleeding. Both recurrent thrombotic episodes and anticoagulant-associated major bleeding carry significant risks of morbidity and mortality. Therefore, the benefits of long-term anticoagulation potentially outweigh the risks of major bleeding in those with a higher risk of recurrent thrombosis, but in others, the risks of long-term treatment potentially outweigh the benefits.

Across all patients receiving anticoagulation for VTE, if anticoagulation is stopped after a minimum of 3 months, the risk of recurrent thrombosis is about 5% in the next year. It has been suggested that if recurrence risk is greater than this, continuation of anticoagulation should be considered, and if the recurrence risk is less than 5%, anticoagulation should stop. If anticoagulation has been continued for at least 3 months, a longer duration of treatment does not appear to reduce the risk of recurrent thrombosis further on stopping anticoagulation.

Higher recurrence risks (~10% at 1 year and 30% at 5 years) are seen in patients with unprovoked VTE episodes and in those with persisting risk factors, including active cancer or antiphospholipid syndrome (APS). Lower recurrence risks are seen in those with transient provoking risk factors, and the stronger the transient provoking factor, the lower the recurrence risk (providing the transient provoking factor is no longer present). Recurrence risks are higher in males than in females, in those with second events, and in those with PE or proximal DVT as opposed to distal DVT. The risk of recurrent PE is higher in those whose initial presentation is with PE rather than with DVT. Because the case fatality rate of PE is higher than that of DVT, this suggests that the threshold for continuing anticoagulation might be lower in patients presenting with PE. Also relevant is the individual patient's cardiopulmonary reserve because those with impaired reserve are less able to compensate for the acute circulatory effects of PE.

Prediction of recurrent venous thromboembolism

Several prediction rules have been devised to assist in determination of risk of VTE recurrence after stopping anticoagulation in those with unprovoked VTE. These consider several clinical features and often the measurement of D-dimer after discontinuation of anticoagulation. Prediction scores for risk of major bleeding on anticoagulant therapy for VTE have also been evaluated, and National Institute for Health and Care Excellence (NICE) suggests the HAS-BLED score (initially developed for patients with atrial fibrillation (AF), though NICE now recommend the ORBIT score in those with AF; see Chapter 17) might be used for assessment, with discontinuation of anticoagulation discussed with those with a score of 4 or more that cannot be modified. However, the decision to continue or stop anticoagulation should not be based solely on existing risk prediction tools. In the UK, long-term anticoagulation is considered in patients with unprovoked proximal DVT or PE and a low bleeding risk.

Patients with risk factors for recurrence should be considered for continuation of anticoagulation if the risk factor persists. For example, the risk of recurrent thrombosis is very high in patients with cancer-associated DVT or PE in whom cancer remains active, but stopping anticoagulation after 6 months might be considered if the cancer is in remission, recurrence is unlikely, and cancer treatment has been completed.

The risk of recurrence is lower after isolated calf vein thrombosis than in proximal DVT or PE. In practice, most patients with symptomatic distal DVTs are treated for 3 months, and anticoagulation is then stopped. In those with a low risk of recurrence (e.g., because of a transient provoking factor that has now resolved and a high risk of bleeding), it may be reasonable to reduce the treatment duration to 6 weeks.

Treatment of patients with upper limb DVT follows the same principles as for those with leg vein thrombosis. Many episodes are associated with indwelling central venous lines; anticoagulation should generally continue for a minimum of 3 months, longer if the line remains *in situ*.

Further reading

Kahn S and de Wit K. Pulmonary embolism. *N Engl J Med* 2022; 387:45–57.
Khan F, Tritschler T, Kahn SR, Rodger MA. Venous thromboembolism. *Lancet* 2021; 398:64–77.
Konstantinides S, Meyer G, Becattini G et al. 2019 ESC guidelines for the diagnosis and management of acute pulmonary embolism. *Eur Heart J* 2020; 41:543–603.
National Institute for Health and Care Excellence. NICE Clinical Guideline ng158: Venous thromboembolic diseases: diagnosis, management and thrombophilia testing. *NICE* 2020: https://www.nice.org.uk/guidance/ng158.

Complications

Death

After VTE is diagnosed and treatment commenced, the short-term mortality rate is lower in patients presenting with DVT than with PE. In those presenting with PE, the mortality rate is around 5% by 1 month, although at least half of these deaths are caused by associated co-morbidities rather directly caused by PE. Nevertheless, compared with those without VTE, the mortality rate remains elevated in both the short and longer term, with observed mortality rates of 10% to 20% at 1 year; cancer is the leading cause of death.

Cancer

Cancer-associated thrombosis, the initial description of which is often attributed to Trousseau in the 19th century (Trousseau's syndrome), is common, and 10% to 20% of all episodes of VTE are diagnosed in people with cancer. The pathogenesis is multifactorial and includes:

- Hypercoagulability caused by cancer
- Surgery or systemic therapy
- Reduced mobility
- Indwelling central venous catheters leading to catheter-associated thrombosis.

The combination of cancer and thrombosis carries a particularly poor prognosis. Up to 5% of patients who present with a seemingly unprovoked episode of VTE are diagnosed with cancer within 12 months. However, current evidence does not support extensive investigations for an unknown cancer unless indicated by history and examination, FBC, liver and kidney function tests, serum calcium, chest X-ray, and urinalysis.

Post-thrombotic syndrome

Post-thrombotic syndrome (PTS) can affect long-term quality of life after VTE. It occurs in about 40% of people after a proximal DVT. It causes symptoms of variable severity in the affected leg, including pain, swelling, heaviness, venous claudication on exercise, itching, and discoloration. Symptoms may be persistent or intermittent and reduce physical functioning and mobility. At its most severe, it leads to permanent skin damage with redness, hyperpigmentation, venous ectasia, and skin fibrosis (lipodermatosclerosis), and 5% to 10% of patients develop leg ulceration that is difficult to treat and tends to recur (Fig. 40.2).

PTS results from:

- Proximal venous occlusion with outflow obstruction
- Damage to the venous valves that normally allow blood flow from superficial to deep and distal to proximal
- Development of a collateral circulation
- Venous hypertension
- Capillary leakage
- Localized inflammation.

The role of knee-length, graduated compression stockings providing at least 24 mmHg of support at the ankle in the reduction of risk of PTS is controversial. Those who develop PTS sometimes benefit symptomatically from graduated compression stockings, but others find them difficult to put on and uncomfortable to wear. A multidisciplinary approach is necessary in patients who develop venous ulceration. With residual ilio-femoral vein occlusion, a vascular surgery opinion should be sought because some patients may benefit from venous stent insertion.

Fig. 40.2 Venous ulceration of the ankle. This patient has an ulcer with local venous-type staining of local skin caused by poor venous drainage from the lower leg. Immobility, previous deep vein thrombosis, and varicose veins contribute to development. *(From Cross SS.* Underwood's Pathology: A Clinical Approach, *7th edn, Elsevier Ltd, 2019, Fig. 13.17.)*

Pulmonary hypertension

After PE, up to 5% of patients remain persistently breathless because of chronic thromboembolic pulmonary hypertension. In this condition, there is incomplete resolution of pulmonary emboli. The diagnosis should be suspected in those with persisting symptoms supported by follow-up perfusion lung scanning or CTPA showing evidence of residual occlusion and echocardiography suggesting pulmonary hypertension. Assessment is undertaken by respiratory specialists. Some patients can be successfully treated surgically with pulmonary endarterectomy, but a much greater proportion have post-PE syndrome with persisting breathlessness, impaired exercise intolerance, and impaired quality of life. Cardiopulmonary exercise testing and assessment by experienced clinicians can be helpful in defining the cause, including the contribution of deconditioning and mental health symptoms, and approaches to management.

INVESTIGATION AFTER VENOUS THROMBOSIS

General investigations

All patients with newly diagnosed, unprovoked VTE should undergo physical examination and have the following general tests to ensure safety of anticoagulation and to look for occult malignancy and other co-morbidities:

- FBC
- Kidney function tests
- Liver function tests
- Serum calcium

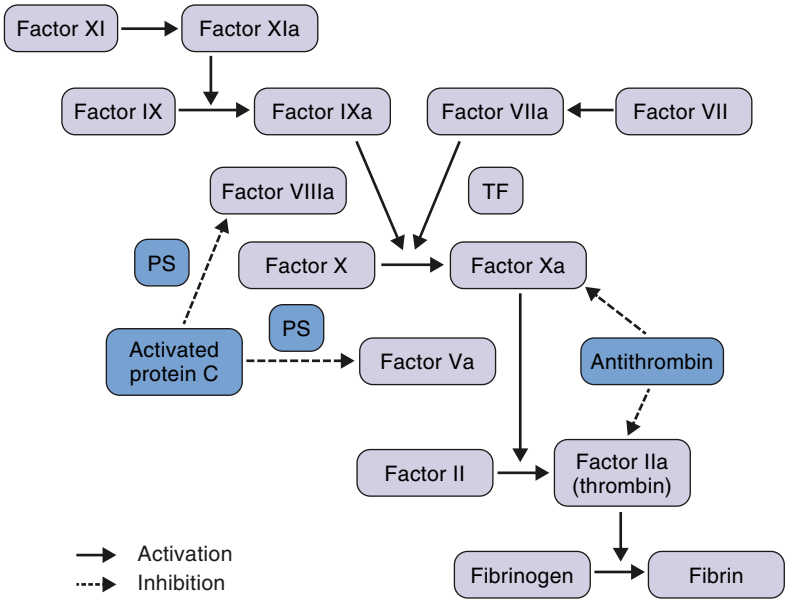


Fig. 40.3 Outline of the coagulation system, highlighting sites of action of the naturally occurring inhibitors of coagulation. Naturally occurring anticoagulants are shown in *blue*. Activated protein C results from activation of protein C by thrombin (not shown). *PS,* Protein S; *TF,* tissue factor. *(From MacCallum P, Bowles L, Keeling D. Diagnosis and management of heritable thrombophilias.* BMJ *2014; 349:g4387.)*

- Clotting screen
- Chest X-ray
- Urinalysis for both proteinuria and microscopic haematuria

Thrombophilia testing

Thrombophilia testing is often considered in patients with VTE, but indications remain controversial. In this context, the term 'thrombophilia' applies to a thrombotic tendency that can be identified in the laboratory. Thrombophilic tendencies can be either heritable or acquired.

Heritable thrombophilias

Heritable thrombophilias can be identified in about 20% of cases of VTE. They do not play a significant role in arterial thrombosis. They can be divided into variants that cause either gain of function of procoagulants or loss of function of anticoagulants.

Gain of function of procoagulants

Factor V Leiden results from a variant in clotting factor V that causes activated clotting factor V (factor Va) to be resistant to inactivation by activated protein C and thereby leads to an increase in thrombin generation. Heterozygous factor V Leiden is present in about 5% of individuals of European or Mediterranean origin but is not found in other ethnic groups such as those from eastern Asia. It increases the risk of VTE fivefold throughout life and interacts synergistically with acquired risk factors such as oestrogen-containing contraceptives and pregnancy. Thus, the risk of thrombosis is increased about 30-fold in heterozygous V Leiden carriers taking the combined oral contraceptive pill compared with non-carriers not taking the combined oral contraceptive pill. However, because the absolute risk of VTE is low in young females, even a 30-fold increase in risk remains significantly below 0.5% per year for any individual. Overall, factor V Leiden is present in about 20% of cases of VTE. Individuals who are homozygous for factor V Leiden are less common, but their risk of thrombosis is about 10-fold higher than those who are heterozygous.

A ***variant in the 3′ untranslated region of the prothrombin gene*** *(G20210A)* causes levels of prothrombin to rise and increases the risk of VTE two- to threefold throughout life. It is present in about 2% of Caucasians and in about 5% of people with VTE.

Loss of function of anticoagulants

The sites of action of the naturally occurring anticoagulants are shown in Fig. 40.3.

Antithrombin is a serine protease inhibitor (serpin) that functions as an anticoagulant by inhibiting predominantly thrombin and factor Xa reactions that are catalysed by heparin. Reduced antithrombin activity therefore leads to increased thrombin generation and activity, thereby predisposing to thrombosis. Heterozygous antithrombin deficiency is rare in the population but increases the risk of VTE about 10-fold. The disorder is identified by measurement of antithrombin activity rather than by genetic testing because multiple genetic mutations (some reducing synthesis and others reducing function) are described. Laboratory diagnosis is not straightforward because low antithrombin concentrations can also be acquired, for example, after acute illness, after surgery, with heparin therapy, in liver disease, and in nephrotic syndrome. Those with antithrombin deficiency often present at a young age and are more likely to have recurrent events. Because antithrombin is required for the action of heparin, those with antithrombin deficiency may be relatively resistant to heparin.

Protein C and protein S are vitamin K–dependent, naturally occurring anticoagulants. Together, they inhibit activated forms of clotting system co-factors, factor Va and factor VIIIa. Deficiency leads to an increase in thrombin generation and a predisposition to thrombosis. Heterozygous deficiencies of protein C and protein S are uncommon in the population and increase the risk of venous thrombosis at least fivefold. As with antithrombin deficiency, multiple genetic mutations have been described that affect either synthesis

or function. Diagnosis is based on measurement of activity or protein concentration and is often difficult because reduced levels can also be acquired because of for example, acute illness; liver disease; warfarin or vitamin K deficiency; or in the case of protein S, pregnancy or oestrogen therapy. Very rarely, homozygous protein C or protein S deficiency may occur and present as neonatal purpura fulminans.

Management of heritable thrombophilias

In general, the more common heritable defects cause a milder clinical picture than those that are rare, such as antithrombin deficiency. The acute management of VTE is not affected by the presence or absence of heritable thrombophilia, and diagnosis of antithrombin, protein C, and protein S deficiencies is problematic in the setting of acute VTE because levels of these proteins are often reduced either by the thrombosis itself or by the anticoagulants used in treatment. Moreover, for the most part, identification of a heritable thrombophilic defect does not influence the long-term management because most defects do not influence the risk of recurrent VTE. Therefore, if a heritable thrombophilia is to be sought, it should preferably be after the acute stage is over, when the patient has completed a minimum of 3 months of anticoagulation, and when their course of treatment either has finished or can be safely interrupted to allow measurement of naturally occurring anticoagulant levels.

Testing for heritable thrombophilias might be considered in:

- Those with unprovoked VTE who are planning to stop anticoagulation, particularly if they have a family history
- Those who present with VTE younger than 40 years of age
- Females of reproductive age for assistance with management of future pregnancies
- Those with unusual-site thromboses.

Acquired thrombophilias

Many risk factors for thrombosis cause an acquired hypercoagulable state, but from a laboratory angle, the main acquired thrombophilia is APS. Other prothrombotic conditions considered in this section are myeloproliferative neoplasm (MPN) and SARS-CoV-2 infection. Heparin-induced thrombocytopenia (HIT) is considered later in this chapter (see 'Heparin-induced thrombocytopenia' section).

Antiphospholipid syndrome

APS is an acquired prothrombotic disorder that has specific clinical and laboratory features. See 'Antiphospholipid syndrome' section in Chapter 32 for a full description, including diagnostic testing.

The diagnosis of APS should be considered in patients presenting with:

- ***VTE if unprovoked, or provoked by a minor risk factor***
- ***Unusual-site venous thrombosis***
- ***Recurrent episodes of venous thrombosis***, particularly if anticoagulants are being given
- ***Arterial thrombosis*** at a young age (younger than 50 years) that is otherwise unexplained
- ***Pregnancy-related morbidity*** with one or more of the following:
  - Three consecutive miscarriages before 10 weeks' gestation after exclusion of other causes such as chromosomal abnormalities
  - One or more otherwise unexplained fetal deaths after 10 weeks' gestation
  - Birth of a morphologically normal baby before 34 weeks' gestation because of severe pre-eclampsia, eclampsia, or placental insufficiency.
- ***Incidentally detected abnormalities of clotting screens.*** The activated partial thromboplastin time (APTT) is typically prolonged; less often, the prothrombin time is prolonged. Although this suggests a bleeding tendency, in practice, a thrombotic tendency is observed.

APS is associated with an increased risk of recurrent thrombotic episodes. This may influence decisions on duration of anticoagulation after an episode of VTE. The oral factor Xa inhibitors appear to be less effective than warfarin in the prevention of recurrent thrombosis, in APS, particularly arterial events such as strokes.

Management of patients with APS presenting with arterial thrombosis is controversial, and there is ongoing debate as to the best approach, including anticoagulation with warfarin at either standard intensity (international normalized ratio [INR], 2.0–3.0) or higher intensity (INR, 3.0–4.0) and/or antiplatelet therapy.

Obstetrical APS in the absence of a thrombotic history is usually managed with low-dose aspirin and a prophylactic dose of low-molecular-weight heparin (LMWH) through the pregnancy.

People with a **myeloproliferative neoplasm (MPN)** (see Chapter 39) have an increased risk of thrombosis, particularly in unusual sites such as splanchnic vein or cerebral venous sinus thrombosis. Hence, when thrombosis occurs in an unusual location or is unexplained with abnormalities in the FBC, an MPN screen is warranted.

Patients admitted to hospital with **SARS-CoV-2** have a high incidence of thromboembolic events, particularly if admitted to intensive care. The pathogenesis of thrombosis in SARS-CoV-2 is complex with coagulopathy (illustrated typically by raised D-dimer concentrations), platelet activation, and endothelial perturbation leading to a thrombo-inflammatory process that includes VTE, arterial thrombosis, and formation of capillary microthrombi. The short-term risk of thrombosis is also increased, albeit to a lesser extent, in those not requiring hospital admission.

Vaccine-induced immune thrombotic thrombocytopenia (VITT) is a rare but life-threatening thrombotic complication of the AstraZeneca and Johnson & Johnson adenoviral vaccines against SARS-CoV-2. Recognition of this side-effect resulted in some national authorities, including those in the UK, recommending alternative vaccines, particularly for those aged under 50 years in whom the risk of VITT was greater. Affected individuals typically present 5 to 30 days after their first vaccine with DVT, PE, unusual-site thrombosis (cerebral venous sinus thrombosis or splanchnic vein thrombosis), or arterial thrombosis and thrombocytopenia. The D-dimer concentration is usually considerably elevated. The diagnosis is confirmed by the presence of antiplatelet factor 4 (PF4) antibodies. Management of VITT includes anticoagulation with a non–heparin-based anticoagulant and immunosuppression with corticosteroids, intravenous immunoglobulin, and/or plasma exchange. The increased risk of thrombosis in SARS-CoV-2 infection is much greater than the approximately 1 in 50 000 or greater risk of VITT.

Further reading

Connors JM. Thrombophilia testing and venous thrombosis. *N Engl J Med* 2017; 377:1177–87.
Klok FA, Pai M, Huisman MV, Makris M. Vaccine-induced immune thrombotic thrombocytopenia. *Lancet Haematol* 2022; 9:73–80.
Knight JS, Branch DW, Ortel TL. Antiphospholipid syndrome: advances in diagnosis, pathogenesis, and management. *BMJ* 2023; 380:e069717.
Knight R, Walker V, Ip S et al. Association of COVID-19 with major arterial and venous thrombotic diseases: a population-wide cohort study of 48 million adults in England and Wales. *Circulation* 2022; 146:892–906.

PREVENTION

Between one-third and half of all episodes of VTE are provoked by surgery or admission to hospital, with most diagnosed in the

3 months after discharge. Studies of routine postoperative screening for venous thrombosis after major orthopaedic or cancer surgery report an asymptomatic DVT incidence as high as 50%. PE is the most common preventable cause of hospital-related death. Clinical trials have shown that thromboprophylaxis to prevent symptomatic VTE in patients undergoing major surgery and in high-risk medical inpatients is beneficial and cost-effective. All adults admitted to hospital in the UK require formal assessment of their risks of VTE and of bleeding. The factors included in the UK Department of Health risk assessment tool are shown in **Box 40.1**.

VTE prophylaxis can be mechanical or pharmacological. Pharmacological prophylaxis is considered for patients at risk of VTE without significant bleeding risk, with or without additional mechanical measures. Those who have a significant bleeding risk are considered for mechanical prophylaxis.

Box 40.1 UK Department of Health risk assessment tool for venous thromboembolism (VTE) and bleeding

Patients at risk of VTE
- Medical patients expected to have ongoing reduced mobility relative to normal state
- Surgical patients

Thrombosis risk factors
- Patient-related
  - Active cancer or cancer treatment
  - Age older than 60 years
  - Dehydration
  - Known thrombophilia
  - Obesity (body mass index >30 kg/m^2)
  - One or more significant medical co-morbidities (e.g., heart failure, acute infectious diseases, inflammatory conditions)
  - Personal history or first-degree relative with a history of VTE
  - Oestrogen containing medication
  - Varicose veins with phlebitis
  - Pregnancy or <6 weeks postpartum
- Admission-related
  - Significantly reduced mobility for ≥3 days
  - Hip or knee replacement
  - Hip fracture
  - Total anaesthetic and surgical time >90 minutes
  - Pelvic or lower limb surgery with total anaesthetic and surgical time >60 minutes
  - Acute surgical admission with inflammatory or intra-abdominal condition
  - Critical care admission
  - Surgery with significant reduction in mobility

Bleeding risk
- Patient related
  - Active bleeding
  - Acquired bleeding disorder (e.g., acute liver failure)
  - Concurrent use of anticoagulants known to increase bleeding risk (e.g., warfarin with international normalized ratio >2)
  - Acute stroke
  - Thrombocytopenia (platelets <75 × 10^9/L)
  - Uncontrolled hypertension (≥230/120 mmHg)
  - Untreated inherited bleeding disorders (e.g., haemophilia, von Willebrand's disease)
- Admission-related
  - Neurosurgery, spinal surgery, or eye surgery
  - Procedure with high bleeding risk
  - Lumbar puncture, epidural, or spinal anaesthesia expected within the next 12 hours
  - Lumbar puncture, epidural, or spinal anaesthesia within the previous 4 hours

Mechanical prophylaxis

Mechanical prophylaxis measures include:
- Early mobilization
- Elevation of the legs
- Use of anti-embolic stockings of knee or thigh length (not in peripheral arterial disease, stroke, or situations when they could result in skin damage)
- Intermittent compression devices during surgery or on bed rest to improve blood flow

Pharmacological prophylaxis

This involves anticoagulants in lower doses than those used in the treatment of VTE. Most often, LMWH is used during hospital admission, and in some groups (e.g., those undergoing major abdominal surgery for cancer), this is continued for a period after discharge. Oral factor Xa and thrombin inhibitors at reduced dose have been approved for VTE prevention after major hip and knee replacement surgery, and these are continued for 2 to 5 weeks after surgery. Though traditionally not thought to be of benefit in VTE, aspirin has been shown to reduce the risk of venous thrombosis after hip or knee replacement surgery.

Prophylactic LMWH should be offered as soon as possible after VTE risk assessment to young people and adults admitted to hospital with SARS-CoV-2. NICE recommend a treatment dose of LMWH is considered in young people and adults who need low-flow oxygen and who do not have an increased bleeding risk.

Further reading

National Institute for Health and Care Excellence. NICE COVID-19 rapid guideline: managing COVID-19. *NICE* 2023: https://www.nice.org.uk/guidance/ng191/resources/covid19-rapid-guideline-managing-covid19-pdf-51035553326.
National Institute for Health and Care Excellence. NICE Guideline ng89: Venous thromboembolism in over 16s: reducing the risk of hospital-acquired deep vein thrombosis or pulmonary embolism. *NICE* 2018; https://www.nice.org.uk/guidance/ng89.

ANTICOAGULANT AGENTS

The physiological purpose of the coagulation system is to generate the enzyme (or serine protease) thrombin rapidly and locally. This converts soluble circulating plasma fibrinogen into insoluble fibrin in response to vascular injury, and secures ***haemostasis***. This is achieved through a complex series of reactions on exposed cell surfaces and activated platelets, in which inactive circulating clotting factors become locally activated in a coordinated and sequential manner. Initiation follows the binding of circulating factor VII to tissue factor, a receptor expressed on subendothelial and adventitial cells, and leads to the generation of activated factor X (traditionally referred to as the ***extrinsic system***; see **Fig. 38.37**). This in turn results in initial thrombin generation and subsequent amplification of the process through activation of the ***intrinsic system***, leading to further activation of factor X and thrombin generation. Venous thrombosis is characterized by the pathological formation of a localized fibrin-rich clot. Anticoagulants work by inhibiting the generation or activity of thrombin in different ways (**Figs. 40.4 and 40.5**).

Parenteral anticoagulants

The main pharmacological properties of parenteral anticoagulants used in the treatment and prevention of VTE are summarized in **Box 40.2**.

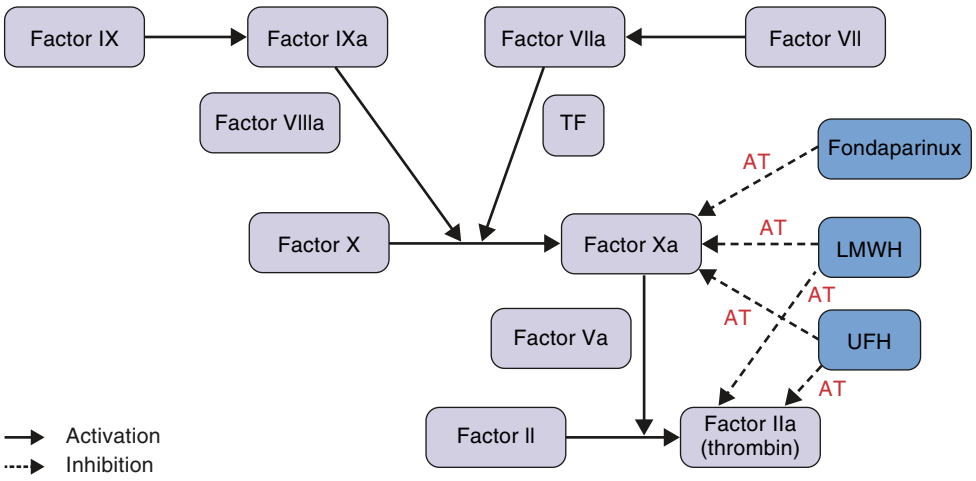


Fig. 40.4 Sites of action of parenteral anticoagulants used in the treatment and prevention of venous thromboembolism. *AT,* Antithrombin; *LMWH,* low-molecular-weight heparin; *TF,* tissue factor; *UFH,* unfractionated heparin.

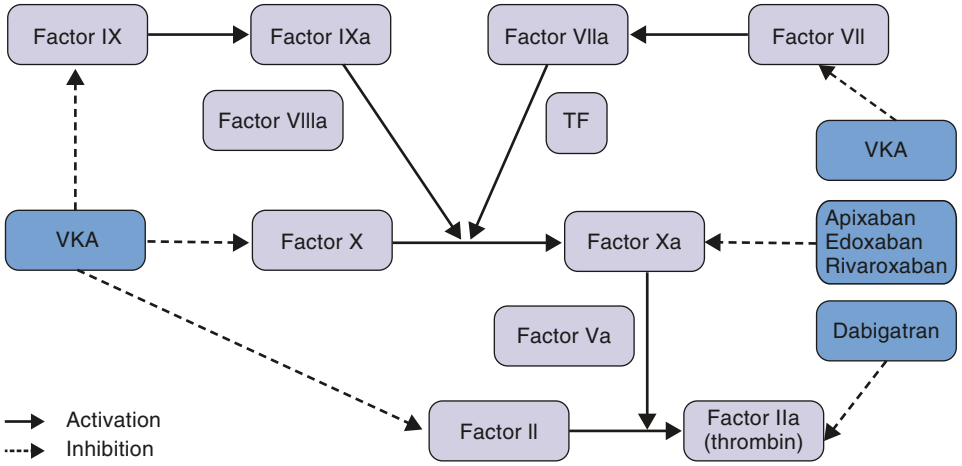


Fig. 40.5 Sites of action of oral anticoagulants used in the treatment and secondary prevention of venous thromboembolism. *TF,* Tissue factor; *VKA,* vitamin K antagonist (e.g., warfarin).

Box 40.2 Pharmacological properties of parenteral anticoagulants

	Unfractionated heparin	Low-molecular-weight heparin	Fondaparinux
Source	Biological	Biological	Synthetic
Molecular weight (Da)	3000–30 000	2000–9000	1500
Target (via antithrombin)	Xa = IIa	Xa > IIa	Xa
Half-life	1 h	4 h	18 h
Monitoring test	APTT	Anti–factor Xa (not usually required)	Anti–factor Xa (not usually required)
Renal excretion	No	Yes	Yes
Antidote	Protamine sulfate	Protamine sulfate (partial reversal)	None

APTT, Activated partial thromboplastin time.

Heparin

Heparin is not a pure substance but a mixture of polysaccharides of different molecular weights of biological origin. In the UK, all heparins are derived from processed porcine gut mucosa. Heparins are destroyed in the stomach and must be administered parenterally. Heparin is an indirect anticoagulant because it does not work directly but by binding through a specific pentasaccharide sequence to naturally occurring antithrombin in plasma and inducing a conformational change that increases the inhibitory activity of antithrombin at least 1000-fold. The result is more rapid inhibition of several of the activated serine protease coagulation factors, particularly thrombin (factor IIa) and factor Xa. Heparins are classed as LMWH or unfractionated heparin (UFH).

Unfractionated heparin

For many years, UFH was the only form of heparin available for the treatment and prevention of VTE. In VTE treatment, it is usually administered as an immediate-acting bolus injection followed by an initially weight-based intravenous infusion. It is difficult to use in practice because it is necessary to monitor its effect on the APTT regularly and to adjust the rate of infusion to maintain the APTT ratio (APTT of patient sample compared to midpoint of the normal range) within a target range: typically, an APTT ratio of 1.5 to 2.5. UFH has a short half-life of 1 hour and, if necessary, can be rapidly reversed with a protamine sulfate. It is not excreted through the kidneys. It carries a small but significant risk of major bleeding and of an uncommon but important immunologically mediated, prothrombotic adverse drug reaction, ***HIT*** (see later). Its use nowadays in the treatment of patients with VTE is largely restricted to particular high-risk settings in which its short half-life, reversibility, and lack of renal excretion are advantageous properties. UFH is rarely used for thromboprophylaxis because its short half-life means that it needs to be given two or three times daily by subcutaneous injection.

Low-molecular-weight heparin

LMWHs are the main type of heparin given because of favourable pharmacokinetic properties. LMWH is produced by the chemical or enzymatic degradation of UFH and results in polysaccharide chains of shorter length and molecular weight than UFH. Whereas long chains with at least 18 saccharides are required for inhibition of thrombin by antithrombin, factor Xa is inhibited by shorter chains; therefore, LMWHs inhibit factor Xa to a greater degree than thrombin compared with UFH. LMWH is administered by subcutaneous injection, and peak activity is seen by 4 hours. The half-life of LMWH is longer than that of UFH at 4 hours. LMWH is renally excreted, and caution is required in kidney impairment. Because of its better bioavailability than UFH, it has the major advantage of fixed, weight-based dosing, without the need for laboratory monitoring. As a result, the majority of patients with DVT and many with PE can be treated as outpatients without hospital admission. LMWH is at least as safe and effective as UFH and carries a lower risk of HIT. However, LMWH is only partially reversed by protamine sulfate.

In VTE treatment, standard practice for many years has been to give an initial subcutaneous injection of LMWH while awaiting confirmation of the radiological diagnosis of DVT or PE. LMWH is then switched to an oral agent for ongoing treatment to avoid the need for daily injection, which is usually self-administered by the patient. Warfarin generally requires continuation of LMWH for at least 5 days until it can provide sufficient anticoagulation for LMWH to be stopped. In certain groups, longer-term LMWH is preferred over oral anticoagulants:

- LMWH is used throughout pregnancy because it does not cross the placenta and is therefore not fetotoxic.
- In patients with cancer-associated thrombosis, LMWH is more effective than warfarin and non-inferior to oral factor Xa inhibitors. If the patient has an increased bleed risk because of having (multiple) surgical interventions or is on an interacting medication, LMWH is preferred over oral options.

LMWH is widely used as thromboprophylaxis in patients admitted to hospital or undergoing surgery and has the advantages over UFH of once-daily administration and lower risk of HIT. LMWH (or an alternative anticoagulant) may also be used to reduce morbidity and mortality in patients with acute coronary syndrome, in whom it is usually given twice daily.

Heparin-induced thrombocytopenia

An uncommon but important adverse effect of heparin administration is HIT. Thrombocytopenia typically increases the tendency to bleed, but in HIT, thrombocytopenia is associated with a strong prothrombotic tendency caused by activation of both platelets and the coagulation system. The disorder usually occurs 5 to 14 days after heparin exposure, with a greater than 30% fall in platelet count occurring over 1 or 2 days. The thrombocytopenia is usually mild to moderate, and it is rare for the platelet count to fall below 20 × 10^9/L. HIT is immunologically mediated, with the development by the patient of immunoglobulin G (IgG) antibodies to a complex of heparin and PF4 that is found in and released from platelet granules. Without prompt diagnosis and treatment, the risk of thrombosis, which can be either venous or arterial, is 50% over the next few days. The diagnosis is based on the clinical features, supplemented by laboratory confirmation of IgG antibodies to the heparin–PF4 complex.

If HIT is suspected or confirmed:

- Specialist advice should be sought.
- All heparins should be discontinued (including that used, e.g., to flush indwelling lines).
- The allergic reaction should be noted on the drug chart.
- An alternative, non-heparin anticoagulant should be administered in view of the high rate of thrombosis.

The most common alternative anticoagulants used in this setting are ***danaparoid*** (a heparinoid), ***argatroban*** (an intravenously administered direct thrombin inhibitor), and ***fondaparinux***. Patients who develop HIT should not be re-exposed to heparin in the future without consultation with a haematologist.

Fondaparinux

Fondaparinux is a synthetic pentasaccharide. It binds to antithrombin, and because of its short chain length, it inhibits only factor Xa and not thrombin. It is administered by subcutaneous injection and has a longer half-life than heparin at around 18 hours. It is renally excreted and cannot be used in those with significant kidney impairment. Its effect is not reversed by protamine sulfate.

Oral anticoagulants

All patients commencing oral anticoagulants should be counselled by a qualified healthcare professional covering topics, including correct dosing, the importance of adherence, bleeding risks and features to look out for, lifestyle factors (including avoidance of contact sports), possible future pregnancy, management around the time of clinical procedures, and the need for periodic checks of kidney function when medications are renally cleared.

Oral factor Xa and thrombin inhibitors

- Three oral factor Xa inhibitors (***apixaban, edoxaban,*** and ***rivaroxaban***) and one oral thrombin (factor IIa) inhibitor ***(dabigatran)*** are currently licensed for the treatment of patients with VTE in the UK (**Box 40.3**). Unlike heparin and warfarin, they directly inhibit their target substrates. Peak concentrations are seen about 2 hours after ingestion. They are variably eliminated through the kidneys and have a half-life of around 12 hours. Pharmacological advantages of these agents include:
- A wider therapeutic index than warfarin
- No interaction with dietary factors or alcohol
- Limited interaction (compared with warfarin) with other drugs

Box 40.3 Pharmacological properties of oral anticoagulants

	Warfarin	Apixaban	Edoxaban	Rivaroxaban	Dabigatran
Target	II, VII, IX, X	Xa	Xa	Xa	IIa
Peak effect	5 days (variable)	2 hours	2 hours	2 hours	2 hours
Approximate half-life	40 hours	12 hours	12 hours	9 hours	15 hours
Renal elimination	No	25%	50%	33%	80%
Drug interactions	Multiple	P-gp, CYP 3A4	P-gp	P-gp, CYP 3A4	P-gp
Diet or alcohol interactions	Yes	No	No	No[a]	No
Monitoring	Yes	No	No	No	No
Measurement	INR	Anti-Xa	Anti-Xa	Anti-Xa	Dilute thrombin time
Reversal	Vitamin K, PCC	Andexanet (licensed for life-threatening GI bleeding)	Andexanet (not currently licensed)	Andexanet (licensed for life-threatening GI bleeding)	Idarucizumab

[a]Rivaroxaban should be taken with food to improve bioavailability.
CYP 3A4, Cytochrome P450 3A4; *GI,* gastrointestinal; *INR,* international normalized ratio; *PCC,* prothrombin complex concentrate; *P-gp,* P-glycoprotein.

- Fixed dosing with no monitoring (a major advantage for patient convenience compared with warfarin).

Clinical trials have shown they are as effective as heparin and warfarin in the treatment and secondary prevention of VTE. Overall, they carry a similar or lower risk of major bleeding, with a reduced risk of intracranial haemorrhage, partly offset by an increased risk of gastrointestinal (GI) haemorrhage. Their use in the acute treatment of VTE has already been outlined. The combination of efficacy, safety, simplified dosing, and avoidance of monitoring has led to a rapid increase in their clinical use for VTE treatment and secondary prevention. Collectively the oral factor Xa and thrombin inhibitors are referred to as DOACs (direct oral anticoagulants) or NOACs (non-vitamin K antagonist oral anticoagulants). For edoxaban and dabigatran, after an initial period of at least 5 days of LMWH, a switch can be made to these oral drugs. For rivaroxaban and apixaban, initial LMWH is not required.

Special circumstances

The oral factor Xa and thrombin inhibitors cross the placenta and should not be used in pregnancy or in those trying to conceive. They are also avoided in breast-feeding because of an absence of lactation safety data. The oral factor Xa inhibitors are licensed for VTE treatment down to a creatinine clearance of 15 mL/min, although data are still limited in patients with a creatinine clearance of less than 30 mL/min, the lower limit for dabigatran. Significant interactions with certain drugs such as protease inhibitors and many antiepileptics preclude their use in some patients. They are also contraindicated in those with mechanical heart valve replacements or moderate or severe mitral stenosis. Recent trials suggest that the oral factor Xa inhibitors are as effective as LMWH in the treatment of patients with cancer-associated thrombosis. However, they increase the risk of GI bleeding, and LMWH remains the preferred option in patients with cancers in whom this is a risk. Oral factor Xa inhibitors are not recommended in patients with thrombotic APS because of an increased risk of arterial thrombosis (especially stroke) compared with warfarin in clinical trials.

Management of bleeding on oral factor Xa and thrombin inhibitors

Management of ***major bleeding*** follows established principles:

- The oral factor Xa or thrombin inhibitor should be stopped.
- Supportive therapy, such as intravenous fluids and blood components, should be administered as appropriate.
- Consideration should be given to targeted intervention (e.g., endoscopy).

In the event of ***life-threatening bleeding***, specific antidotes are emerging as additional options. In the case of dabigatran, a monoclonal antibody, ***idarucizumab***, that rapidly binds and completely reverses dabigatran is now available and can also be used to reverse dabigatran before urgent surgery. An antidote to apixaban and rivaroxaban, ***andexanet alfa***, is approved by NICE for use in those with life-threatening or uncontrolled GI bleeding. Other national regulators have licensed its use for reversal of oral factor Xa inhibitors in intracranial haemorrhage. In the absence of an antidote, prothrombin complex concentrate (PCC) administration should be considered.

Vitamin K antagonists

These agents are indirect anticoagulants that inhibit the final stage of the synthesis of vitamin K–dependent proteins in the liver: clotting factors II (prothrombin), VII, IX, and X and the naturally occurring anticoagulants protein C and protein S. The main vitamin K antagonist used in the UK is ***warfarin***. The full anticoagulant effect of warfarin takes at least 5 days because it affects only synthesis of new proteins; those already circulating or fully formed in hepatocytes are unaffected and decline as a function of their half-lives, which vary from 6 hours (factor VII) to 60 hours (prothrombin). Therefore, in the setting of VTE treatment, in which an immediate anticoagulant effect is desired, it is necessary to commence with a quick-acting anticoagulant such as heparin and to continue until warfarin is providing sufficient anticoagulation on its own.

The anticoagulant effect of vitamin K antagonists is measured using the ***prothrombin time (PT)***, which measures the extrinsic and common pathways of the coagulation system. Because different laboratories use different reagents with different sensitivities and normal ranges to measure the PT, a system has been devised to standardize assessment of the degree of anticoagulation by the vitamin K antagonists. In this system, the PT for each patient test (compared with the normal PT of the general population not on anticoagulants) is converted to a standardized ratio, the ***INR***. This enables comparison between results in different laboratories, and target ranges defined in clinical trials can be adopted into clinical

practice. The higher the INR, the greater the anticoagulation. For most patients taking warfarin for VTE treatment, the target INR is 2.0 to 3.0. A small proportion of patients (e.g., those who have a recurrent VTE episode while the INR is in the target range or with high-risk mechanical heart valve replacements) need to be more intensively anticoagulated, and the target range may be set at 3.0 to 4.0, reflecting greater prolongation of the PT.

Warfarin is metabolized in the liver and has a half-life of about 36 hours. It has a narrow therapeutic index and a wide range of interactions with dietary factors, alcohol, and other medications. Drug interactions can be either ***pharmacokinetic*** or ***pharmacodynamic***. The former include drugs that either induce or inhibit warfarin metabolism by the cytochrome P450 system, thereby reducing or increasing warfarin concentrations and the resulting INR. The latter are exemplified by aspirin and clopidogrel, which do not affect warfarin concentrations but increase the bleeding risk of warfarin through antiplatelet effects. For reasons that are partly genetic, there is considerable variation between individuals in the warfarin dose required for the same effect on the INR.

Monitoring warfarin therapy

During the initial period of anticoagulation with warfarin and heparin, it is necessary to measure the INR frequently (almost daily) so that the dose of warfarin can be adjusted and heparin can be discontinued when the INR is in the target range for 2 consecutive days. After this initial period, the frequency of INR monitoring can be reduced, but it remains necessary on a long-term basis at least every 8 to 12 weeks in patients whose INR control is stable and more frequently in those with unstable INRs to allow appropriate dose adjustment. that is, a dose increase if the INR is below target and a dose reduction and/or omission if the INR is too high. INR monitoring is usually done either in hospital-based anticoagulant clinics or in general practice, with emerging availability for self-testing of INRs at home using point-of-care coagulometers.

Pregnancy and breast-feeding

In general, warfarin should not be used in pregnancy because it crosses the placenta and is teratogenic between the 6th and 12th weeks of pregnancy, as well as causing fetal bleeding later in pregnancy. It does not cross into breast milk and therefore is safe during breast-feeding.

Risks of warfarin therapy

The main risk of warfarin is bleeding. Most bleeds are not major. The risk of major bleeding is around 2% per year, and about 0.25% patients per year will have a fatal bleed, usually caused by intracranial haemorrhage. Most bleeds occur when the INR is within the therapeutic range, but the higher the INR is, the greater the risk of a major bleed. Management of high INRs and of bleeding on warfarin is shown in **Box 40.4**.

Peri-procedural management of anticoagulation

Given the increasing number of people taking oral anticoagulants for AF, VTE treatment and secondary prevention, and mechanical heart valve replacements, management around the time of operations and procedures is a common clinical concern. The British Society for Haematology and other national specialty societies have produced guidelines, and most hospitals have adapted these for local use.

Anticoagulant management around the time of procedures depends on the bleeding risk of the procedure (high, low, or negligible), the thrombotic risk during anticoagulant interruption (high or low), the anticoagulant used (warfarin vs an oral factor Xa or thrombin inhibitor), and whether the procedure is elective or emergency.

Box 40.4 Management of high international normalized ratios (INRs) and bleeding in patients taking warfarin

INR >5.0, no bleeding

1. Withhold warfarin.
2. Reduce maintenance warfarin dose.
3. Investigate cause of elevated INR.

INR >8.0, no bleeding or minor bleeding

1. Withhold warfarin.
2. If no bleeding, give 1–5 mg ***p.o.*** vitamin K.
3. If minor bleeding, give 1–3 mg ***i.v.*** vitamin K.
4. Recheck INR within 24 hours.

Major bleeding

1. Withhold warfarin.
2. Give four-factor PCC 25–50 u/kg. (Give FFP 15 mL/kg only if PCC is not available.)
3. Give 5 mg i.v. vitamin K.
4. Recheck INR after administration of PCC.
   In the event of bleeding, investigate a local anatomical cause.

FFP, Fresh-frozen plasma; *i.v.,* intravenous; *PCC,* prothrombin complex concentrate; *p.o.,* oral; *u,* units.
From Keeling D, Baglin T, Tait C et al. and British Committee for Standards in Haematology. Guidelines on oral anticoagulation with warfarin—fourth edition. Br J Haematol *2011; 154:311–324.*

Elective procedures

Some procedures have a negligible bleeding risk, and anticoagulation can continue uninterrupted. Others carry a significant bleeding risk, and anticoagulation must be stopped in advance and only reintroduced when safe to do so after the procedure.

Because of their short half-life, this is generally straightforward for the oral factor Xa and thrombin inhibitors. Oral factor Xa inhibitors need to be discontinued at least 48 hours before procedures with a high bleeding risk (72 hours if creatinine clearance <30 mL/min) and at least 24 hours before procedures with low bleeding risk (48 hours if creatinine clearance <30 mL/min). Because of the predominant renal excretion of the oral thrombin inhibitor dabigatran, these times are extended in those with moderately impaired kidney function (creatinine clearance 30–50 mL/min) and even more so in those with significantly impaired kidney function (creatinine clearance <30 mL/min). The re-introduction of oral factor Xa and thrombin inhibitors after the procedure may have to be delayed by 2 to 3 days if the bleeding risk is high. Standard thromboprophylaxis should be considered in the interim.

Because of its long half-life, warfarin needs to be stopped for 5 days before procedures associated with a bleeding risk, assuming the INR is in the therapeutic range at the outset. Patients at high risk of thrombosis during this interruption are considered for bridging therapy with therapeutic dose of LMWH starting 48 to 72 hours after the last warfarin dose and with the last dose given at least 24 hours before the procedure to give it time to clear. LMWH and warfarin are resumed when safe to do so after the procedure, generally at least 48 hours later in those with a high bleeding risk with thromboprophylaxis considered in the interim. Patients who are at high thrombotic risk and considered for bridging during warfarin interruption include those with:

- A VTE episode within the previous 3 months or recurrent episode on anticoagulants
- AF with a stroke or transient ischaemic attack within the previous 3 months or with a CHA_2DS_2-VASc score of 5 or more

- A mechanical heart valve replacement (excluding a bileaflet aortic valve and no other risk factors).

Consideration should be given to deferring elective procedures in those with a VTE episode within the previous 3 months.

Emergency procedures

Multidisciplinary discussion is often necessary when patients taking oral anticoagulants present needing urgent surgery. For patients taking oral factor Xa inhibitors, which have no available antidote, a decision is required on whether it is preferable to proceed with emergency surgery or delay by 12 to 24 hours or more to allow the drug to clear. The time the drug was last taken, measurement of kidney function, and drug concentration (if available) should be considered. PCC should be available if surgical bleeding occurs.

For patients taking dabigatran, idarucizumab is licensed for reversal before emergency surgery or urgent procedures. A multidisciplinary discussion should therefore involve whether to proceed to emergency surgery with or without idarucizumab administration or delay until the drug had cleared naturally.

Because of its long half-life, warfarin reversal is usually necessary before urgent surgery. The INR should be checked. PCC reverses warfarin coagulopathy almost immediately by replacing the vitamin K–dependent clotting factors. The effect wears off over a number of hours, and a small dose of vitamin K is generally given to extend the period of warfarin reversal. If surgery is urgent but will wait 12 to 24 hours, it may be reasonable to administer vitamin K alone to reverse the coagulopathy. After surgery, warfarin can be re-introduced when safe to do so, generally with administration of LMWH at prophylactic or treatment dose until the INR reaches a therapeutic level when warfarin can continue alone.

Further reading

Chan N, Sobieraj-Teague M, Eikelboom JW. Direct oral anticoagulants: evidence and unresolved issues. *Lancet* 2020; 396:1767–76.
Greinacher A. Heparin-induced thrombocytopenia. *N Engl J Med* 2015; 373:252–261.
Keeling D, Tait RC, Watson H, on behalf of the British Committee for Standards in Haematology. *Br J Haematol* 2016; 175:602–13.

Questions for this chapter can be found on Elsevier eBooks+.

41 Malignant disease

Hamoun Rozati, Emily Scott, Piers Nicholas Plowman, and Jonathan Shamash

Clinical skills 1376
Background 1378
Principles of oncology 1378
The biology of cancer 1378
Aetiology and epidemiology 1380
Screening and investigations 1382
Principles of cancer treatment 1386
Aims of treatment 1386
Principles of chemotherapy 1387
Principles of endocrine therapy 1392
Principles of biological and targeted therapy 1392
Principles of radiation therapy 1394
Common solid tumour treatment 1396
Lung cancer 1396
Non–small-cell lung cancer 1397
Small-cell lung cancer 1398
Mesothelioma 1399
Metastases in the lung 1399
Breast cancer 1399
Management of early disease 1399
Management of advanced disease 1400
Upper gastrointestinal cancers 1402
Oesophageal cancer 1402
Gastric cancer 1402
Gastrointestinal stromal tumours 1403
Lower gastrointestinal cancers 1403
Small intestinal cancer 1403
Colorectal cancer 1403
Anal cancer 1405
Hepatobiliary and pancreatic cancers 1405
Liver 1405
Biliary tract 1405
Pancreas 1405
Urological cancers 1406
Renal cancer 1406
Urothelial (bladder and ureteric) cancer 1407
Prostate cancer 1407
Penile cancer 1409
Testicular and ovarian germ cell tumours 1409
Gynaecological cancers 1410
Cervical cancer 1410
Endometrial cancer 1410
Ovarian cancer 1411
Brain tumours 1412
Brain metastases 1412
Primary brain tumours 1412
Head and neck cancer 1413
Malignant melanoma 1414
Metastatic cancer of unknown primary origin 1415

CORE SKILLS AND KNOWLEDGE

In high-income countries, nearly half of the population will develop cancer at some time during their lives, and it is second only to cardiovascular disease as a cause of death. Many advances have been made in recent years in terms of both the molecular understanding and the targeted treatments of different cancers.

Patients are diagnosed with cancer in a number of different settings: by screening, via their general practitioners, as an emergency admission, or in a routine hospital outpatient or inpatient encounter. Many professionals are involved in the care of patients with cancer, and management decisions are made at multidisciplinary team (MDT) meetings (usually tumour site specific), with input from surgeons, oncologists, radiologists, pathologists, palliative care physicians, clinical nurse specialists, and allied health professionals.

Oncologists are trained to deliver radiotherapy and systemic therapy. They must combine clinical skill and knowledge with empathy and excellent communication skills because they often see patients when they are at their most vulnerable.

Key learning outcomes in malignant disease include:

- Understanding the biology of cancer, including its presentation and natural history
- Understanding the goals of treatment (cure, prolonging life expectancy, and palliation) and the principles underlying the main classes of cancer intervention (surgery, chemotherapy, radiotherapy, hormonal therapy, targeted therapy, and immunotherapy)
- Being able to assess patients with malignant diseases, understanding where patients are in the course of their disease and treatment, and communicating likely outcomes
- Managing emergencies caused by the cancers themselves (e.g., spinal cord compression) or by anticancer treatment (e.g., neutropenic sepsis).

Opportunities for learning about malignant disease include reviewing patients undergoing urgent investigation for malignant disease (e.g., in urgent referral outpatient clinics), attending cancer MDT meetings, visiting chemo- and radiotherapy day units, attending clinics with treating clinicians, and shadowing clinical nurse specialists.

CLINICAL SKILLS

History

The history needs to be tailored to the clinical setting: in a patient being investigated for malignant disease, diagnostic questions are crucial, as is patient's functional status.

In a patient already undergoing active therapy, questions about treatment-related symptoms and general fitness are more relevant. At all stages, it is necessary to consider how the patient is coping with the emotional challenges of confirmed or suspected malignancy and to adopt an empathetic and caring approach.

The questions in **Box 41.1** are grouped into different sections that can be brought together according to the clinical setting. Patients often present with non-specific systemic symptoms such as weight loss, malaise and lethargy, pains all over, and 'not feeling great'; for these patients, it can be worth using 'review of systems' questions to slowly tease out specific system-based problems.

Investigations

Cancers require both radiological and histological evaluation for a complete diagnosis (**Box 41.2**). The findings are used to establish histological grade and radiological staging, according to internationally agreed systems specific to each type of malignancy.

Formulating a management plan

A diagnosis of cancer can have different implications for a patient. After seeing patients with a confirmed or suspected malignancy, it is crucial to establish where they are in the disease process. Are they previously fit and well but recently diagnosed with a potentially curable disease? Or are they into their third cycle of chemotherapy for metastatic malignancy, with few additional treatment options left? Asking the questions in **Box 41.3** will shed light on the patient's perspective, and the answers can then be supplemented by reviewing medical records for investigation results and details of MDT discussions.

i Box 41.1 Taking a history in malignant disease

General symptoms of malignancy

- Weight loss
- Anorexia, loss of appetite
- Malaise, weakness, and tiredness
- Paraneoplastic syndromes

Common site-specific symptoms

- Brain: seizures, any focal neurological symptoms, personality change
- Lung: cough, haemoptysis, chest wall pain (if invasive), breathlessness
- Oesophagus and stomach: mechanical dysphagia, early satiety, regurgitation, symptoms of anaemia caused by occult blood loss
- Hepatobiliary tract: upper abdominal pain and heaviness, jaundice
- Colon and rectum: altered bowel habit, rectal bleeding, tenesmus, symptoms of anaemia from blood loss
- Breast: breast lump, skin changes, nipple discharge
- Prostate: obstructive lower urinary tract symptoms
- Ovary: abdominal discomfort and bloating
- Cervix and endometrium: postmenopausal, intramenstrual, or postcoital bleeding
- Thyroid: neck swelling, thyrotoxicosis
- Urinary tract: painless haematuria
- Bone: dull bony pain, pathological fractures, back pain
- Mouth and throat: non-healing ulcer, voice change, hoarseness
- Skin: progressive moles or ulcers
- Metastases: may present with any of the above symptoms and are most common to the long bones, spine, liver, lung, and brain

Risk factors

- Smoking
- Obesity
- Air pollution
- Industrial exposure: asbestos, chemicals, dyes, dusts
- Viral infection: hepatitis viruses, human immunodeficiency virus, human papillomavirus, Epstein–Barr virus
- Lack of engagement with screening: breast, cervix, bowel

Common chemotherapy related side-effects

- Nausea and vomiting
- Oral mucositis
- Anorexia, loss of appetite
- Diarrhoea
- Peripheral neuropathy
- Fever: if present, always consider neutropenic sepsis, which is a medical emergency

i Box 41.2 Investigation of malignant disease

Radiology

- Typically, a CT scan of the chest, abdomen, and pelvis is performed to assess the local tumour site for the size of the tumour and for signs of localized invasion.
- ^{18}FDG-PET/CT scanning may be used if there is evidence of lymphadenopathy to assess whether lymph nodes are metabolically active, which may suggest malignant spread.
- Other forms of imaging are used if symptoms are suggestive of malignancy (e.g., a contrast-enhanced CT scan of the head to look for brain metastases, a radiolabelled technetium bone scan to assess for bony metastases).

Histology

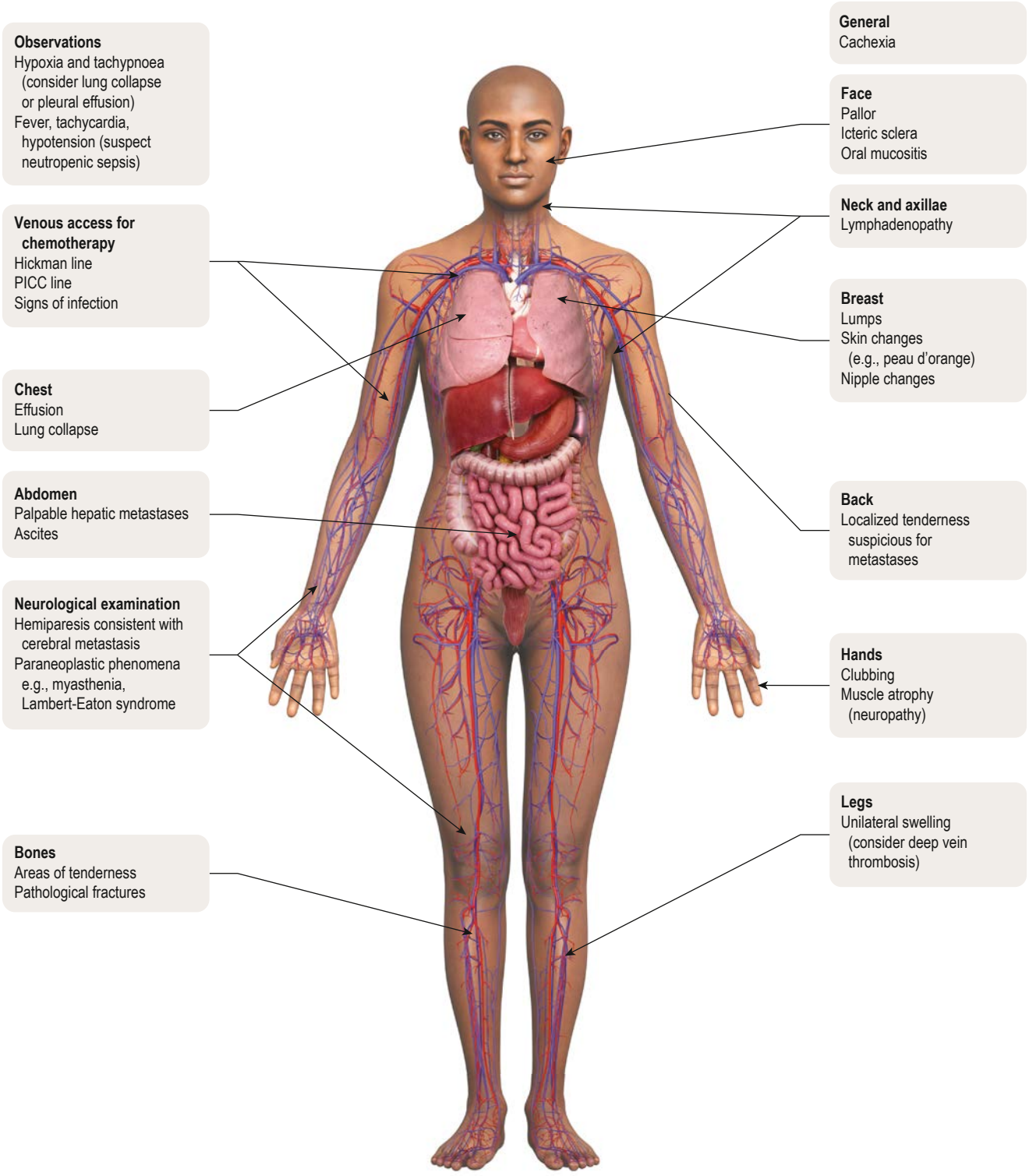
- Tissue biopsy is generally required for all forms of malignancy before treatment is considered.
- Basic histology confirms the tumour type (classified by the original cell type (e.g., adenocarcinoma from glandular epithelium), the degree of differentiation (poor differentiation is a bad prognostic sign), and local neurovascular invasion.
- More advanced techniques may look for expression of cell-surface molecules predictive of response to treatment (e.g., the oestrogen receptor in breast cancer).
- Genetic sequencing of tumour material has the potential to allow therapeutic regimens to be designed that target the particular genetic abnormalities in the malignant cells present.

CT, Computed tomography; *^{18}FDG*, ^{18}fluorodeoxyglucose; *PET*, positron emission tomography.

Box 41.3 Implications of a cancer diagnosis for the patient

- When was the patient diagnosed?
- What do they understand about the condition?
- Have they been offered surgery, chemotherapy, radiotherapy, or other forms of treatment?
- What do they understand about the aims of the treatment—cure or palliation?
- What do they understand about the likely outcomes of their disease?
- How do they feel about further courses of treatment?
- Who is supporting them through their disease, and what other sources of support (e.g., social networks, religious faith) do they have?
- Do they have late effects from their curative treatment?

Examination



BACKGROUND

Cancer is on the rise globally, driven by a range of factors, including obesity, urbanization, air pollution, tobacco smoking, and increasing life expectancy in the developing world. The Global Cancer Observatory predicts that the worldwide incidence of cancer will rise by 56% between 2020 and 2040. In the UK, the number of new cases of cancer continues to increase, with more than 375 000 new cancers registered between 2016 and 2018, equivalent to 1000 new cases a day. Of these, more than 50% were accounted for by breast (15%), prostate (14%), lung (13%), and colorectal (11%) malignancies, although globally, there is significant variation in the pattern of malignant disease, reflecting differences in environmental risk factors. The global burden of different types of cancer is shown in Fig. 41.1. Cancer is a disease of older people, with those older than 75 years accounting for 36% of total registered cancers.

Data from the UK suggests that although most patients with cancer are diagnosed after presenting to primary care with symptoms of malignancy, about 25% are diagnosed in the context of an emergency presentation requiring hospital attendance or admission. Patients diagnosed in an emergency situation tend to have worse outcomes because their cancers are more aggressive or more advanced at presentation.

An appreciation of the genetic and biological basis of cancer is crucial in understanding how the disease progresses and how different modalities of treatment work.

PRINCIPLES OF ONCOLOGY

THE BIOLOGY OF CANCER

Most human neoplasms are clonal in origin: that is, they arise from a single population of precursor or cancer stem cells. This process is typically initiated by genetic aberrations within this precursor cell that may be inherited (germline) or acquired (somatic). Cancer

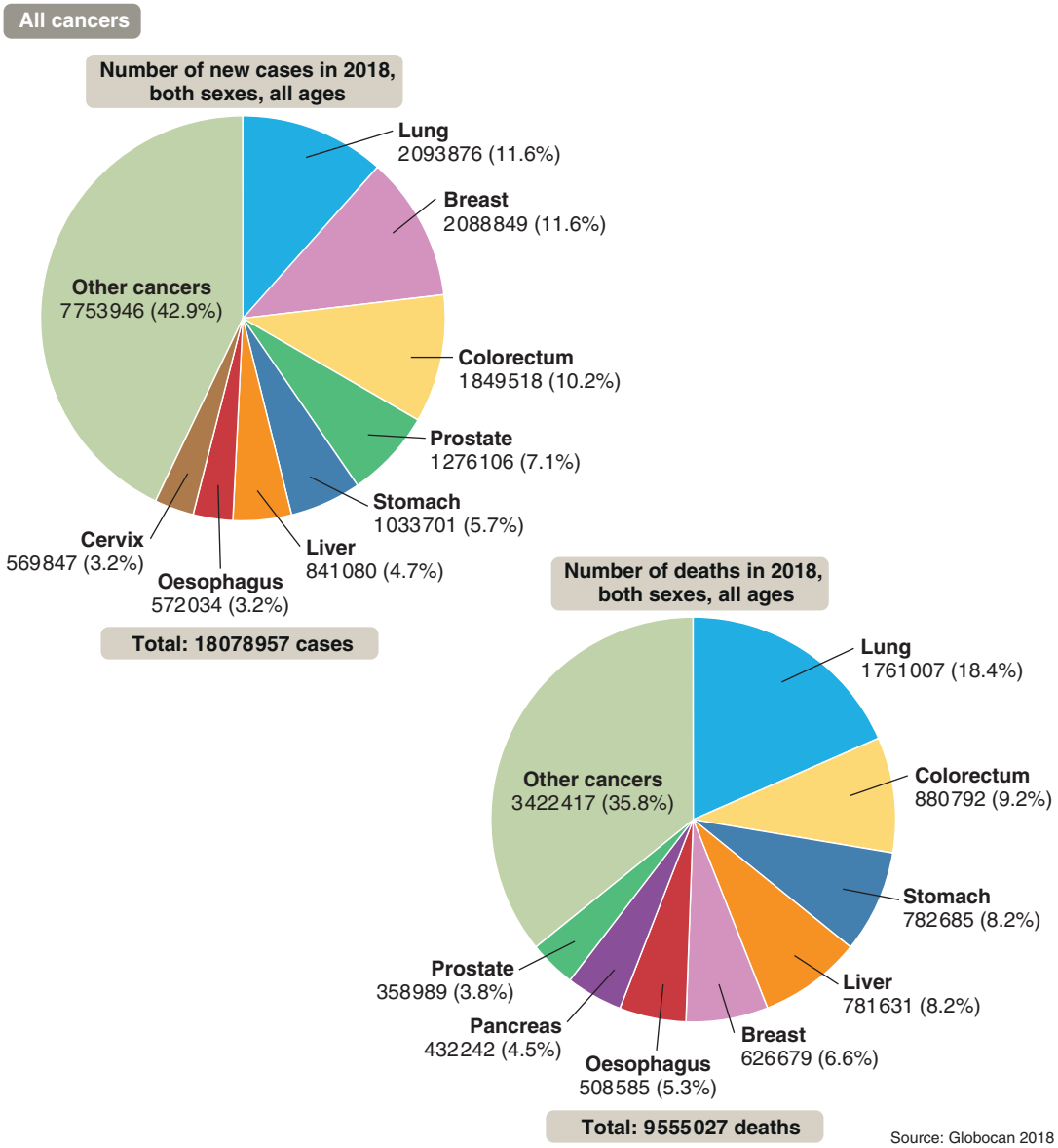


Fig. 41.1 The worldwide burden of cancer: new cases and deaths in 2018. (*GLOBOCAN 2018 Graph production: IARC [http://gco.iarc.fr/today]. World Health Organization.*)

becomes increasingly common with increasing age and can be related to a time-dependent accumulation of DNA damage that is not repaired by the normal mechanisms of genome maintenance, damage tolerance, and checkpoint pathways. The hallmark areas of loss or gain in function in developing cancer are shown in **Fig. 41.2**.

Whole-genome sequencing experiments are now identifying significant clonal genetic heterogeneity within tumours, including 'passenger' (inactive) and 'driver' (active) mutations. This heterogeneity is a potential source of treatment-resistant clones and is temporally variable, so that, at relapse, the predominant founder clone that has not been eradicated is seen to have acquired new mutations on its return; alternatively, a subclone, selected out by treatment pressure, has gained further mutations and led to a recurrence. A fuller description of the genetic basis of cancer is found in Chapter 5.

Evading growth suppression

Tumour cells are usually not recognized and killed by the immune system for two main reasons:

- There is a failure to express molecules such as human leucocyte antigen and co-stimulatory B7 molecules that are required for activation of cytotoxic, or 'killer', T lymphocytes.
- Tumours may also actively secrete immunosuppressive cytokines and cause a generalized immunosuppression.

Successful strategies for tumour vaccines that overcome these obstacles are developing in renal cancer and prostate cancer. The monoclonal antibody (MAb) ipilimumab, which works against the inhibitory cytotoxic T lymphocyte–associated antigen 4 (CTLA4) molecule expressed after T-cell activation, is used in patients with melanoma and renal cell carcinoma. Multiple tumours respond to immunotherapy targeting the programmed death pathway (PD/PD-ligand), a checkpoint inhibitor (see Fig. 6.8) that negatively affects the immune response. Blocking of either programmed death protein 1 (PD-1) (on T cells) or its ligand, programmed death ligand 1 (PDL-1), which is present in some tumours, has been associated with an immune response capable of producing durable tumour shrinkage. Common tumours, such as breast, lung, head and neck, gastrointestinal (GI), and transitional cell carcinomas (TCCs), have all shown response to PD-1– or PD-L1–targeted immunotherapy. This class of anticancer therapies, commonly known as immunotherapies, now has a role in tumours from common endometrial cancers to rare Merkel cell carcinomas and are likely to be commonly encountered in the acute medical setting as their use increases. They present a unique set of challenges because their side-effect profiles differ from those of conventional chemotherapy agents.

Inducing angiogenesis

As a tumour grows larger, its rate of growth progressively slows down. This occurs for many reasons, but outgrowing the blood supply to the tumour is a crucial factor. New vessel formation (angiogenesis) is stimulated by a variety of peptides produced both by tumour cells and by host inflammatory cells, such as basic fibroblast growth factor, angiopoietin 2, and vascular endothelial growth

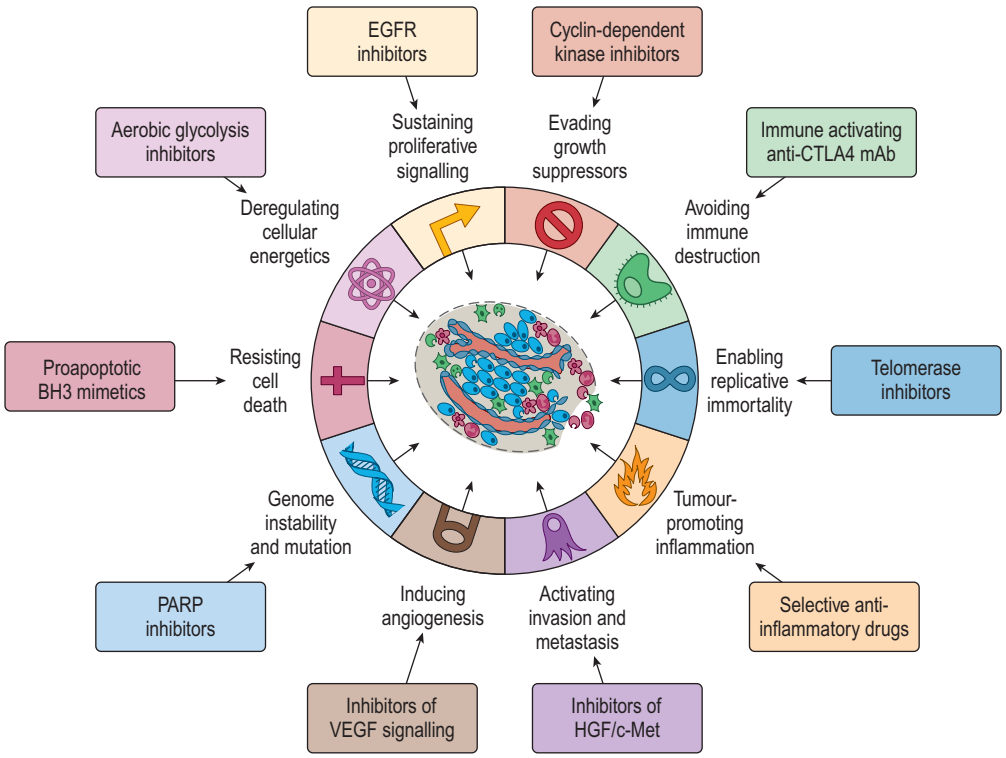


Fig. 41.2 The hallmarks of cancer. Ten biologically damaging capabilities are acquired during the multistep development of human tumours. Different colours indicate different actions of malignant cells, as depicted in the circle. Each of these can be targeted by anticancer drugs, as shown by the different-coloured rectangular boxes. In multitargeting, toxicity to normal healthy tissue is a major challenge. *BH3,* Member of bcl2 family of proteins/homologue BH3; *CTLA4,* cytotoxic T lymphocyte–associated antigen; *EGFR,* epidermal growth factor receptor; *HGF/c-Met,* hepatocyte growth factor/c-Met receptor complex; *PARP,* poly-ADP ribose polymerase; *VEGF,* vascular endothelial growth factor. (*From Hanahan D. The cancer wars 2: rethinking the war on cancer.* Lancet *2014; 383:558–563, with permission. Adapted from Hanahan D, Weinberg PA. The hallmarks of cancer: the next generation.* Cell *2011; 144:646–474*).

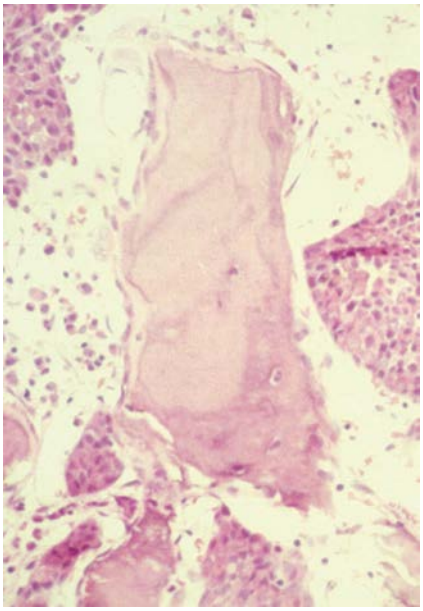


Fig. 41.3 Bone metastasis with osteoclast and osteoblast activation.

factors (VEGFs), which are stimulated by hypoxia. The anti-VEGF receptor MAb bevacizumab has had some success in the treatment of GI, cervical, and ovarian cancers. Alternatively, small molecules that target post-receptor response (tyrosine kinase inhibitors [TKIs]) may be used; examples include sunitinib and pazopanib.

Invasion and metastasis

Solid cancers spread by both local invasion and distant metastasis (Fig. 41.3), through the vessels of the blood and lymphatic systems. Infiltration into surrounding tissues is associated with loss of cell–cell cohesion, which is mediated by active homotypic cell adhesion molecules. Epithelial cadherin (E-cadherin) is expressed by many carcinomas and mutated in some, such as familial gastric carcinoma.

Invasion is also determined by the balance of activators and inhibitors of proteolysis. The matrix metalloproteinases and their tissue inhibitors are involved in tumour growth, invasion, metastasis, and angiogenesis and are being targeted by new therapeutic drugs for cancer treatment.

Dissemination of tumour cells occurs through intravasation into the vascular and lymphatic vessels and dissemination to distant sites, partly by chance but also because of specific interactions between receptors and cytokines found on stromal and tumour cells, such as tumour necrosis factor (TNF), interleukin 6 (IL-6), and chemokines.

Further reading

Greaves M, Maley CC. Clonal evolution in cancer. *Nature* 2012; 481(7381):306–13.
Hanahan D. Hallmarks of cancer: new dimensions. *Cancer Discov* 2022; 12(1):31–46.
Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA Jr, Kinzler KW. Cancer genome landscapes. *Science* 2013; 339(6127):1546–58.
Waldman AD, Fritz JM, Lenardo MJ. A guide to cancer immunotherapy: from T cell basic science to clinical practice. *Nat Rev Immunol* 2020; 20(11):651–668.

AETIOLOGY AND EPIDEMIOLOGY

In most patients, the cause of cancer is unknown, probably representing a multifactorial interaction between individual genetic predispositions and environmental factors.

? Box 41.4 Some causative factors associated with the development of cancer

Causative factor	Neoplasms
Smoking	Mouth, pharynx, oesophagus, larynx, lung, bladder, lip
Obesity	Breast, bowel, pancreas, gynaecological, oeseophageal, thyroid
Alcohol	Mouth, pharynx, larynx, oesophagus, colorectal
Iatrogenic	
Alkylating agents	Bladder, bone marrow
Oestrogens	Endometrium, vagina, breast, cervix
Androgens	Prostate
Radiotherapy (e.g., mantle radiotherapy)	Breast, bronchus
Diet	
High-fat diet, obesity	Colorectal
Environment, occupation	
Air pollution	Lung
Vinyl chloride	Liver (angiosarcoma)
Polycyclic hydrocarbons	Skin, lung, bladder
Aromatic amines	Bladder
Asbestos	Lung, mesothelium
Ultraviolet light	Skin, lip
Radiation	Leukaemia, thyroid cancer
Aflatoxin	Liver
Biological agents	
Hepatitis B virus	Liver (hepatocellular carcinoma)
Hepatitis C virus	Liver (hepatocellular carcinoma)
Epstein–Barr virus	Nasopharyngeal carcinoma
Human papillomavirus types 16 and 18	Cervix
	Oral (type 16)
Schistosoma japonicum	Bladder
Helicobacter pylori	Stomach

Aetiology

Genetic factors

These are outlined in Chapter 5 ('The genetic basis of cancer' section).

Environmental factors

A wide range of environmental factors has been identified as being associated with the development of malignancy (Box 41.4) and may be amenable to preventive action, such as smoking cessation, dietary modification, and antiviral immunization (Box 41.5). Environmental factors interact with genetic predisposition. For example, subsequent generations of people moving from countries with a low incidence of breast or colon cancer to those with a high incidence acquire the cancer incidence of the country to which they have moved, and Northern Europeans exposed to strong ultraviolet (UV) radiation have the highest risk of developing melanoma.

Tobacco

Since the mid-1990s, the incidence of lung cancer, in both males and females, has increased dramatically worldwide but is falling in many high-income countries. The association of smoking with lung cancer is indisputable, and causative mechanisms have been identified;

Box 41.5 Key messages for a healthy lifestyle and the prevention of cancer

- Stop smoking.
- Consume alcohol in moderation.
- Maintain a healthy weight.
- Take moderate exercise.
- Eat healthily (fruit and vegetables; high fibre; low fat, salt, and sugar).
- Limit sun exposure.
- Minimize occupational risk.
- Minimize radiation exposure.
- Vaccinate against hepatitis B and human papillomavirus.

cigarette tobacco is responsible for one-third of all deaths from cancer in the UK. Smoking not only causes lung cancer but also is associated with cancer of the head and neck, GI tract, kidney, cervix, and bladder. Smoking cessation is discussed in Chapter 20 ('Smoking cessation' section).

Obesity

Obesity has been identified as the second largest cause of cancer in the UK after tobacco smoking. It has been associated with increased incidents of many types of tumour, including tumours of the GI and biliary tract, breast, uterus, ovary, and thyroid. Cancer risk is related to the degree and duration of obesity and reduces with successful weight loss. Causative mechanisms identified include DNA damage through chronic inflammation, metabolic stress as a result of insulin resistance, and increased growth stimulus on cells mediated by high serum levels of growth hormone and sex steroids.

Alcohol

Alcohol is associated with cancers of the upper respiratory and GI tracts, and it also interacts with tobacco in the aetiology of these tumours. It may be associated with an increased risk of breast cancer.

Diet

One-third of cancer deaths have been attributed to dietary factors, although it is often difficult to differentiate these from other epidemiological causes. For example, whereas the incidence of stomach cancer is particularly high in the Far East, breast and colon cancers are more common in Western high-income countries. Many associations have been observed without a causative link being identified between the incidence of cancer and the consumption of food constituents, including dietary fibre, red meat, saturated fats, salted fish, vitamin E, vitamin A, and many others.

Ultraviolet light

UV light is known to increase the risk of skin cancer (basal cell, squamous cell, and melanoma). The incidence of melanoma is therefore particularly high in the White Anglo-Celtic populations of Australia, New Zealand, and South Africa, where exposure to UV light is combined with a genetically predisposed population.

Arsenic contamination

Arsenic contamination of water supplies has been linked to a high incidence of lung and colon cancers in South-east Asia, particularly where boreholes are the main water source.

Occupational factors

In 1775, Percival Pott described the association between carcinogenic hydrocarbons in soot and the development of scrotal epitheliomas in chimney sweeps. The principal causes nowadays are ***asbestos*** (lung and mesothelial cancer) and ***polycyclic hydrocarbons*** from fossil fuel combustion (skin, lung, and bladder cancers). ***Organic chemicals***, such as benzene, may cause the development of bone marrow conditions, such as myelodysplastic syndrome and acute myeloid leukaemia.

Infectious agents

The geographical distribution of some rare malignancies led to the discovery that cancer may be caused by infective agents. Chronic persistent infection provides growth stimulation, and many viruses contain transforming viral oncogenes.

T-cell leukaemia, seen almost exclusively in residents of the southern island of Japan and in the West Indies, is caused by infection with the locally endemic retrovirus human T-cell leukaemia (or lymphotropic) virus type 1 and integration of the oncogene *TAX* into the cellular genome.

Hepatocellular carcinoma (HCC) occurs in patients with hepatitis B and C virus infections, and Burkitt's lymphoma and nasopharyngeal carcinoma are associated with the Epstein–Barr virus (EBV). EBV is also linked with Hodgkin lymphoma (see Chapter 39).

Patients with human immunodeficiency infection or immunosuppression from organ transplantation have an increased incidence of EBV-related lymphoma and herpesvirus-8-associated Kaposi's sarcoma.

Human papillomavirus (HPV) infection types 16 and 18, for which an effective vaccine is available, is an established cause of a rise in cervical cancer among sexually active females; it has also been associated with an increase in head and neck cancers. Bacterial infection with *Helicobacter pylori* predisposes to the development of gastric cancer and gastric lymphoma, and *Schistosoma japonicum* infection predisposes to the development of squamous cell carcinomas (SCCs) in the bladder.

Medication

Oestrogens have been implicated in the development of vaginal, endometrial, and breast carcinoma. Certain cytotoxic drugs, such as those given for Hodgkin lymphoma (see Chapter 39), are associated with an increased incidence of secondary acute myeloid leukaemia and bladder and lung cancers. Androgens have been linked with both benign and malignant liver tumours.

Radiation

Accidental

The nuclear incidents at Hiroshima, Nagasaki, and Chernobyl led to an increased incidence of leukaemia after 5 to 10 years in the exposed population, as well as greater incidences of thyroid and breast cancer. Radiation workers are at a higher risk of malignancy because of occupational exposure unless precautions are taken to minimize the danger by using personal and environmental shielding and to record and limit the amount of personal exposure.

Therapeutic

Long-term survivors who have been treated with radiotherapy, such as those with Hodgkin lymphoma, for which patients were historically treated with wide radiation fields, have an increased incidence of cancer, particularly at the radiation field margins. These patients should have regular surveillance follow-up, for example, for the early identification of breast cancer in the previously treated Hodgkin lymphoma cohort of patients.

Diagnostic

Imaging procedures involving radiation exposure are associated with an increased risk of cancer. This risk is cumulative, dose dependent, and time dependent, so children are at higher risk than

Box 41.6 Radiation exposure from common diagnostic radiological procedures[a]

Procedure	Radiation exposure (mSv)
Chest X-ray	0.02
Chest CT	7
Abdomen CT	8–10
Whole-body CT	20
Percutaneous coronary intervention	15
Myocardial perfusion imaging	15.6

[a]UK background radiation is 2.6 mSv per year. 1 mSv carries a lifetime cancer risk of 1 in 17 500, and 5 mSv carries a risk of 1 in 3500.
CT, Computed tomography.
Modified from Smith-Bindman R, Lipson J, Marcus R et al. Radiation dose associated with common computed tomography examinations and the associated lifetime attributable risk of cancer. Arch Intern Med *2009; 169:2078–2086 and Fazel R, Krumholz HM, Wang Y et al. Exposure to low-dose ionizing radiation from medical imaging procedures.* N Engl J Med *2009; 361:849–857.*

adults. The cancer risk of various common investigations is shown in **Box 41.6**. All doctors should strive to minimize diagnostic exposure to radiation where possible by using alternative modalities such as ultrasonography or magnetic resonance imaging (MRI). Good documentation of radiation doses is required, particularly in children and pregnant females.

Epidemiology

Geographical distribution

The incidence of cancer across the world depends on local environmental factors, diet, and the genetics of the local population (see interactive maps at the International Agency for Research on Cancer: https://gco.iarc.fr/today/en/dataviz/pie?mode=population&group_populations=0, and **Fig. 41.4**). Age is also a factor because most cancers occur in those older than the age of 65 years, who make up 3.3% of the population in Africa compared with 15.2% in Europe. Reproductive patterns influence the incidence of breast cancer. Migrating individuals often take on the risks associated with local environmental factors.

Other factors

Incidence and mortality rates are closely linked for cancers for which treatment has yet to make significant improvements, such as lung, pancreas, stomach, and liver cancers. However, in countries with effective screening programmes, there are increasing incidences but decreasing mortality rates for breast, cervix, bowel, and prostate cancers.

Further reading

Global Burden of Disease 2019 Cancer Collaboration. Cancer incidence, mortality, years of life lost, years lived with disability, and disability-adjusted life years for 29 cancer groups from 2010 to 2019: a systematic analysis for the global burden of disease study 2019. *JAMA Oncol* 2022; 8(3):420–444.
Tomasetti C, Li L, Vogelstein B. Stem cell divisions, somatic mutations, cancer etiology, and cancer prevention. *Science* 2017; 355(6331):1330–1334.

SCREENING AND INVESTIGATIONS

Asymptomatic detection through screening

Most common cancers start as focal microscopic clones of transformed cells, and diagnosis becomes likely only when sufficient tumour bulk has accumulated to cause symptoms or signs. To make an earlier diagnosis and enhance the curative possibility, an increasing number of screening programmes are being developed that either target the asymptomatic or pre-invasive stages of the cancer, as in cervix, breast, lung, and colon cancers, or use serum tumour markers, as in prostate cancer. 'Circulating tumour DNA' (ctDNA) is a subtype of serum cell-free DNA derived from cancer cells that contains mutations known to associate with malignancy. It is a promising biomarker that may allow early diagnosis of cancer from a single blood test.

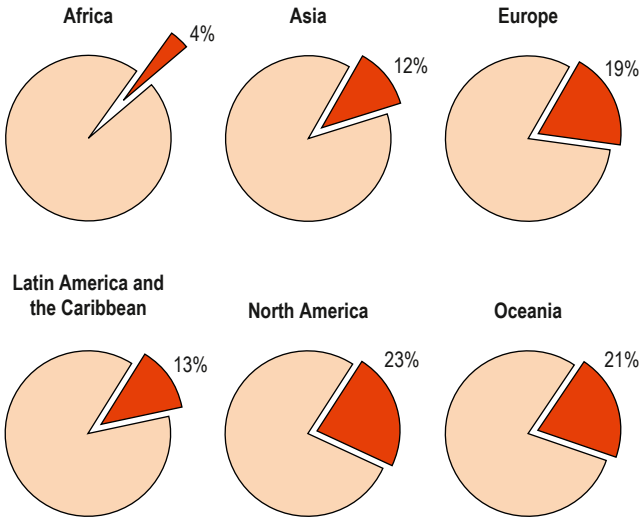


Fig. 41.4 Percentage of all deaths from cancer in different regions of the world. *(From https://gco.iarc.fr/today/en/dataviz/pie?mode=population&group_populations=0.)*

Genetic screening can be used to target screening to those groups at most risk of developing cancer, such as *BRCA1*-positive individuals who may develop breast or ovarian cancer (**Box 41.7**).

The aim of screening programmes is to improve individual or population survival by detecting cancer in its very early stages when patients are asymptomatic. This strategy depends on finding tests that are sufficiently sensitive and specific, using detection methods that identify cancer before it has spread, and having curative treatments that are practical and consistent with maintenance of a normal lifestyle and quality of life.

In the UK, population screening is provided for breast, cervical, and colon cancers. Lung cancer screening is available in certain parts of the UK and is gaining momentum. Prostate cancer can be detected using a simple blood test available to individuals via annual check-ups, but a screening programme has not been instituted in the UK because of a lack of evidence that the benefit outweighs the risks.

Unfortunately, earlier diagnosis does not necessarily mean longer survival times, and randomized trials of screening programmes are necessary to prove benefit. With lead-time bias, patients diagnosed via screening may merely be treated at an earlier date, so survival appears longer; death may still occur at the same time from the point of genesis of the cancer (**Fig. 41.5**). With length-time bias, a greater number of slowly growing tumours are detected when asymptomatic individuals are screened, which may lead to a false impression of an improvement in survival times.

The characteristics of effective screening programmes are shown in **Box 41.8**.

Symptomatic patients with cancer

Patients may offer information about predisposing conditions and family history that alerts the clinician to the likelihood of a cancer

Box 41.7 Familial cancer syndromes for which genetic screening may be relevant

Syndrome	Gene	Neoplasms
Autosomal dominant		
Retinoblastoma	*RB1*	Eye
Wilms' tumour	*WT1*	Kidney
Li–Fraumeni	*p53*	Sarcoma, brain, leukaemia
Neurofibromatosis type 1	*NF1*	Neurofibromas, leukaemia
Familial adenomatous polyposis	*APC*	Colon
Lynch's syndrome (previously hereditary non-polyposis colon cancer)	*MLH1* and *MSH2*	Colon, endometrium
Hereditary diffuse gastric cancer syndrome	E-cadherin	Stomach
Breast, ovary families	*BRCA1*	Breast, ovary
	BRCA2	
	p53	
Melanoma	*p16*	Skin
Von Hippel–Lindau	*VHL*	Renal cell carcinoma, haemangioblastoma
Multiple endocrine neoplasia type 1	*MEN1*	Pituitary, pancreas, parathyroid
Multiple endocrine neoplasia type 2	*RET*	Thyroid, adrenal medulla
Autosomal recessive		
Xeroderma pigmentosa	*XP*	Skin

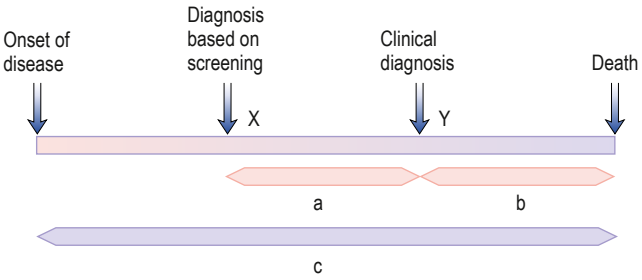


Fig. 41.5 Lead-time bias. Earlier diagnosis, at X, made by screening tests before the clinical diagnosis, at Y, suggests an increased survival time of a + b. The actual survival time (c) remains unchanged.

diagnosis. Many patients present with a history of tumour site–specific symptoms, such as pain, and physical signs, such as a mass, which readily identify the primary site of the cancer (Box 41.9). However, some patients seek medical attention only when more systemic and non-specific symptoms occur, such as weight loss, night sweats, fever, fatigue, recurrent infections, and anorexia. These usually indicate a more advanced stage of disease except in some paraneoplastic and ectopic endocrine syndromes. Other patients are diagnosed only on discovery of established metastases, such as the abdominal distension of ovarian cancer, the back pain of metastatic prostatic cancer, or the liver enlargement of metastatic GI cancer.

Box 41.8 Characteristics of an effective screening programme

- It must be affordable for the healthcare system.
- It must be acceptable to all social groups so they attend for screening.
- It must have a good discriminatory index between benign and malignant lesions.
- It must show a reduction in mortality rate from the cancer.

Paraneoplastic syndromes are indirect effects of cancer (Box 41.10 and Fig. 41.6) that are often associated with specific types of cancer and may be reversible with treatment of the cancer. The effects and mechanisms can be very variable. For example, in Lambert–Eaton syndrome (see Chapter 28, 'Lambert–Eaton myasthenic syndrome' section), there is cross-reactivity between tumour antigens and the normal tissue acetylcholine receptors at neuromuscular junctions.

The ***coagulopathy of cancer*** may present with thrombophlebitis, deep vein thrombosis, and pulmonary emboli, particularly in association with cancers of the pancreas, endometrium, lung, stomach, kidney, brain, and bladder. Some 18% of patients with recurrent pulmonary embolus will be found to have an underlying cancer, and many patients with cancer are at increased risk of venous thromboembolism (VTE) after diagnosis. All patients with active cancer admitted to hospital are at high risk of VTE and should be given prophylaxis with, for example, subcutaneous low-molecular-weight heparin in the absence of any contraindications (see Chapter 40).

Other cancer symptoms may be related to peptide or hormone release, such as in carcinoid or Cushing's syndrome.

Cachexia of advanced cancer is thought to be attributable to systemic effects, including the release of chemokines such as TNF, or local pressure effects, in which certain cancers may impair the patient's ability to swallow or may cause early satiety. The unexplained loss of more than 10% of a patient's body weight should always act as a red flag and be investigated.

Cancer-associated immunosuppression can lead to reactivation of latent infections such as herpes zoster (Fig. 41.7) and tuberculosis.

Serum tumour markers

Tumour markers are intracellular proteins or cell-surface glycoproteins released into the circulation and detected by immunoassays. Examples are given in Box 41.11. Values in the normal range do not necessarily equate with the absence of disease and a positive result must be corroborated by histology because these markers can be seen in many benign conditions. They are most useful in the serial monitoring of response to treatment. MicroRNA may be more tumour specific and offer the possibility of increased sensitivity compared with classical tumour markers.

Cancer imaging

Radiological investigation is required at various stages, including at initial diagnosis and staging of the disease, during the monitoring of treatment efficacy, at the detection of recurrence, and for the diagnosis and treatment of complications. See Chapter 8 for a fuller discussion of the role of imaging in malignant disease.

Contrast agents are used for increased structural discrimination in cross-sectional imaging and can be further enhanced with functional specificity for metabolically active tissue with ^{18}fluorodeoxyglucose (^{18}FDG) positron emission tomography (PET)/computed tomography (CT) scan, although these specialist scans are generally not done for early-stage disease and should generally be requested only after recommendation by a cancer specialist or MDT meeting.

Box 41.9 Symptoms and signs of malignant disease

Degree of spread	Anatomical location	Examples of clinical problems
Local	Mass	Thyroid nodule, pigmented naevus, breast lump, abdominal mass, testicular mass
	Local infiltration of the skin	Dermal nodules, peau d'orange, ulceration
	Local infiltration of a nerve	Neuropathic pain and loss of function, Horner's syndrome, cord compression, Pancoast tumour, focal CNS deficit, hypopituitarism
	Local infiltration of a vessel	Venous thrombosis, tumour emboli, haemorrhage (e.g., gastrointestinal)
	Obstruction of the viscera or duct	Small or large bowel obstruction, dysphagia, SVC obstruction, obstructive uropathy, acute kidney injury, urinary retention, stridor, lobar collapse, pneumonia, cholestatic jaundice
Nodal	Peripheral	Supraclavicular fossa, Virchow's node, lymphoedema
	Central	Mediastinum: SVC obstruction; porta hepatis: obstructive jaundice, para-aortic nodes, back pain
Metastatic	Lung	Pleuritic pain, cough, shortness of breath, lymphangitis and respiratory failure, recurrent pneumonia
	Liver	RUQ pain, anorexia, fever, raised serum liver enzymes, jaundice
	Brain	Headache and vomiting of raised intracranial pressure, focal deficit, coma, seizure
	Bone	Bone pain, cord compression, fracture, hypercalcaemia
	Pleura	Effusion, pain, shortness of breath
	Peritoneum	Ascites, Krukenberg tumours
	Adrenal	Addison's disease (hypoadrenalism)
	Umbilicus	Sister Mary Joseph's nodule

CNS, Central nervous system; *RUQ,* right upper quadrant; *SVC,* superior vena cava.

Box 41.10 Paraneoplastic syndromes

Syndrome	Tumour	Serum antibodies
Neurological		
Lambert–Eaton syndrome	Lung (small-cell) carcinoma	Anti-VGCC
Peripheral sensory neuropathy	Lung (small-cell), breast, ovary, lymphoma	Anti-Hu
Cerebellar degeneration	Lung (particularly small-cell) lymphoma	Anti-Yo
Opsoclonus or myoclonus	Breast, lung (small-cell)	Anti-Ri
Stiff person syndrome	Breast, lung (small-cell)	Anti-amphiphysin
Limbic, hypothalamic, or brainstem encephalitis	Lung	Anti-Ma protein
	Testicular	Anti-NMDAR
Endocrine or metabolic		
SIADH	Lung (small-cell)	
Ectopic ACTH secretion	Lung (small-cell)	
Hypercalcaemia	Renal, breast, myeloma, lymphoma	
Fever	Lymphoma, renal	
Musculoskeletal		
Hypertrophic pulmonary osteoarthropathy	Lung (non–small-cell)	
Clubbing	Lung	
Skin		
Dermatomyositis or polymyositis	Lung and upper gastrointestinal	
Acanthosis nigricans	Mainly gastric	
Velvet palms	Gastric, lung (non–small-cell)	
Hyperpigmentation	Lung (small-cell)	
Haematological		
Erythrocytosis	Renal cell carcinoma, hepatocellular carcinoma, cerebellar haemangioblastoma	
Thrombocytosis	Ovarian cancer	
Migratory thrombophlebitis	Pancreatic adenocarcinoma	
DVT	Adenocarcinoma	
DIC	Adenocarcinoma	
Renal		
Membranous nephropathy (frequently presenting with nephrotic syndrome)	Lung, stomach, and colon cancers	
Amyloidosis (presenting with nephrotic syndrome)	Myeloma	

ACTH, Adrenocorticotrophic hormone; *DIC,* disseminated intravascular coagulation; *DVT,* deep vein thrombosis; *NMDAR,* N-methyl-D-aspartate receptors; *SIADH,* syndrome of inappropriate antidiuretic hormone secretion; *VGCC,* voltage-gated calcium channels.

Fig. 41.6 Paraneoplastic peripheral neuropathy. There is wasting of the small muscles of the hands.

Fig. 41.7 Reactivated herpes zoster (shingles).

Radionuclide imaging of sentinel lymph nodes is used to guide lymphatic surgery in many cancers. Tumour-targeted contrast agents can improve detection rates; examples include the radiolabelled MAb rituximab for lymphoma and radiolabelled small molecules such as octreotide for neuroendocrine tumours. Research into the use of reporter agents that become visible only on activation within the tumour environment holds the promise of greater sensitivity and specificity in the future.

Biopsy and histological examination

A diagnosis of cancer may be suspected by both the patient and doctor, but advice about treatment can usually be given only on the basis of a tissue diagnosis. This may be obtained by endoscopic, radiologically guided, or surgical biopsy or sometimes by cytology (e.g., lung cancer may be diagnosed by sputum cytology). Malignant lesions can be distinguished morphologically from benign ones by the pleomorphic nature of the cells; increased numbers of mitoses; nuclear abnormalities of size; chromatin pattern and nucleolar organization; and evidence of invasion into surrounding tissues, lymphatics, or vessels. The degree of differentiation of the tumour has prognostic significance: generally speaking, well-differentiated tumours have a better prognosis than poorly differentiated or anaplastic ones. In some tumours, diagnosis and treatment may be combined within a single surgical procedure, with the exact nature of the procedure being carried out varying depending on the presence of malignancy. An intraoperative provisional histological opinion can be rapidly obtained with a tissue sample processed using 'frozen section' techniques, rather than having the sample paraffin embedded, which takes more than a day.

Immunocytochemistry using MAbs against tumour antigens is very helpful in differentiating between lymphoid and epithelial tumours and between some subsets of these (see **Box 41.18**). However, there is much overlap in the expression of many of these ***tissue tumour markers***, and some adenocarcinomas and squamous carcinomas do not bear any distinctive immunohistochemical markers that are diagnostic of their primary site of origin.

Molecular markers of genetic abnormalities have long been available in haematological cancers and are increasingly available in solid cancers. For example, ***fluorescence in situ hybridization*** (see Chapter 39) can be used to look for characteristic chromosomal translocations, as in lymphoma and leukaemia, as well as deletions or amplifications, as in breast cancer (see Chapter 47, 'Breast cancer' section). The development of the new targeted MAbs and TKIs has been guided by increasing use of mutation analysis of common oncogenes, such as *RAS* and *RAF* in colon cancer and melanoma. ***Tissue microarrays*** can identify patterns of multiple genomic alterations and single nucleotide polymorphisms, as in breast cancer and lymphoma (see Chapter 39, 'Lymphomas' section), and ***RNA assays*** with reverse transcriptase polymerase chain reaction can be used to identify tissue of origin with prognostic and predictive relevance.

Genomics and proteomics have been used to guide the use of novel therapies, such as imatinib in gastrointestinal stromal tumours (GISTs); trastuzumab, pertuzumab, and lapatinib in breast cancer;

Box 41.11 Serum tumour markers

Serum tumour marker	Neoplasms
α-Fetoprotein (AFP)	Hepatocellular carcinoma and non-seminomatous germ cell tumours of the gonads
β-Human chorionic gonadotrophin (β-hCG)	Choriocarcinomas, germ cell tumours (testicular), and lung cancers
Prostate-specific antigen (PSA)	Carcinoma of the prostate
Carcinoembryonic antigen (CEA)	Gastrointestinal cancers
CA-125	Ovarian cancer
CA19-9	Gastrointestinal cancers, particularly pancreatic cancer
CA15-3	Breast cancer
Osteopontin	Many cancers, including mesothelioma
Calcitonin	Medullary carcinoma of the thyroid
Thyroglobulin	Papillary and follicular thyroid cancer

osimertinib and crizotinib in lung cancer; and cetuximab and bevacizumab in colon cancer.

Further reading

Baracos VE, Martin L, Korc M, Guttridge DC, Fearon KCH. Cancer-associated cachexia. *Nat Rev Dis Primers* 2018; 4:17105.
Fenton JJ, Weyrich MS, Durbin S, Liu Y, Bang H, Melnikow J. Prostate-specific antigen-based screening for prostate cancer: evidence report and systematic review for the us preventive services task force. *JAMA* 2018; 319(18):1914–1931.
Khorana AA, Mackman N, Falanga A et al. Cancer-associated venous thromboembolism. *Nat Rev Dis Primers* 2022; 8(1):11.

PRINCIPLES OF CANCER TREATMENT

Aims of treatment

Optimal cancer treatment is delivered by an MDT that coordinates the delivery of the appropriate anticancer treatment, supportive and symptomatic care, and psychosocial support.

Establishment of agreed patient pathways has enabled more effective and timely delivery of care and post-treatment rehabilitation. Central to this endeavour is the involvement of the patient through education about the nature of their disease and the treatment options available.

A curative approach

For many solid tumours, local control is necessary, but not sufficient, for cure because of the presence of systemic (microscopic) disease. Improvement in the rate of cure of most cancers is thus dependent on earlier detection to increase the success of local treatment and effective systemic response. The likelihood of cure of the systemic disease rests on the type of cancer and its expression of appropriate treatment targets, its drug sensitivity, and tumour bulk (microscopic or clinically detectable). A few rare cancers are so chemosensitive that even bulky metastases can be cured (e.g., gonadal germ cell tumours and choriocarcinoma). For most common solid tumours, such as lung, breast, and colorectal cancers, there is no current cure of bulky (clinically detectable) metastases, but micrometastatic disease treated by adjuvant systemic therapy after surgery can be cured in 10% to 20% of patients.

Adjuvant therapy for solid tumours

Micrometastatic spread by lymphatic or haematological dissemination often occurs early in the development of the primary tumour and can be demonstrated by molecular biological methods capable of detecting the small numbers (1 in 10^6) of circulating cells. Studies correlating prognosis with histological features of the primary cancer, such as differentiation, invasion of blood vessels or regional lymph nodes, and molecular markers such as human epidermal growth factor receptor 2 (HER2) in breast cancer, enable risk stratification and increasing individualization of therapy.

- ***Adjuvant therapy*** is defined as treatment given, in the absence of macroscopic evidence of metastases, to patients at risk of recurrence from micrometastases after treatment given for the primary lesion.
- ***Neoadjuvant therapy***, alternatively, is given before primary surgery, both to shrink the tumour to improve local excision and to treat any micrometastases as soon as possible.

The success of adjuvant treatment across many tumour types relies on careful selection of patients according to defined risk criteria and reduction of treatment toxicity to reach a balanced risk-to-benefit ratio. However, the majority of patients who receive such treatment do not benefit, either because they were already cured or because the cancer is resistant to the treatment. Better tests, such as gene arrays and circulating tumour cells, are being developed to identify those with micrometastases, who really do benefit from treatment.

A palliative approach

When cure is not possible, palliation—that is, relief of tumour symptoms, preservation of quality of life, and prolongation of life—is possible in many cancers in proportion to their drug and radiation sensitivity. There is, on average, a 2- to 18-month prolongation in median life expectancy with treatments for solid tumours; individuals with the most responsive tumours experience the greatest benefit. In addition, through early assessment during treatment, it is possible to stop if no evidence of benefit is demonstrable early on so as to minimize exposure to toxic and unsuccessful treatment.

Assessment before treatment

Staging

Before a decision about treatment can be made, not only does the type of tumour need to be established but also its extent and distribution (**Box 41.2**). Various 'staging investigations' are therefore performed before a treatment decision is made. To be useful clinically, the staging system must subdivide the patients into groups with different prognoses, which can guide treatment selection. The TNM (tumour, node, metastases) classification (see **Box 41.12**) can be adapted for application to most common cancers.

Performance status

In addition to anatomical staging, the person's age and general state of health need to be taken into account when planning treatment. The latter has been called 'performance status' and is of great prognostic significance for all tumour types (**Box 41.13**). Performance status reflects the effects of the cancer on the patient's functional capacity. An alternative performance rating scale is provided by Karnofsky. With a performance status of 2 or more, response to and survival after treatment are greatly reduced for most tumour types.

Assessment of the benefits of treatment

A measurable response to treatment can serve as a useful early surrogate marker when deciding whether to continue a given treatment for an individual patient. Trials to assess response to treatment in advanced disease have identified active agents for use in the more curative setting of adjuvant treatment of patients with early-stage disease.

Response to treatment can be subjective or objective.

- ***A subjective response*** is one perceived by the patient in terms of, for example, relief of pain and dyspnoea or improvement in appetite, weight gain, or energy. Such a subjective response is a major aim of most palliative treatments. Quantitative measurements of these subjective symptoms (patient-reported outcome measures) form a part of the assessment of response to chemotherapy, especially in situations when cure is not possible and when the aim of treatment is to provide prolongation of good-quality life. In these circumstances, measures of quality of life enable an estimate of the balance of benefit and side-effects to be made.
- ***An objective response*** to treatment is assessed clinically and radiologically (**Fig. 41.8**). The term 'remission' is often used

Box 41.12 Staging and survival of colorectal cancers: TNM and Dukes' systems

TNM classification	Description	Notation	Modified Dukes' classification	5-year survival rate (%)
Stage I (N0, M0)	Tumours invade submucosa Tumours invade muscularis propria	T1 T2	A	90
Stage IIA (N0, M0) **Stage IIB**	Tumours invade into subserosa Tumours invade directly into other organs	T3 T4	B	70 65
Stage III (M0) **Stage IIIB**	T1, T2 + one to three regional lymph nodes involved T3, T4 + one to three regional lymph nodes involved	N1 N1	C	60 35
Stage IIIC	Any T + four or more regional lymph nodes	N2	C	25
Stage IV	Any T, any N + distant metastases	M1	D	7

M, Metastasis; *N,* node; *T,* tumour.

Box 41.13 Eastern Cooperative Oncology Group performance status scale

Status	Description
0	Asymptomatic, fully active, and able to carry out all pre-disease performance without restrictions
1	Symptomatic, fully ambulatory but restricted in physically strenuous activity and able to carry out performance of a light or sedentary nature (e.g., light housework, office work)
2	Symptomatic, ambulatory, and capable of all self-care but unable to carry out any work activities; up and about >50% of waking hours; in bed <50% of the day
3	Symptomatic, capable of only limited self-care, confined to a bed or chair >50% of waking hours but not bedridden
4	Completely disabled, cannot carry out any self-care, totally bedridden

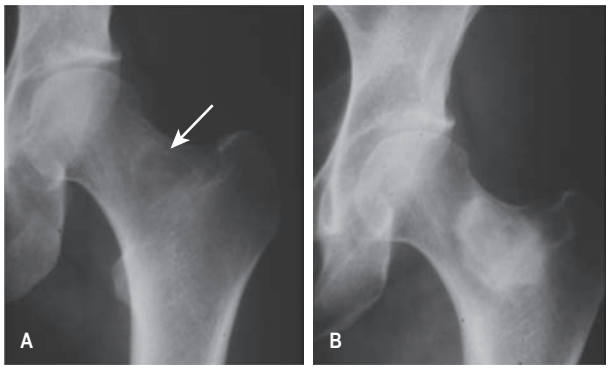


Fig. 41.8 Radiological response to chemotherapy and radiotherapy. X-rays showing response to chemotherapy and radiotherapy. **(A)** The *arrow* points to lytic bone metastasis. **(B)** Bone metastasis showing response to chemotherapy and radiotherapy with sclerotic new bone formation.

synonymously with 'response', which, if complete, means an absence of detectable disease without necessarily implying a cure of the cancer. The terms used to evaluate the responses of tumours are defined in Box 41.14. For a complete response, all previous clinical abnormalities should have resolved, and this needs to be confirmed by clinical examination or sampling of the primary disease site or by detection of serum tumour markers. Radiological response can be quantified through use of the Response Evaluation Criteria in Solid Tumours convention, although this is mainly used for patients in clinical trials.

The increasing use of PET/CT has meant that metabolic responses (i.e., reduction in tracer avidity with or without a reduction in the size of the mass) may predict long-term outcome more accurately and is the imaging modality of choice for surveillance in some cancers.

The final assessment of treatment outcome is the impact of the therapy on remission duration and survival, that is, the **cure rate**. Such survival figures are increasingly incorporating quality-of-life assessments and a health economic assessment to calculate the number of quality-adjusted life years gained for the cost of the treatment to guide judgements about health system affordability (see Chapter 9, 'Health economics' section).

Further reading

Bai JW, Qiu SQ. & Zhang GJ. Molecular and functional imaging in cancer-targeted therapy: current applications and future directions. *Sig Transduct Target Ther* 2023; 8:89.

Koh DM, Papanikolaou N, Bick U et al. Artificial intelligence and machine learning in cancer imaging. *Commun Med* 2022; 2:133.

Principles of chemotherapy

Cytotoxic chemotherapy uses systemically administered drugs that directly damage cellular DNA (and RNA). It kills cells by promoting apoptosis and sometimes frank necrosis. Different cytotoxic drugs work at different stages in the cell cycle (Fig. 41.9).

There is a narrow therapeutic window between doses necessary for effective treatment of the cancer and those that cause normal tissue toxicity because cytotoxic drugs are not cancer specific (unlike some of the targeted biological agents), and the increased proliferation in cancers is often not much greater than in normal tissues. The dose and schedule of the chemotherapy are limited by normal tissue tolerance, especially in the more proliferative tissues of the bone marrow and GI tract mucosa. All tissues can be affected, however, depending on the pharmacokinetics of the drug and its affinity for particular tissues (e.g., heavy metal compounds for kidneys and nerves).

The therapeutic effect on the cancer is achieved by a variety of mechanisms that seek to exploit differences between normal and transformed cells. Although most drugs in the past (e.g., alkylating agents) have been derived by empirical testing of many different compounds, new molecular biology is leading to targeted approaches to particular genetic defects in the cancer (see 'Targeted therapies' section).

Toxicity to normal tissue can be mitigated in some instances by concurrent measures, including supplying growth factors such as granulocyte colony-stimulating factor (G-CSF). The use of more specific targeted biological agents with relatively weak pro-apoptotic effects, in combination with general cytotoxics, has also

Box 41.14 RECIST criteria for assessing response to treatment in solid tumours

Criterion	Description
Complete response	Disappearance of all target lesions
Partial response	≥30% decrease in the baseline sum of the longest diameters of all target lesions
Progression	≥20% increase in the sum of the longest diameters of all target lesions compared with the smallest recorded sum since treatment started, or the appearance of any new lesions
Stable disease	All values between partial response and progressive disease
Target lesions	Measurable lesions up to 5 per organ and 10 in total, selected on size and replicable measurement
Non-target lesions	Non-measurable lesions recorded as present or absent at baseline and on follow-up
Measurable	Lesions with the longest diameter in one dimension of ≥2.0 cm or ≥1.0 cm if assessed by spiral CT scan
Non-measurable	e.g., Bone lesions, meningeal disease, ascites, pleural effusion, inflammatory breast cancer, lymphangitis cutis and pulmonis, cystic lesions

CT, Computed tomography; *RECIST,* Response Evaluation Criteria in Solid Tumours.

improved the therapeutic ratio (e.g., trastuzumab and breast cancer; see 'Breast cancer' section). Certain cytotoxic therapies may also be administered into the pleural space, peritoneum or cerebrospinal fluid (CSF), or into the arterial supply of a tumour.

Most tumours rapidly develop resistance to single agents given on their own through changes in membrane transport and DNA repair pathways. For this reason, the principle of intermittent combination chemotherapy was developed. Several drugs are combined, chosen on the basis of differing mechanisms of action and non-overlapping toxicities. These drugs are given over a period of a few days followed by a rest of a few weeks, when the normal tissues have the opportunity for regrowth. If the normal tissues are more proficient at DNA repair than the cancer cells, it may be possible to deplete the tumour while allowing the restoration of normal tissues between chemotherapy cycles (Fig. 41.10).

In many experimental tumours, it has been shown that there is a log-linear relationship between drug dose and number of cancer cells killed, and that the maximum effective dose is very close to the maximum tolerated dose at which dose-limiting toxicity is reached. With a chemosensitive tumour, relatively small increases in dose may have a large effect on tumour cell kill. It is therefore apparent that when cure is a realistic option, the dose administered is critical and may need to be maintained despite toxicity. When cure is not a realistic possibility and palliation is the aim, a sufficient dose to exceed the therapeutic threshold but not cause undue toxicity is required because the short-term quality of life becomes a major consideration.

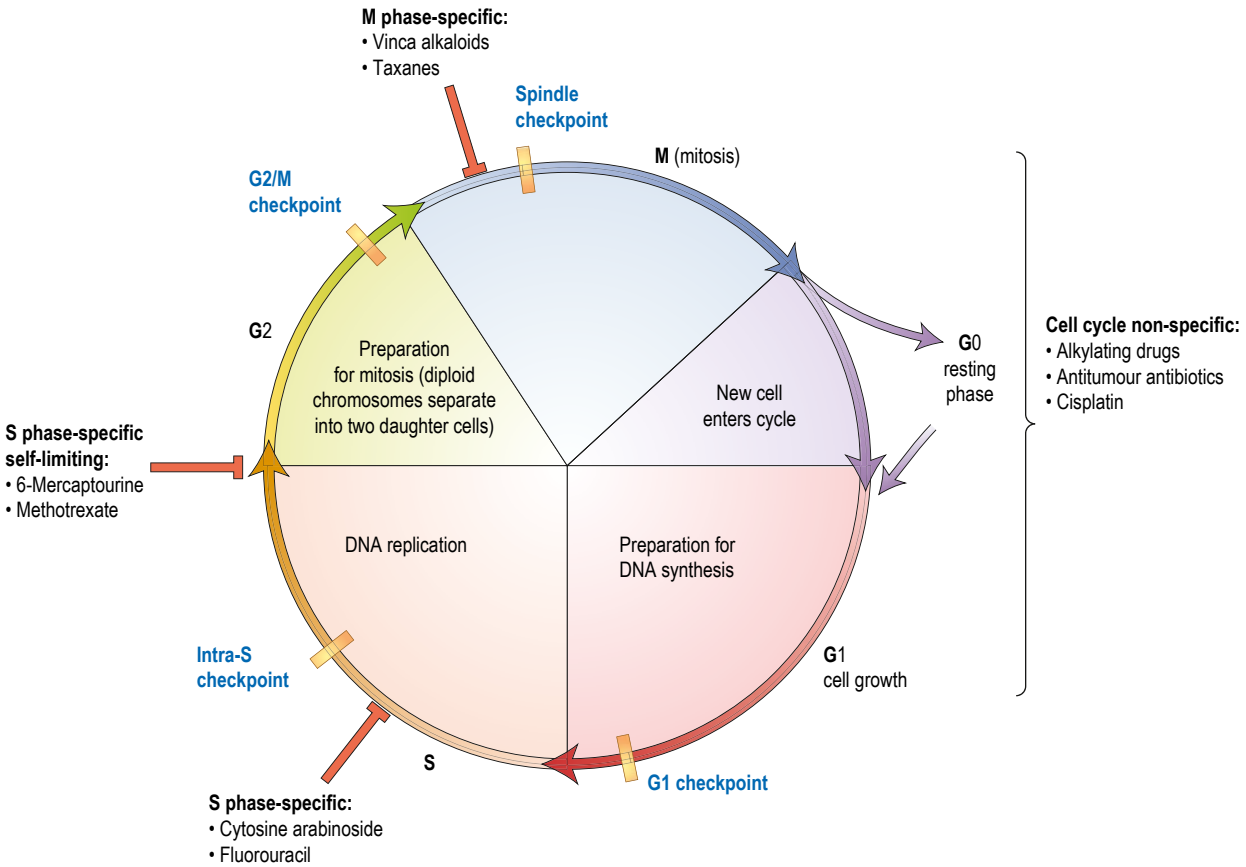


Fig. 41.9 **Cell cycle and chemotherapy drugs.**

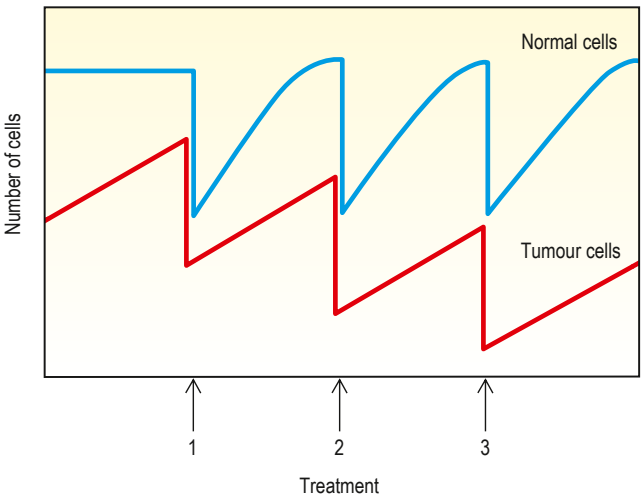


Fig. 41.10 Effects of multiple courses of cytotoxic chemotherapy. The decrease in the number of cells with each course is shown.

Classification of cytotoxic drugs

See **Box 41.15**.

DNA-damaging drugs

Alkylating agents

These act by covalently binding alkyl groups. Their major effect is to cross-link DNA strands, interfering with DNA synthesis and causing strand breaks. Examples include bendamustine and cyclophosphamide.

Platinum compounds

Cisplatin, carboplatin, and oxaliplatin cause interstrand cross-links of DNA and are often regarded as non-classical alkylating agents. They have transformed the treatment of patients with testicular cancer (cisplatin) and play a major role against many other tumours.

Antimetabolites

Antimetabolites are usually structural analogues of naturally occurring metabolites that interfere with normal synthesis of nucleic acids by falsely substituting purines and pyrimidines in metabolic pathways.

Folic acid antagonists

This class includes methotrexate, which is structurally very similar to folic acid and binds preferentially to dihydrofolate reductase, the enzyme responsible for the conversion of folic acid to folinic acid. It is used widely in the treatment of solid tumours and haematological malignancies. Folinic acid is often given to 'rescue' normal tissues from the effects of high doses of methotrexate. A related drug, pemetrexed, targets the enzyme thymidylate synthase, as well as glycinamide ribonucleotide formyltransferase.

Pyrimidine antagonists

5-Fluorouracil (5-FU) consists of one uracil molecule with a substituted fluorine atom. It acts by blocking thymidylate synthase, which is essential for pyrimidine synthesis. Oral capecitabine is metabolized to 5-FU, as is tegafur with uracil.

Arabinosides

Arabinosides inhibit DNA synthesis by inhibiting DNA polymerase. Cytarabine is used almost exclusively in the treatment of patients with acute myeloid leukaemia, in which it remains the backbone of therapy.

Purine antagonists

This class includes 6-mercaptopurine and 6-tioguanine, which are both used almost exclusively in the treatment of patients with acute leukaemia.

DNA repair inhibitors

Epipodophyllotoxins

These are semisynthetic derivatives of podophyllotoxin, which inhibit the topoisomerase enzymes that allow unwinding and uncoiling of supercoiled DNA and maintain DNA strand breaks.

Cytotoxic antibiotics

The anthracyclines act by intercalating adjoining nucleotide pairs on the same strand of DNA and by inhibiting topoisomerase-II DNA repair. They have a wide spectrum of activity in haematological and solid tumours.

Antitubulin agents

Vinca alkaloids

Drugs such as vincristine, vinblastine, and vinorelbine act by binding to tubulin and inhibiting microtubule formation during mitosis (see **Fig. 41.10**). They are associated with neurotoxicity because of their antimicrotubule effect and must never be given intrathecally because this is lethal.

Taxanes

Paclitaxel and docetaxel bind to tubulin dimers and prevent their assembly into microtubules.

Side-effects of chemotherapy

Chemotherapy carries many potentially serious side-effects and should be used only by trained practitioners; however, an appreciation of its common potential side-effects is necessary for all general physicians who are called to see patients with cancer who

Box 41.15 Chemotherapy: some cytotoxic drugs

DNA-damaging agents
- Free radicals:
  - Alkylators (e.g., cyclophosphamide, melphalan, chlorambucil, ifosfamide)
  - Nitrosoureas (e.g., carmustine [BCNU], bendamustine, lomustine [CCNU], busulfan)
  - Tetrazines (e.g., dacarbazine, temozolomide)
- DNA cross-linking—platinum (e.g., cisplatin, carboplatin, oxaliplatin)

Antimetabolites
- Pyrimidine synthesis (e.g., 5-fluorouracil, capecitabine, cytarabine, fludarabine, gemcitabine, cladribine, clofarabine, nelarabine, azacitidine)
- Purine synthesis (e.g., mercaptopurine, tioguanine)
- Antifolates (e.g., methotrexate, pemetrexed)

DNA repair inhibitors
- Topoisomerase-I inhibitors (e.g., irinotecan)
- Topoisomerase-II inhibitors (e.g., etoposide)
- Anthracyclines (e.g., daunorubicin, doxorubicin, epirubicin)

Antitubulin agents
- Tubulin-binding: alkaloids (e.g., vincristine, vinorelbine)
- Taxanes (e.g., paclitaxel, docetaxel)
- Eribulin, epothilones

Others
- Bleomycin, mitomycin

Box 41.16 Side-effects of chemotherapy

Common

- Nausea and vomiting
- Hair loss
- Myelosuppression
- Mucositis
- Fatigue

Drug specific

- Cardiotoxicity (e.g., anthracyclines)
- Progressive pulmonary fibrosis (e.g., bleomycin, busulfan)
- Neurotoxicity (e.g., cisplatin, vinca alkaloids, taxanes)
- Nephrotoxicity (e.g., cisplatin)
- Skin (e.g., anti-EGFR TKIs, immune checkpoint inhibitors)
- Skin: plantar–palmar dermatitis (e.g., 5-fluorouracil, liposomal doxorubicin)
- Gut: diarrhoea, mucositis (e.g., fluorouracil, methotrexate, etoposide, TKIs, immune checkpoint inhibitors)
- Sterility (e.g., alkylating agents, anthracyclines)
- Secondary malignancy (e.g., alkylating agents, epipodophyllotoxins, anthracyclines)
- Central nervous system (e.g., TKIs)

EGFR, Epidermal growth factor receptor; *TKI*, tyrosine kinase inhibitor.

Box 41.17 Some common chemotherapy regimens

Cancer	Abbreviation[a]	Regimen
Breast	AC	Adriamycin (doxorubicin) and cyclophosphamide
Lung	PE	Cisplatin and etoposide
Stomach	ECF	Epirubicin, cisplatin, and 5-fluorouracil
Colorectal	FOLFOX	Oxaliplatin, 5-fluorouracil, and folinic acid

[a]Some abbreviations are related to trade names.

are receiving chemotherapy. The five most common side-effects are vomiting, hair loss, tiredness, myelosuppression, and mucositis (Box 41.16). Side-effects are much more directly dose related than anticancer effects, and it has been the practice to give drugs at doses close to their maximum tolerated dose, although this is not always necessary to achieve their maximum anticancer effect. Common combination chemotherapeutic regimens are shown in Box 41.17.

Common side-effects

Extravasation of intravenous drugs

Cytotoxic drugs should be given only by trained personnel. They cause severe local tissue necrosis if leakage occurs outside the vein. If this occurs, the infusion should be stopped immediately, and local measures should be instituted; chemotherapy agents have specific extravasation protocols that should be followed. Antihistamines and corticosteroids may give symptomatic relief.

Nausea and vomiting

The severity of these common side-effects varies with the cytotoxicity and are particular problems with platinum analogues. They can be eliminated in 75% of patients by using modern antiemetics. A stepped policy, with antiemetics such as metoclopramide and domperidone, followed by 5-HT_3 (5-hydroxytrypamine, serotonin) antagonists (e.g., ondansetron, granisetron) combined with dexamethasone, should be used to match the emetogenic potential of the chemotherapy. ***Aprepitant***, a neurokinin receptor antagonist, is helpful in preventing acute and delayed nausea and vomiting. It is used with dexamethasone and a 5-HT_3 antagonist. Drugs such as cyclizine, olanzapine and levomepromazine, and benzodiazepines can be used to control persistent nausea.

Fig. 41.11 Nail growth cessation with each drug cycle (Beau's lines).

Effects on hair, skin, and nails

Many, but not all, cytotoxic drugs are capable of causing hair loss. Scalp cooling can sometimes be used to reduce hair loss, but in general, this side-effect can be avoided only by selection of drugs when it is possible. Hair regrows on completion of chemotherapy. Nails may demonstrate banding, reflecting periods of cessation of growth during each chemotherapy cycle (Fig. 41.11), and skin toxicity may be particularly pronounced with 5-FU, capecitabine, and docetaxel (Fig. 41.12).

Fatigue

Fatigue is often significant and may continue beyond completion of therapy. Other problems, such as anaemia or depression, may exacerbate fatigue. Attention should be paid to nutrition, hydration, sleep hygiene, gentle exercise, task prioritization, pacing, realistic target setting, and scheduling of rest within the day.

Bone marrow suppression and immunosuppression

Suppression of the production of red blood cells, white blood cells, and platelets occurs with most cytotoxic drugs and is a dose-related phenomenon (Fig. 41.13). Severely myelosuppressive chemotherapy may be required if treatment is to be given with curative intent, despite the potential for rare but fatal infection or bleeding. Anaemia and thrombocytopenia are managed by red cell or platelet transfusions. The risk of infective problems can be ameliorated by the use of prophylactic antimicrobials, such as ciprofloxacin, or the use of G-CSF as primary prophylaxis in chemotherapy regimens that carry a significant risk of febrile neutropenia (or in patients on less intensive therapies who are at higher risk because of age or co-morbidity).

Fig. 41.12 **Hands showing palmar–plantar skin toxicity.**

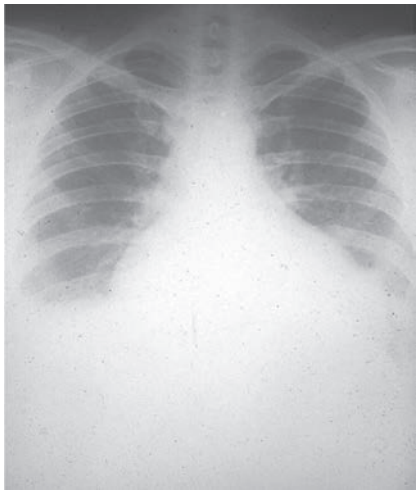


Fig. 41.14 **Anthracycline dilated cardiomyopathy with congestive heart failure.**

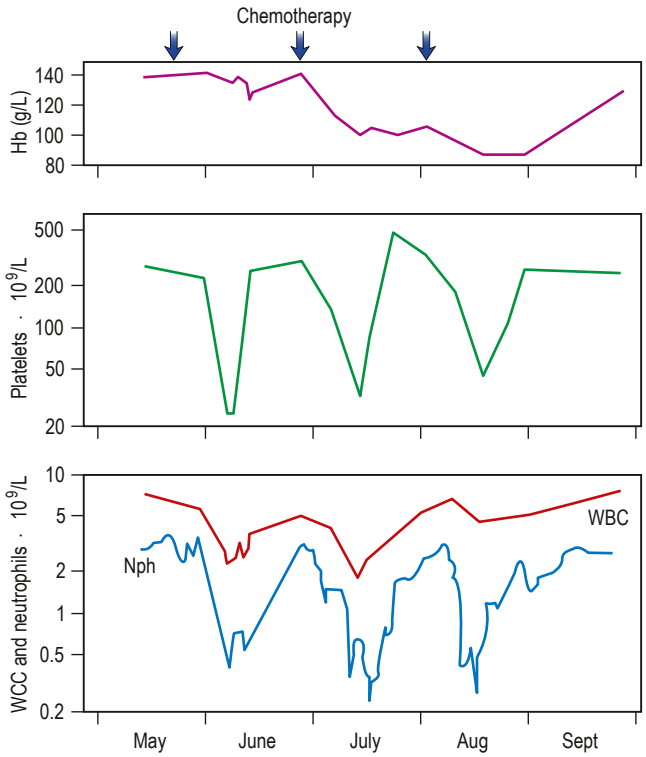


Fig. 41.13 **Cyclical myelosuppression from chemotherapy *(arrow)*.** *Hb,* Haemoglobin; *Nph,* neutrophils; *WCC,* white blood cell count.

Mucositis

This common side-effect of chemotherapy reflects the sensitivity of the mucosa to antimitotic agents. Mucositis causes severe pain in the oropharyngeal region and problems with swallowing and nutrition. It can be generalized throughout the intestinal tract and can cause life-threatening diarrhoea. Treatment is with antiseptic and anti-candidal mouthwash; if mucositis is severe, fluid and antibiotic support is required because the mucosa is a portal for entry of enteric organisms.

Other toxicities

Cardiotoxicity

This is a rare side-effect of chemotherapy and is usually associated with anthracyclines such as doxorubicin; effects can present as an acute arrhythmia during administration or cardiac failure due to cardiomyopathy after chronic exposure (Fig. 41.14). This effect is dose related and can largely be prevented by restricting the cumulative total dose of anthracyclines within the safe range (equivalent to 450 mg/m^2 body surface area cumulative doxorubicin dose). The risk of anthracycline cardiomyopathy is also dependent on other treatments, such as trastuzumab or radiotherapy, as well as other cardiac risk factors, such as hypertension, smoking, and hypercholesterolaemia. Cardiotoxicity can also be reduced by using the analogue epirubicin or by reducing peak drug concentrations through delayed-release preparations, such as liposomal doxorubicin. 5-FU and its prodrug, capecitabine, can cause cardiac ischaemia.

Neurotoxicity

This occurs predominantly with the vinca alkaloids, taxanes, and platinum analogues (but not carboplatin). It is dose related and cumulative. Chemotherapy is usually stopped before the development of a significant polyneuropathy, which, when established, is only partially reversible. *Vinca alkaloids, such as vincristine, must never be given intrathecally* because the neurological damage is progressive and fatal.

Nephrotoxicity

Cisplatin (but not oxaliplatin or carboplatin), methotrexate, and ifosfamide can potentially cause renal damage. This can usually be prevented by maintaining an adequate diuresis during treatment to reduce drug concentration in the renal tubules and by monitoring renal function.

Sterility and premature menopause

Some anticancer drugs, particularly alkylating agents but also anthracyclines and docetaxel, may cause gonadal damage, resulting in sterility and, in females, the loss of ovarian oestrogen production, which may be irreversible.

In males, storage of sperm before chemotherapy should be offered to the patient when chemotherapy is given with curative intent.

In females, collection of oocytes to be fertilized *in vitro* and cryopreserved as embryos for subsequent implantation is most successful; however, it is also possible to collect and freeze by vitrification of unstimulated oocytes. Cryopreservation of ovarian tissue and retrieval of viable oocytes for subsequent fertilization are still experimental. The recovery of gonadal function is dependent on the status before treatment; in females, this is mostly related to age since menarche, with those younger than 40 years of age having significantly more ovarian reserve.

Secondary malignancies

Anticancer drugs have mutagenic potential, and the development of secondary malignancies, predominantly acute leukaemia, is an uncommon but particularly unwelcome long-term side-effect in patients otherwise cured of their primary malignancies. The alkylating agents, anthracyclines, and epipodophyllotoxins are particularly implicated.

Principles of endocrine therapy

Oestrogens are capable of stimulating the growth of breast and endometrial cancers, and androgens can stimulate the growth of prostate cancer. Removal of these growth factors by manipulation of the hormonal environment may result in apoptosis and regression of the cancer. Endocrine therapy can be curative in a proportion of patients treated for micrometastatic disease in the adjuvant setting for breast and prostate cancer and provides a minimally toxic, non-curative (palliative) treatment in advanced or metastatic disease. The presence of detectable cellular receptors for the hormone is strongly predictive of response. However, this is also modified by the many molecular interactions between the activation pathways of, for example, epidermal growth factor receptor (EGFR), the androgen receptor (AR [see 'Endocrine therapy' section]) and oestrogen receptor (ER [see 'Oestrogen receptor–positive disease' section]).

Principles of biological and targeted therapy

Immunotherapy

Certain antigens that are specific to cancer cells, such as sequences of tumour immunoglobulin from melanoma antigens, have been used as tumour vaccines, together with manipulation of the immune system to overcome tolerance. Engineered chimeric antigen receptors that direct T cells to specific antigenic targets are emerging as a powerful potential tool in the treatment of a wide range of cancers.

Immune checkpoint inhibitors

These agents directly interfere with the relationship between T cells and antigen-presenting cells or tumour cells to overcome the common problems of immune tolerance and anergy (see **Fig. 6.8**).

Ipilimumab is a MAb that blocks the CD152 (CTLA4) receptor responsible for downregulating T-cell responses and is active in the treatment of melanoma. The inhibitors of the PD-1 receptor or its ligands, PD-L1 and PD-L2, on activated T cells, further reactivate tumour-specific cytotoxic T cells, so ipilimumab plus nivolumab (a PD-1 receptor blocker) double the response rate from 20% to 40% in metastatic melanoma. Many clinical trials are now examining these agents in the context of a variety of other cancers.

As the use of immunotherapies increases, physicians are likely to be increasingly exposed to patients receiving this therapy. Immunotherapies can present with many side-effects that are quite distinct to those of conventional chemotherapy, relating to excess immune system activity against healthy body tissues. Relatively mild symptoms can be a precursor to imminent life-threatening complications. Many of the side-effects of immunotherapy respond to high dose corticosteroid therapy. However, supportive medications such as antibiotics and corticosteroids may counter the beneficial effect of immunotherapies in cancer treatment and should ideally be avoided when safe to do so. Close liaison with oncology support services is required for patients on immunotherapy who present acutely unwell.

Common side-effects associated with immunotherapies include:

- Colitis
- Skin toxicities
- Hepatitis
- Endocrinopathies
- Neuropathies
- Nephritis

Immunomodulatory drugs: thalidomide, lenalidomide, and pomalidomide

This family of immunomodulatory drugs (IMIDs) is increasingly being used in a range of malignancies, such as myeloma, chronic lymphocytic leukaemia, and myelodysplastic syndrome, as well some solid tumours. They have anti-angiogenic functions, as well as effects on cytokine production, tumour–stromal interactions, and T-cell co-stimulatory functions. They are all considered teratogenic and should be avoided in females of reproductive potential unless extra precautions are taken against conception. There is increasing evidence that IMIDs work through binding of the E3 ligase protein cereblon, which acts to degrade two specific B-cell transcription factors, IKZF1 and IKZF3.

Interferons

Interferons are naturally occurring cytokines that mediate the cellular immune response. Interferon-alpha (IFN-α) has been used to treat patients with advanced melanoma and renal cell carcinoma.

Treatment with IFN has side-effects, most commonly, influenza-like symptoms, which tend to diminish with time, and fatigue, which generally does not and can be treatment limiting. IFN used to be given as a daily subcutaneous injection, but conjugation with polyethylene glycol (has led to a reduction in frequency of injection and severity of side-effects.

Interleukins

Interleukin 2 (IL-2), a recombinant protein, is used to activate T-cell responses, often in conjunction with IFN-stimulated B-cell activation. Anti-tumour activity has been observed in renal cell carcinoma and melanoma, with responses in 10% to 20% of patients, occasionally for prolonged periods. Toxicity is common; acutely, this includes the capillary leak syndrome with hypotension and pulmonary oedema; autoimmune thyroiditis and vitiligo occur later.

Targeted therapies

The fast-developing field of targeted therapies includes MAbs, receptor kinase inhibitors, and small molecule receptor blockers. Expression of a targeted molecule is insufficient to indicate that use of the corresponding drug will be therapeutic because it is also necessary for the drug to be able to interrupt a growth-critical pathway in that cancer.

Monoclonal antibodies

MAbs (**Box 41.18**) directed against tumour cell-surface antigens are 'humanized' by being genetically engineered to have a range of functions, including:

- Acting as a carrier molecule to target toxins or radioisotopes to the tumour cells (e.g., trastuzumab or trastuzumab emtansine for breast cancer)
- Acting as anti-growth factor agents added to chemotherapy by inhibiting dimerization of the extracellular receptor molecules.

Side-effects are those of hypersensitivity to the foreign protein and specific cross-reactivities, for example, trastuzumab for the myocardium, bevacizumab for the mucosa and renal tubule, and cetuximab for the skin follicles.

Intracellular signal inhibitors

Many cancer cells are transformed by the activity of the protein products of oncogenes that signal growth by phosphorylation of tyrosine residues on the intracellular portion of growth factor receptors. Small molecule inhibitors (**Box 41.19**) have many pharmacokinetic advantages over the MAb inhibitors. The first example was the TKI imatinib, which specifically inhibits the BCR-ABL fusion oncoprotein and c-Kit. This compound is an extremely effective treatment for chronic myeloid leukaemia and GISTs, which are characterized by the presence of the c-Kit target. Lapatinib, which inhibits HER2, has increased survival in breast cancer. The less specific TKIs sunitinib and pazopanib, which inhibit signalling by EGFR and vascular endothelial growth factor receptor (VEGFR), have proved effective in metastatic renal cancer, and afatinib, erlotinib, and gefitinib have shown activity in lung cancer. Vemurafenib is a kinase inhibitor and has a specificity for BRAF-V600, as used in patients with malignant melanoma (see 'Malignant melanoma' section). Many other similar molecules are in pre-clinical or early clinical development, but heterogeneity within a single tumour occurs, and genomic anomalies may not all be represented in a single biopsy specimen. Vandetanib, a combined VEGF and C-RET inhibitor, is useful in the treatment of patients with medullary thyroid cancer.

Proteasome inhibitors

The proteasome degrades redundant or damaged proteins that have been labelled by a process called ubiquitination. Such proteins include cyclins and cyclin-dependent kinases, as well as factors in the nuclear factor kappa-B pathway. Inhibition of the proteasome leads to apoptosis in cancer cells and is synergistic with other treatments such as steroids and chemotherapy. Bortezomib is the first of such inhibitors to reach clinical practice and is used in myeloma, as well as some types of non-Hodgkin lymphoma. There are also recent additions to this class, including carfilzomib and ixazomib.

i Box 41.18 Monoclonal antibodies in cancer

Antigen	Mutation/ amplification	Tumour type	Monoclonal antibody
EGFR	*cERBB2/HER2* amplification	Breast, gastric	**Trastuzumab**
	cERBB2/HER2 amplification	Breast	**Pertuzumab**
	RAS/BRAF wt	Colorectal, oropharyngeal	**Cetuximab**
	RAS/BRAF wt	Colorectal	**Panitumumab**
CTLA-4	+	Melanoma	**Ipilimumab**
VEGFR	+	Colorectal, ovarian	**Bevacizumab**
PD-1	+	Melanoma, Hodgkin lymphoma, NSCLC	**Nivolumab**
RANKL	+	Bone metastases	**Denosumab**

CTLA-4, Cytotoxic T-lymphocyte associated protein 4; *EGFR,* epidermal growth factor receptor; *NSCLC,* non–small-cell lung cancer; *PD-1,* programmed cell death protein 1; *RANKL,* receptor activator of nuclear factor kappa-B ligand; *VEGFR,* vascular endothelial growth factor receptor.

Gene therapy

Antisense oligonucleotides are short sequences of DNA bases that specifically inhibit complementary sequences of either DNA or RNA. As a result, they can be generated against genetic sequences that are specific for tumour cells. Their clinical development has been hampered by poor uptake by tumour cells and rapid degradation by natural endonucleases. However, one antisense sequence directed against the *Bcl-2* oncogene has been shown to have an anti-tumour effect in patients with non-Hodgkin lymphoma. Viral vectors for the transfection of tumour cells *in vivo* are being tested as a way of delivering specific replacement gene therapy in head and neck cancers.

Further reading

Hahn WC, Bader JS, Braun TP et al. An expanded universe of cancer targets. *Cell* 2021; 184(5):1142–1155.
Sterner RC, Sterner RM. CAR-T cell therapy: current limitations and potential strategies. *Blood Cancer J* 2021; 11:69.
Waldman AD, Fritz JM, & Lenardo M.J. A guide to cancer immunotherapy: from T cell basic science to clinical practice. *Nat Rev Immunol* 2020; 20:651–668.

i Box 41.19 Intracellular signal inhibitor therapies in cancer

Gene	Genetic alteration	Tumour type	Therapeutic agent
Receptor tyrosine kinase			
EGFR	Mutation, amplification	Lung	**Gefitinib, erlotinib, afatinib**
ERBB2	Amplification	Breast	**Lapatinib**
PDGFRA	Mutation	GIST	**Sunitinib, imatinib, nilotinib**
PDGFRB	Translocation	Chronic myelomonocytic leukaemia	**Sunitinib, sorafenib, imatinib**
ALK	Mutation or amplification	Lung, neuroblastoma	**Crizotinib, alectinib**
c-Met	Amplification	Gefitinib-resistant NSCLC, gastric	**Crizotinib**
c-Kit	Mutation	GIST	**Sunitinib, imatinib**
RET	Mutation, translocation	Thyroid medullary carcinoma	**Vandetanib**
Serine–threonine–lipid kinase			
BRAF	Mutation (V600E)	Melanoma	**Vemurafenib**
mTOR	Increased activation	Renal cell carcinoma, breast	**Everolimus**
DNA damage or repair			
BRCA1 and *BRCA2*	Mutation (synthetic lethal effect)	Breast, ovarian	**Olaparib (PARP inhibitors)**

GIST, Gastrointestinal stromal tumour; *NSCLC,* non–small-cell lung cancer; *PARP,* poly-ADP ribose polymerase.

Principles of radiation therapy

Radiation is an important modality for the treatment of patients with cancer and can be used with radical, adjuvant, or palliative intent. More than 50% of patients with a cancer diagnosis are treated with radiotherapy.

Radical radiotherapy aims to deliver a tumouricidal treatment dose to a well-defined target volume with curative intent, sparing the surrounding normal tissues as much as possible. It is usually given as a fractionated course over a number of weeks, with patients receiving daily treatment. Carcinoma of the cervix, for example, can be cured with 5.5 weeks of daily radiotherapy delivered Monday to Friday with concurrent cisplatin chemotherapy.

Adjuvant radiotherapy is used to reduce the risk of tumour recurrence after primary surgery. The aim of treatment is to eradicate occult micrometastatic disease that cannot be demonstrated on imaging, as occurs in a significant percentage of patients. In breast cancer, for example, a patient may have radiotherapy after wide local excision surgery, and this reduces the risk of recurrence to the same level as if they had had a mastectomy.

Palliative radiotherapy can be used to alleviate symptoms of local disease (e.g., haematuria) or distant metastases (e.g., bone pain). Treatment is usually given with a small number of fractions, and effective results can also be achieved with a single fraction of radiotherapy.

Theoretical background

Radiation, used therapeutically, exerts its effects through the ionization of intracellular molecules; the radiation can be electromagnetic or particulate. The two types of ***electromagnetic radiation*** used in radiotherapy are X-rays and gamma rays, and both of these can also be described by the term 'photon'. Gamma rays are produced from the nuclear decay of radioactive isotopes, whereas X-rays are produced artificially by electrical means, usually in a linear accelerator, accelerating electrons to a high energy and then stopping them abruptly in a heavy metal target. ***Particulate radiation*** consists of subatomic particles, including alpha particles, protons, and neutrons.

Interactions of X-rays with matter

The process by which photons are absorbed depends principally on their energy and the chemical composition of the absorbing material. The radiation effect depends on the intensity of the radiation source, measured as the linear energy transfer or frequency of ionizing events per unit of path, which is subject to the inverse square law as the energy diminishes with distance from the source. The depth of penetration of biological tissues by the photons depends on the energy of the beam. Low-energy photons from an 85-kV source are suitable for superficial treatments, whereas high-energy 15-mV sources produce a beam with deeper penetration, lower dose at the initial skin boundary (skin-sparing), sharper edges, and less absorption by bone. Superficial radiation may also be delivered by electron beams from a linear accelerator, from which the target electrode that generates the X-rays has been removed. The radiation dose is measured in Gray (Gy), with 1 Gy being equal to 1 joule (J) absorbed per kilogram of absorbing tissue. There are differing processes of energy absorption, depending on the energy of the photons delivered, but all share the fact that the photon causes excitation of electrons within the targeted tissue.

Biological effect

The resultant biological damage is caused in either an indirect or a direct way. X-rays and gamma rays mainly interact with target biological molecules in an indirect way. The interaction is initially with other atoms and molecules within the cells, producing free radicals, which are highly reactive molecules generated from water and oxygen. These free radicals exert a biological effect through the generation of single- and double-strand DNA breaks, inducing apoptosis of cells as they progress through the cell cycle.

The biological effect is dependent on the dose rate, duration, volume irradiated, and tissue sensitivity. Sensitivity to photon damage is greatest during the G2/M phase of the cell cycle (see **Fig. 41.10**) and is also dependent on the DNA repair capacity of the cell.

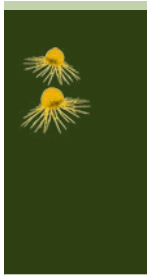
Radiobiology is the study of the actions of ionizing radiation on living things. The effects of radiation on normal and malignant tissues evolve through a series of steps. Classical radiobiology describes the 'five Rs' of the radiation response:

- ***Repair.*** This is the ability of cells to repair DNA damaged by ionizing radiation using different DNA repair mechanisms. Delivery of small daily doses of radiotherapy enables repair of DNA in the cells of normal tissues.
- ***Redistribution.*** When radiation is delivered, cells in the G2/M phase of the cell cycle are most radiosensitive. Cells in the more radioresistant phases (e.g., S phase) survive and progress through the cell cycle and are then in the more radiosensitive phase when the next fraction of radiotherapy is delivered.
- ***Reoxygenation.*** Oxygenated cells are more susceptible to DNA damage by radiation. Because of irregular blood supplies, tumour cells may be hypoxic. After radiation, the more oxygenated cells die, and the surviving hypoxic cells are reoxygenated as oxygen diffuses into these areas, making these cells more susceptible to radiation with the next fraction of treatment.
- ***Repopulation.*** During a course of fractionated radiotherapy, viable tumour cells continue to divide and repopulate the tumour. In some cancers (e.g., those of the head and neck or cervix), it is important not to extend the overall treatment time if there are gaps in treatment because the tumour will repopulate.
- ***Radiosensitivity.*** Some tumours are more sensitive to radiation than others. For example, SCCs and lymphomas are radiosensitive, whereas melanomas and brain tumours are radioresistant.

Clinical application of radiation therapy

Modern radiotherapy allows highly targeted radiation to maximize tumour dose while sparing normal tissue. Radiotherapy planning involves both detailed physics of the applied dose and detailed knowledge of the biology of cancer; it is also necessary to establish whether the intention is to treat the tumour site alone or to include the likely loco-regional patterns of spread. Normal tissue tolerance will determine the extent of the side-effects and therefore the total achievable dose to the tumour. A balanced decision is made according to the curative or palliative intent of the treatment and the likely early or late side-effects. The clinician outlines on the planning scan the areas that they wish to treat using all the clinical information available. They also delineate the critical structures that they do not wish to receive a dose. The physicist then uses planning algorithms to enable the best delivery of dose to the tumour, avoiding critical structures as far as possible. Depending on the site of the tumour, the patient is immobilized to allow reproducibility of position for each fraction of treatment. For example, if the brain is being treated, the patient will wear a mask that keeps them still on the treatment couch.

The cancers for which radiotherapy is usually used in a primary curative approach, when the tumour is anatomically localized, are listed in **Box 41.20**, along with those in which radiotherapy has



Box 41.20 Curative radiotherapy treatment

Primary modality	Adjuvant to primary surgery
• Retina	• Lung
• Central nervous system	• Breast
• Skin	• Uterus
• Pharynx and larynx	• Bladder
• Cervix and vagina	• Rectum
• Prostate	• Testis
• Lymphoma	• Sarcoma

Box 41.21 Palliative benefits of radiotherapy

- Pain relief (e.g., bone metastases)
- Reduction of headache and vomiting in raised intracranial pressure from central nervous system metastases
- Relief of obstruction of the bronchus, oesophagus, ureter, and lymphatics
- Preservation of skeletal integrity from metastases in weight-bearing bones
- Reversal of neurological impairment from the spinal cord or optic nerve compression by metastases

curative potential when added to surgery (adjuvant radiotherapy). Palliative treatments are frequently used to provide relief of symptoms and to improve the quality, if not duration, of survival (Box 41.21). Palliative treatment is usually given in as few fractions as possible over as short a time as possible.

Radiotherapy planning with CT guidance has been complemented by the introduction of three-dimensional planning and volumetric-modulated arc therapy, which can deliver curved dose distributions to improve the therapeutic ratio. This allows a greater differential in dose between the tumour and critical normal structures, in turn permitting dose escalation or a reduced risk of toxicity.

Four-dimensional radiotherapy planning is also becoming widely available. It varies the radiation dose over time (e.g., the respiratory cycle during lung cancer treatment).

Types of radiation therapy

Fractionated external-beam radiotherapy uses photons generated in a linear accelerator. The radiation dose is delivered in increments separated by at least 4 to 6 hours to try to exploit any advantage in DNA repair between normal and malignant cells.

Stereotactic radiosurgery or radiotherapy (Gamma Knife, CyberKnife, or linear accelerator–based treatment) can concentrate X-rays or gamma radiation from multiple sources on to a small volume to generate an ablative dose of radiation with rapid dose fall-off, minimizing the dose to surrounding tissues. Patients with small lung cancers can be treated in this way, allowing for the more aggressive treatment of cancer that has spread to the brain.

Proton-beam therapy is an alternative type of external-beam radiotherapy that uses charged proton particles. Because of their larger mass, the protons have little side scatter and so offer a much more focused beam with less dose to surrounding structures. Proton therapy is particularly recommended for the treatment of children, in whom radiotherapy can affect normal growth and development, and for tumours in close proximity to structures of the central nervous system.

Brachytherapy is the use of radiation sources in close contact with the tissue to provide intense exposure over a short distance to a restricted volume. Such techniques have been used to treat localized breast, prostatic, and cervical carcinoma.

Box 41.22 Side-effects of radiotherapy

Acute temporary side-effects depending on region being treated

- Anorexia, nausea, malaise
- Mucositis, oesophagitis, diarrhoea
- Alopecia
- Myelosuppression

Late side-effects

- Skin: ischaemia, ulceration
- Bone: necrosis, fracture, sarcoma
- Mouth: xerostomia, ulceration
- Bowel: stenosis, fistula, diarrhoea
- Bladder: fibrosis, frequency, incontinence
- Vagina: dyspareunia, stenosis
- Lung: fibrosis, breathlessness
- Heart: pericardial fibrosis, cardiomyopathy, vasculopathy
- Central nervous system: myelopathy
- Gonads: infertility, menopause
- Secondary malignancies (e.g., leukaemia, thyroid cancer)
- Other: carotid artery stenosis

Systemic radionuclides, such as ^{131}iodine- (in thyroid cancer) or radioisotope-labelled MAbs (e.g., anti-CD20 for lymphoma) and hormones (e.g., somatostatin for carcinoid tumours), can be administered by intravenous or intracavitary routes to provide radiation targeted to particular tissue uptake via surface antigens or receptors.

Radiotherapy in combination with systemic therapy

The local efficacy of radiotherapy can be increased by the simultaneous but not serial addition of chemotherapy with agents such as cisplatin, mitomycin, and 5-FU for cancers of the head and neck, lung, oesophagus, stomach, rectum, anus, cervix, and bladder. Reduced local recurrence rates have translated into survival benefits, and further research is investigating the concurrent use of biological agents (e.g., EGFR inhibitors) and immunotherapy with radiation. The ***abscopal effect*** occurs when localized radiotherapy induces the regression of non-irradiated metastatic lesions at a distance from the primary site of irradiation by a systemic immune-mediated response, which may be augmented by giving immunotherapy with the radiotherapy. This is an exciting area of oncology that is being explored in a number of tumour sites.

Side-effects of radiotherapy

Early radiotherapy side-effects may occur within days to weeks of treatment; at this point, they are usually self-limiting but associated with general systemic disturbance (Box 41.22). The side-effects depend on tissue sensitivity, fraction size, and treatment volume and are managed with supportive measures until normal tissue repair occurs. The toxicity may also be enhanced by exposure to other radiation-sensitizing agents, especially some cytotoxics, such as bleomycin, actinomycin, anthracyclines, cisplatin, and 5-FU.

Later side-effects occur months to years later and are unrelated to the severity of the acute effects because they have a different mechanism. Late effects reflect both the loss of slowly proliferating cells and a local endarteritis that produces ischaemia and proliferative fibrosis. The risks of late side-effects are related to the fraction size and total dose delivered to the tissue. Growth may be arrested if bony epiphyses are not yet fused and are irradiated, leading to distorted skeletal growth in later life.

Secondary malignancies after radiotherapy may appear 10 to 20 years after cure of the primary cancer. Haematological

malignancies tend to occur sooner than solid tumours from the irradiated tissues. The latter are very dependent on the status of the tissue at the time of treatment; for example, the pubertal breast is up to 300 times more sensitive to malignant transformation than the breast tissues of a female in her 30s. Patients who smoke are more liable to develop lung cancer. Treatment of these secondary cancers can be successful, provided there is normal bone marrow to reconstitute the haematopoietic system, or the whole tissue at risk can be resected (e.g., thyroid after mantle radiotherapy for lymphoma).

COMMON SOLID TUMOUR TREATMENT

Common cancer mortality rates are listed in **Box 41.23**; improvements in survival in recent years have stemmed from advances in prevention, diagnosis, and treatment. The presentation, diagnosis, and natural history of the common cancers are described in the relevant systems-based chapters of this book; staging and anticancer treatment is covered here. Haematological malignancy is presented in Chapter 39.

The decision to treat and the aim of the treatment, whether palliation or cure, require knowledge of the natural history of the disease, prognostic and predictive factors, the patient's performance status, and the potential efficacy of treatment. Management should be carried out by an MDT, which is usually led by an oncologist.

Box 41.23 Index of net survival of all cancers at 5 years after diagnosis

Cancer type	5-year survival rate (%)	
	Males	Females
Pancreas	3.6	3.1
Lung	8.4	11.6
Oesophagus	15.6	14.7
Stomach	19.5	17.7
Brain	17.8	19.5
Multiple myeloma	50	43.8
Ovary		46.4
Leukaemia	53.3	49.1
Kidney	56.7	55.6
Colon	59.2	57.3
Rectum	56.6	59.8
Non-Hodgkin lymphoma	68.1	69.7
Prostate	84.8	
Bladder	56.5	45.3
Cervix		67.5
Melanoma	87.8	92.4
Breast		86.7
Hodgkin lymphoma	84.1	86.3
Testis	98	
All cancers	49.2	59.2

From Quaresma M, Coleman MP, Rachet B. 40-year trends in an index of survival for all cancers combined and survival adjusted for age and sex for each cancer in England and Wales, 1971–2011: a population-based study. Lancet *2015; 385:1206–1218, with permission.*

LUNG CANCER

The presentation and diagnosis of lung cancer are covered more fully in Chapter 20. Current treatment reflects the fact that the majority of patients are diagnosed at an advanced stage with a poor prognosis; there are therefore many trials assessing different screening strategies to try to achieve an earlier diagnosis that would permit more curative treatment.

Prognosis

Lung cancer histology is divided into two main types (**Box 41.24**):

- ***Small-cell lung cancers (SCLCs)***, which derive from adenocarcinomas but develop neuroendocrine differentiation and tend to rapidly metastasize. They are sensitive to both chemotherapy and radiotherapy but frequently relapse after initially appearing to respond well.
- ***Non–small-cell lung cancers (NSCLC)***, which are more likely to be diagnosed in a localized form. As treatments have improved, it has become clear that what previously was simply called NSCLC to be separated into individual pathological entities because the management varies, and some therapies are effective only in particular subtypes.

Tumour stage and patient performance status are used in selecting treatment and predicting response and prognosis (**Box 41.25**). Although the overall 5-year survival rate has remained approximately 10%, treatment is beginning to have an impact in selected groups.

Box 41.24 Histological subtypes of lung cancer

Histological subtype	Percentage of all lung cancers
Small-cell lung cancer	15
Non–small-cell lung cancer	85
Adenocarcinoma	40
Squamous cell carcinoma	35
Large cell carcinoma	10
Other (including carcinoid tumours and sarcomatoid carcinomas)	3

Adapted from Wistuba I, Brambilla E, Noguchi M. Classic anatomic pathology and lung cancer. In: Pass HI, Ball D, Scagliotti GV, eds. IASLC Thoracic Oncology, *2nd edn. Elsevier; 2018:143–146.*

Box 41.25 Survival in non–small-cell cancer based on clinical stage[a]

Stage[a]	5-year survival rate (%)
IA1	92
IA2	83
IA3	77
IB	68
IIA	60
IIB	53
IIIA	36
IIIB	26
IIIC	13
IVA	10
IVB	0

[a]See **Box 20.35**.

Based on data from Goldstraw P, Crowley J, Chansky K et al. The IASLC Lung Cancer Staging Project: proposals for the revision of the TNM stage groupings in the forthcoming (8th) edition of the TNM classification of malignant tumours. J Thorac Oncol *2016; 2:706–714.*

Staging

Where the technology is available, staging is normally based on PET/CT (Fig. 41.15). In the absence of distant metastases, further evaluation of local and regional lymph nodes is frequently undertaken if lymph nodes are enlarged but of low glucose avidity or lymph nodes do not appear enlarged but do show significant glucose avidity.

In these cases, confirmatory tests are required to yield histological samples, for example, endobronchial ultrasound and biopsy or, in the case of central nodes closer to the oesophagus, endoscopic ultrasound–based biopsy.

General approach to management

This is based on the stage of disease; the histological subtype; the presence of particular genetic mutations; the patient's performance status (see Box 41.13); and, crucially, the patient's own wishes.

- For ***stage 1 and 2 disease***, the treatment of choice is surgical resection by lobectomy or pneumonectomy, with stereotactic ablative radiotherapy when surgery is contraindicated. Increasingly, segmentectomies are offered and have been shown to reduce morbidity without an increase in local relapse rate. The use of neoadjuvant chemo-immunotherapy in patients with tumours larger than 40 mm in diameter with or without nodal involvement has been shown to reduce the risk of recurrence; increase the complete remission rate; and in patients with nodal involvement, to improve overall survival. It should be considered in these cases, and normally three cycles of platinum-based chemotherapy with a taxane, gemcitabine, or pemetrexed with nivolumab should be given before surgery.
- Selected patients with ***stage 3A disease*** can be managed surgically, but the alternative approach is a combination of chemotherapy and radiotherapy directed by histological subtype and molecular testing.
- For patients with ***more advanced disease***, a variety of agents can prolong survival, and specific palliative measures can relieve symptoms and improve quality of life.

Non–small-cell lung cancer

Staging of NSCLC is classified according to the TNM system (see Box 41.12 and Box 20.35); this system divides the disease into local, locally advanced, and advanced stages, with 5-year survival rates varying from 55% to 67%, 23% to 40%, and 1% to 3%, respectively.

Management

Surgery

Surgery can be curative in NSCLC (T1, N0, M0) but only 5% to 10% of all patients are suitable for resection; about 70% of these patients survive for 5 years. Preoperative assessment includes lung function testing and walking oximetry, which is used to predict postoperative potential. An active life after pneumonectomy is unlikely if the gas transfer is reduced below 50%.

Adjuvant radiotherapy and chemotherapy after surgery can improve prognosis in patients of good performance status, as shown by the International Adjuvant Lung Cancer Trial and a meta-analysis of 12 randomized controlled trials. Cisplatin-based combination chemotherapy induced a response in 60% of patients and produced a relative risk reduction of 11% with an absolute improvement in 5-year survival rate for stage II and IIIa disease of 4%, from 40.4% to 44.5%. This should be offered if patients have not had neoadjuvant chemo-immunotherapy. Patients who have high PDL-1 expression may be offered 1 year of adjuvant atezolizumab, and those with EGFR-mutated lung cancer should be offered 3 years of adjuvant osimertinib, which has been shown to improve survival.

Radiation therapy for cure

In patients who are fit and who have a stage 1 NSCLC, high-dose radiotherapy (65 Gy or 6500 rads) can result in a 27-month median survival time and a 22% 5-year survival rate. It is the treatment of choice if surgery is not appropriate; however, poor lung function can also be a relative contraindication to radiotherapy. ***Radiation pneumonitis*** (defined as an acute infiltrate precisely confined to the radiation area and occurring within 3 months of radiotherapy) develops in 10% to 15% of patients. ***Radiation fibrosis***, a fibrotic

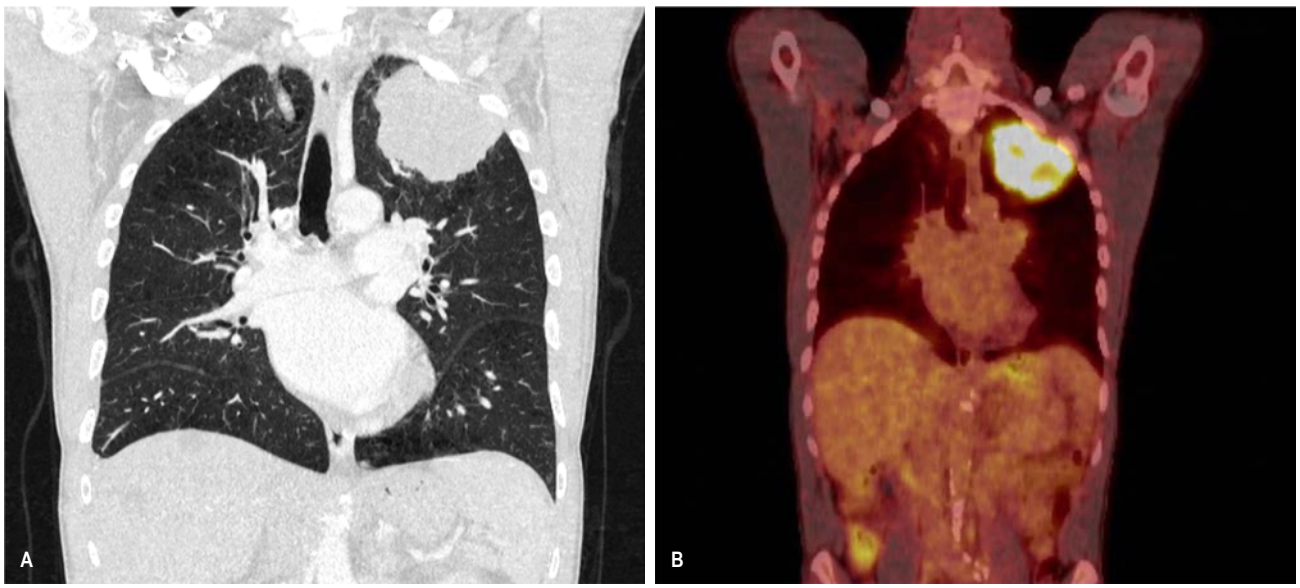


Fig. 41.15 **(A)** Coronal contrast-enhanced computed tomography (CT) scan showing a large left upper lobe primary lung malignancy. **(B)** Corresponding positron emission/CT scan show significant avidity with central necrosis. These findings are typical for a primary lung malignancy.

change occurring within 1 year or so of radiotherapy and not precisely confined to the radiation area, occurs to some degree in all patients but is usually asymptomatic.

Genetic profiling and tyrosine kinase inhibitors

Increasingly, tissue obtained at biopsy is investigated using next-generation sequencing (NGS; see Chapter 5, 'Next-generation sequencing' section) to seek the presence of specific oncogenic (activating) mutations (**Box 41.26**). These are more common in non-smokers, although they account for less than 5% of cases in unselected series. Where mutations are identified, targeting these driver mutations with an appropriate TKI is the first-line treatment (see **Box 41.19**); however, if the cancer is localized, standard radical options (surgery, radiotherapy) may still be preferred. In patients with more advanced disease, targeted drugs are preferable to cytotoxic chemotherapy in certain situations.

For EGFR mutants (with the exception of exon 20 insertions), osimertinib should be offered as first-line therapy. Currently, on progression, further treatment consists of combination chemotherapy, most often platinum and pemetrexed based with maintenance pemetrexed.

Other commonly targeted mutations are *ALK* and *ROS*. Alk mutations are most often treated with alectinib in the first instance with lorlatinib a useful second-line drug, and for ROS-1, mutations crizotinib is often the drug of first choice.

In other cases, cytotoxic chemotherapy is chosen first, and targeted drugs are used in the second-line setting (e.g., in *KRAS* mutations, in which the G12C variant is seen, sotorasib is a useful option). In RET fusions selpercatinib may be used, and in *NTRC* mutations, larotrectinib may be chosen. In all these cases, response rates of 40% to 70% may be seen with response durations between 5 and 12 months.

Immunotherapy

PDL-1 and its receptor, PD-1, are often overexpressed in patients with NSCLC, allowing cancer cells to evade destruction by the body's immune system. In patients who are PDL-1 positive (>50%) but lack the driver mutations described earlier, pembrolizumab (a MAb that acts as a PD-1 antagonist) is used alone, and in those with lower levels of expression, it may be combined with chemotherapy, with significant improvements in survival. The PDL-1 antagonist durvalumab, used every 2 weeks for 1 year alongside cisplatin-based chemotherapy, has been shown to raise the median time to progression or death from 16 to 28 months in patients with NSCLC in the setting of chemoradiotherapy.

Conventional chemotherapy

This is used in the absence of specific driver mutations. Cisplatin (or carboplatin) plus gemcitabine for squamous carcinoma and cisplatin and pemetrexed for adenocarcinoma remain the standard treatments. (An alternative, particularly in older patients, is carboplatin and paclitaxel.) Single-agent docetaxel with or without the targeted drug nintedanib may be offered as second-line therapy if patients are fit.

Box 41.26 Activating mutations seen in non–small-cell lung cancer

	EGFR	*ALK*	*ROS-1*	*PDL-1*	*BRAF*	*NTRK*
Adenocarcinoma/ large-cell neuroendocrine tumours	Y	Y	Y	Y	Y	Y
Squamous cell carcinoma in never smokers	Y	Y		Y		Y
Squamous cell carcinoma in smokers				Y		

Small-cell lung cancer

Prognosis

SCLC is an aggressive neuroendocrine tumour. It is staged in the same way as NSCLC, and patients with stage 1 tumours may benefit from surgery. For non-operable cases, disease is classified as limited- or extensive-stage disease.

Management

Limited-stage disease is confined to one anatomical area and is present in approximately 30% of patients. It is best treated with concurrent chemotherapy and radiotherapy using either cisplatin or carboplatin, with either etoposide or irinotecan, which increases the survival rate at 5 years from 15% to 25% compared with radiotherapy alone. Recurrence is usual within 6 to 12 months; for those who remain fit, further chemotherapy is appropriate. A similar degree of improvement can also be achieved with hyperfractionated radiotherapy. Prophylactic whole-brain radiation to prevent cerebral metastases can reduce symptomatic central nervous system disease and improve the overall survival rate by 5%.

For ***extensive disease***, concurrent chemoradiotherapy with carboplatin, etoposide, and the anti–PDL-1 MAb atezolizumab over four cycles followed by atezolizumab maintenance for up to 2 years is considered optimal. This can increase the median survival time from 10 to 12 months and the 18-month survival rate from 21% to 34% compared with chemotherapy alone. Second-line chemotherapy can provide further palliation for patients with good performance status.

Symptomatic care for patients with lung cancer

The prognosis for the majority of patients remains poor because the disease is diagnosed at an advanced stage, and the co-morbidity from other smoking-related diseases compromises treatment. Therefore, much of the treatment, whether symptomatic or anticancer, is delivered with palliative intent. (General palliative care is discussed in Chapter 50.) Specific issues for patients with lung cancer are the relief of bronchial obstruction and breathlessness. Options for dealing with these problems include:

- A ***neodymium-Yag laser*** passed through a fibreoptic bronchoscope can be used to vaporize inoperable fungating intraluminal carcinoma involving short segments of trachea or main bronchus. Benign tumours, strictures, and vascular lesions can also be treated effectively with immediate relief of symptoms.
- ***Cryotherapy*** is an endobronchial technique in which a cryoprobe is passed through the bronchoscope. The cryoprobe repeatedly freezes and thaws the tumour, which enables parts of it to be excised, restoring airway patency without causing bleeding. The excised tissue can be sent for histological analysis.
- ***Tracheobronchial stents*** made of silicone or in the form of expandable metal springs are available for insertion into strictures caused by tumour, external compression, or weakening and collapse of the tracheobronchial wall.
- Local radiotherapy can be used for the alleviation of local pain related to malignant infiltration of structures such as the pleura or chest wall.

Mesothelioma

Diagnosis of mesothelioma is discussed in Chapter 20 (see 'Mesothelioma' section). The treatment of patients with mesothelioma is currently not curative, although long-term survival is possible. There is no routine role for surgery either in the form of extrapulmonary pneumonectomy or radical decortication. Immunotherapy using checkpoint inhibitors (e.g., nivolumab and ipilimumab) as first-line therapy has been shown to improve survival compared with cisplatin and pemetrexed chemotherapy, with the median survival time increasing from 14 to 18 months and the 2-year rate survival from 27% to 41%.

Chemotherapy can be offered after treatment failure or when immunotherapy is contraindicated.

Metastases in the lung

Metastases are very common and usually present as round shadows (1.5–3.0 cm in diameter). They are usually detected on chest X-ray in patients already diagnosed as having cancers at other sites but may be the initial presentation of disease. Typical sites for the primary tumour include the kidney, prostate, breast, bone, GI tract, cervix, or ovary.

Metastases usually develop in the parenchyma or pleura and are often relatively asymptomatic even when the chest X-ray shows extensive disease. Rarely, metastases may develop within the bronchi and present with haemoptysis.

Metastatic carcinomas (particularly of the stomach, pancreas, and breast) can involve the mediastinal lymph nodes and spread along the lymphatics of both lungs ***(lymphangitis carcinomatosa)***, leading to progressive and severe breathlessness. Chest X-ray signs of hilar lymphadenopathy and basal shadowing are unreliable compared with the characteristic signs on CT scan of irregular thickening of the interlobular septa in a polygonal pattern around a thick-walled central vessel.

The differential diagnosis includes:

- Primary bronchial carcinoma
- Tuberculoma
- Benign tumour of the lung
- Hydatid cyst

Single pulmonary metastases can be removed surgically, but because CT scans usually show the presence of small metastases that remain undetected on chest X-ray, detailed imaging, including PET scanning and assessment, is essential before undertaking surgery.

Further reading

Forde PM, Spicer J, Lu S et al. Neoadjuvant nivolumab plus chemotherapy in resectable lung cancer. *N Engl J Med* 2022; 386(21):1973–1985.
Gandhi L, Rodríguez-Abreu D et al., KEYNOTE-189 Investigators. Pembrolizumab plus chemotherapy in metastatic non-small-cell lung cancer. *N Engl J Med* 2018; 378:2078–2092.
Peters S, Scherpereel A, Cornelissen R et al. First-line nivolumab plus ipilimumab versus chemotherapy in patients with unresectable malignant pleural mesothelioma: 3-year outcomes from CheckMate 743. *Ann Oncol* 2022; 33(5):488–499.
Soloman BJ, Mok T, Kim DW et al. First-line crizotinib versus chemotherapy in ALK-positive lung cancer. *N Engl J Med* 2014; 371:2167–2177.
Soria JC, Ohe Y, Vansteenkiste J et al., FLAURA Investigators. Osimertinib in untreated EGFR-mutated advanced non-small-cell lung cancer. *N Engl J Med* 2018; 378:113–125.

BREAST CANCER

Breast cancer is the most common cancer in females in the UK. Breast cancer screening has improved the detection of early and curable breast cancer and is a model for other screening programmes. The diagnosis of breast cancer is covered in Chapter 47.

Improvements in multimodality treatment have increased overall cure rates. Breast-conserving surgery for selected patients with early disease has greatly ameliorated the psychosexual impact of the disease (Fig. 41.16).

When therapy is being planned, decisions are based on:

- The extent of the disease (staging)
- Histological features (grading)
- The presence or absence of three crucial cell-surface receptors: the ER, progesterone receptor (PR), and HER2.

Triple-negative disease tends to be the most aggressive disease, often requiring use of systemic chemotherapy regardless of stage and grade. ER positivity allow use of hormonal treatment such as tamoxifen, while HER2 is the target of trastuzumab. Younger patients and those with a family history of the disease should be genotyped for the *BRCA* gene because those who are positive may elect to undergo bilateral mastectomy to reduce risk of subsequent disease.

Management of early disease

Surgery

In early disease, surgical options include wide local excision and segmental mastectomy with breast conservation; simple mastectomy, with or without reconstruction, is used for larger, inflammatory, or multicentric tumours. The choice is dictated by the location, extent of the breast mass, and patient preference. Clear histological margins around the cancer are important for cure.

In the absence of clinical or ultrasound evidence of lymphadenopathy, surgery to the axilla can be minimized if sentinel lymph node–guided sampling excludes local spread; otherwise, axillary nodal dissection is required if there are clinically involved nodes to gain local control and provide prognostic information to guide adjuvant treatment. The greater the amount of axillary surgery, the greater the risk of postoperative lymphoedema.

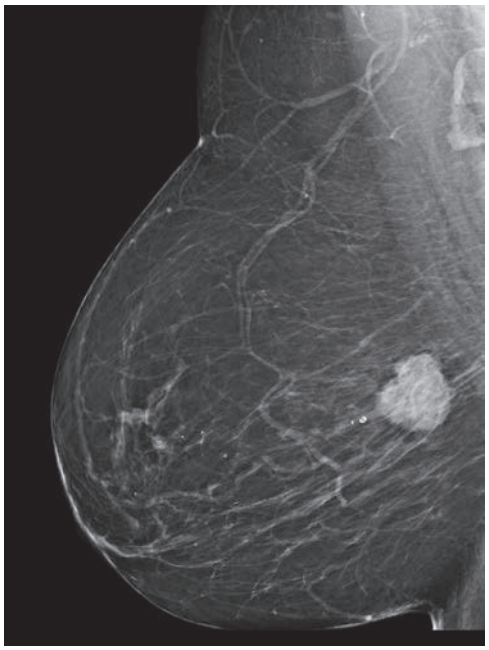


Fig. 41.16 A mammogram showing a 30-mm irregular mass in the upper-outer quadrant of the right breast, which is suspicious for carcinoma.

Box 41.27 Indications for adjuvant radiotherapy in breast cancer

- Breast-conserving surgery
- Large, high-grade primary tumour
- Proximity to surgical margins
- Lymph node metastases

Radiotherapy

All patients who undergo breast-conserving surgery for early breast cancer require postoperative radiotherapy to the breast to provide the security of local control equivalent to that conferred by mastectomy (Box 41.27). Those undergoing mastectomy who have disease close to resection margins also require radiotherapy to the chest wall. The axillary nodes are spared during radiotherapy if the axilla has been dissected (to avoid worsened lymphoedema), but supraclavicular nodes are included if there are four or more positive axillary nodes or one to three with high-risk features. Internal mammary node irradiation is occasionally given in selected higher-risk cases.

Systemic therapy

For locally advanced breast cancers, systemic therapy ***(neoadjuvant therapy)*** is sometimes given to downstage the cancer before locoregional treatment with surgery and radiotherapy. This may allow safe breast-conserving therapy in a patient with disease who, at presentation, would have required full mastectomy or to allow time for genetic testing. Neoadjuvant chemotherapy is also given in HER2-positive breast cancer, and patients with triple-negative breast cancer are treated with a combination of chemotherapy and immunotherapy. Pathological treatment response at surgery gives prognostic information for both HER2 and triple-negative breast cancer and directs adjuvant treatment options.

The term ***adjuvant therapy*** applies to systemic treatment that is given after definitive locoregional therapy in a patient who is perceived to have a significant risk of future relapse.

The aim of chemotherapy is to destroy microscopic residual disease that may have disseminated elsewhere in the body, and trials have shown benefit from both endocrine treatment and chemotherapy. Typical chemotherapy regimens are discussed later in the section 'Management of advanced disease'. In patients with triple-negative breast cancer, immunotherapy is often given to maintain treatment response after surgery if there has been a complete treatment response to neoadjuvant treatment.

Targeted therapy

A range of agents can be used, depending on the receptor status of the underlying disease.

HER2-positive disease

HER2 is overexpressed in 20% of all breast cancers and acts as an oncogene, 'driving' the cancer to grow. HER2-positive cancers tend to behave aggressively but respond to treatment with HER2 inhibitors such as trastuzumab and pertuzumab; endocrine therapy is used subsequently if tumours are also ER positive. If indicated, radiotherapy is usually given after systemic therapy, the logic being that systemic therapy is most likely to cure any microscopic disease when the tumour burden is lowest and that systemic therapy also treats regional disease and so should be given as soon as possible.

Oestrogen receptor–positive disease

For cancers that present a lesser risk and are ER positive, adjuvant endocrine therapy is delivered, for between 5 to 10 years after surgery. There are two strategies for endocrine therapy:

- ***ER blockade*** using tamoxifen (a competitive inhibitor of the oestrogen receptor) or fulvestrant (which also causes degradation of the oestrogen receptor)
- ***Oestrogen deprivation*** using aromatase inhibitors such as letrozole (which blocks non-ovarian oestrogen synthesis); aromatase inhibitors are not active in premenopausal females

In a head-to-head study, a slight advantage was proven for use of aromatase inhibitors as adjuvant therapy in postmenopausal females with ER-positive breast cancer compared with tamoxifen; consequently, this is now first-line endocrine therapy in this group. In premenopausal females with higher-risk ER-positive disease, there is an advantage for complete ovarian suppression (laparoscopic oophorectomy or luteinizing hormone-releasing hormone agonist therapy) plus aromatase inhibition.

In ER-positive breast cancer, the effector pathway that stimulates cellular proliferation is mediated via a cyclin D/cyclin-dependent kinase phosphorylating holoenzyme, which regulates the RB protein (a suppressor of G1- to S-phase progression). In patients with a high risk of recurrence, the addition of cyclin inhibitor (abemaciclib) to endocrine treatment significantly enhances the response rate and duration of this endocrine-based therapy in ER-positive breast cancer.

Bisphosphonates

For postmenopausal females at high-risk of recurrence, adjuvant bisphosphonate infusions given every 6 months for 3 to 5 years have been shown to reduce the incidence of metastatic breast cancer recurrence in bone.

Follow-up

After the end of therapy (or during ongoing adjuvant endocrine, anti-HER2 therapy, or immunotherapy), high-risk patients are reviewed clinically every 3 to 6 months by examination and enquiry into general wellbeing. Health economic studies have not supported the usefulness of proactive scanning in patients who have negative follow-up results based on these clinical assessments. However, an annual mammogram, both for the index breast and the contralateral side (where, irrespective of the genetics, there is an increased risk of a new cancer), is recommended for at least 5 years of follow-up. Low-risk patients are also followed up with 5 years of mammograms to both breasts.

Management of advanced disease

The management of patients presenting with metastatic disease depends on the extent and location of disease (Fig. 41.17).

Widespread visceral disease

Patients are advised to have systemic chemotherapy (with or without HER2-directed therapy) followed by endocrine therapy if the cancer is ER positive. Treatment is unlikely to be curative but may well improve prognosis by years, during which quality of life may be good.

The chemotherapy advised for this situation (and, indeed, the adjuvant therapy discussed earlier) is based on combinations of drugs, thereby reducing resistance (the same premise as is applied with multidrug antibiotic therapy for tuberculosis infection).

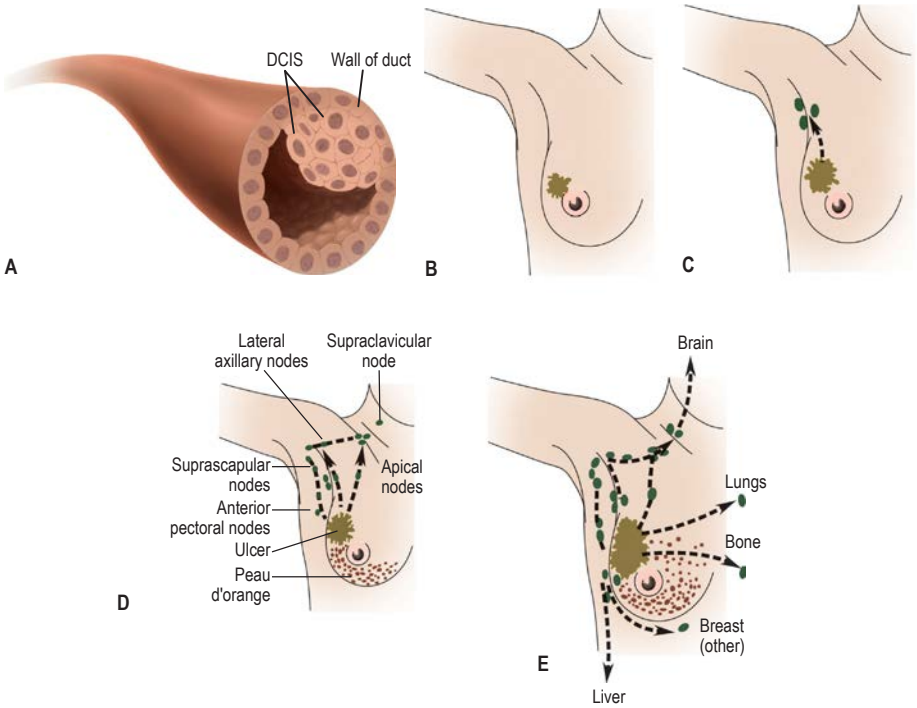


Fig. 41.17 Staging of breast cancer. **(A)** Carcinoma in situ. **(B)** Stage I: tumour is smaller than 2 cm with no nodal involvement. Stage IIa: cancer cells found in axillary lymph nodes, tumour is 2 cm or smaller with metastasis to axillary lymph nodes, or tumour is larger than 2 cm but not larger than 5 cm with no metastasis to axillary lymph nodes. **(C)** Stage IIb: tumour is larger than 2 cm but not larger than 5 cm with positive axillary lymph node involvement, or tumour is larger than 5 cm with negative axillary lymph node involvement. **(D)** Stage IIIa: tumour is 5 cm or smaller with metastasis to axillary lymph nodes that are attached to each other or to other structures or has spread to lymph nodes behind breastbone, or tumour is larger than 5 cm and has spread to axillary lymph nodes that are alone or attached to each other or to other structures or has spread to lymph nodes behind the breast bone. Stage IIIb: tumour of any size that has grown into the chest wall or skin of the breast, causing swelling of the breast or nodules in breast skin. **(E)** Stage IV: breast cancer that has spread to other parts of the body such as the bones, liver, or lung. *DCIS,* Ductal carcinoma *in situ*. (*From* Alexander's Care of the Patient in Surgery, *Seventeenth Edition*.)

Combinations of drugs are selected for their anti-tumour activity in breast cancer and their non-overlapping toxicities. Anthracyclines, taxanes, vinca alkaloids, capecitabine, and alkylating agents all feature in first-line regimens. For triple-negative breast cancer, immunotherapy and platins also have a role. Interestingly, the triple-negative group tends to carry a high mutational load, and this is the only group of breast cancers that respond well to immunotherapy.

Limited metastatic disease

For patients who present with one or only a few sites of metastatic disease (oligometastatic disease), there may still be a chance of cure. Aggressive treatment with ablative focal therapy to those oligometastatic sites (e.g., radiofrequency ablation, stereotactic radiation, or surgical metastasectomy) may be worth pursuing in a fit patient.

Indolent metastatic disease that is not immediately life-threatening

If the cancer is ER-positive, primary endocrine therapy may hold the disease at bay for long periods; usually with a CDK 4/6 inhibitor (e.g., ribociclib, abemaciclib, or palbociclib). Chemotherapy may be delayed until the cancer progresses despite this endocrine-based therapy. ER-positive breast cancers with a phosphoinositide 3-kinase escape pathway may also benefit from alpelisib when the cancer progresses.

Metastatic disease confined to regional lymph nodes

An aggressive regimen of neoadjuvant chemotherapy, surgery to breast and axillary nodes, regional radiotherapy, and long-term endocrine or HER2-directed therapy offers a chance of cure.

Relapsed disease after previous systemic therapy

The systemic therapy previously received has to be carefully considered before new therapy is instituted. The timing between the previous therapy and relapse is important because re-introduction of agents may be sensible only if the cancer has not been exposed to a particular group of drugs for a number of years. New agents (i.e., ones to which the cancer has not been exposed) are preferable, and eribulin is a chemotherapy agent that has value in this situation. In *BRCA*-driven cancer, the poly-ADP ribose polymerase (PARP) inhibitors (DNA repair inhibitors that seem to augment the defect of *BRCA* mutation in a lethal way) may enhance the efficacy of certain drugs, such as the platins, or be active as single agents. It is sometimes worth obtaining a biopsy at the time of relapse for

further genomic analysis of the 'new' tumour because the constant mutational events in a progressing cancer throw up new driver mutations, some of which can be blocked.

Further reading

Baselga J, Cortés J, Kim SB et al. Pertuzumab plus trastuzumab plus docetaxel for metastatic breast cancer. *N Engl J Med* 2012; 366:109–119.
Cortes J, Cescon D, Rugo H et al. Pembrolizumab plus chemotherapy versus placebo plus chemotherapy for previously untreated locally recurrent inoperable or metastatic triple-negative breast cancer (KEYNOTE-355): a randomised, placebo-controlled, double-blind, phase 3 clinical trial. *Lancet* 2020; 396(10265):1817–1828.
Donker M, van Teinhoven G, Straver ME et al. Radiotherapy or surgery of the axilla after a positive sentinel node in breast cancer (EORTC 10981-22023 AMROS): a randomized, multicenter, open-label, phase 3 non-inferiority trial. *Lancet Oncol* 2014; 15:1303–1310.
Early Breast Cancer Trialists' Collaborative Group (EBCTCG). Adjuvant bisphosphonate treatment in early breast cancer: meta-analyses of individual patient data from randomised trials. *Lancet* 2015; 386(10001):1353-1361. Erratum in: Lancet. 2016 Jan 2;387(10013):30. Erratum in: Lancet. 2017 Jun 24;389(10088):2472.
Gao J, Cheng J, Bloomquist, E et al. CDK4/6 inhibitor treatment for patients with hormone receptor-positive, HER2-negative, advanced or metastatic breast cancer: a US Food and Drug Administration pooled analysis. *Lancet Oncol* 2020; 21(2):250. Epub 2019 Dec 16.
Johnston S, Harbeck N, Hegg R et al. Abemaciclib combined with endocrine therapy for the adjuvant treatment of HR+, HER2-, node-positive, high-risk, early breast cancer. *J Clin Oncol* 2020; 38(34):3987.
Schmid P, Cortes J, Pusztai L et al. Pembrolizumab for early triple-negative breast cancer. *N Engl J Med* 2020; 382:810–821.

UPPER GASTROINTESTINAL CANCERS

Oesophageal cancer

Presentation and diagnosis are described in Chapter 22 (see 'Cancer of the oesophagus' section). Early diagnosis, as pioneered in Japan, where there is a particularly high incidence, has shown that it is possible to improve the prognosis of a disease that otherwise is typically diagnosed only when local metastases have already occurred.

Prognosis and management

Pretreatment histology, stage, age, and performance status are critical prognostic factors for treatment decisions, which should be made by an MDT in designated units with the surgical expertise to minimize treatment mortality. Although oesophageal SCC and adenocarcinoma are undoubtedly two different disease processes with independent tumour biology, the majority of trial data do not discriminate between the two. The influence of histology on treatment is therefore unclear and varies around the globe.

Treatment is dependent on the age and performance status of the patient and the stage of the disease, with approximately 40% of all patients still alive 1 year after diagnosis. The 5-year survival rates are 80% for stage 1 disease (T_1/T_2, N0, M0), 30% for stage 2, 18% for stage 3, and 4% for stage 4. Some 70% of patients present with stage 3 or greater disease, with 50% having distant metastases at the time of diagnosis and the majority of the remainder having regional spread into adjacent mediastinal structures, so the overall survival rates are 40% at 1 year and around 15% at 5 years.

Staging with endoscopic ultrasonography, CT, and PET/CT has improved the selection of patients of good performance status with truly localized disease, for whom curative treatment may be attempted.

Treatment of early-stage disease

Surgery provides the best chance of cure, particularly for tumours in the distal oesophagus, but it should be used as the primary modality only when imaging has shown that the tumour has not infiltrated outside the oesophageal wall (stage 1). Fewer than 40% of patients have potentially resectable disease at the time of presentation. Endoscopic mucosal resection may be appropriate for those with early mucosal disease.

Poor outcome data from surgery alone have challenged its role as monotherapy, and patients with locally advanced disease that is potentially resectable are recommended to have neoadjuvant chemotherapy with cisplatin/5-FU before surgery. Response to neoadjuvant therapy is a strong prognostic factor. Radical chemoradiotherapy alone is an alternative approach for patients with SCC of the thoracic oesophagus and avoids the morbidity associated with surgery. Those with adenocarcinoma of the oesophago-gastric junction can be treated with perioperative chemotherapy in a combination of docetaxel, oxaliplatin, fluorouracil, and leucovorin (FLOT).

Treatment of advanced or metastatic disease

This can be palliated with 5-FU or capecitabine chemotherapy in approximately 30%, increasing to 45% to 55% with the addition of oxaliplatin or irinotecan for a median duration of 6 to 8 months.

Palliative approaches

Distressing symptomatic problems with dysphagia can be partially relieved by endoscopic techniques, although dilation alone is only of short-term benefit, and the perforation risk is higher than for benign strictures. Combination of endoscopic dilation with laser or brachytherapy (see 'Types of radiation therapy' section) prolongs luminal patency and gives as good, if not better, functional results than stenting. Insertion of an expanding metal stent allows liquids and soft foods to be eaten, and percutaneous endoscopic gastrostomy tubes support liquid enteral feeding. The patient and family need considerable support and explanation to enable them to understand that feeding, including parenteral methods, does not improve survival beyond that dictated by the underlying cancer and may introduce its own complications with adverse effects on quality of life.

Gastric cancer

The presentation and diagnosis of gastric cancer are described in Chapter 22 (see 'Gastric tumours' section).

As with all cancers, treatment is discussed with an MDT.

Prognosis and management

The majority of patients are still diagnosed at an advanced stage, except in Japan, which has an active surveillance policy. Thus, in the West, the overall prognosis has not improved above 10% survival at 5 years. Selected groups may do much better and the histological grade and staging, with respect to the presence of serosal involvement (T3), nodal involvement (N1–2) and performance status, are the main factors in determining prognosis and selecting treatment.

Treatment of early-stage disease

Early non-ulcerated, mucosal lesions can be removed endoscopically by either endoscopic mucosal resection or endoscopic submucosal dissection.

Otherwise, surgery remains the most effective form of treatment if the patient is considered 'operable' in terms of lack of

direct invasion and nodal spread: selection based on operability has reduced the numbers undergoing surgery and significantly improved the overall surgical 5-year survival rates to around 30% or even 50% in those undergoing 'curative' operations. Perioperative chemotherapy with FLOT improves the 3-year overall survival rate in this patient group from 35% to 57%. If a patient has not had neoadjuvant treatment, adjuvant chemotherapy or chemoradiotherapy can be used.

Treatment of advanced disease

This may be palliated with chemotherapy, such as oxaliplatin and capecitabine combined with cisplatin and infusional 5-FU for patients with SCC, with a response in 40% to 50% for a median of 8 to 12 months in patients with good performance status (see 'Performance status' section). The addition of trastuzumab a and pertuzumab for tumours overexpressing HER2 (see 'HER2-positive disease' section) has improved the median time survival from 11 to 16 months. For patients with deficient mismatch repair, high levels of microsatellite instability (MSI), or PDL-1 overexpression, immune checkpoint inhibitors targeting PD-1 (pembrolizumab) have activity in this tumour type and may be considered in the future for patients who have relapsed. Trifluridine/tipiracil, which interferes with tumour DNA synthesis, may also have activity in the relapsed setting.

Palliative approaches

Supportive care for patients with upper GI cancers, more than cancers in any other site, must include attention to nutrition with the use of endoscopic stents to relieve obstruction, nasojejunal and percutaneous gastrostomy feeding tubes, and occasionally parenteral nutrition. When the disease progresses and active anticancer treatment is no longer appropriate, management of distressing obstructive symptoms can include octreotide to reduce secretions and, if necessary, a venting gastrostomy.

Gastrointestinal stromal tumours

GISTs are rare, slow-growing neoplasms that may arise in the stomach or small or large intestine. The prognosis depends on size, site, Ki-67 proliferation index (see Chapter 22, 'Gastrointestinal stromal tumours' section), and association with neurofibromatosis type 1 (see Chapter 28, 'Neurofibromatosis type 1 [von Recklinghausen's disease]' section). Immunohistochemistry for CD117 and DOG-1 (discovered on GIST-1) is usually positive; if negative, the diagnosis can be reached by analysis of the known mutational sites in *c-Kit* or *PDGFRA* oncogenes. These mutations predict response to TKIs such as imatinib (see Chapter 39, 'Chronic myeloid leukaemia and the myeloproliferative neoplasms' section). Surgery is potentially curative for patients with localized disease, and imatinib adjuvant therapy for 3 years prolongs survival. In metastatic disease, imatinib treatment should continue until relapse and can induce remissions in 50% to 80% of patients for a median of 2 years; further responses are possible with alternative TKIs such as sunitinib and nilotinib.

Further reading

Al-Batran SE, Homann N, Schmalemberg H et al. Perioperative chemotherapy with docetaxel, oxaliplatin, and fluorouracil/leucovorin (FLOT) versus epirubicin, cisplatin, and fluorouracil or capecitabine (ECF/ECX) for resectable gastric or gastroesophageal junction (GEJ) adenocarcinoma (FLOT4-AIO): a multicenter, randomized phase 3 trial. *J Clin Oncol* 2017; 35(suppl; abstr 4004).
Bolger JC, Donohoe CL, Lowery M et al. Advances in the curative management of oesophageal cancer. *Br J Cancer* 2022; 126:706–717.
Judson I, Bulusu R, Seddon B et al. UK clinical practice guidelines for the management of gastrointestinal stromal tumours (GIST). *Clin Sarcoma Res* 2017; 7:6.

LOWER GASTROINTESTINAL CANCERS

Small intestinal cancer

Predisposing factors for small intestinal cancer and its presentation are described in Chapter 22. The principles of treatment are determined by extrapolation from management of the more common colorectal cancers (**Figs. 41.18 and 41.19**).

Colorectal cancer

The presentation and diagnosis of colorectal cancer are described in Chapter 22 (see 'Colorectal carcinoma' section).

Prognosis

The site of the disease (above or below the pelvic peritoneal reflection), TNM stage (see **Box 41.11**), surgical margins, and the patient's performance status (see **Box 41.2**) are the main clinical prognostic factors (**Box 41.28**). Gene expression profiling using Oncotype DX and Coloprint, for example, can identify prognostic subgroups but is not predictive of treatment benefit. High MSI predicts a better prognosis and can be used to identify low-risk patients for whom adjuvant therapy is unlikely to be of benefit. Mutations in *BRAF* occur in 6% to 8% of patients and indicate a poor prognosis, as do mutations with predictive effect in *KRAS* in 40% and *NRAS* in 10% to 15% of cancers.

Management

Management should be undertaken by MDTs working in specialist units. About 80% of patients with colorectal cancer undergo surgery (often laparoscopically), although fewer than half of them survive for more than 5 years. The operative procedure depends on the cancer site. Long-term survival is only likely when the cancer is completely removed by surgery with adequate clearance margins and regional lymph node clearance.

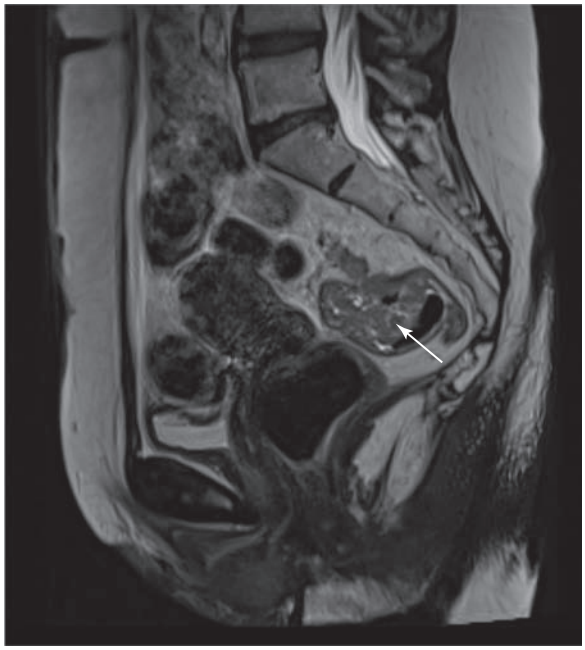


Fig. 41.18 A coronal contrast-enhanced T1-weighted magnetic resonance imaging scan showing a 5.5-cm rectosigmoid tumour, indicated by the *arrow*.

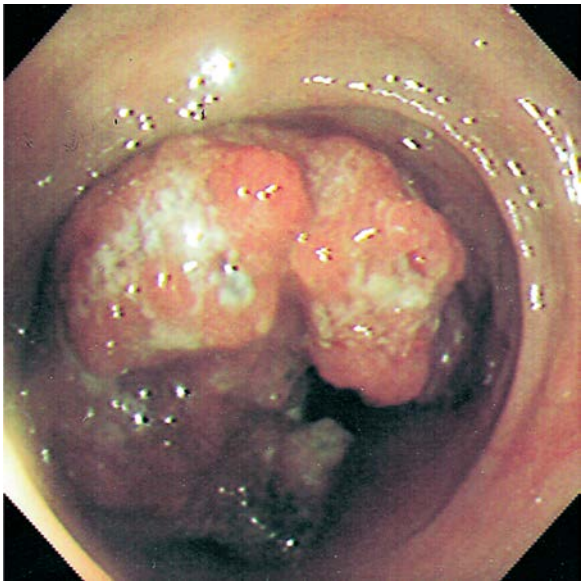


Fig. 41.19 Colon cancer. *(From* Medical Assistant: Digestive System, Nutrition, Financial Management and First AidR—Module C, *Second Edition. B9780323242561000024/u02-02-9780323242561.)*

Box 41.28 Prognostic and predictive factors for colorectal cancer

- Site above or below pelvic peritoneal reflection
- TNM stage and performance status
- Surgical margins
- Favourable mutational profile:
  - Low microsatellite instability
  - *BRAF* wild type
  - *RAS* wild type

M, Metastasis; *N,* node; *T,* tumour.

Surgery

- ***Total mesorectal excision (TME)*** is required for rectal cancers and removes the entire package of mesorectal tissue surrounding the cancer. A low rectal anastomosis is then performed. Abdomino-perineal excision, which requires a permanent colostomy, is reserved for very low tumours within 5 cm of the anal margin. TME combined with preoperative radiotherapy reduces local recurrence rates in rectal cancer to around 8% and improves survival. Pre- or postoperative chemotherapy reduces local recurrence rates but had no effect on survival in a recent study.
- ***Segmental resection*** and restorative anastomosis, with removal of the draining lymph nodes as far as the root of the mesentery, is used for cancer elsewhere in the colon. Surgery in patients with obstruction carries greater morbidity and mortality. When technically possible, preoperative decompression by endoscopic stenting with a mesh-metal stent relieves obstruction, so surgery can be elective rather than emergency and is probably associated with a decrease in morbidity and mortality.
- ***Local transanal surgery*** is very occasionally used for early superficial rectal cancers.
- ***Surgical or ablative treatment of liver and lung metastases*** prolongs life when treatment is technically feasible and the patient is fit enough to undergo the treatment.
- ***Radiotherapy is not helpful*** for colonic cancers proximal to the rectum because of difficulties in delivering a sufficient dose to the tumour without excess toxicity to adjacent structures, particularly the small bowel.
- ***Adjuvant postoperative chemotherapy*** improves disease-free survival and overall survival in patients with stage III colon cancer. Those with stage II tumours and advanced features such as vascular invasion may also benefit. Genetic testing for *KRAS* mutations can help target appropriate chemotherapy regimens.

Neoadjuvant chemotherapy and radiotherapy

Neoadjuvant chemoradiation treatment of locally advanced rectal cancers, using 5-FU or capecitabine with 5 weeks of radiotherapy and then a 2-month wait, has increased the proportion of locally advanced tumours that can be resected with clear surgical margins and has improved the local control rate but has not had a beneficial effect on distant metastatic relapse and overall survival. Preoperative, rather than postoperative, treatment has reduced toxicity to the other pelvic structures because of preservation of the normal anatomical relations. Alternatives include short-course radiotherapy over 1 week with immediate or delayed surgery and total neoadjuvant treatment with chemotherapy before or after chemoradiotherapy in patients most at risk of metastatic disease, although this has not been proven in large randomized trials. In cases in which it is desirable to avoid radiation, preoperative chemotherapy alone can achieve similar response rates. In 20% to 30% of patients, there is a complete clinical and radiological response to treatment after chemoradiotherapy, and in these individuals, it may be possible to avoid surgery with stringent follow-up.

Adjuvant chemotherapy and radiotherapy

Adjuvant chemotherapy with oxaliplatin and 5-FU or capecitabine chemotherapy increases survival in patients with node-positive colorectal cancer, with a 5-year overall survival rate of 73%. Recent data have demonstrated that in those with lower-risk stage III disease, it is possible to give 3 months of adjuvant therapy rather than 6 months without reducing survival, sparing patients the neuropathy that is associated with oxaliplatin. Less benefit is seen with adjuvant treatment for stage II, node-negative patients. Those with higher-risk features can be considered for adjuvant single-agent capecitabine. If the tumour is deficient in one of the mismatch repair genes, adjuvant chemotherapy in this patient cohort has been shown not to be effective. Immunotherapy could be considered for these individuals in a trial setting.

Follow-up

In all patients who have surgery, a total colonoscopy should be performed before surgery to look for additional lesions. If total colonoscopy cannot be achieved before surgery, a second 'clearance' colonoscopy within 6 months of surgery is essential. Patients with stage II or III disease should be followed up with regular colonoscopy and carcinoembryonic antigen (CEA) measurements; rising levels of CEA suggest recurrence. CT scanning to detect operable liver metastases should be performed for up to 5 years after surgery.

Metastatic colorectal cancer

Patients with colorectal cancer may present with liver metastases. With appropriate selection of patients who have a good performance status and in whom MRI and PET/CT scans do not demonstrate disease elsewhere, local treatment can prolong good-quality survival. A variety of methods are used, including surgical resection, stereotactic body radiotherapy, radiofrequency ablation, cryoablation, or hepatic artery embolization. Small lesions can be ablated, but larger ones are best managed by partial hepatectomy or a

combination approach, so that embolization is followed by hepatic regeneration before final resection. Patient selection is critical; long-term survival without recurrence is reported in up to 20% at 5 years with a single lesion smaller than 4 cm amenable to resection presenting more than 1 year from initial diagnosis. More patients can be rendered suitable with perioperative chemotherapy, such as FOLFOX (oxaliplatin, 5-FU, and folinic acid) or FOLFIRI (irinotecan and 5-FU) with cetuximab if *RAS* is wild type. A similar approach is used for patients with lung metastases.

Untreated advanced colorectal cancer has an overall survival time of 5 to 6 months. With sequential chemotherapy regimens (including FOLFOX, CAPOX [oxaliplatin and 5-FU or capecitabine], FOLFIRI and LONSURF [trifluridine–tipiracil]), this can increase to more than 2 years. Patients without mutations in the *KRAS* gene can be considered for treatment with EGFR-targeted agents such as cetuximab and panitumumab. Immunotherapy can be of use in patients with loss of function of one or more of the mismatch repair genes.

Anal cancer

Patients with anal SCC are radically treated with chemoradiotherapy with mitomycin C and 5-FU or capecitabine. This is organ-preserving treatment, enabling the patient to avoid surgery that would result in a colostomy. Salvage abdominoperineal resection is possible if the tumour recurs. Patients with metastatic anal cancer are treated with carboplatin and paclitaxel chemotherapy. Immune checkpoint inhibitors may have activity in these tumours, as they do in SCC of the head and neck.

Further reading

Biller LH, Schrag D. Diagnosis and treatment of metastatic colorectal cancer: a review. *JAMA* 2021; 325(7):669–685.
Boland CR, Goel A. Prognostic subgroups among patients with stage II colon cancer. *N Engl J Med* 2016; 374:277–278.
Sinicrope FA. Increasing incidence of early-onset colorectal cancer. *N Engl J Med* 2022; 386(16):1547–1558.

HEPATOBILIARY AND PANCREATIC CANCERS

Liver

Hepatocellular carcinoma

HCC (see 'Infectious agents' section) arises in cirrhotic livers in 95% of cases and can be screened and detected by cross-sectional imaging and a rise in serum α-fetoprotein (AFP).

Prognosis and management

Surgical resection of isolated lesions of less than 5 cm in diameter or up to three lesions of less than 3 cm in diameter is associated with a median survival of 5 years, although the remaining liver remains at risk of further recurrence. Liver transplantation offers the only opportunity for cure for patients with a small primary tumour, but is often limited by the underlying cause of the hepatitis and cirrhosis. Other focal treatments include transarterial embolization and chemoembolization, radioembolization, radiofrequency ablation, and stereotactic body radiotherapy in patients with small primaries and adequate liver function; these prolong survival, though less successfully than surgery.

In progressive disease, a combination of the immunotherapy atezolizumab and the VEGF inhibitor bevacizumab has shown median overall survival outcomes of 19.2 months and is considered the standard of care for appropriate patients. Second-line therapies include sorafenib and lenvatinib.

Biliary tract

Cancer of the biliary tract may be intra- or extrahepatic. These malignancies represent approximately 1% of all cancers. A number of associations have been identified, such as choledochal cyst and chronic infection of the biliary tract with, for example, *Clonorchis sinensis*. There are also associations with autoimmune disease processes, such as primary sclerosing cholangitis and inflammatory bowel disease.

Cholangiocarcinoma

Cholangiocarcinoma usually presents with jaundice and is detected by imaging: initially ultrasonography, and most important, magnetic resonance cholangiopancreatography. Spread is usually by local lymphatics or local extension. Cholangiocarcinoma of the common bile duct may be resectable at presentation, but local extension precludes such management in the majority of more proximal lesions. Localized disease justifies an aggressive surgical approach, including partial hepatic resection. Hepatic transplantation for selected stage 1 and 2 disease has achieved an 80% 5-year survival rate.

Palliative chemotherapy for patients with good performance status and advanced disease using gemcitabine and cisplatin effects a response in 50% with a median survival of 12 months. The addition of immunotherapy extends the median overall time survival by an additional 1.3 months and is considered in select cases. Chemoradiation has been used to treat patients with localized small hilar cholangiocarcinomas, and radiotherapy can provide good analgesia.

Pancreas

Pancreatic adenocarcinoma

The presentation and initial investigation of pancreatic cancer are covered in Chapter 24 (see 'Pancreatic cancer' section).

Prognosis

The 5-year survival rate for carcinoma of the pancreas is approximately 2% to 5%, with surgical intervention representing the only chance of long-term survival and 90% of patients having a cause of death directly related to the disease. Approximately 20% of all cases have a localized tumour suitable for resection, but in an older adult population, many individuals have co-morbid factors that preclude such major surgery.

Surgery

Surgical intervention represents the only chance of long-term survival. Assessment of the primary tumour and involvement of local vessels, including the coeliac artery, superior mesenteric artery and vein, portal vein, and hepatic artery, is critical in determining resectability. When treatment is considered to be of curative intent, a pancreatoduodenectomy ***(Whipple procedure)*** (Fig. 41.20) is required to remove tumours of the head and neck of the pancreas. Tumours of the body and tail are resected as part of a distal pancreatectomy, which is increasingly carried out by a laparoscopic approach.

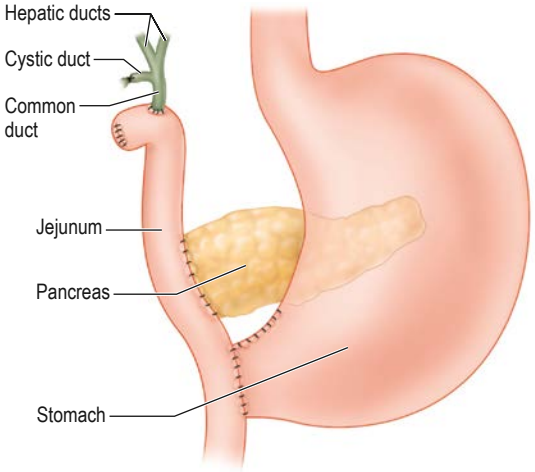


Fig. 41.20 Diagram of the aftermath of a Whipple's procedure, in which the pancreas, stomach, and jejunum are joined back together after a pancreaticoduodenectomy. (*From* Saunders Comprehensive Review for the NCLEX-PN® Examination, *Ninth Edition pii:B9780443112874000045X/f45-04-9780443112874*.)

Borderline-resectable tumours can be considered for neoadjuvant therapy to downstage the tumour, and a combination of chemotherapy followed by chemoradiotherapy before surgery has shown an improvement in overall survival over upfront surgery and the previous standard-of-care regimen of adjuvant gemcitabine and cisplatin chemotherapy. More intensive adjuvant chemotherapy in the form of combination fluorouracil, oxaliplatin, irinotecan, and leucovorin (FOLFIRINOX) has also led to better overall survival. The decision over whether to treat patients with a neoadjuvant or a more intensive adjuvant approach is individualized.

Advanced disease

In the majority of cases, management is palliative. Palliative chemotherapy for advanced disease can be managed with combination chemotherapy. First-line treatment for patients with a good performance status is FOLFIRINOX. This combination treatment was found to lead to a median overall survival time of 11.1 months. An alternative regimen for patients who are not fit enough for this treatment is gemcitabine and nab-paclitaxel (paclitaxel as albumin-bound nanoparticles).

Palliative interventions

- ***Pain*** is a debilitating feature in many cases. Management is best directed by dedicated pain teams. Opiates are a mainstay. There is some evidence that early intervention with coeliac axis block may have benefits.
- ***Nutritional deficit*** is a frequent feature at presentation and may be exacerbated by chemotherapy. Early dietetic support may alleviate this. Malabsorption is common with pancreatic cancer and can be managed by the introduction of pancreatic enzyme supplementation. The adverse metabolic effects of diabetes can be avoided by early detection and management.
- ***Obstructive jaundice*** will occur at some stage in 70% of cases and is frequently associated with anorexia and nausea, as well as pruritus. Endoscopic placement of endoprostheses (stents) offers excellent palliation (see Chapter 50).
- ***Duodenal obstruction*** is seen in up to 20% of cases, particularly when the tumour is situated in the head or uncinate process of the pancreas. Endoscopic stenting is the treatment of choice with excellent palliation. Surgical bypass is an option in selected cases.

Further reading

Conroy T, Hammel P, Hebbar M et al. FOLFIRINOX or gemcitabine as adjuvant therapy for pancreatic cancer. *N Engl J Med* 2018; 379(25):2395–2406.
Valle JW, Kelley RK, Nervi B, Oh DY, Zhu AX. Biliary tract cancer. *Lancet* 2021; 397(10272):428–444.
Vogel A, Meyer T, Sapisochin G, Salem R, Saborowski A. Hepatocellular carcinoma. *Lancet* 2022; 400(10360):1345–1362.

UROLOGICAL CANCERS

Renal cancer

Clinical features and diagnosis

See Chapter 26 (see 'Renal cell carcinoma' section).

Investigations and management

At diagnosis, renal cancers are staged using CT, MRI, or both to assess operability and the presence or absence of metastases.

Disease confined to the kidney

A nephrectomy is performed unless bilateral tumours are present, in which case conservative surgery may be indicated. Preservation of as many nephrons as possible is beneficial in the long term, and a partial nephrectomy should be undertaken when possible.

Metastatic disease

For patients with metastatic disease, prognosis is predicted on the basis of the International Metastatic Renal Cell Carcinoma Database Consortium criteria. One point is scored for each of the following poor prognostic features:

- Impaired fitness (<80% on the Karnofsky performance status score)
- Haemoglobin below the lower limit of normal
- Neutrophils above the upper limit of normal
- Platelets above the upper limit of normal
- Serum calcium above the upper limit of normal
- Less than 12 months from diagnosis to the requirement for systemic chemotherapy

Those with no adverse prognostic factors have a 2-year survival rate of more than 75% and are considered to have a good prognosis, those with one or two factors are regarded as having an intermediate prognosis and have an approximately 50% chance of surviving for 2 years, and those with three factors or more have a poor prognosis and a 2-year survival rate of 7%. Patients with a good prognosis may benefit from a debulking palliative nephrectomy, which is sometimes followed by regression of metastases. Adjuvant therapy reduces the risk of progression, but the benefits must be weighed against toxicity.

Systemic therapy

The TKIs are the mainstay of therapy for metastatic disease. When the prognosis is good, either sunitinib or pazopanib is given as first-line treatment. For patients with an intermediate or poor prognosis but preserved performance status, the combination of the checkpoint inhibitor pembrolizumab (which targets PD-1) and the TKI axitinib increases survival compared to sunitinib (hazard ratio [HR], 0.53). Similar results have been seen with nivolumab and

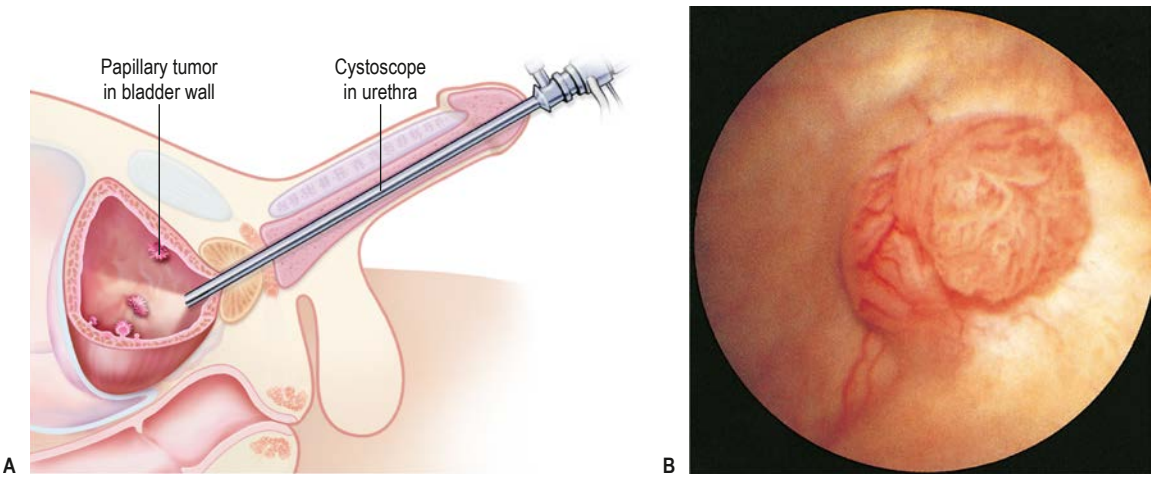


Fig. 41.21 Imaging of bladder cancer. **(A)** Cystoscope in typical male bladder. **(B)** Cystoscopic view of a cancerous growth (a transitional cell carcinoma) on the bladder wall. (The Human Body in Health & Disease, *Eighth Edition. pii:B9780323734165000293/f20-14-9780323734165*.)

cabozantinib and lenvatinib and pembrolizumab. The response rate to these regimens is 60% to 70%. An alternative in these patients is the combination of ipilimumab and nivolumab (HR, 0.69 compared with sunitinib for overall survival). This combination has a lower response rate than TKI based combinations but a greater proportion of durable responses. Good-risk patients may benefit from high-dose IL-2, which appears to increase the survival rate by 5% to 10%. Sequential use of other drugs, such as cabozantinib, lenvatinib, and everolimus, can produce further responses and improvement in disease control. Isolated metastases may be amenable to surgery or stereotactic irradiation.

Urothelial (bladder and ureteric) cancer

These cancers are usually TCCs, associated with smoking and exposure to a range of chemical agents, although SCCs are common in areas where schistosomiasis is endemic (Fig. 41.21). Their clinical presentation is described in Chapter 26 ('Tumours of the kidney and genitourinary tract' section).

Superficial cancers

Patients with superficial TCC of the bladder without invasion through the basement membrane may be treated by transurethral resection or local diathermy. The risk of recurrence varies according to the degree of differentiation, and follow-up check cystoscopies and cytological examination of the urine are required. Patients with recurrent superficial TCCs can be treated with bladder installation of the mycobacterium bacille Calmette–Guérin, which activates immune destruction of cancer cells; alternatively, a cytotoxic drug may be used (either gemcitabine or mitomycin).

Invasive bladder cancer

For invasive cancers, the strongest evidence is for neoadjuvant chemotherapy followed by either surgery (cystectomy) or chemoradiotherapy.

- ***Neoadjuvant chemotherapy*** is usually cisplatin based, most commonly with gemcitabine. It improves absolute survival by 10%, the prognosis ranging from a 5-year survival of 80% to 90% for lesions not involving bladder muscle to 5% to 10% for those presenting with metastases. If no neoadjuvant therapy has been offered, then postoperative adjuvant treatment is recommended.
- ***Surgical treatment*** involves cystectomy, generally followed by ileal conduit and stoma formation; in some cases, however, a neobladder can be constructed, allowing more normal micturition. This very much depends on the position of the tumour and the absence of extensive carcinoma *in situ.*
- ***Bladder preservation strategies*** for those with T1 with grade 3 histology or T2–T4 tumours can alternatively be offered with chemoradiotherapy using mitomycin and fluorouracil. In many patients, this can achieve complete disease response, bladder preservation with a good quality of life, and comparable long-term survival (30%–50% at 5 years) to cystectomy.

Metastatic urothelial cancer

Patients with metastatic disease may be palliated by chemotherapy (cisplatin or carboplatin based) or immunotherapy (using a PD-1/PDL-1 checkpoint inhibitor such as pembrolizumab or atezolizumab). Responses are seen in approximately 25% of patients with PDL-1 expression. In patients who have completed chemotherapy, maintenance avelumab for up to 2 years is indicated regardless of the degree of PDL-1 expression. Encouraging responses have been seen with the Mab enfortumab vedotin targeted against nectin 4. This has been shown to improve overall survival compared with chemotherapy (13 vs 9 months). Combining it with pembrolizumab may be particularly effective.

Upper tract tumours

Patients with early-stage tumours are treated by nephroureterectomy. The addition of adjuvant therapy with cisplatin and gemcitabine in the management of non-metastatic T2 tumours has been shown to reduce the risk of recurrence by half.

Prostate cancer

Clinical features and diagnosis

The pathogenesis, presentation, diagnosis, and initial investigation of prostate cancer are discussed in Chapter 47 (see 'Prostate cancer' section).

Prognosis

The histological appearances are graded and accorded a Gleason score (see Fig. 47.12). Together with the serum prostate-specific antigen (PSA) level, and accurate staging of the local extent of disease with pelvic MRI (dynamic contrast) and transrectal ultrasound examination, this score can identify prognostic groups. Treatment decisions must balance the age and performance status of the patient with the predicted behaviour of the cancer, which is becoming more objectively identifiable with advances in gene profiling. This allows the selection of patients with a good prognosis for no active treatment; they may reasonably choose to be kept under surveillance and die with, but not because of, their prostate cancer, as do 75% of males older the age of 80 years. In experienced hands, pre-biopsy MRI may allow targeted biopsy of the prostate and prediction of those who will have clinically significant prostate cancer (Fig. 41.22).

Management

Disease localized to the prostate

This can be managed by surgical resection, brachytherapy, radiotherapy, or watchful waiting with regular surveillance. The optimum approach depends on both features of the disease and the wider patient context.

Locally advanced disease

Patients with locally advanced prostate cancer (T3, N0) without distant metastases are best treated with combined androgen deprivation (see later) and radiotherapy; this improves the 10-year survival rate, compared with endocrine treatment alone, from 61% to 71%.

Endocrine therapy

Prostate cancer is the most hormone-sensitive malignancy, with the AR playing a critical role. However, prostate cancer tissue is able to trap circulating androgens so that tissue levels of androgens are maintained despite very low levels of circulating androgens. This means that resistance to endocrine therapy often develops over time, commonly involving receptor super-sensitivity and altered ligand binding, allowing other steroids to act as agonists and eventually bypassing the AR. A number of strategies are used to reduce or abolish androgen stimulation of cancer cells:

- ***Gonadotrophin-releasing hormone (GnRH) agonists*** (e.g., goserelin and leuprorelin) are as effective as orchidectomy at lowering circulating androgens and inducing responses in patients with prostate cancer. However, in the first week, GnRH agonists produce increases in luteinizing hormone and testosterone, which can result in a tumour flare in metastatic disease; they must therefore be combined with an antiandrogen (e.g., flutamide) in the initial phases. An alternative is to use a gonadorelin antagonist (e.g., degarelix). Physical castration, achieved by bilateral orchidectomy, achieves very effective androgen deprivation but is invasive, often unacceptable to patients, and generally unnecessary.
- ***Androgen receptor–targeted inhibitors*** (ARTIs): These include drugs such as bicalutamide and the more potent antagonists enzalutamide and darolutamide. They directly block the action of testosterone on prostate cancer cells.
- ***Androgen synthesis inhibitors***, such as abiraterone, act by inhibiting CYP17, a key enzyme in the testosterone production pathway.
- ***Corticosteroids and oestrogens*** may also be helpful in disease that has become refractory to castration.

Hormonal agents may also be used after chemotherapy for prostate cancer, and both enzalutamide and abiraterone have been shown to prolong life in this situation.

Non-hormonal therapy

Non-hormonal approaches are frequently used for metastatic disease:

- ***Conventional chemotherapy.*** The use of docetaxel in the early management of metastatic castration-sensitive disease has led

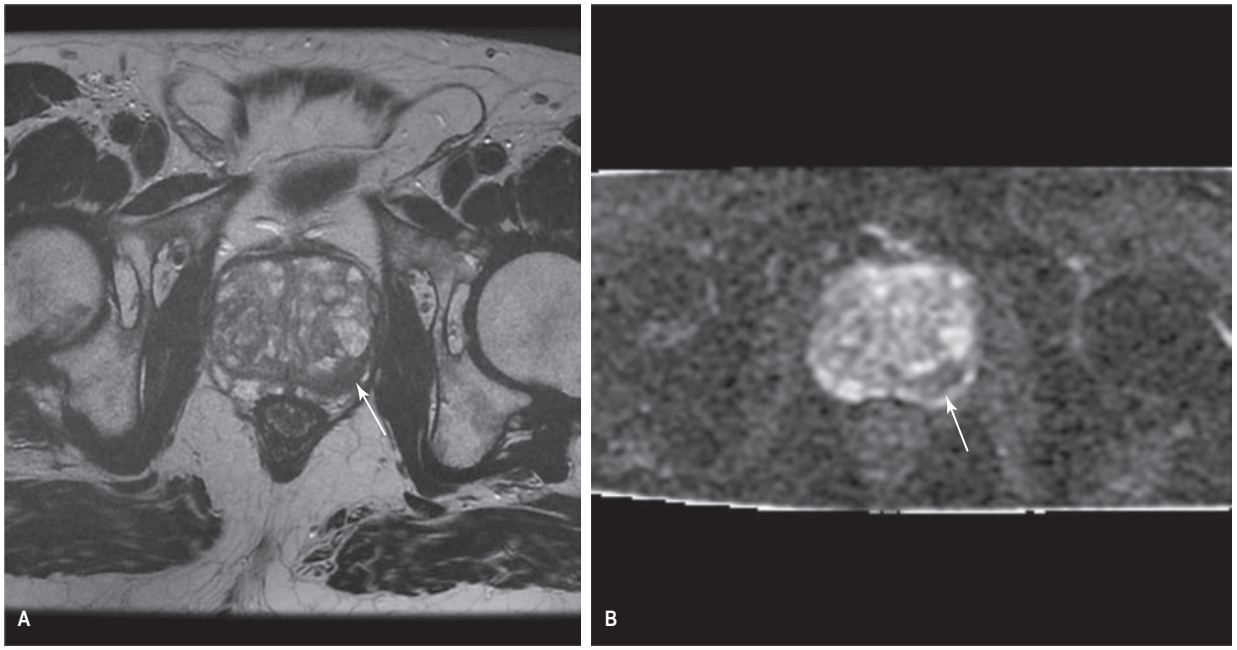
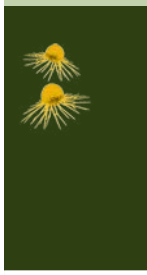


Fig. 41.22 **(A)** An axial slice of a T2-weighted magnetic resonance imaging scan demonstrating reduced signal intensity in the right peripheral zone midgland. **(B)** The corresponding diffusion-weighted sequence. The tumour is shown by the *arrow*.



to an improvement in overall survival. The HR for benefit from docetaxel is constant, but when it is used early at a point when the survival rate is greater, this translates into a greater absolute benefit in overall survival than when it is used after failure of hormonal therapy. The timing of chemotherapy versus hormonal treatment is currently not clear in prostate cancer: early use of the androgen synthesis inhibitor abiraterone, as an alternative to docetaxel, in the management of metastatic castration-sensitive disease seems to confer similar benefits. If chemotherapy fails, a return to hormonal therapy may prolong disease control. Currently, there is no consensus as to whether chemotherapy should be deferred until all available hormonal therapy has failed or be used earlier in disease control.

- ***PARP inhibitors.*** It is clear that a proportion of prostate cancers have deficits in DNA repair, often associated with carriage of the *BRCA2* gene. These patients seem to respond to PARP inhibitors such as olaparib, which prevent repair of DNA strand breaks and cause the death of cancer cells.
- ***Bone-targeted therapy.*** Drugs such as alpharadin (^{223}radium), zoledronic acid, and denosumab, alongside palliative radiotherapy, can be used to target painful metastases.
- ***Radioligand therapy.*** Recently developed therapy using MAbs with a radionuclide payload (e.g., lutetium-labelled antibodies against prostate-specific membrane antigen [PSMA]) has been shown to be effective post docetaxel and an ARTI in metastatic, castration-resistant disease.

General approach to treatment

Profound androgen deprivation using ARTIs from the start (darolutamide or enzalutamide or abiraterone) leads to improved survival rates in the initial treatment of prostate cancer compared with docetaxel and a gonadorelin analogue alone and may become the treatment of choice for younger patients (and particularly those with high-volume disease). After the failure of an AR-targeted drug and docetaxel, alternatives include further hormone treatments (oestrogens and corticosteroids) or further chemotherapy (e.g., cabazitaxel).

Penile cancer

Penile cancer is described in Chapter 47 (see 'Penile cancer' section).

Testicular and ovarian germ cell tumours

The diagnosis of these conditions is dealt with in Chapter 47.

Management

- ***In males***, surgery involves radical inguinal orchidectomy to avoid spillage of highly metastatic tumour in the scrotum.
- ***In females***, surgery for diagnosis and staging should always be conservative, compared with the approach in epithelial ovarian cancer, with preservation of fertility because of the efficacy of chemotherapy.

Seminomas

Seminomas (in males) are very radiosensitive and chemosensitive. They are associated with raised serum lactate dehydrogenase (LDH) but only rarely mildly raised β-human chorionic gonadotropin (β-hCG) and never raised AFP. Stage I disease limited to the gonad is associated with a 10% to 30% 5-year risk of recurrence with surgery alone, so adjuvant therapy, with either chemotherapy or radiotherapy to the para-aortic lymph nodes, is preferred because it leads to a cure rate of more than 95% in patients with early-stage disease. Carboplatin is the first choice because of convenience, reduced acute side-effects, and absence of the long-term risks of secondary malignancy that are associated with radiotherapy. Alternatively, post-surgical surveillance can be undertaken, with treatment reserved for those who relapse; this approach carries an equally high cure rate because combination chemotherapy (e.g., cisplatin, etoposide and bleomycin [BEP]) cures more than 90% of those with visible metastatic disease. In females, the equivalent of the seminoma is the dysgerminoma, which is morphologically identical to seminoma and is managed similarly.

Non-seminomas

Early disease

The risk of relapse after surgery for stage I disease varies from 5% to 50%, depending on the prognostic factors of histological differentiation, presence of embryonal elements, and extent of local and vascular invasion. For those at moderate to high risk, adjuvant chemotherapy with a single cycle of BEP carries a 97% cure rate.

Metastatic disease

This commonly involves para-aortic lymph nodes and lungs but may spread rapidly (especially if trophoblastic, β-hCG–producing elements are present) and cause life-threatening respiratory or other organ failure. In males, rapid diagnosis can be made in the presence of gynaecomastia and a positive urinary pregnancy test result, allowing the institution of potentially life-saving treatment. About 80% of teratomas express either β-hCG or AFP, and almost all cases of metastatic disease are associated with an elevation of the less specific serum marker, LDH.

The percentage cure rate depends on the height of the tumour markers (AFP >10 000 ng/mL and β-hCG >100 000 IU/L) and the sites of metastases; it ranges from 60% to 95%. The most established regimen is BEP; three or four cycles are usually given. More intensive regimens appear to yield higher progression-free survival rates and may be preferred in some patients.

Relapsed disease

Patients who relapse after initial chemotherapy may well be cured by subsequent chemotherapy; this is most common with further cisplatin-based chemotherapy or high-dose chemotherapy, usually with carboplatin and etoposide, and a stem cell transplant. The timing of high-dose chemotherapy is uncertain: some clinicians prefer to defer it and use it only after standard-dose salvage treatment has failed. Others believe that the best results are seen if it is used on first relapse.

Although approximately 20% of males are infertile because of azoospermia at the time of diagnosis, the majority of the remainder retain their fertility after chemotherapy and are able to father children normally. Similarly, most females retain their fertility, although less is known about the association with infertility at presentation owing to the much lower frequency of germ cell tumours in females.

Further reading

Davis ID, Martin AJ, Stockler MR, et al; ENZAMET Trial Investigators and the Australian and New Zealand Urogenital and Prostate Cancer Trials Group. Enzalutamide with standard first-line therapy in metastatic prostate cancer. *N Engl J Med* 2019; 381(2):121–131.

de Bono J, Mateo J, Fizazi K et al. Olaparib for metastatic castration-resistant prostate cancer. *N Engl J Med* 2020; 382(22):2091–2102.

Hamdy FC, Donovan JL, Lane JA et al. Fifteen-year outcomes after monitoring, surgery, or radiotherapy for prostate cancer. *N Engl J Med* 2023; 388(17):1547–1558.

GYNAECOLOGICAL CANCERS

The main gynaecological cancers are those of the cervix, the body of the uterus (endometrium), and the ovary. The presentation and investigation of these cancers are introduced in Chapter 47. All are important, and enormous progress has been made in therapy.

Cervical cancer

A significant reduction in the incidence of cervical cancer has been achieved in high-income countries through vaccination against HPV and cervical screening, which may detect pre-malignant carcinoma intraepithelial neoplasm (CIN) (**Fig. 41.23**).

Management and prognosis

CIN is classified into three grades, according to the thickness of the cervical epithelium involved: CIN1, CIN2, and CIN3. Removal of abnormal cells is recommended in most CIN2 and all CIN3 disease and is most commonly achieved by a colposcopic loop electrosurgical excision procedure, also known as a large loop excision of the transformation zone.

Localized disease

Patients with early-stage cancer may be treated using limited surgical techniques that preserve the function of the cervix and are compatible with future fertility. More deeply invasive disease that remains non-metastatic is still highly curable by more major resection of the uterus, ovaries, and pelvic nodes (which are the first site of metastatic disease in most cases); this is termed the 'Wertheim hysterectomy'. When disease is more advanced or surgery is contraindicated for reasons of age or co-morbidity, pelvic radiotherapy (usually with concomitant chemotherapy with cisplatin and a transvaginal radiation brachytherapy boost to the cervix) is advised; this confers a similarly high chance of cure.

Metastatic disease

There are no curable options, but platin-based chemotherapy, often together with a taxane, with or without a VEGF Mab (bevacizumab), is recommended for fitter patients and delays progression for some time in many patients. In addition, immunotherapy can also be added in to improve disease control.

Endometrial cancer

Clinical features are discussed in Chapter 47. Endometrial cancers have been broadly categorized into two types: type I is the more

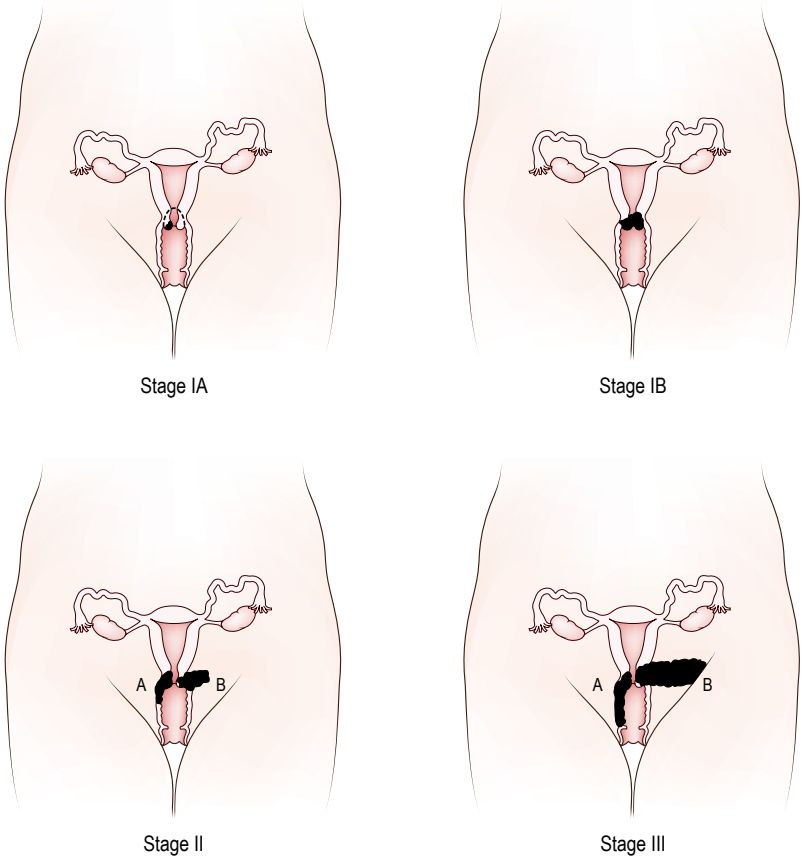


Fig. 41.23 Staging of cervical cancer: findings an examiner may appreciate on bimanual and rectovaginal examination. Stage IA: only microscopic cancer present. Stage IB: cancer is confined to the cervix. Stage II: the cancer extends beyond the cervix to involve nearby structures such as the vagina but is still confined to the pelvis. Stage III: the tumour extends to the lower part of the vagina or reaches the pelvic side wall. The *dotted line* in the stage IA diagram illustrates the line of resection for a cold-knife cone biopsy. The A and B substages for stages II and III disease are represented on the patient's right and left sides, respectively. (*From* Colposcopy Principles and Practice: An Integrated Textbook and Atlas, *Second Edition. pii:B9781416034056100149/gr13.*)

commonly seen endometrioid adenocarcinoma (80%–90%), and type II are non-endometrioid subtypes such as serous, clear-cell, and undifferentiated carcinomas (10%–20%).

Early disease

When the cancer is diagnosed at an early stage and is confined to the uterus, hysterectomy and bilateral salpingo-oophorectomy is the recommended treatment. The 5-year survival rate exceeds 90%.

Advanced and metastatic disease

When disease is more advanced, postoperative pelvic radiotherapy lowers the risk of local recurrence when the cancer has penetrated deep into the myometrium. For those with disease outside the uterus, chemotherapy (typically a doublet regimen including a platinum agent) may delay progression for some time. Progesterones have a small suppressive role for those with advanced disease and those who are too frail to benefit from systemic chemotherapy.

Ovarian cancer

Ovarian cancer arising from the surface epithelial cells ('epithelial ovarian carcinoma') is discussed in this section; other rarer cancers arising from the ovary include stromal cell tumours and germ cell tumours and are discussed earlier. The clinical features of all ovarian tumours are described in Chapter 47.

Prognosis

Histological subtype (clear cell and mucinous are worse), grade and differentiation, stage, extent of residual disease after surgery (macroscopic vs microscopic or none), and performance status are all significant independent prognostic factors for survival (**Box 41.29**).

BRCA mutation status should be assessed in patients with a positive family history (**Box 41.30**). *BRCA* status is also a predictive marker for increased sensitivity to platinum and anthracycline cytotoxics and to PARP inhibitors such as olaparib.

Management

Early disease

Curative treatment inevitably requires radical surgery (laparotomy with bilateral salpingo-oophorectomy, hysterectomy, and omentectomy). Because of the high risk of peritoneal dissemination, all but patients with very early, ovary-confined cancers benefit from postoperative adjuvant chemotherapy, which is given despite complete macroscopic resection of all visible disease. Clinical trials have shown that such postoperative chemotherapy reduces the chance of subsequent relapse when the subsequent relapse risk is high.

Advanced disease

In patents presenting with widespread peritoneal disease or metastatic disease beyond the abdomen and pelvis, neoadjuvant chemotherapy is usually used first to shrink the cancer down to minimal proportions before debulking surgery is attempted. Among the chemotherapy options used for treating patients with ovarian cancer (adjuvantly or for advanced disease), the platinum group includes the best single agents, and a taxane is usually the partnering cytotoxic. A good majority of advanced cancers respond to these agents. The VEGF inhibitor bevacizumab may be usefully added in cases of advanced disease.

Box 41.29 Survival rates in invasive epithelial ovarian cancer

Stage	Relative 5-year survival rate (%)
I	89
IA	94
IB	91
IC	80
II	66
IIA	76
IIB	67
IIC	57
III	34
IIIA	45
IIIB	39
IIIC	35
IV	18

From Heintz AP, Odicino F, Maisonneuve P et al. Carcinoma of the ovary. FIGO 26th Annual Report on the Results of Treatment in Gynecological Cancer. Int J Gynaecol Obstet *2006; 95(Suppl 1):S161–192, with permission.*

Box 41.30 Familial patterns of breast and ovarian cancer with increased risk of inherited *BRCA* mutations

- A mother or sister diagnosed with breast cancer before the age of 40 years
- Two close relatives from the same side of the family diagnosed with breast cancer; at least one must be a mother, sister, or daughter
- Three close relatives diagnosed with breast cancer at any age
- A father or brother diagnosed with breast cancer at any age
- A mother or sister with breast cancer in both breasts, the first cancer diagnosed before the age of 50 years
- One close relative with ovarian cancer and one with breast cancer diagnosed at any age; at least one must be a mother, sister, or daughter
- A close relative of Ashkenazi Jewish origin with breast or ovarian cancer
- A close relative with pancreatic cancer and breast or ovarian cancer

Relapsed disease

Response to therapy may be assessed clinically, by scanning and by measurement of the useful ovarian tumour marker CA-125: a rising CA-125 suggests relapse. Drugs used for patients with relapsed disease include other cytotoxics, but new agents are also making an impact on therapy. The realization that ovarian cancer is driven by DNA repair faults and that the protein PARP1 is significant in this repair has led to the use of PARP inhibitors to treat patients with advanced ovarian cancer, particularly those with cancers that have a known DNA repair deficit such as *BRCA* or *PALB2* mutations. PARP inhibitors used in combination with a DNA 'toxin' that causes strand breaks, such as platin chemotherapy, achieves a 'synthetic lethality', a new concept in cancer therapy that made its first inroads in ovarian cancer. It is also predicted that immunotherapy will have a role for patients with advanced ovarian cancer.

Palliative approaches

Up to 30% of those with metastatic disease may be alive after 5 years, although this falls to 5% to 10% if it is not possible to debulk the cancer at operation or if it has spread outside the peritoneal cavity. Bulky disease in the peritoneum commonly causes progressive

bowel obstruction, which may require palliative surgery or expert palliative care support to manage the terminal phases of the illness.

Further reading

Kuroki L, Guntupalli SR. Treatment of epithelial ovarian cancer. *BMJ* 2020; 371:m3773.

BRAIN TUMOURS

The most common intracranial tumours are brain metastases, which can be solitary or multifocal. Primary brain tumours account for 3% of total cancer cases in the UK. The clinical presentation and initial investigation of brain tumours are discussed in Chapter 28 (see 'Brain tumours' section) (**Fig. 41.24**).

Brain metastases

Any cancer can metastasize to the brain, but the most common tumours to do so are those arising from the following:

- Lung
- Breast
- Skin (melanoma)
- Kidney

Metastases may occur anywhere in the brain and may be single or multiple.

Cerebral oedema commonly surrounds brain metastases; this oedema usually responds to steroid therapy, which is often the first measure to ameliorate the symptoms.

Many systemic therapies for the primary tumour do not cross the blood–brain barrier and therefore do not control brain metastases. In recent years, with better systemic control of cancers, brain metastases have become an increasingly important threat to life as a site of uncontrolled disease. There are various treatment approaches. For large metastases causing raised intracranial pressure, surgery is considered with urgency. An alternative, particularly for smaller or multiple metastases and where the metastases are not abutting radiosensitive structures (e.g., the optic chiasm, brainstem), focal radiation therapy (stereotactic radiation therapy) is commonly used for therapy, including Gamma Knife, CyberKnife. and linear accelerator–based therapy. In these techniques, focused high-dose radiotherapy is delivered to the small-volume metastases within the brain via multiple pencil-beam radiation portals from three dimensions, coinciding at the target. Hence, a very high obliterative dose is deposited on each metastasis with a steep external dose gradient such that the radiation dose 'falls off' very quickly at the edge of the targeted metastasis, thus sparing adjacent normal brain. There is still a role for whole-brain radiotherapy when there are large numbers of brain metastases or they are too large to treat with stereotactic radiotherapy. Occasionally, hydrocephalus causes raised intracranial pressure, and a shunt may need to be inserted before radiotherapy commences (particularly because the radiation may transiently worsen cerebral oedema notwithstanding anti-oedema therapy with high-dose steroids).

An increasing number of systemic therapies are active against brain metastases. Immunotherapy (checkpoint inhibitors) is often active against melanoma brain metastases, and osimertinib (a TKI active against lung cancer brain metastases driven by an activating mutation of the *EGFR* gene) has considerable success in treating the metastases in this subtype of lung cancer.

Primary brain tumours

Glioma

Gliomata arise from astrocytes, ependymal cells, or oligodendroglial cells. They can be low or high grade; this has an impact on prognosis, although low-grade tumours nearly always transform into high-grade tumours eventually. Survival is very limited for patients with high-grade glioma, and the worst prognosis of all is for the

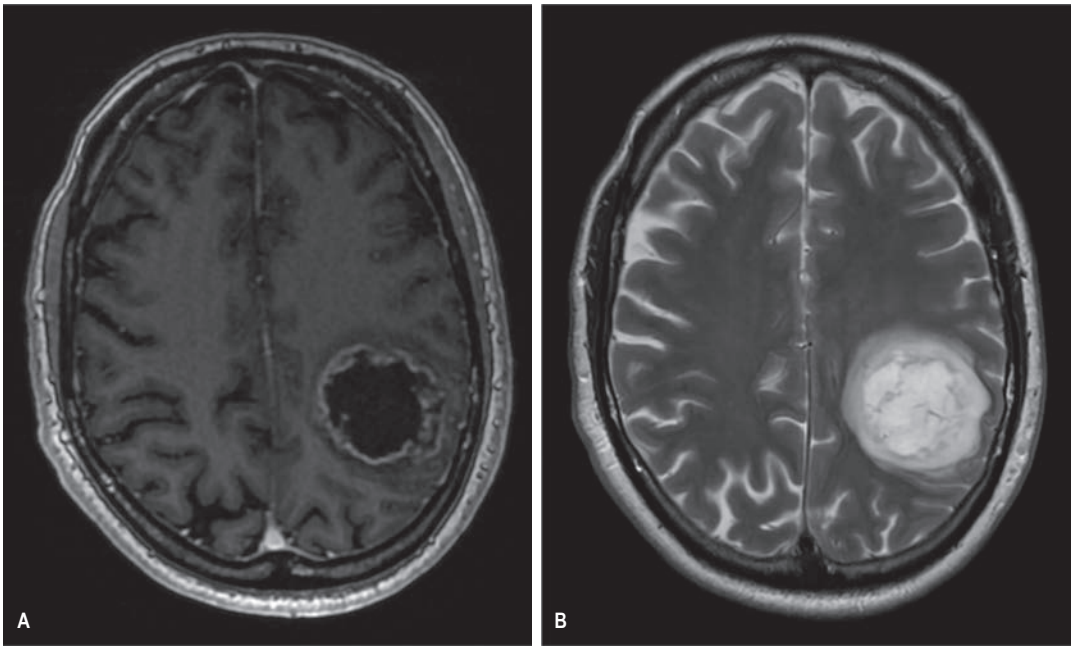


Fig. 41.24 **(A)** A T1-weighted, contrast enhanced magnetic resonance imaging scan with a large tumour in the left fronto-parietal lobe. **(B)** The corresponding T2-weighted sequence. Images are typical of a primary glioblastoma.

glioblastoma (grade 4) population of patients in whom survival (even after treatment with surgery, radiotherapy, and chemotherapy) is rarely beyond 2 years. However, the survival of patients with low-grade glioma may be well into the second decade from diagnosis.

Knowledge of molecular pathology can inform prognosis and predict response to treatment and may direct therapy in the future:

- ***Mutations in isocitrate dehydrogenase 1 or 2 (IDH1 or IDH2)*** indicate that a tumour may be slower growing. If they are found in a patient with a high-grade tumour, it may indicate that the tumour has transformed from a low-grade one rather than arising *de novo.*
- ***Loss of heterozygosity of* 1p *and* 19q *genes*** is associated with oligodendrogliomas and a better prognosis and response to treatment.
- ***Promoter methylation of the O6-methylguanine-DNA methyl-transferase* (MGMT) *gene*** causes inactivation of the gene and predicts a better response to temozolomide chemotherapy, conferring a better prognosis.

Management

The management of gliomas depends on the grade, the molecular pathology (if known), the genomics and epigenetics of the tumour, and the tumour site; the patient's performance status and their wishes (regarding the risks they are willing to accept) are also critical. Options include:

- ***Surveillance.*** Low-grade tumours were traditionally followed up with serial imaging alone, but a more proactive approach with earlier intervention is now also considered an option.
- ***Surgery.*** Craniotomy and maximum debulking surgery are carried out, if possible. This depends on the site of the glioma because it is not possible to resect a tumour in an area of the brain that would cause significant morbidity. Because gliomas spread along white matter tracts, full resection of a high-grade glioma is never possible. A stereotactic biopsy via a burr hole to inform diagnosis may be a preferable option to debulking surgery for critically situated tumours.
- ***Adjuvant treatment.*** Treatment after surgery can include radiotherapy with or without chemotherapy. The chemotherapy must cross the blood–brain barrier (e.g., temozolomide, an oral alkylating agent, which is commonly used both during radiotherapy in the case of high-grade tumours and after radiotherapy).

At relapse, further surgery, radiotherapy, or chemotherapy (temozolomide or procarbazine, lomustine and vincristine [PCV]) may all play a part. The role of immunotherapy in the treatment of patients with gliomas is still under investigation, with vaccine trials and other immunotherapy methods the subject of current trials. Targeted therapies include agents targeting *IDH1* (e.g., vorasidenib). Indeed, a recent trial of prescribing vorasidenib to patients with low-grade *IDH*-mutant gliomas has shown positive results in extending progression-free survival. PARP inhibitors may have a future role in therapy. More extensive genomics will be of always interest in the search of other targeted therapies and *BRAF, EGFR, H3,* and *VHL/HIF* are genes of interest in intracranial tumours; the quest is for inhibitors of actionable driving mutations.

Anti-angiogenic agents may also have a role in the relapse setting, particularly when there is a lot of widespread oedema resistant to steroids, but have not been shown to improve survival.

Meningioma

Meningioma arises from the arachnoid fibroblast and is usually a benign tumour with a slow growth rate, but it can cause damage because of pressure effects. Most are curable by surgical excision, Although the majority of meningiomata are low grade and as curable as just stated, not all behave in a 'benign' manner because they can cause significant disability to the patient, for instance, by causing mass effect, eroding adjacent structures, or obstructing the flow of CSF. Indeed, the rare grade 3 meningioma is a malignant cancer of sarcoma type. Methylation of meningioma tumour DNA also affects grading; this epigenetic phenomenon (as in other tumours) alters the growth of the tumour.

Surgery is the definitive treatment but may not be feasible because of the sensitive location of the tumour. There is a role for fractionated external-beam radiotherapy and stereotactic radiosurgery in the management of inoperable, recurrent, and residual meningiomas or in lieu of surgery. Meningiomas often carry hormone receptors, particularly PRs, and grow at an accelerated rate in pregnancy, and anti-hormonal drugs may be useful at retarding growth.

Neurofibroma and schwannoma

These benign tumours occur with different frequencies; the eighth or vestibular nerve (particularly in neurofibromatosis type 2) is the commonest site of origin, and patients present with ipsilateral hearing loss and often tinnitus. Neurofibromata arising from the other cranial nerves are less common Although surgery is curative if complete excision can be achieved, stereotactic radiosurgery is commonly used to treat patients with these tumours and prevent hearing loss.

HEAD AND NECK CANCER

The term 'head and neck cancer' generally describes SCCs arising from the mucosa lining the aerodigestive tract. See Chapter 21 for details on how these tumours may present and be diagnosed. Thyroid cancer is discussed in Chapter 34 (see 'Thyroid cancer' section).

Staging

Squamous head and neck cancer describes a group of cancers that spread 'logically': first to regional lymph nodes (the horizontal group—occipital, post-auricular, pre-auricular, facial, submandibular, and submental, or the vertical deep cervical chain) before spreading elsewhere, often to the lung. A considerable time frequently elapses before disease spreads beyond the primary site and nodes. This provides a greater chance for cure in the form of local and regional therapy, such as surgical resection with local radiotherapy. Most cases of early head and neck cancer are curable (Fig. 41.25).

Biopsy-proven cancer requires staging using the TNM system. Whereas PET scanning is often useful for staging distant disease, MRI is best for obtaining anatomical detail of local spread and depth of invasion. Viral status (HPV or EBV) may have implications for therapy. Genomic analysis and predictive tests for immunotherapy have roles in relapsed patients.

Management

Early disease

When early cases are easily operable with margins clear of disease, surgery is usually curative. Elective dissection of the deep cervical chain is sometimes indicated to treat 'one step ahead of the cancer's spread', that is, resecting the next most likely site to harbour microscopic disease. When surgical resection of the primary may lead to great morbidity (e.g., laryngectomy for early laryngeal

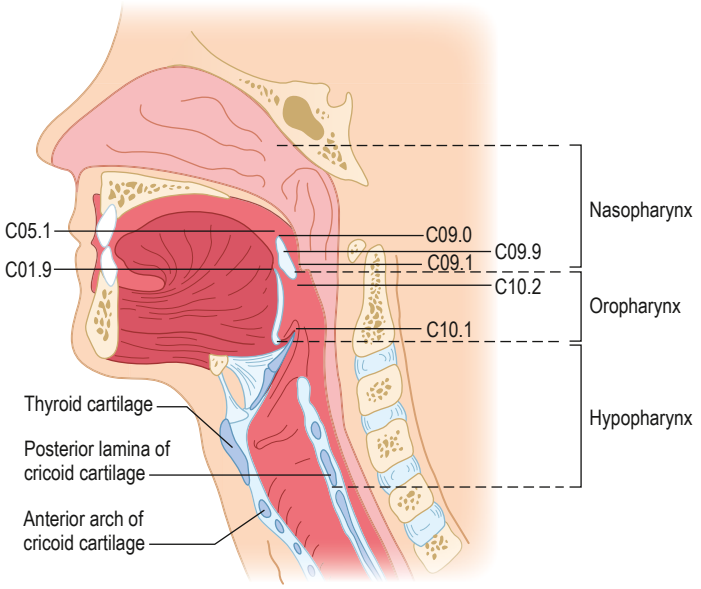


Fig. 41.25 Anatomy of head and neck cancers, showing subdivisions into distinct anatomical areas (nasopharynx, oropharynx, and hypopharynx). More than 95% of head and neck cancers arise from the mucosal surfaces of the upper aerodigestive tract. Specific anatomical sites are associated with unique aetiologies, including environmental and viral risk factors, and underpin modern staging systems and treatment paradigms. *(From* Goldman-Cecil Medicine, *Twenty Seventh Edition B9780323930383001763/f176-01-9780323930383.)*

cancer), radiotherapy (to which this type of cancer is sensitive) may have an equivalent chance of cure and be the preferred treatment choice. When the cancer is invading cartilage or bone, it is considered to be more radiotherapy resistant, and surgery may be the better option.

Locally advanced disease

In more locally advanced cases for which radiotherapy is chosen, chemotherapy (usually cisplatin) in conjunction with the radiotherapy enhances the radiation effect. The EGFR inhibitor cetuximab can be used to inhibit signal transduction within cancer cells, again in conjunction with radiotherapy, although this is generally reserved for patients unsuitable for chemoradiotherapy.

Metastatic disease

The lungs and bones are the most common sites for spread after local lymph node groups. Immunotherapy agents, such as the PD-1 inhibitor pembrolizumab, may have significant efficacy and can induce remarkable remissions in some patients, presumably because of the high mutational load and PDL-1 expression of these squamous cancers. The use of pembrolizumab is licensed as first-line therapy for metastatic head and neck cancers provided PDL-1 expression is positive. Chemotherapy agents such as 5-FU and cisplatin are reserved as second-line.

MALIGNANT MELANOMA

Management

In the UK, all people with suspected melanoma should be referred urgently (2-week wait pathway) to the local hospital's multidisciplinary skin cancer team. Clinical features and diagnosis are discussed in Chapter 45.

Surgery remains the definitive treatment for patients with primary melanoma, and early wide excision of thin tumours is usually curative. Suspicious lesions should be completely and promptly excised with a narrow (2-mm) margin for review by a dermatopathologist experienced in melanoma diagnosis. The pathological staging of the primary melanoma depends on the Breslow thickness of the melanoma (i.e., the depth of dermal invasion) and whether the tumour is ulcerated (Fig. 41.26).

Staging

The eighth ***American Joint Committee on Cancer (AJCC) melanoma staging system***, published in 2017, defines four broad categories of tumour, T1 to T4, ranging from T1a (<0.8 mm thickness without ulceration) to T4b (>4.0 mm with ulceration). These categories guide the margins of the subsequent wide local excision and the likely benefit from a sentinel lymph node biopsy (SLNB) to detect occult metastases in the adjacent lymph nodes. The latest National Institute for Health and Care Excellence guidelines recommend that SLNB should be offered to patients with stage 1B to 2C melanoma with a Breslow thickness of more than 1 mm. Patients with clinically evident lymphadenopathy or positive SLNB are further evaluated with PET/CT to detect distant metastases.

Treatment for patients with metastatic melanoma (stages 3 and 4) has been revolutionized in the past few years with the advent of targeted therapy. This includes the oral TKIs vemurafenib and dabrafenib and the MEK inhibitor trametinib for patients whose melanomas demonstrate the *BRAF V600* mutation. They are more effective than conventional chemotherapy when given in combination, but toxicity and tumour resistance are limitations. Immunotherapy with immune checkpoint-blocking agents, including CTLA4 antibody (ipilimumab) and PD-1 antibody (pembrolizumab, nivolumab), has led to significant improvements in survival for patients with advanced melanoma with tolerable toxicity.

Fig. 41.26 Typical appearance of a malignant melanoma on the upper limb. *(From* Medical Assistant: Integumentary, Sensory Systems, Patient Care and Communication—Module A. *B9780323242554000011/f01-25-9780323242554.)*

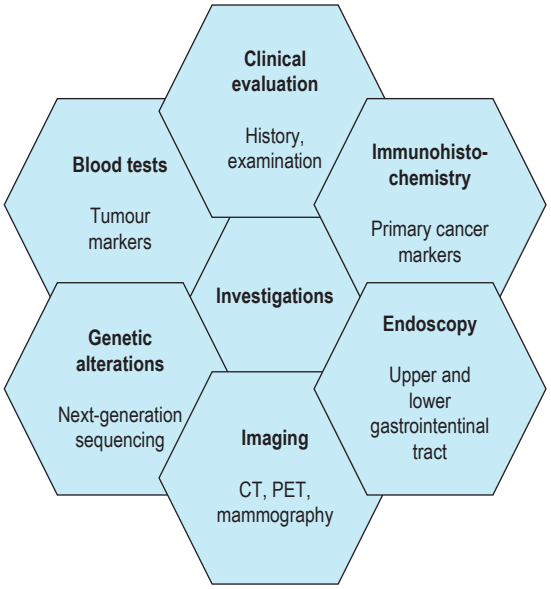


Fig. 41.27 Numerous tests may be indicated for patients presenting with a carcinoma of unknown primary. A non-exhaustive selection of these investigations is shown. *CT,* Computed tomography; *PET,* positron emission tomography.

METASTATIC CANCER OF UNKNOWN PRIMARY ORIGIN

Patients presenting with symptoms of metastases or with an incidental finding on imaging without a clinically obvious primary cancer after investigation represent a well-recognized clinical problem and make up 5% to 10% of patients in a specialist oncological centre. Clinical guidelines are based on several systematic studies, some with postmortem follow-up. Patients predicted to have a poor prognosis (based on performance status, histology, and site and extent of incurable disease) can be spared the discomfort of intensive investigation and be given appropriate palliative care (Fig. 41.27).

Diagnosis and investigation

The diagnosis requires histology first and foremost because it will lead to the identification of several distinct groups. If clinical examination does not give an obvious clue to the probable primary origin of the cancer, then a whole-body scan (usually FDG-PET body scan) is performed and usually gives the clue as to the site of origin. A biopsy follows for histological definition, whereafter staging and appropriate therapy is instituted.

- ***Squamous cancers*** may present with neck lymphadenopathy; 80% are associated with an occult head and neck primary, the majority remainder arising from the lung. Inguinal nodes usually point to a primary of the genital tract or anal canal. Treatment with radiotherapy and chemotherapy may have curative potential, especially in the head and neck area and even in the absence of an identifiable primary on pan-endoscopy.
- ***Poorly differentiated or anaplastic cancers*** include a number of potentially curable cancers, such as high-grade lymphomas and germ cell tumours, and should be suspected in all young patients with midline masses. They are sometimes identifiable by their immunocytochemistry and tumour markers. Gene markers, such as i12p for germ cell tumours in cancer of unknown primary, and BCL-2 for lymphomas, are increasingly available to aid diagnosis. Treatment and prognosis are as outlined for lymphoma (see Chapter 39) and germ cell tumours.
- ***Adenocarcinomas*** form the majority of cases of metastatic cancer of unknown primary. Their investigation should be guided by the desire to identify the most treatable options and the knowledge that the largest proportion will have arisen from the lung or pancreas, with relatively poor treatment prospects.

Investigations

Investigations should always start with a thorough review of the histology and appropriate imaging, such as FDG-PET scan (or indeed a tissue-specific PET, e.g., PSMA PET for prostate cancer and Dotatate PET for neuroendocrine tumours). CTs of the chest, abdomen, and pelvis are anatomic scans compared with PET scans, which assess function. In males, further investigations should include serum PSA and pelvic MRI to identify prostate cancers; in females, mammography and breast MRI and ultrasound examination should assess for breast cancer; pelvic MRI is used when gynaecological cancer is suspected (Box 41.31).

Circulating tissue-specific serum markers can be helpful (e.g., PSA, thyroglobulin, CEA, Ca153, CA-125, Ca 19-9, HCG, AFP) (Box 41.32), and increasingly, tumour genetics are able to identify the primary site of origin by gene expression microarray and polymerase chain reaction. Genomics of cancers are also very important with therapeutics in mind to search for actionable mutations that are driving the cancer's growth beyond physiologic or homeostatic cell reproduction. Not infrequently, mutations of transduction pathways are associated with cancer, and *RAS, RAF, PI3K,* and *mTOR* mutations are important in this regard because there are now specific inhibitors of these oncogenic driver mutation pathways. The

Box 41.31 Adenocarcinoma of unknown primary site: where to search for the primary is of therapeutic benefit

- Breast (e.g., isolated axillary lymphadenopathy)
- Ovary (e.g., peritoneal carcinomatosis)
- Prostate (e.g., pelvic lymphadenopathy)
- Colon (e.g., liver metastases)

Box 41.32 Carcinoma of unknown primary site

Immunohistochemistry marker	Most probable (but not exclusive) tissue of origin
Epithelial membrane antigen (EMA)	Epithelial
Leucocyte common antigen (LCA)	Lymphoid
Cytokeratin 7+ 20+	Pancreas, 65%; cholangiocarcinoma, 65%
	Gastric, 40%; transitional cell, 65%; ovarian mucinous, 90%
Cytokeratin 7+ 20–	Ovarian (except mucinous), 100%; breast, 90%; lung adenocarcinoma, 90%; uterus or endometrioid, 85%; transitional cell, 35%; pancreas, adenocarcinoma, 30%; cholangiocarcinoma, 30%; thyroid, 100%; mesothelioma, 65%
Cytokeratin 7– 20+	Colorectal adenocarcinoma, 80%; gastric adenocarcinoma, 35%; Merkel cell, 70%
Cytokeratin 7– 20–	Hepatocellular, 80%; carcinoid, 80%; lung small-cell and squamous, 75%; prostate, 85%; renal adenocarcinoma, 80%; adrenal, 100%; germ cell, 95%; squamous cancer of the head and neck and oesophagus, 70%; mesothelioma, 35%
Thyroid transcription factor 1 (TTF-1)	Lung, thyroid
Thyroglobulin	Thyroid
Oestrogen receptor (ER), progesterone receptor (PR), and HER2 receptor	Breast
Prostate-specific antigen (PSA)	Prostate, breast
Carcinoembryonic antigen (CEA)	Gastrointestinal
CDX2	Colorectal, small intestinal
Villin	Gastrointestinal
CA-125 peritoneal antigen	Ovarian, fallopian tube and primary peritoneal, and breast
WT1	Ovarian serous cancer, mesothelioma, desmoplastic tumours, Wilms' tumours
S100, melanin, and HMB45	Melanoma
Myosin, desmin, and factor VIII	Soft tissue sarcoma
Chromogranin and neurone-specific enolase (NSE)	Neuroendocrine
α-Fetoprotein (AFP)	Germ cell tumours, hepatocellular carcinoma
β-Human chorionic gonadotrophin (β-hCG)	Germ cell and trophoblastic tumours
CD117	Gastrointestinal stromal tumours

predictive test for success of immunotherapy (PDL-1 expression combined positive score [CPS] score), is also relevant to the work-up of many metastatic cancers, whatever their tissue of origin.

Immunohistochemical analysis of cytokeratin (CK) proteins may discriminate the likely site of origin of a carcinoma. For example, CK20 positivity is a strong indicator of a GI origin of a cancer, whereas CK7 positivity alone suggests urothelial origin. Thyroid transcription factor-1 (TTF-1) positivity correlates with either lung or thyroid cancer origin.

Management

If investigations have not identified a primary site, then surgery may be considered for unifocal adenocarcinoma of unknown primary metastases.

In females, an isolated axillary lymph node metastasis (biopsy demonstrating the characteristics of breast cancer) should be treated as for lymph node–positive breast cancer. Cytology-positive (adenocarcinoma) ascites in females should be treated with a trial of chemotherapy, as for primary peritoneal, fallopian tube, or epithelial ovarian cancer. The serum marker CA-125 is almost always very high in these instances. The prognosis for those responding to such a therapeutic trial is similar to that in disease of known primary origin.

For patients with metastatic disease at multiple sites yet no primary organ of origin is found, no genomics suggestive of containing actionable driver mutations and negative CPS score for PDL-1 expression, then chemotherapy is offered to patients with reasonable performance status.

Such cancers are often poorly differentiated but may still display some response to chemotherapy, although the overall prognosis is poor. When there is a targetable mutation or a high predictive CPS/PDL-1 score, immunotherapy or checkpoint inhibitor therapy may be worth a trial and offer improved survival.

Further reading

Rassy E, Pavlidis N. Progress in refining the clinical management of cancer of unknown primary in the molecular era. *Nat Rev Clin Oncol* 2020; 17:541–55.

Questions for this chapter can be found on Elsevier eBooks+.

42 Infectious disease

Michael Brown, Philip Gothard, and William L. Irving

Clinical skills 1418
History 1418
Background 1421
Clinical approach to the patient with a suspected infection 1421
Diagnostic assessment 1421
Clinical examination 1427
Investigations 1428
Management of infectious diseases 1432
Epidemiology and prevention of infections 1433
Acquisition of infection 1434
Routes of transmission 1434
Classification of outbreaks 1436
Infection prevention and control 1437
Principles and mechanisms of infection 1441
Host–organism interactions, including the human microbiome 1441
Symptoms of infection and the host response 1442
Specificity of microorganisms 1442
Pathogenesis 1442
Metabolic and immunological consequences of infection 1443
Viral infections 1444
Virus infections of the skin and mucous membranes 1444
Vesicular viral rashes 1445
Maculopapular viral rashes 1449
Virus infections of the respiratory tract 1451
Upper respiratory tract infections 1451
Lower respiratory tract infections 1451
Systemic viral infections 1454
Virus infections of the nervous system 1458
Virus infections of the gastrointestinal tract 1461
Viral hepatitis 1462
Viruses and malignant disease 1462
Viral haemorrhagic fevers 1463
Antiviral drugs 1465
Transmissible spongiform encephalopathies (prion diseases) 1468
Bacterial infections 1468
Bacterial infections of the skin and soft tissues 1469
Toxin-mediated skin disease 1470
Bacterial infections of the respiratory tract 1470
Upper respiratory tract infections 1470
Lower respiratory tract infections 1472
Bacterial infections of the gastrointestinal tract 1472
Bacterial infections of the cardiovascular system 1478
Bacterial infections of the nervous system 1478
Bacterial meningitis 1478
Toxin-mediated infections 1479
Bacterial bone and joint infections 1481
Bacterial infections of the urinary tract 1481
Systemic and multisystem bacterial infections 1481
Bacterial infections seen in low- and middle-income countries 1483
Skin, soft tissue, and eye diseases 1483
Gastrointestinal infections 1487
Systemic infections 1489
Other bacterial infections 1492
Antibiotic therapies 1493
Fungal infections 1493
Systemic fungal infections 1493
Subcutaneous fungal infections 1495
Superficial fungal infections 1496
Antifungal drugs 1496
Protozoal infections 1497
Blood and tissue protozoa 1497
Malaria 1497
Trypanosomiasis 1502
Leishmaniasis 1503
Other protozoal diseases of the blood and tissues 1505
Gastrointestinal protozoa 1506
Helminth infections 1508
Nematodes 1509
Tissue-dwelling worms 1509
Human intestinal nematodes 1511
Zoonotic nematodes 1513
Trematodes 1513
Water-borne flukes 1513
Food-borne flukes 1515
Cestodes 1516
Arthropod ectoparasites 1517

CORE SKILLS AND KNOWLEDGE

Humans have evolved to have robust defences against infections, and this interplay has had a profound effect on our development. Consequently, most infections are self-limiting and either managed at home by the patient or cared for by community services. Hospital-based infection services manage patients with complicated or severe infections that require expertise from trained infection specialists in settings that vary from outpatient clinics (e.g., those that treat patients with chronic infections such as tuberculosis) to critical care units. Outpatient parenteral antimicrobial therapy (OPAT) services are now used to bridge the gap for patients who need intravenous therapy and close specialist oversight but are stable enough to be at home.

There are two main clinical infection subspecialties:

- ***Clinical microbiologists and virologists***, who predominantly provide advice on patients in the care of other doctors and manage the microbiology laboratory
- ***Infectious disease specialists***, who are often trained as general physicians and treat patients directly under their care or provide a 'liaison infection service' to other teams.

These differences are increasingly blurred because most UK trainees now gain skills in both bench and bedside medicine. A small number of tropical medicine specialists are concentrated in tertiary centres, such as the Hospital for Tropical Diseases in London, and many infection specialists are research active and hold joint academic appointments in the UK and overseas.

Key skills to master in infectious diseases include:

- Taking a comprehensive history from patients to draw up a thorough account of their presentation and risk factors
- Understanding the geographical distribution of different types of infectious agent and their varied clinical presentations
- Being familiar with the role of laboratory-based experts in interpreting complex microbiological investigations.

The best way to learn about infectious diseases is by following patients as the stories of their illnesses unfold. Assess patients in the clinical environment where they present (often the emergency department), review available and subsequent investigations, and discuss them with the healthcare professionals caring for these individuals. Reflect on whether the initial assessment, investigations, and management were appropriate in light of any problems that emerged (e.g., adverse drug reactions or the need to intensify therapy) and review the subsequent processes of care (e.g., de-escalation of antibiotic therapy) and the final infection diagnosis.

CLINICAL SKILLS

History

Many patients with infectious diseases present with symptoms that are obviously suggestive of a particular type of infection. Others do not and may present with non-specific symptoms (fever, malaise, vomiting, or a rash). In addition, various types of immunodeficiency alter the way that infections present in these groups of patients. Taking a full history is critical in reaching a diagnosis (**Box 42.1**). A detailed description of how to ask about different risk factors is given later in this chapter. As put simply by Sir Eldryd Parry, a pioneer of training in infectious diseases, *'why has this person, from this place, presented in this way, at this time?'* The figure a couple pages away shows pathogens commonly affecting different organ systems, allowing a move from a clinical syndrome to a likely microbiological diagnosis.

Box 42.1 Taking a history in infectious diseases

Patient

- Age
  - Children and very old adults are more susceptible than young adults to some types of infectious disease.
- Sex
  - Epididymo-orchitis or prostatitis in males.
  - Pelvic inflammatory disease or tampon-related toxic shock syndrome in females.
- Occupation
  - Contact with animals raises the possibility of zoonotic infections.
  - Healthcare workers may acquire infectious diseases from their patients.
- Co-morbidity and immunosuppression
  - Some infectious diseases (e.g., bacterial meningitis or pneumococcal pneumonia) can affect even previously fit and healthy individuals.
  - Most infectious diseases are more common in immunosuppressed individuals, although the presentation might be blunted or less severe.
  - Certain co-morbidities raise the risk of specific types of infection (e.g., incomplete bladder emptying in neurological disease raises the risk of urinary tract infection).
  - Significant immunosuppression (e.g., in organ transplant recipients or patients with significant bone marrow disease) increases the risk of infection by a wide range of atypical (opportunistic) pathogens, including fungi, low-pathogenicity viruses and non-tuberculous mycobacteria.
- Socio-economic status
  - Malnutrition, poor sanitation, homelessness, and overcrowding increase the risk of a number of 'diseases of poverty'.
- Activities
  - People who inject drugs are at risk of blood-borne viruses and bacterial bloodstream infection.
  - Unprotected sex presents the risk of developing body-fluid-borne viruses and various sexually transmitted infections (see Chapter 43).
  - Freshwater swimming is a risk factor for leptospirosis and (in certain regions) schistosomiasis.
  - Domestic pets can transmit disease, including cat-scratch fever or toxoplasmosis.
- Close contacts
  - Diseases that occur in outbreaks (e.g., food poisoning) may also present in household or other close contacts.
- Uncommon exposures
  - When the diagnosis is unclear, taking time to revisit the history and ask about any unusual activities or exposures is sometimes helpful. Activities that may not be common in a particular setting (e.g., caving, cleaning aquariums, art restoration, repairing air-conditioning systems) all carry risks for developing particular infectious diseases. Unless you ask, you may never find out.

Place

- A detailed travel history is essential
  - Travel to tropical and subtropical countries may be a risk factor for a wide range of infectious diseases, including malaria, dengue, and typhoid.
  - Chronic infectious diseases, such as tuberculosis and viral hepatitis, may have been acquired during a childhood spent in endemic regions and then present in adulthood.
  - Recent geographically localized epidemics, such as Middle Eastern respiratory syndrome or Ebola (Central Africa), should be borne in mind.
- Contact with healthcare settings
  - Nursing or residential home residents are at risk of developing infectious diseases that are usually associated with hospital inpatients (e.g., 'hospital-acquired pneumonia' or *Clostridioides difficile* infection).
- There are significant global differences in antibiotic resistance rates, so patients who have been inpatients in or travelled to other countries may present with multi-resistant organisms not usually seen in their country of residence.

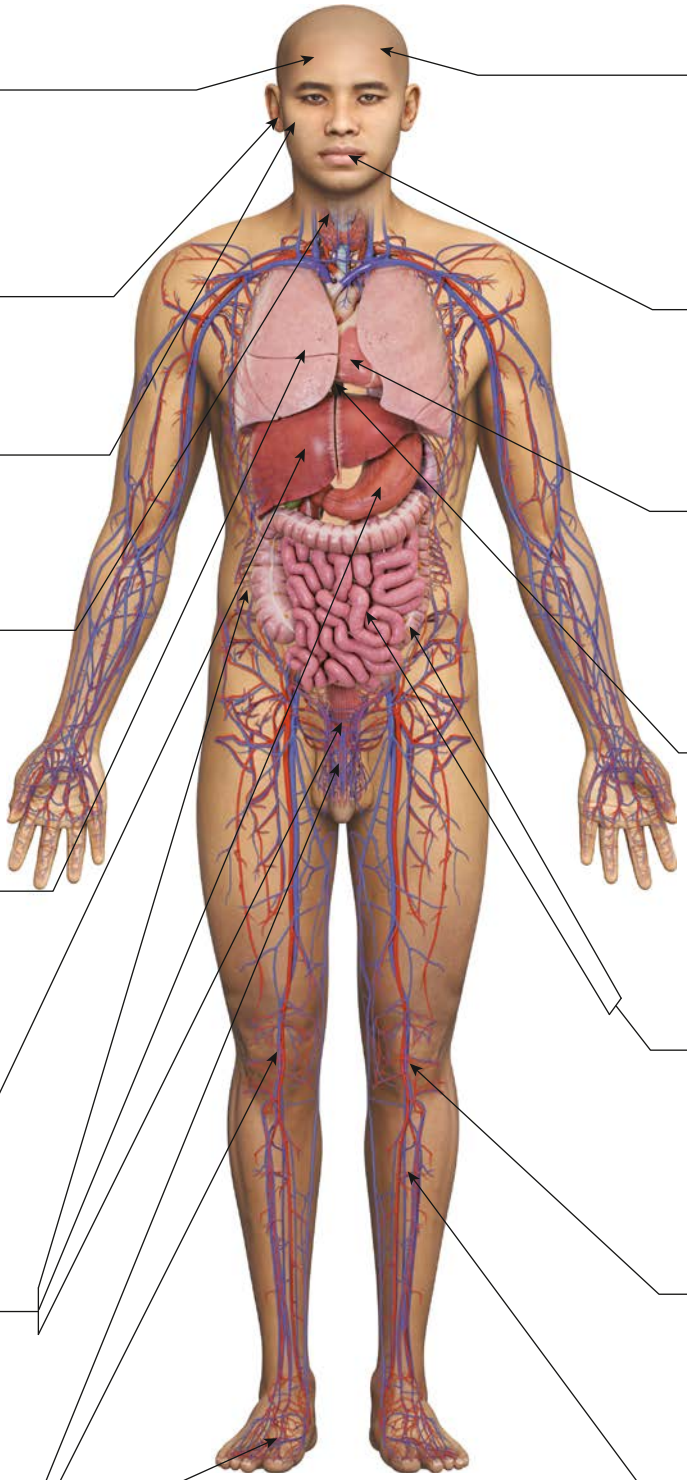
Time

- It is often helpful to produce a timeline describing the development (and perhaps resolution) of different symptoms over time.
- Symptoms of many acute viral and bacterial infectious diseases emerge over several days.
- Symptom duration of more than 2 weeks suggests chronic infection (e.g., tuberculosis, leishmaniasis, brucellosis, human immunodeficiency virus) or a non-infectious (e.g., autoimmune or malignant) disease process.
- An abrupt onset of severe symptoms suggests a highly virulent organism and may imply the need for urgent and empirical treatment before a diagnosis is made.
- Some diseases have characteristic time courses (e.g., the biphasic febrile response in dengue, or the initial 'Katayama fever' followed by haematuria seen in *Schistosoma haematobium* infection).

Symptoms

- Producing a clear account of the different general and organ-specific symptoms experienced by the patient is crucial.
- Sepsis is a physiological state that must not be missed because it always requires empirical antibiotic therapy (see Chapter 12).
- Some symptoms suggest the involvement of a particular anatomical site or organ system, such as localized pain, jaundice, or the bloody diarrhoea of dysentery.
- Other symptoms, such as headache and vomiting or watery diarrhoea, often occur non-specifically in various infectious and non-infectious diseases and should not be presumed to imply neurological or gastrointestinal disease, respectively.
- Some infectious diseases have unique pathological features that, if properly described, are highly suggestive of a particular diagnosis (e.g., the 'erythema migrans' rash of Lyme disease, the progressive phases of typhoid).

Common pathogens affecting different sites of the body



Sinuses/nose
Strep. pneumoniae
H. influenzae
Staphylococci (MSSA/MRSA)
Rhinoviruses
Polymicrobial infection
RSV
Other viruses

Outer ear
Staphylococci
Streptococci
Pseudomonas

Inner ear
Strep. pneumoniae
H. influenzae

Throat
Candida
Strep. pyogenes
MRSA (colonization)
Viruses, e.g., adenovirus
Anaerobic bacteria

Lungs
Strep. pneumoniae
H. influenzae
Staph. aureus
Gram-negative organisms
Tuberculosis
SARS-CoV-19

Liver
Viral hepatitis
Amoebiasis
Schistosomiasis
Strep. milleri

Kidney/bladder
Gram-negative organisms (from gut), especially *E. coli*

Cannula sites / skin
Staphylococci (MSSA/MRSA)
Streptococci
Candida

CNS
Meningococcus
Strep. pneumoniae
H. influenzae
Viruses, e.g., enterovirus, Herpes simplex
Listeria monocytogenes
Mycobacterium tuberculosis

Mouth
Staphylococci
Streptococci

Heart valves (endocarditis)
α-Haemolytic streptococci
Staph. aureus (e.g., right-sided valves in i.v. drug users)
Yeasts

Bloodstream infection (bacteraemia)
Staph. aureus
E. coli
Meningococcus
Strep. pneumoniae

Large and small intestine
Gram-negative organisms, e.g., *E. coli, Klebsiella*
Campylobacter
Salmonella
C. difficile
Bacteroides / anaerobes

Joints
Staphylococci
Streptococci
Neisseria gonorrhoeae
Haemophilus in children
Kingella in young children (≤5 years old)

Bones
Staphylococci
Salmonella
Streptococci
Gram-negatives

BACKGROUND

'Infection' is defined as the process of pathogenic organisms invading and multiplying in or on a host's tissues. The term should be reserved for situations in which this results in harm, or 'disease', rather than when a microorganism simply colonizes the host without ill effect. In some cases, pathogenic organisms have a dormant phase when they do not cause harm, for example, with latent TB or some herpes viruses. In fact, most microbes cause no harm, and the relationship between colonizing microorganisms and human health is increasingly being recognized in studies of the human microbiome. Some microbes, however (including some colonizing bacteria such as *Staphylococcus aureus*) have the potential to cause disease in the right circumstances.

Infectious agents

Novel infectious diseases continue to emerge around the world, and known diseases sometimes cause outbreaks in new areas (**Fig. 42.1**), making a patient's travel history a crucial risk factor in infectious disease. Up-to-date information on new outbreaks is maintained by the International Society for Infectious Diseases at http://promedmail.org.

The causative agents of infectious diseases can be divided into four groups:

Prions are the simplest infectious agents, consisting of a single protein molecule. They contain no nucleic acid and therefore no genetic information; their ability to propagate within a host relies on inducing the conversion of endogenous prion protein PrP^{c} into an abnormal protease-resistant isoform referred to as PrP^{Sc} or PrP^{Res}. Prion diseases are described in the 'Transmissible spongiform encephalopathies (prion diseases)' section.

Viruses contain both protein and nucleic acid, and so carry the genetic information for their own reproduction. However, they lack the apparatus to replicate autonomously, relying instead on 'hijacking' the cellular machinery of the host. They are small (usually <300 nm in diameter), and each virus possesses only one type of nucleic acid (either RNA or DNA).

Bacteria are usually larger than viruses. They are enclosed by a cell membrane and most have a cell wall (although some, e.g., *Mycoplasma* spp. have only a cell membrane, making them naturally resistant to cell-wall-acting antibacterial agents such as β-lactams).

Eukaryotes are the most sophisticated infectious organisms and include unicellular protozoa, fungi (which can be unicellular or filamentous) and multicellular parasitic worms. Other higher classes, such as the insects and the arachnids, also contain species that can envenom, injure or parasitise humans and cause disease.

Further reading

https://www.bhiva.org
https://www.rstmh.org
https://www.britishinfection.org
https://bsac.org.uk
https://www.idsociety.org
https://isid.org
http://promedmail.org
https://www.escmid.org
https://www.gov.uk/government/organisations/uk-health-security-agency
https://www.jrcptb.org.uk/specialties/infectious-diseases-and-tropical-medicine
https://www.rcplondon.ac.uk/education-practice/advice/specialty-spotlight-infectious-diseases

CLINICAL APPROACH TO THE PATIENT WITH A SUSPECTED INFECTION

Fever is the cardinal feature of infection, but not all febrile illnesses are infections and not all infections present with a fever. **Fig. 42.2** provides an outline of the approach to the patient with acute (undifferentiated) fever of unclear cause. Fever is usually intermittent, so may not be present at the time of presentation, and occurs less commonly in elderly and newborn patients. Infection can also present with hypothermia (temperature <36°C), which is a poor prognostic sign. The biological behaviour of the pathogen and the consequent host responses are responsible for the clinical expression of disease, which sometimes allows for clinical recognition. Knowledge of the incubation period following exposure can be helpful (e.g., up to 21 days for chickenpox). The site and distribution of a rash may be diagnostic (e.g., shingles), while symptoms of cough, sputum and pleuritic chest pain may suggest lobar pneumonia. Fever and neck stiffness characterize meningitis but do not inform about the microbial cause or management.

The three critical points in the assessment and management of a patient with fever, suspected infection or sepsis are:

- ***Physiological assessment*** (severity of illness), predominantly based on systematic examination of key physiological markers at initial presentation and subsequently.
- ***Diagnostic assessment***, predominantly based on history-taking, examination and interpretation of those diagnostic tests available at presentation and subsequently performed in the context of the history and examination.
- ***Management assessment***, based on linking physiological and diagnostic assessments to key interventions. Decisions will be based on the seriousness of the underlying infection and the likely cause. Interventions may include:
  - The use of antimicrobials
  - Intravenous fluids or oxygen
  - Infection source control (e.g., drainage of an abscess)
  - Infection prevention and control measures
  - Use of advanced physiological support in a critical care setting.

A detailed history can be taken in a physiologically stable patient, but may need to be delayed until after a patient is stabilized in the presence of sepsis, or taken after isolating the patient if a highly transmissible infection is suspected. The recognition and management of sepsis are considered in Chapter 12.

Diagnostic assessment

History-taking, examination and investigations aim to identify the site(s) of infection and also the likely causative organism(s). **Boxes 42.2 to 42.4** outline the links between symptoms, risk factors, examination findings and potential infection diagnoses. **Box 42.5** shows the different components of an infection diagnosis and how and why these are ascertained.

History

The history is the most important component of diagnostic assessment. A detailed, systematic history is taken with specific questions about the nature, timing and periodicity of symptoms (see **Box 42.1**), in particular in relation to any infection exposures and other risk factors (see **Box 42.3**). If symptoms started after the upper limit of the incubation period of any potential transmissible infection exposure identified, the diagnosis is unlikely. Consideration of the incubation

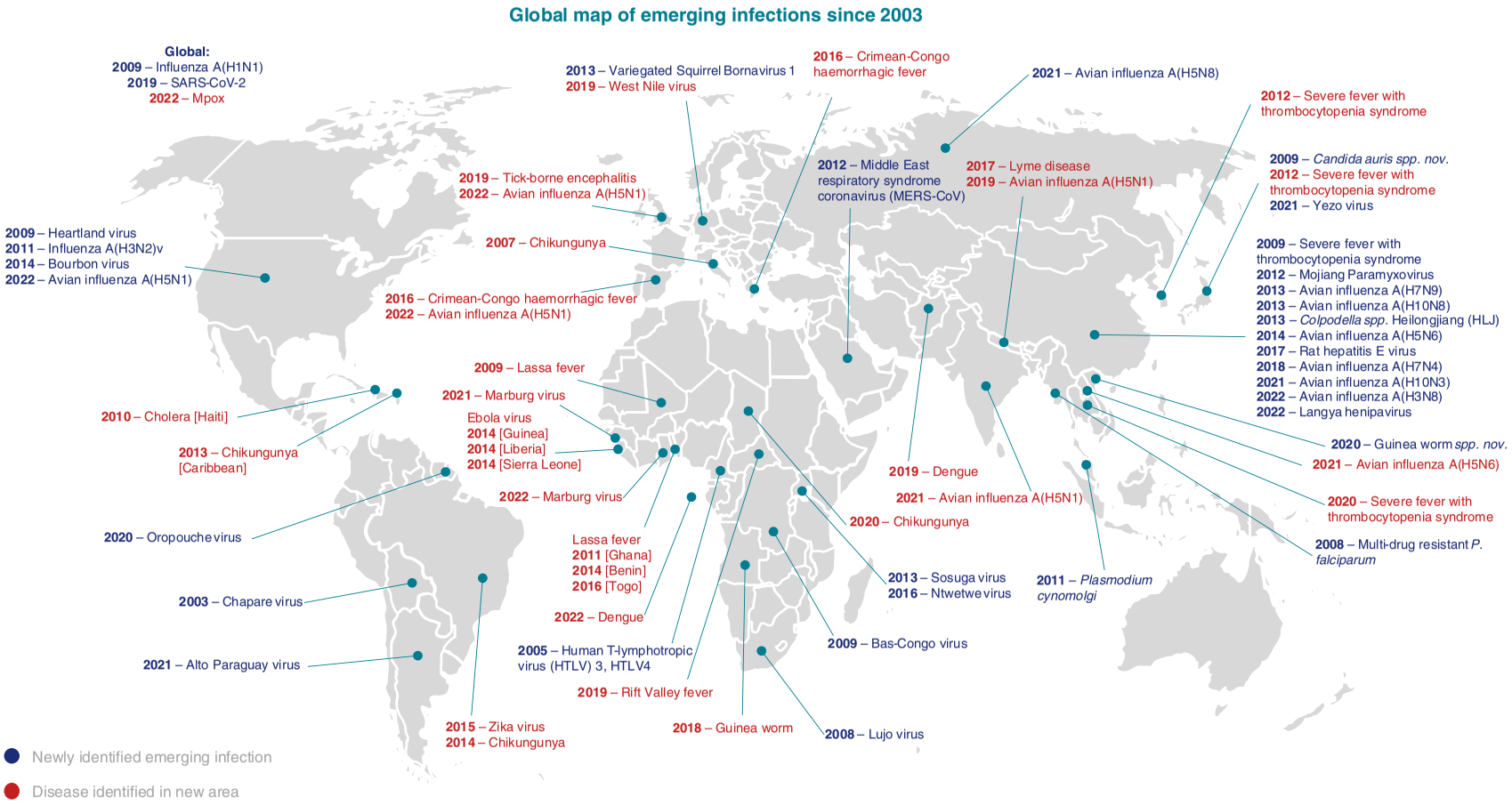


Fig. 42.1 Global map of newly identified emerging infections and notable outbreaks of diseases in humans in new areas between 2003 and 2022. Map lines delineate study areas and do not necessarily depict accepted national boundaries. *(UK Health Security Agency.)*

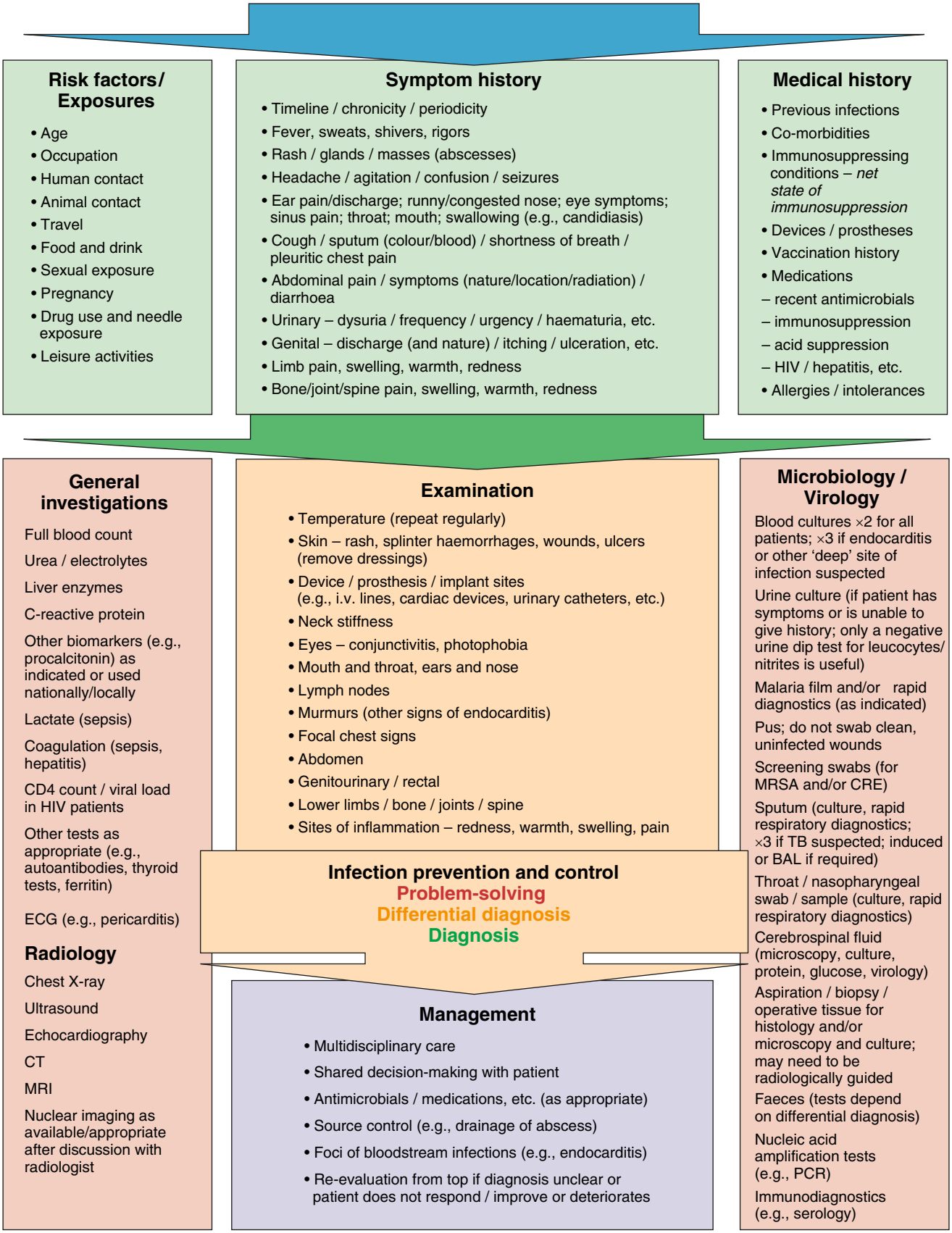


Fig. 42.2 Clinical approach to the patient with fever of unclear cause but without sepsis.
For patients with sepsis, see Chapter 8. *BAL,* Bronchoalveolar lavage; *CRE,* carbapenem-resistant Enterobacteriaceae; *CT,* computed tomography; *ECG,* electrocardiogram; *HIV,* human immunodeficiency virus; *i.v.,* intravenous; *MRE,* magnetic resonance enterography; *MRI,* magnetic resonance imaging; *MRSA,* meticillin-resistant *Staphylococcus aureus*; *PCR,* polymerase chain reaction; *TB,* tuberculosis.

Box 42.2 Linking symptoms elicited from the patient history to infection diagnoses

Element(s) in the symptom history of a patient with acute fever or possible infection	Example infection diagnoses
Acute headache, non-blanching rash, neck stiffness, photophobia	Meningococcal meningitis
Acute headache, neck stiffness, photophobia	Meningitis (bacterial, viral, fungal)
Acute agitation, confusion	Encephalitis
Ear pain with pus-like discharge	Otitis externa
Maxillary sinus area pain	Sinusitis
Sore throat, large neck glands (lymph nodes)	Tonsillitis, glandular fever syndromes
Sore, ulcerated mouth or throat with rash on the hands or feet	Hand, foot, and mouth (viral) infection; other viral infections (remember syphilis in patients with rashes on their hands or feet)
Cough, dyspnoea, pleuritic chest pain	Lower respiratory tract infection, pneumonia
Fever, shivers, rigors	Bloodstream infection (usually bacterial), malaria, viral infections (e.g., influenza), endocarditis, causes of fever or pyrexia of unknown origin (if history is ≥2 weeks); all may or may not be associated with other symptoms
Non-specific abdominal pains or cramps and diarrhoea	*Clostridioides difficile* infection, gastroenteritis
Right upper quadrant pain	Hepatitis, biliary tract infection, liver abscess, pneumonia adjacent to diaphragm
Left iliac fossa pain	Diverticulitis, left ovarian abscess
Renal angle pain with urinary symptoms	Pyelonephritis, renal abscess
Frequency, urgency, pain on micturition	Urinary tract infection
Pelvic pain	Pelvic inflammatory disease, ovarian abscess (if pain is unilateral)
Penile or vaginal discharge	Sexually transmitted infections, retained tampon (females)
Back pain	Discitis (with or without epidural abscess)
Red, swollen, hot joint	Septic arthritis (non-infection diagnoses such as gout and reactive arthritis should be considered; see Chapter 32)
Red, swollen, hot, painful limb	Cellulitis, soft tissue infection
Pain, redness, and discharge at the site of a prosthesis or invasive device	Prosthesis or invasive device infection

Box 42.3 Linking risk factors elicited from the patient history to infection diagnoses

Element(s) in the risk factor history of a patient with acute fever or possible infection	Example infection diagnoses
Alcohol abuse	Spontaneous bacterial peritonitis (if ascites), invasive pneumococcal disease (*Streptococcus pneumoniae*), tuberculosis, aspiration pneumonia
Diverticular disease	Diverticulitis, diverticular abscess
Gallstones	Biliary tract infection, pancreatitis
Immunosuppression or HIV	Opportunistic bacterial (e.g., nocardiosis), fungal (e.g., cryptococcosis) and viral infections (e.g., cytomegalovirus)
Intravenous drug use	Bloodstream infection, especially *Staphylococcus aureus*, blood-borne infections (e.g., HIV, hepatitis B and C)
Lung disease, smoker	Respiratory tract infections or pneumonia, e.g., with *Haemophilus influenzae* (chronic obstructive pulmonary disease) or *Pseudomonas aeruginosa* (in cystic fibrosis and bronchiectasis)
Prostheses or invasive devices	Prosthesis or invasive device infection
Recent abdominal surgery	Intra-abdominal infection or peritonitis, abscess(es), wound infection
Recent influenza	Secondary bacterial infection (e.g., with Lancefield group A streptococci or meningococcus)
Recent hospitalization	Healthcare-associated infections, e.g., pneumonia and urinary tract, surgical site and bloodstream infections; at higher risk of an antimicrobial-resistant pathogen
Recent new sexual partner	Chlamydia, gonorrhoea, HIV, syphilis, other sexually transmitted infections
School or university student, military service	Meningococcal disease, mumps, other infections based on local contemporaneous epidemiology
Splenectomy	Pneumococcal, *Haemophilus* spp., and meningococcal infections (vaccination and antibiotic prophylaxis reduce but do not eliminate risk)
Travel abroad	E.g., dengue, malaria, legionnaires' disease, leptospirosis, typhoid (depends on travel history)
Urinary tract problems (e.g., renal stone, benign prostatic hypertrophy)	Urinary tract infection, pyelonephritis

HIV, Human immunodeficiency virus.

period is irrelevant for many infections, however, for example, when the invading pathogen is often a colonizer of the host, such as in *S. aureus* bloodstream infections; when a specific exposure history cannot be identified; or when the patient's history and examination signs are not diagnostic.

Symptoms are often non-specific (e.g., a 3-day history of lethargy, loss of appetite, and shivers) but may lead to a diagnosis or differential diagnosis through careful interpretation and problem-solving (by linking the patient's symptoms to their exposures and risk factors and then to potential clinical infection diagnoses and causal pathogens; examination findings and any available test results also need to be considered).

In patients with sepsis, a full history should be obtained after, or at the same time as, initiating emergency medical management (see

Box 42.4 Linking examination findings to infection diagnoses

Examination finding(s) in a patient with fever or possible infection	Site of infection or diagnosis
Meningism, petechial rash	Meningitis (most likely meningococcal)
Cervical lymphadenopathy, enlarged tonsils	Tonsillitis (viral or bacterial), glandular fever syndromes (including HIV)
Tachypnoea, coarse focal crackles (can be bilateral)	Lower respiratory tract infection, pneumonia
Temperatures ≥37.8°C, new-onset regurgitant systolic murmur	Endocarditis
Right upper quadrant tenderness	Biliary tract or liver infection
Abdominal tenderness with rebound or guarding	Intra-abdominal infection (peritonitis)
Left iliac fossa tenderness	Diverticulitis or diverticular abscess, tubulo-ovarian abscess in females
Right iliac fossa tenderness	Appendicitis or appendix abscess, tubulo-ovarian abscess in females
Suprapubic tenderness(± catheter)	Urinary tract infection, pelvic inflammatory disease in females
Renal angle tenderness	Pyelonephritis, renal abscess
Red, hot, swollen limb	Cellulitis, soft tissue infection (consider necrotizing fasciitis if pain is out of keeping with examination appearances or if there is severe pain, rapidly progressing history and 'odd' examination appearance or findings)
Intravascular devices ± erythema, pus	Device infection, bloodstream infection, endocarditis

HIV, Human immunodeficiency virus.

Box 42.5 Components of the infection diagnosis and how and why these are ascertained

Diagnostic component	How?	Why?
Body organ(s) or system(s) involved	History (most important) Examination Non-microbiological investigations (e.g., WCC, CRP, CXR, other radiology)	Causal pathogens are relatively predictable (e.g., UTI vs pneumonia) Required intervention (e.g., drainage of abscess) Prognostic (e.g., mortality rate of meningitis vs cellulitis)
Microbiological or virological	Diagnostics (blood and urine culture, lumbar puncture, rapid diagnostics) Epidemiologically if diagnosis confirmed in other patients (e.g., outbreak of norovirus in a nursing home) (Clinical assessment if pathognomonic signs)	To avoid antibiotics (e.g., viral vs bacterial RTI) To target pathogen with optimal therapy Required investigation or intervention (e.g., need for TTE in *Staphylococcus aureus* BSI) Prognostic (e.g., viral meningitis vs bacterial) Epidemiology (for disease and resistance surveillance)
Prognostic (severity)	Physiology Age, co-morbidity, and so on Clinical or microbiological diagnosis (see earlier) if available	Dictates level of required intervention (i.e., sepsis vs self-limiting infection) Informs discussion with patient and relatives

BSI, Bloodstream infection; *CRP,* C-reactive protein; *CXR,* chest X-ray; *RTI,* respiratory tract infection; *TTE,* transthoracic echocardiogram; *UTI,* urinary tract infection; *WCC,* white blood cell count.

Chapter 12) to avoid therapeutic delay. If the patient is confused, the history may have to be obtained from a relative or friend. The time course of symptoms is one of the most relevant parts of the history:

- With some exceptions, patients with short, acute symptom histories are more likely to have infections caused by microbes that tend to cause acute presentations (e.g., *Streptococcus pneumoniae* infection in a patient with a 3-day history of shivers, cough with production of green sputum, and pleuritic chest pain).
- Patients with longer symptom histories are more likely to have infections caused by microbes that tend to cause chronic presentations (e.g., pulmonary TB in a patient with a 2-month history of sweats, weight loss, and cough productive of blood-stained sputum; a patient who has recently returned from living in sub-Saharan Africa).

Although this may seem obvious, it is often overlooked when it comes to requesting a battery of investigations.

Key diagnostic information in the patient's history should include the following.

Age

Older patients are more prone to some infections (e.g., *Clostridioides difficile* infection or pneumonia). Some infections (e.g., glandular fever, chickenpox) are less common in older adults but are sometimes more severe.

Sex

Some diagnoses are sex-specific, such as epididymo-orchitis, prostatitis, pelvic inflammatory disease, and tampon-related toxic shock syndrome (TSS).

Occupational history

The potential links between occupation and infection are sometimes significant. Healthcare workers may be exposed to certain infections, for example, by air-borne (e.g., chickenpox, TB), person-to-person (e.g., rotavirus), and needle-stick injury (e.g., blood-borne viruses) routes of transmission. Teachers may be exposed to the common infections of childhood. People who work with animals may be exposed to zoonoses.

Symptoms

See **Box 42.2** for examples of how symptoms can be linked to a diagnosis or differential diagnosis.

Human infection contact

Contact with others who have an infection diagnosis or symptoms of an infectious disease should be discussed. A patient with a new vesicular rash, for example, is asked about contact with chickenpox or shingles within the previous 21 days (the upper limit of the

incubation period). A patient with gastroenteritis is asked about contact with other people with similar symptoms, when that contact was with respect to the start of symptoms, and the social context of that contact. A patient with recurrent skin abscesses is asked about contact with other persons with recurrent skin infections (e.g., to assess the risk of transmission of Panton–Valentine–leukocidin (PVL) producing *S. aureus* skin infections).

Animal contact

Domestic (e.g., cats and *Toxoplasma gondii* infection), farm (e.g., sheep and *Coxiella burnetii* infection) and wild animal (e.g., bats and rabies) contact may be relevant.

Travel history

Some diseases are more prevalent in certain geographical locations. A detailed travel itinerary, including any flight stopovers, should be taken from anyone who is unwell after arriving from another country. Previous travel should also be covered because some infections may be chronic or recurrent (e.g., *Plasmodium vivax* malaria). It is necessary to find out not only which countries were visited but also the type of environment and length of stay; a prolonged stay in a remote rural village carries much higher health risks (for some conditions) than a short holiday in an air-conditioned coastal holiday resort, even in the same country. Food and water consumption, bathing and swimming, animal and insect contact, and contact with human illness need to be established. Enquiry should be made about sexual contacts, drug use, and medical treatment (especially parenteral). Patients who have spent time in hospitals elsewhere may be colonized with carbapenem-resistant organisms or *Candida auris*.

In some parts of the world, a high-proportion of professional sex workers are human immunodeficiency virus (HIV) positive, and hepatitis B and C (HBV and HCV) are common in some parts of Africa and Asia. Special tests may be needed, depending on the epidemiological risks and clinical signs, and a blood film is mandatory in anyone who is unwell up to 1 year after being in a malaria-endemic area. A reliable resource such as Travel Health Pro (see Further reading) or https://promedmail.org is consulted regarding current infectious disease alerts and endemic risks for the countries visited. Some of the more common causes of a febrile illness in returning travellers are listed in **Box 42.6**.

Food and drink history

Exposure to unclean water and uncooked, unpasteurized, or contaminated foods or drinks (e.g., *Listeria monocytogenes* infection from soft cheeses) should be asked about, and illness in other people also exposed.

Sexual activity

The sexual history is often omitted or poorly taken but is essential if diagnoses such as HIV infection are not to be missed. Sexual exposure is common in travellers, and routine testing of patients for blood-borne viruses is a simple and cost-effective solution. Sexually transmitted infections (STIs) are discussed in Chapter 42.

Pregnancy

Pregnant females may be more susceptible to some infections (e.g., malaria, HIV infection, and listeriosis) and may be at higher risk of severe infections (e.g., disseminated herpes simplex virus [HSV] infection, hepatitis E, influenza, malaria, measles).

Intravenous drug use and needle exposure

As well as blood-borne viruses (mostly HIV, HBV, and HCV), people who inject drugs are susceptible to a wide variety of bacterial and fungal infections, in particular those caused by *S. aureus* and *Streptococcus pyogenes* (group A streptococci [GAS]) as a result of contamination of the injecting equipment or inoculation from the patient's colonized skin. Abscesses (e.g., in the psoas or other muscles) and soft tissue infections at the site of injection are common, especially in the groin, and may involve adjacent vascular structures (forming pseudoaneurysms) and bones (osteomyelitis). Bloodstream and haematogenously spread infections, including endocarditis (often right-sided), spinal, and bone and joint infections, are also common. Other needle exposures, such as tattooing, body piercing, needle-stick injuries, and receipt of blood products are also risk factors for blood-borne infections, although this is much less common because the risks are better recognized and addressed.

Leisure activities

Certain pastimes (e.g., open-water swimming and leptospirosis) may predispose to water-borne infections or zoonoses.

Previous infections

Previous infections are reviewed, especially for the prior 12 months, for any pathogens isolated and associated antimicrobial susceptibilities. This is particularly useful in patients with a history of colonization or infection by meticillin-resistant *S. aureus* (MRSA) or other

? Box 42.6 Examples of causes of febrile illness in travellers returning from the tropics and other countries

The World Health Organization advises that fever occurring in a traveller 1 week or more after entering a malaria risk area and up to 3 months after departure is a medical emergency. A reliable travel health source is always consulted.

Low- and middle-income countries (mostly)

- Malaria
- Typhoid
- Leptospirosis
- Dengue
- Chikungunya
- Dysentery
- Hepatitis A
- Amoebiasis
- Schistosomiasis
- Tick typhus

Examples of specific geographical areas (see text)

- Ebola, Marburg, and Lassa fever
- Middle Eastern respiratory syndrome
- Zika
- Histoplasmosis
- Brucellosis

Worldwide

- Multidrug- and extensively drug-resistant Gram-negative infections
- Upper respiratory tract infection
- Pneumonia
- Influenza and other common viral infections
- Traveller's diarrhoea
- Tuberculosis

resistant microbes or of recurrent respiratory or urinary tract infection (UTI). Some infections, such as chickenpox, may rarely occur for a second time because of waning immunity, although chickenpox may recrudesce from latency in the form of shingles (zoster). Patients who are infected with dengue virus for a second time are at higher risk of developing severe forms of the infection.

Co-morbidities

The potential association between co-morbid illnesses and infections is considered (e.g., haematological malignancy and medication-related immune suppression are risk factors for invasive pneumococcal disease and opportunistic infections, such as *Pneumocystis jirovecii* infection).

Devices, prostheses, and implants

Symptoms relating to any recently inserted or longstanding devices, prostheses, or implants are discussed (e.g., increasing pain in a total hip replacement; swelling and warmth over a pacemaker insertion site).

Immunization history

Patients who have not received certain childhood (e.g., measles, mumps) or pre-travel (e.g., rabies) vaccines are at higher risk if exposed.

Medications

Some medications increase the risk of certain infections (e.g., proton pump inhibitors and histamine$_2$ antagonists may predispose to gastrointestinal [GI] infections, including, for example, *Salmonella* spp. gastroenteritis and *C. difficile* infection). Other drugs may be potent suppressors of immune function (e.g., ciclosporin or corticosteroids and cellular immunity; see later). The patient's recent antimicrobial exposure should also be ascertained; it is well known, for example, that individuals recently exposed to an antimicrobial are more likely subsequently to harbour microorganisms resistant to that antimicrobial (and possibly to other antimicrobials, too), and this may influence prescribing decisions.

Allergies and intolerances

The detailed nature of any prior allergies or intolerances should be determined with respect to antimicrobials or infection-related drugs (e.g., HIV medications). Most patients who report a penicillin allergy without clear recollection of events turn out not to be allergic to penicillin-based drugs. Some hospitals have introduced penicillin testing services, often run by infection specialists, to ensure a safe approach to penicillin allergy delabelling.

Infections in immunocompromised hosts

Increasing age and co-morbidity, as well as advances in medical treatment over recent decades, have led to a huge increase in the number of patients living with immunodeficiency states of varying degrees. Cancer chemotherapy; immunosuppressive drugs, including the increasing use of monoclonal antibody-based biologic therapy (e.g., infliximab); and the worldwide HIV epidemic have all contributed.

A patient's immunosuppression status is often multifactorial because of age, co-morbidity, and therapy, and an assessment of the patient's ***net state of immunosuppression*** is an important concept in determining the nature of the infections from which a patient is at risk. For example, an otherwise well 60-year-old patient taking low-dose maintenance prednisolone for polymyalgia rheumatica is considerably less immunosuppressed and hence at lower risk of infections than an 82-year-old patient with diabetes, rheumatoid arthritis, and a past history of a renal transplantation who is taking a variety of immunosuppressant and disease-modifying medications. The time course of immunosuppression is also relevant: patients within 6 months of solid organ transplantation are at higher risk of infection than those further out because this period corresponds to the most intense time of immunosuppression. Patients who do not have a spleen (or a functional spleen) are at risk of infections caused by capsulated bacteria (i.e., pneumococci, *Haemophilus* spp., and meningococci).

Clinical presentations may be atypical in immunocompromised patients, with few, if any, localizing symptoms or signs. Infection can be caused by organisms that are not usually pathogenic, including environmental bacteria and fungi (opportunistic microbes). The normal physiological responses to infection (e.g., fever, neutrophilia) may be diminished or absent. The onset of symptoms may be sudden and the course of the illness fulminant. A high index of suspicion for atypical infections in immunosuppressed individuals is required, with early investigation (e.g., computed tomography [CT] of the thorax for invasive pulmonary aspergillosis) and rapid intervention. Common causes of infection in immunosuppressed patients are shown in **Box 42.7**. For neutropenic sepsis, see Chapter 11 ('The unwell cancer patient' section).

Clinical examination

A careful and systematic examination covering all systems is required to avoid missing subtle signs, such as splinter haemorrhages in patients with endocarditis or the petechial rash of early meningococcal disease. The examination is used to support or refute likely differential diagnoses based on the history and may also help guide the approach to investigations; new differential diagnoses may also emerge based on examination findings. The fever pattern may occasionally be helpful but in most patients fever is non-specific and not diagnostic.

- Clothes and dressings should be removed to inspect the underlying skin, wounds (including historic wounds), and invasive devices (e.g., pacemaker sites).
- Skin rashes are commonly associated with infection (**Box 42.8**) and range from a diffuse, blanching maculopapular rash seen in many viral infections to the purpuric, non-blanching rash associated with meningococcal bacteraemia (**Fig. 42.3**).
- Lymphadenopathy is a common feature of infectious diseases but can also occur in patients with non-infectious conditions, such as lymphoma.
- The ears (necrotizing otitis externa), nose (common cold), eyes (Roth's spots in endocarditis), mouth (Koplik's spots in measles), and throat (tonsillitis) should all be inspected carefully.
- Auscultation of the heart for murmurs is critical in patients with fever of unknown cause, bloodstream infections, and other potential signs of endocarditis.
- Likewise, examination of the musculoskeletal system, including palpation of the spine (discitis), is performed in patients with bloodstream infection, especially those caused by *S. aureus*, and fever or pyrexia of unknown cause.
- Abdominal examination, including the renal angles (pyelonephritis; see Chapter 26, 'Urinary tract infection' section), should be performed in patients with Gram-negative bloodstream infection (e.g., that caused by *Escherichia coli*, *Klebsiella pneumoniae*, *Proteus mirabilis*) because the urinary tract (most commonly), intra-abdominal sites, and the liver and biliary tract (liver enzymes are often elevated) are the most common underlying foci of infection.
- Rectal, vaginal, and penile examinations are required in those at risk of STIs (see Chapter 43).

Box 42.4 outlines the potential links between some examination findings and diagnoses of infection.

Box 42.7 Common causes of infection in immunocompromised patients

Deficiency	Causes	Organisms (examples)
Neutropenia	Chemotherapy Myeloablative therapy Immunosuppressant drugs	*Escherichia coli* *Klebsiella pneumoniae* *Staphylococcus aureus* *Staphylococcus epidermidis* *Aspergillus* spp. *Candida* spp.
Cellular immune defects	HIV infection Lymphoma Myeloablative therapy Congenital syndromes	Respiratory syncytial virus Cytomegalovirus Epstein–Barr virus Herpes simplex and zoster *Salmonella* spp. *Mycobacterium* spp. (esp. *Mycobacterium avium* complex infections) *Cryptococcus neoformans* *Candida* spp. *Cryptosporidium parvum* *Pneumocystis jirovecii* *Toxoplasma gondii*
Humoral immune deficiencies	Congenital syndromes Chronic lymphocytic leukaemia Corticosteroids	*Haemophilus influenzae* *Streptococcus pneumoniae* Enteroviruses
Terminal complement deficiencies (C5–C9)	Congenital syndromes	*Neisseria meningitidis* *Neisseria gonorrhoeae*
Splenectomy	Surgery Trauma	*Streptococcus pneumoniae* *N. meningitidis* *H. influenzae* Malaria

HIV, Human immunodeficiency virus.

Investigations

The results of investigations are used to support, confirm, or refute likely differential diagnoses based on the history and examination findings and may also lead to new differential diagnoses and the need for further investigations.

In some infections, such as chickenpox, the clinical presentation is so distinctive that no investigations are necessary to confirm the diagnosis. In other mild and self-limiting infections, such as those frequently managed in primary and ambulatory care settings (e.g., minor skin infections, upper respiratory tract infections [URTIs], cystitis), extensive investigations will not change management and are therefore unwarranted because the diagnosis and treatment are clear.

Some patients, especially when they are being managed in secondary care or when an antibiotic prescription may be avoidable, require further tests to determine the organ, system, or site of infection and the microbiological cause to guide required intervention(s). A shared decision-making approach should be taken with the patient to avoid tests, treatments, and procedures that are unlikely to be of benefit.

General investigations

Box 42.9 shows general investigations for a patient with acute fever or suspected infection of unclear cause. It is vital to review at presentation any tests that have been performed previously, including prior microbiological tests. Treating a patient known to harbour resistant microorganisms with an antibiotic to which those organisms are

Fig. 42.3 Non-blanching urticarial rash associated with meningococcal bacteraemia and other bloodstream infections.

Box 42.8 Infections commonly associated with a rash

Macular or maculopapular
- Measles
- Rubella
- Enteroviruses
- Human herpesvirus 6
- Epstein–Barr virus
- Cytomegalovirus
- Human erythrovirus (parvovirus) B19
- Human immunodeficiency virus
- Dengue
- Typhoid
- Secondary syphilis
- Rickettsial spotted fevers

Vesicular
- Chickenpox (varicella zoster virus)
- Shingles (varicella zoster virus)
- Herpes simplex virus
- Hand, foot, and mouth disease (enteroviruses, enterovirus 71)
- Herpangina (enteroviruses)

Petechial or haemorrhagic
- Meningococcal septicaemia
- Any septicaemia with disseminated intravascular coagulation
- Rickettsiae
- Viruses (see Box 42.37)

Erythematous
- Scarlet fever
- Lyme disease (erythema chronicum migrans)
- Toxic shock syndrome
- Human erythrovirus (parvovirus) B19

Urticarial
- *Toxocara* spp.
- *Strongyloides* spp.
- *Schistosoma* spp.
- Cutaneous larva migrans

Others
- Tick typhus (eschar)
- Primary syphilis (chancre)
- Anthrax (ulcerating papule)

Box 42.9 General investigations and possible causes in suspected infection

Investigation	Possible causes
Full blood count	
Neutrophilia	Bacterial infection Inflammation
Neutropenia	Viral infection Brucellosis Typhoid Typhus Severe sepsis
Lymphocytosis	Viral infection
Lymphopenia	HIV infection (not specific) Other viral infections
Atypical lymphocytes	Infectious mononucleosis
Eosinophilia	Invasive parasitic infection
Thrombocytopenia	Sepsis Malaria Some viral infections
Erythrocyte sedimentation rate or C-reactive protein	Elevated in many infections
Serum creatinine and electrolytes	Potentially deranged in severe illness from any cause
Procalcitonin	Elevated particularly in bacterial infection
Beta-D-glucan and galactomannan	Maybe elevated in fungal infection
Liver enzymes	
Minor elevation of transferases	Non-specific feature of many infections Mild viral hepatitis
High transferases, elevated bilirubin	Viral hepatitis (usually A, B, or E)
Coagulation	May be deranged in hepatitis and sepsis
Serum lactate	Used in the assessment of sepsis

HIV, Human immunodeficiency virus.

resistant may be unwise, depending on clinical circumstances; for example, a patient presenting with pyelonephritis should not be treated with co-amoxiclav if *E. coli* resistant to co-amoxiclav were recently grown from a midstream urine specimen. The exact tests performed will vary, depending on clinical presentation and differential diagnoses, severity of illness, and available resources:

General blood tests

These provide clues to the diagnosis or organ or system involved (e.g., raised liver enzymes in liver or biliary tract infection) but in general are non-specific and non-diagnostic. Helpful tests may include:

- ***The full blood count and differential white cell count (WCC).*** In adults, lymphopenia is commonly associated with various viral infections, whereas neutrophilia, particularly if marked, is more likely to be associated with certain bacterial infections, although such associations are far from precise.
- ***The erythrocyte sedimentation rate, plasma viscosity, and C-reactive protein (CRP)*** may be raised in infection and other inflammatory states.
- ***Liver biochemistry and function.***
- ***Serum creatinine and electrolytes.*** These assess for infection-related acute kidney injury.
- ***Coagulation screen***, ***fibrinogen***, and ***D-dimer*** (to exclude disseminated intravascular coagulation [DIC] in severe sepsis or liver dysfunction in viral hepatitis).
- ***Serum lactate*** (in suspected sepsis).

CRP is a non-specific marker of inflammation and is raised in many different infections and in acute and chronic non-infection conditions associated with inflammation (e.g., rheumatoid arthritis). It has a short plasma half-life (19 hours) and is more useful in monitoring response to treatment than in making a diagnosis, although it is sometimes used as a point-of-care test in primary or ambulatory care to guide the need for antibiotic therapy, for example, in acute bronchitis, using cut-offs of the following:

- 100 mg/L for an antibiotic prescription
- 20 to 100 mg/L for a delayed prescription
- Less than 20 mg/L for no antibiotic.

The exact role of ***procalcitonin*** in the diagnostic and prognostic assessment of bacterial infections is controversial; it may have a useful role in antibiotic stewardship (see Chapter 12, 'Treatment of infection caused by multidrug-resistant organisms' section) when used within algorithms, particularly for respiratory tract infections and in the critical care environment, to determine when it is safe to start or stop antibiotic therapy.

Radiological imaging

X-ray (e.g., pneumonia), ultrasonography (e.g., abscess), echocardiography (endocarditis), CT (e.g., intra-abdominal infection), and magnetic resonance imaging (MRI; e.g., bone and spinal infections) are used to identify and localize infections. Positron emission tomography (PET) and single-photon emission computed tomography have also proved useful in localizing infection, especially when combined with CT scanning, when the focus of infection is unclear (**Fig. 42.4**). Biopsy or aspiration of tissue for microbiological examination is often facilitated by ultrasound or CT guidance. Indium- or technetium-labelled white blood cell (WBC) scans (using WBCs harvested from the patient before the test) may also help to localize infection, but their role has largely been replaced by PET scans. When possible, cases should be discussed with a radiologist before ordering tests to ensure that patients receive the optimal investigations in the context of their individual clinical presentation and prior investigations.

Microbiological investigations

It is often helpful to discuss the clinical problem with an infection specialist to ensure that appropriate tests are performed to identify the causative organism and that specimens are collected and transported correctly. Accurate completion of the request form, including clinical information, antimicrobial history, and epidemiological and risk factor details (e.g., onset and duration or symptoms, recent travel) enable laboratory staff to add or delete tests appropriately.

Microscopy

The most widely used microscopic technique is the Gram-stain, which determines the shape (cocci or bacilli) and Gram status (positive or negative) of bacteria. This allows most pathogenic bacteria to be classified into four groups: Gram-positive cocci, Gram-negative cocci, Gram-positive bacilli, and Gram-negative bacilli. To some extent, this then dictates empirical antibacterial therapy, if appropriate. **Box 42.10** shows some key bacterial causes of bloodstream infections in the UK, along with their shape and Gram status, and how bacterial identity links to potential foci of infection.

Culture

Specimens for culture should, ideally, be taken before administration of or a change in antibiotic therapy. After antibiotics are commenced, however, there will be an impact on the host's resident

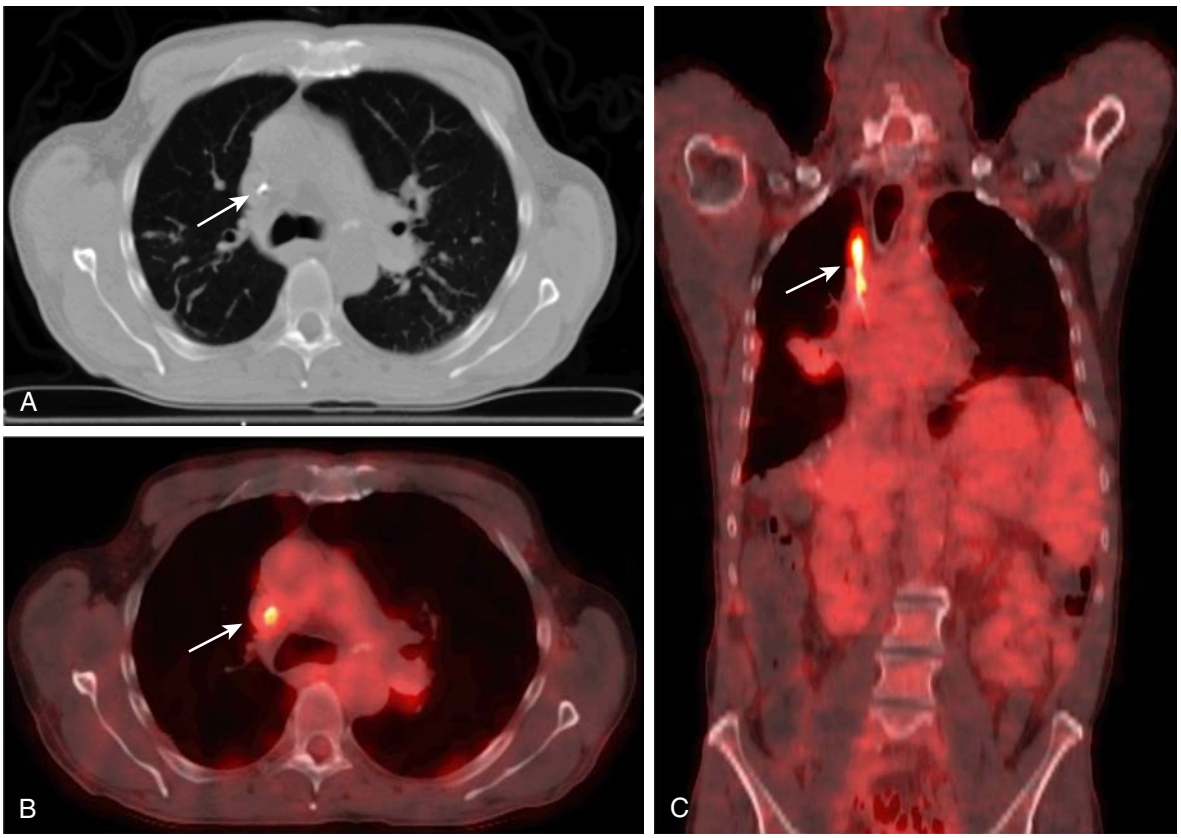


Fig. 42.4 ^{18}Fluorodeoxyglucose positron emission tomography (^{18}FDG-PET) in infection. Axial computed tomography images **(A)** of a 67-year-old male with a history of colon cancer and longstanding chest port *(arrow)* for chemotherapy, presenting with recurrent fevers and rigors of unknown origin with negative blood cultures, unresponsive to antibiotics. Axial **(B)** and coronal **(C)** fused ^{18}FDG-PET scans demonstrate metabolic uptake *(arrow)* along the catheter fibrin sheath. Explant of the port revealed colonization by *Stenotrophomonas maltophilia*, a multidrug-resistant organism known to colonize medical devices. Patient symptoms resolved after removing the port. *(From CPET.)*

microbiome, including infecting pathogens, resulting in potentially misleading results. Culture techniques can be applied to a wide variety of bacteria, fungi, protozoa, and viruses. However, some organisms are difficult to grow and may require special culture media or conditions. Viruses are particularly difficult (and, in many cases, impossible) to culture in a laboratory.

Antimicrobial susceptibility testing is performed on isolates thought to be clinically significant. Increasingly, traditional culture techniques are combined with new technologies, such as fluorescence *in situ* hybridization, matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectroscopy, morphokinetic cellular analysis, polymerase chain reaction (PCR) testing of identifier genes, and whole-genome sequencing of pathogenic organisms. This reduces the time to identification of pathogens, informs subsequent antimicrobial susceptibility testing (including whether a pathogen produces an extended-spectrum β-lactamase or carbapenemase resulting in multidrug resistance in Gram-negative bacteria), and enables linkage of bacteria and patients in outbreaks. In the future, new technologies may be applied directly to blood samples, possibly replacing the current culturing process.

Specimens sent for microscopy or culture (**Box 42.11**) are described as follows:

- ***Blood*** is sent for bacterial culture in hospitalized patients if systemic infection is suspected, regardless of whether fever is present at the time of presentation. Additional blood cultures are taken if fever subsequently occurs. In patients with suspected endocarditis or cardiac device infection or other 'deep' foci of infection (e.g., psoas abscess, discitis), at least three sets of blood cultures are taken because bacteraemia may be intermittent. Whole-blood samples may also be useful for genome detection, usually by PCR, of a range of viruses present in the bloodstream.
- ***Urine*** is sent for microscopy and culture in patients with urinary tract symptoms, and also in patients unable to provide a clear history, when UTI or sepsis is suspected. Urine culture is not indicated in asymptomatic patients able to provide a clear history regardless of whether they are catheterized.
- ***Cerebrospinal fluid (CSF)*** (meningitis, encephalitis), ***sputum*** (e.g., pneumonia, TB), ***articular fluid*** (e.g., septic arthritis, prosthetic joint infection), ***pus*** (e.g., abscess), and ***biopsy*** (e.g., of a lymph node) specimens are sent for microscopy and culture when clinically indicated.
- ***Swabs*** of clean wounds, without pus or surrounding cellulitis, are avoided because of the risk of identifying colonizing bacteria and encouraging inappropriate antibiotic prescribing. Tissues, aspirates, and fluids are usually preferred to swabs. Vaginal swabs are useful, however, to identify *Candida* spp., *Trichomonas vaginalis*, and other bacterial causes of vaginitis; nucleic acid amplification techniques (NAATs; see next section) are more commonly used to screen for *Neisseria gonorrhoeae*, *Chlamydia* spp., and HSV.
- ***Special culture techniques or prolonged culture*** is required for *Mycobacteria* spp.; slow-growing bacteria, such as *Actinomyces* spp. and *Brucella* spp.; and sometimes fungi. The laboratory must be informed of patients with suspected unusual infections or infections caused by fastidious microbes.

Box 42.10 Key bacterial and fungal causes of bloodstream infections in the UK

Common bacterial causes of bloodstream infection	Potential sources (foci) of bloodstream infection
Gram-negative bacilli (rods)	
Escherichia coli and other coliforms	Urinary tract (most commonly) Biliary tract (common) Other intra-abdominal sources, e.g., diverticular abscess Healthcare-associated pneumonia; community-acquired pneumonia if *Klebsiella pneumoniae* Other HAIs Bone, joint, or spine Meningitis (especially neonates)
Pseudomonas spp.	HAIs, especially pneumonia and i.v. or urinary catheters (most common) Neutropenic mucositis Urinary tract Skin or soft tissue, wounds (less common) Bone, joint, spine (less common)
Bacteroides spp.	Intra-abdominal Wounds
Gram-positive cocci	
Staphylococcus aureus (MSSA and MRSA)	Skin or soft tissue, wounds Bone, joint, spine Endocarditis Intravenous drug use, including infected deep venous thrombosis and septic pulmonary emboli Vascular infections, including grafts i.v. and urinary catheters Other HAIs Abscesses, especially psoas, epidural, and pyomyositis Meningitis (rare)
Group A *(Streptococcus pyogenes)*, B, C, and G streptococci	Skin or soft tissue, wounds, necrotizing fasciitis Bone, joint, spine Genitourinary (especially group B streptococci) Neonatal sepsis or meningitis (group B streptococci) Tonsillitis
Viridans group streptococci (e.g., *Streptococcus mitis* and *mutans* groups)	Endocarditis Abscesses, especially liver and brain Bone, joint, spine (rare)
Streptococcus pneumoniae (pneumococcus)	Upper or lower respiratory tract Pneumonia Meningitis Endocarditis (rare)
Coagulase-negative staphylococci	Usually skin contamination but consider prosthesis or device infection if repeat multiple blood cultures are positive for same bacteria with same antibiotic susceptibilities
Enterococcus spp.	HAIs, especially i.v. and urinary catheters (common) Urinary tract Intra-abdominal Endocarditis
Gram-negative cocci	
Neisseria meningitidis (meningococcus)	Meningitis Bloodstream infection Others, including septic arthritis and pneumonia, are rare
Fungi	
Candida spp.	HAIs, especially i.v. catheters and other devices (common) Intra-abdominal (especially patients who have had abdominal surgery and prolonged hospital stay or are or have been in critical care) Urinary tract i.v. drug use Endocarditis

HAI, Healthcare-associated infection; *i.v.,* intravenous; *MRSA,* meticillin-resistant *Staphylococcus aureus*; *MSSA,* meticillin-susceptible *Staphylococcus aureus.*
Gram positive rods are rarely a cause of human disease, and when cultured often represent sample contaminants.

- ***Faecal culture*** should be reserved for cases of suspected bacterial gastroenteritis or dysentery, especially when there may be public health implications. Faecal culture for viruses is not helpful; antigen or nucleic acid detection techniques (see next section) are more appropriate, especially in the investigation of an outbreak of diarrhoea and vomiting (e.g., norovirus). Protozoa (e.g., amoebiasis, giardiasis) may be a cause of diarrhoea in returning travellers, immunocompromised patients, toddlers, males who have sex with males, farm workers, and in any case of prolonged unexplained diarrhoea. Antigen or PCR-based tests are useful adjuncts to traditional stool microscopy for ova, cysts, and parasites. Detection of a specific clostridial toxin is a more reliable test for *C. difficile* infection than culture.

Box 42.11 Specimens and indications for microscopy, culture, and other microbiological tests

Specimen	Investigation	Indication
Blood	Giemsa stain for malaria	Any symptomatic traveller returning from a malarious area
	Malaria rapid antigen detection test	
	Stains for other parasites	Specific tropical infections
	Culture	All suspected bacterial infections
	PCR (viruses)	Fever, blood-borne viruses, HIV
Urine	Microscopy and culture	All suspected bacterial infections
	TB culture	Suspected TB Unexplained leucocytes in urine
	NAAT	STIs
Faeces	Microscopy ± iodine stain	Suspected protozoal diarrhoea, helminth infection
	Culture	All unexplained diarrhoea
	PCR, antigen detection	Suspected viral diarrhoea in children
	Clostridioides difficile toxin	Diarrhoea after hospital stay or antibiotic treatment
Throat swabs	Culture	Suspected bacterial tonsillitis and pharyngitis
	PCR, antigen detection	Viral respiratory infections
Sputum	Culture	Pneumonia, respiratory tract infections
	Auramine stain, TB culture (liquid culture; see Chapter 20, 'Tuberculosis' section)	Suspected TB
	Other special stains, cultures	Immunocompromised patients Suspected fungal infections
Cerebrospinal fluid	Microscopy and culture	Suspected meningitis
	Auramine stain, TB culture	Suspected TB, meningitis
	Other special stains, cultures	Immunocompromised patients Suspected fungal infections
	PCR	Suspected encephalitis or viral or bacterial meningitis
Rash aspirate		
Petechial	Microscopy and culture	Meningococcal disease
Vesicular	PCR, antigen detection, viral culture	Herpes simplex, zoster

HIV, Human immunodeficiency virus; *NAAT,* nuclear acid amplification test; *PCR,* polymerase chain reaction; *STI,* sexually transmitted infection; *TB,* tuberculosis.

Nucleic acid detection

Specific genes from many pathogenic microorganisms have been cloned and sequenced. Nucleic acid probes can be designed to detect these sequences, identifying pathogen-specific nucleic acid in body fluids or tissue. The use of NAATs, such as PCR, has increased the power of these tests to detect very small quantities of microbial material. Such techniques not only are exquisitely sensitive but also enable quantitation (e.g., viral load testing in HIV, HBV, and HCV) and subspeciation (e.g., at the genotype level). NAAT tests can be multiplexed (i.e., designed to detect a range of different pathogens in the same reaction mixture). Such assays have been widely used for viral infections (e.g., a respiratory panel may contain primers and probes able to detect up to 20 different viruses), but increasingly, PCR-based bacterial identification is also used in clinical practice. This allows 'screening' of clinical samples from sterile sites for a wider range of organisms, but because of the risk of detecting contaminating, colonizing, or multiple bacteria, results must be interpreted carefully in the context of clinical findings.

Immunodiagnostic tests

These can be divided into two types:

- Tests that detect ***microbial components*** (antigens) using a polyvalent antiserum or a monoclonal antibody (e.g., hepatitis B surface antigen, galactomannan in *Aspergillus* spp. infections)
- Tests that detect an ***immune response*** to infection (serological tests for antibodies and lymphocyte stimulation assays such as interferon-gamma (IFN-γ) release assays in TB for cellular immunity).

These investigations are valuable in the identification of organisms that are difficult to culture, especially viruses, atypical bacteria, and fungi, and they can also be helpful when antibiotics have been administered before samples were obtained. Elevated antibody titres on a single occasion (especially of immunoglobulin G (IgG) are rarely diagnostic, and it may be difficult to distinguish between past and acute infection. Specific assays for pathogen-specific IgM antibodies are useful for diagnosing an acute infection but may be poorly specific (e.g., for *Mycoplasma pneumoniae*). Paired serological tests taken a few weeks apart to observe IgG seroconversion are useful retrospectively to confirm the diagnosis but rarely change the acute management of the patient. Avidity testing may also enable recent infection (low-avidity antibodies) and historical infection (high-avidity antibodies) to be distinguished. Numerous serological tests are available; they should be used only in light of the clinical presentation and exposure and risk history.

Management of infectious diseases

Multidisciplinary infection care

It has increasingly been recognized that complex infectious diseases cannot be managed singlehandedly by a lone specialist clinician regardless of experience. Many infections require multidisciplinary teams (MDTs) to provide optimal care and clinical outcomes. A good example is that of orthopaedic infections. These require the expertise of multiple doctors (e.g., orthopaedic surgeons, plastic surgeons, infection doctors, and radiologists), nurses (e.g., specialist OPAT staff), pharmacists, and professionals allied to medicine (e.g., physiotherapists). Care plans are discussed at MDT meetings or clinics, which all attend. Skills and technologies developed during the COVID pandemic have led to many MDT meetings now being held online. There are many other infections and situations for which the MDT approach is necessary, such as HIV infection, TB, hepatitis, diabetic foot infection, OPAT, infection prevention and control, and antimicrobial stewardship. A weekly multidisciplinary infection meeting to discuss outpatients or inpatients with complicated infections, when decisions taken are recorded, is a standard of good departmental practice for adequately resourced infectious disease departments.

Shared decision-making

Infection practice is often 'grey'. Patients frequently present with non-specific histories and examination findings, and tests are often not diagnostic. The degree of investigation and intervention is determined by the severity or seriousness of the clinical presentation. A clear diagnosis is not always possible, and it is therefore important to use a shared decision-making approach to avoid tests, treatments, and procedures that are unlikely to be of benefit. Discussing the following four questions with patients may be useful in determining appropriate investigation and management strategies:

- What are the benefits?
- What are the risks?
- What are the alternatives?
- What if I do nothing?

Confidentiality

Many infections (e.g., hepatitis, HIV, TB) continue to be associated with social stigma. It is important to maintain patient confidentiality about the diagnosis at all times. Case notes should not be left lying around, and electronic patient records should not be visible in public areas. The patient's permission is always sought before discussion with spouses, partners, relatives, or friends. Communication with their primary care physicians is always encouraged to facilitate high-quality integrated care. Confidentiality is broken only when there is an overwhelming individual or public health risk, although multidisciplinary discussion and legal advice are usually sought beforehand.

Re-evaluation

Patients with sepsis and other serious infections need to be evaluated regularly, sometimes several times a day. Whether as an outpatient or inpatient, those who remain symptomatic without a clear diagnosis are re-evaluated by retaking the history and performing repeat examinations and appropriate investigations using a shared decision-making approach. If the patient's clinical condition deteriorates or does not respond to therapy or if adverse effects occur on intervention, careful re-evaluation is essential. Such patients are discussed at the hospital's multidisciplinary infection meeting.

Further reading

Bhargava A, Ralph R, Chatterjee B et al. Assessment and initial management of acute undifferentiated fever in tropical and subtropical regions. *BMJ* 2018; 363:k4766.
Chapman ALN, Patel S, Horner C et al. Updated good practice recommendations for outpatient parenteral antimicrobial therapy (OPAT) in adults and children in the UK. *JAC Antimicrob Resist* 2019; 1(2).
Drago F, Ciccarese G, Gasparini G et al. Contemporary infectious exanthems: an update. *Future Microbiol* 2017; 12:171–193.
Osthoff M, Gürtler N, Bassetti S et al. Impact of MALDI-TOF-MS-based identification directly from positive blood cultures on patient management: a controlled clinical trial. *Clin Microbiol Infect* 2017; 23(2):78–85.
Roberts MB, Fishman JA. Immunosuppressive agents and infectious risk in transplantation: managing the 'net state of immunosuppression'. *Clin Infect Dis* 2021; 73(7):e1302–e1317.

EPIDEMIOLOGY AND PREVENTION OF INFECTIONS

Infectious diseases continue to be major causes of morbidity and mortality in humans, particularly in low- and middle-income countries (LMICs), where they are associated with poverty, overcrowding, and poor healthcare access. In high-income countries (HICs), increasing prosperity and hygiene (most important) along with high immunization coverage and prompt antimicrobial treatment have reduced the burden of infectious disease. However, antimicrobial-resistant and some difficult-to-treat microorganisms are increasingly prevalent in all settings, and diseases such as HIV infection, TB, malaria, and Gram-negative bacterial sepsis remain problematic. Major respiratory viral pathogens such as influenza, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), and measles, as well as arthropod-borne viruses such as Zika and Chikungunya and highly pathogenic zoonoses such as Ebola have demonstrated significant potential for severe epidemic spread. Highly pathogenic avian influenza is top of the international watch list for transforming into a disease that could transmit to and between humans with devastating consequences.

Increased population mobility

The phenomenon of ever-increasing global mobility, whether driven by war, civil unrest, climate change, habitat disruption, or tourism, has aided the spread of infectious diseases and allowed previously localized pathogens to establish themselves across much wider territories (e.g., New Delhi metallo-β-lactamase–producing Gram-negative bacteria). Greater movement of livestock and animals, and encroachment of humans into new ecosystems has enabled the spread of zoonotic diseases. Additionally, changes in farming and food processing methods have contributed to an increase in the incidence of food- and water-borne disease. Deteriorating social conditions in major cities have facilitated the resurgence of TB and other infections. Refugee camps, where large numbers of people are forced to live in close proximity, also continue to provide breeding grounds for devastating epidemics such as diphtheria and cholera. There are ongoing concerns about the deliberate release of infectious agents, such as anthrax, by terrorist groups or national governments.

Changing patterns of disease

In LMICs, there have been notable successes such as:

- The eradication of smallpox and near eradication of polio and Guinea worm infection (dracunculiasis)
- Reductions in the incidence of HIV infection
- Reductions in the mortality rates of TB and malaria, as well as in many of the 'neglected tropical diseases'.

These have been balanced or outweighed by new problems to the extent that communicable diseases were still responsible for nearly half of all deaths in low-income countries (LICs) in 2019. Although the leading causes of death globally are increasingly non-communicable diseases, in LICs, malaria (sixth), TB (eighth), and HIV (ninth), as well as lower respiratory tract infections (LRTIs) and diarrhoeal diseases feature prominently (**Box 42.12**).

- The COVID-19 pandemic stalled progress in reducing TB incidence, with 10.6 million new cases and 1.6 million deaths in 2021. About 450 000 people developed multidrug resistant disease.
- Progress in reducing malaria also stalled during the pandemic, with an estimated 234 million cases in 2021, 95% in the Africa region. There were more than 600 000 deaths, 75% being in children younger than 5 years of age. Resistance to the key antimalarial (artemisinins) is rising with associated treatment failure in >10% cases.
- Some 360 million people were estimated to be living with hepatitis B or C in 2019, although the incidence of hepatitis B in those younger than 5 years old has decreased from 4.7% in the pre-vaccine era to less than 1% in 2022.
- In 2022, approximately 630 000 people still died of HIV-related disease (compared with 1.9 million in 2005) despite the

Box 42.12 Leading causes of mortality worldwide, 2019

Cause of death	Mortality	
	Millions (annual)	Crude mortality (/100 000)
Ischaemic heart disease	8.9	126
Stroke	6.2	77
Chronic obstructive pulmonary disease	3.2	41
Lower respiratory infections	2.6	40
Neonatal conditions	2.0	
Trachea, bronchus and lung cancers	1.8	
Alzheimer's disease and other dementias	1.6	27
Diarrhoeal diseases	1.5	
Diabetes mellitus	1.5	21
Kidney diseases	1.3	19
Liver cirrhosis	1.3	17
Road injury	1.3	
Tuberculosis	1.2	

Data from World Health Organization, http://ww.who.int.

availability of effective therapy; globally, only 76% living with HIV receive anti-retroviral therapy. The global incidence of HIV has fallen, however, from 0.4 per 1000 uninfected population in 2005 to 0.17 per 1000 uninfected population in 2022 but remains high in Africa (0.57 per 1000 population).

- Neglected tropical diseases, such as filariasis, onchocerciasis, and trachoma remain problematic, with 1.65 billion persons requiring mass or individual treatment in 2021, of whom 80% were in 16 LICs.

Factors limiting control of infectious diseases in low- and middle-income countries

Most infectious diseases affecting LMICs are preventable or treatable, but they continue to thrive because of a lack of resources and political commitment and disruption by conflict. The surveillance of infectious diseases has often been poor in many regions, resulting in a slow response to emerging problems. Distrust of immunization programmes in conflict zones (especially Pakistan and Afghanistan) has been key in delaying the eradication of polio. It seems likely that the ongoing emergence of highly drug-resistant Gram-negative bacterial infections will affect poor and vulnerable populations disproportionately. The impact of global warming on the spread and incidence of infections remains uncertain, but it is clearly driving transmission of infectious diseases through direct influence on the spread of insect vectors and by population movement and migration in response to climate change.

Goals in international development

The World Health Organization (WHO) has established a portfolio of targets to combat disease burden (see Chapter 10). The 17 Sustainable Development Goals (SDGs) include SDG 3, which is 'to ensure healthy lives and promote wellbeing for all at all ages', with the specific commitment to 'end the epidemics of AIDS [acquired immunodeficiency syndrome], tuberculosis, malaria and neglected tropical diseases, and combat hepatitis, waterborne diseases and other communicable diseases by 2030'. Although a lot has been achieved by public–private partnerships and other charitable funders, including major technological breakthroughs in terms of antimicrobial developments and new vaccinations, as well as improved reporting systems, it is clear that there is still much to be done.

Acquisition of infection

Many infections are acquired from other people, who may be symptomatic or asymptomatic carriers. Some bacteria, like the meningococcus, are common transient commensals but cause invasive disease in a small minority of those colonized, sometimes after other infections such as influenza. Infections with organisms such as HIV and hepatitis B viruses (HBVs) are followed in most cases by a carrier state that has potential to transmit while the patient is asymptomatic, and *M. tuberculosis* infection can become transmissible after reactivation after even prolonged periods of asymptomatic 'latent' infection.

Zoonoses are infections that can be transmitted from wild or domestic animals to humans. Infection can be acquired in a number of ways:

- Direct contact with the animal
- Ingestion of meat or animal products
- Contact with animal urine or faeces, including aerosol inhalation
- Via an arthropod vector
- By inoculation of saliva in a bite wound.

Many zoonoses can also be transmitted from person to person. Some zoonoses are listed in **Box 42.13**.

Most microorganisms do not have a vertebrate or arthropod host but are free living in the environment. The vast majority of these environmental organisms are non-pathogenic, but some can cause human disease (**Box 42.14**). Person-to-person transmission of these infections is rare. Some parasites may have a stage of their life cycle that is environmental (e.g., the free-living larval stage of *Strongyloides stercoralis* and the hookworms) even though the adult worm requires a vertebrate host. Other pathogens can survive for periods in water or soil and are transmitted from host to host via this route (see later).

Routes of transmission

Endogenous infection

The body's own endogenous flora (often called the ***microbiome***) can cause infection if the organism gains access to a usually sterile site. Non-specific host defences are breached, for example, by cutting or scratching the skin and allowing surface commensals to gain access to deeper tissues; this is the aetiology of staphylococcal wound infections and folliculitis. Similarly, cystitis and ascending UTI occur when colonic bacteria enter the female urinary tract through the urethra. There may be more serious defects in host immunity owing to disease or chemotherapy, allowing normally harmless skin and bowel flora ('opportunist infections') to produce invasive disease.

Air-borne spread

Many respiratory tract pathogens are spread from person to person by aerosol or droplet transmission. Secretions containing the infectious agent are coughed, sneezed, or breathed out and are then inhaled by a new host. Some enteric viral infections may also be spread by aerosols of faeces or vomit. Environmental pathogens, such as *Legionella pneumophila*, and zoonoses such as histoplasmosis are also acquired by aerosol inhalation.

Box 42.13 Zoonotic infections

Disease	Pathogen	Animal reservoir	Mode of transmission
Prions			
vCJD	Prion protein	Cattle	Ingestion (CNS tissue)
Viruses			
Lassa fever	Arenavirus	Multimammate rats	Direct contact
Avian influenza	Influenza H5N1	Birds	Inhalation
Japanese encephalitis	Flavivirus	Pigs	Mosquito bite
Rabies	Rhabdovirus	Dogs and other mammals	Saliva, faeces (bats)
Yellow fever	Flavivirus	Primates	Mosquito bite
Mpox	Orthopox virus	Rodents, small mammals	Uncertain
SARS	Coronavirus	Unsure, most likely bats	Droplet
MERS	Coronavirus	Camels	Droplet
SARS-CoV-2	Coronavirus	Unsure, most likely bats	Droplet
Bacteria			
Gastroenteritis	*Escherichia coli* *Salmonella enteritidis* and others *Campylobacter jejuni*	Cattle, chickens Chickens, cattle Various, e.g., chickens	Ingestion (contaminated food) Ingestion (meat, eggs) Ingestion (meat, milk, water)
Leptospirosis	*Leptospira interrogans*	Rodents	Ingestion (urine)
Brucellosis	*Brucella abortus* *Brucella melitensis*	Cattle Sheep, goats	Contact; ingestion of milk or cheese
Anthrax	*Bacillus anthracis*	Cattle, sheep	Contact; ingestion
Lyme disease	*Borrelia burgdorferi*	Deer	Tick bite
Cat-scratch disease	*Bartonella henselae*	Cats	Flea bite
Plague	*Yersinia pestis*	Rodents	Flea bite
Typhus	Various *Rickettsia* spp.	Various	Arthropod bite
Ornithosis (psittacosis)	*Chlamydia psittaci*	Psittacine and other birds	Aerosol
Others			
Toxoplasmosis	*Toxoplasma gondii*	Cats and other mammals	Ingestion (meat, faeces)
Hydatid disease	*Echinococcus granulosus*	Dogs	Ingestion (faeces)
Trichinosis	*Trichinella spiralis*	Pigs, bears	Ingestion (meat)
Toxocariasis	*Toxocara canis*	Dogs	Ingestion
Cutaneous larva migrans	*Ancylostoma caninum*	Dogs	Penetration of skin by larvae
Leishmaniasis	*Leishmania* spp.	Dogs	Ingestion

CNS, Central nervous system; *SARS-CoV-2,* severe acute respiratory syndrome coronavirus 2; *vCJD,* variant Creutzfeldt–Jakob disease.

Faecal–oral spread

Transmission of organisms by the faecal–oral route occurs by direct transfer, by contamination of clothing or household items or, most commonly, via contaminated food or water. Human and animal faecal pathogens enter the food supply at any stage. Raw sewage is used as fertilizer in many parts of the world, contaminating growing vegetables and fruit. Poor personal hygiene can result in contamination during production, packaging, preparation, or serving of food. In HICs, centralization of the food supply and increased processing of food have allowed the potential for relatively minor episodes of contamination to cause widely disseminated outbreaks of infection, such as *Salmonella* from the egg industry and *Listeria* from refrigerated sandwiches and other prepared foods.

Water-borne faecal–oral spread is usually the result of inadequate access to clean water and safe sewage disposal and is common throughout the developing world. Worldwide, 74% of the global population have access to safely managed drinking water, but only 54% have access to safe sanitation. There are major regional differences with 80% those without safe drinking water living in rural areas.

Vector-borne disease

Many tropical infections, including malaria, are spread from person to person or from animal to person by an arthropod vector. Vector-borne diseases are also found in temperate climates but are relatively uncommon, although some are becoming more prevalent because of climate change, most notably tick- and mosquito-borne pathogens in Europe. In most cases, part of the life cycle takes place within the body of the arthropod and each species requires a specific vector. Some vector-borne diseases are shown in **Box 42.15**.

Direct person-to-person spread

Organisms can be passed on directly in a number of ways. STIs are dealt with in Chapter 42. Skin infections such as PVL toxin-expressing *S. aureus* folliculitis, mpox among males who have sex with males, and ectoparasites such as scabies and head lice, can be spread by simple skin-to-skin contact. Other organisms are passed on by transmission from blood (or, occasionally, some other body fluid) to blood. Blood-to-blood transmission can occur during sexual contact; from mother to infant, either transplacentally or in the peripartum period; between intravenous drug users sharing any

Box 42.14 Environmental organisms that can cause human infection

Organism	Disease (most common presentations)
Bacteria	
Burkholderia pseudomallei	Melioidosis
Burkholderia cepacia	Lung infection in cystic fibrosis
Pseudomonas spp.	Various
Legionella pneumophila	Legionnaires' disease (pneumonia)
Bacillus cereus	Gastroenteritis
Listeria monocytogenes	Various
Clostridium tetani	Tetanus
Clostridium perfringens	Gangrene, septicaemia
Non-tuberculous mycobacteria	Pulmonary infections
Fungi	
Candida spp.	Local and disseminated infection
Cryptococcus neoformans	Meningitis, pulmonary infection
Histoplasma capsulatum	Pulmonary infection
Coccidioides immitis	Pulmonary infection
Mucor spp.	Mucormycosis (rhinocerebral, cutaneous)
Sporothrix schenckii	Lymphocutaneous sporotrichosis
Blastomyces dermatitidis	Pulmonary infections
Aspergillus fumigatus	Pulmonary infections

part of their injecting equipment; when infected medical or other (e.g., tattoo needles) equipment is reused; if contaminated blood or blood products are transfused; and in any sporting or accidental contact when blood is spilled. Ingestion of infected breast milk is another route of person-to-person spread for some infections (e.g., HIV).

Indirect person-to-person spread

Many organisms can be spread from person to person indirectly by contamination by fomites (e.g., on door handles), which are subsequently touched by another person. Common examples include respiratory viruses, such as influenza, and healthcare-associated pathogens, such as MRSA and the spores of *C. difficile*. This is sometimes an overlooked route of transmission; minimization requires, for example, hand-washing immediately after sneezing or coughing and when moving from ward to ward in a hospital, and optimal cleaning regimens in institutions.

Direct inoculation

Infection can occur when organisms breach the normal mechanical defences by direct inoculation from the environment (e.g., contamination of bone after an open fracture, leading to subsequent osteomyelitis, or as seen in tetanus and certain fungal and mycobacterial infections). Some of the circumstances in which this can occur are covered under endogenous infection and blood-to-blood transmission, described earlier. Rabies virus is inoculated by the bite of an infected animal.

Consumption of infected material

Although many food-related zoonotic infections are caused by contamination of food with animal faeces (and are thus spread by the faecal–oral route), several diseases are transmitted directly in animal products. These include *Listeria* (from soft cheeses and pates among other foods), some strains of *Salmonella* (eggs, chicken meat), brucellosis (unpasteurized milk), *E. coli*, and the prion diseases kuru and variant Creutzfeldt–Jakob disease (CJD; ingested neural tissue).

Box 42.15 Infections transmitted by arthropod vectors

Vector	Disease	Microorganism
Mosquito	Malaria	*Plasmodium* spp.
	Lymphatic filariasis	*Wuchereria bancrofti*, *Brugia malayi*
	Yellow fever	Flavivirus
	West Nile fever	Flavivirus
	Dengue	Flavivirus
	Zika	Flavivirus
Sandfly	Leishmaniasis	*Leishmania* spp.
	Bartonellosis	*Bartonella bacilliformis*
Blackfly	Onchocerciasis	*Onchocerca volvulus*
Tsetse fly	Sleeping sickness	*Trypanosoma brucei*
Flea	Plague	*Yersinia pestis*
	Endemic typhus	*Rickettsia typhi*
Reduviid bug	Chagas' disease	*Trypanosoma cruzi*
Louse	Epidemic typhus	*Rickettsia prowazekii*
	Louse-borne relapsing fever	*Borrelia recurrentis*
Hard tick	Lyme disease	*Borrelia burgdorferi*
	Typhus (spotted fever group)	*Rickettsia* spp.
	Babesiosis	*Babesia* spp.
	Tick-borne relapsing fever	*Borrelia duttonii*
	Tick-borne encephalitis	Flavivirus
	Congo–Crimean haemorrhagic fever	*Orthonairovirus* (Nairoviridae)
Trombiculid mite	Scrub typhus	*Orientia tsutsugamushi*

Classification of outbreaks

The type of outbreak has a bearing on the public health measures that need to be instituted for its control.

- ***Person-to-person*** is when infection is passed from one infected individual to another and outbreaks of infection are separated by the incubation period. Control measures may include hand and personal hygiene (a sensible early control measure in any outbreak), immunization (e.g., influenza, measles), and antimicrobial agents (e.g., anti-retroviral therapy).
- ***Point source*** is when there is a single source of infection (e.g., food eaten at a social function). All those infected develop symptoms at the same time, around the expected incubation period. If the point source is temporary or removed and there is no person-to-person or other mode of transmission, new cases decline after the incubation period.
- ***Common source*** is when there is a single source of infection but infections are spread over a period of time (e.g., a contaminated food preparation facility or restaurant). Many people may be exposed over a potentially long period of time. Recent large-scale epidemics of enteric fever in sub-Saharan Africa have been

associated with street-vended beverages from contaminated water supplies, and other outbreaks (most famously 'Typhoid Mary') have been associated with transmission from asymptomatic carriers via food preparation. After the source is identified, control measures may include cleaning and disinfecting the facility, closing the facility or restaurant, testing and excluding asymptomatic carriers from the workforce, recalling food, and informing the public on how to make food safe (e.g., by cooking it) or to avoid or throw food away.

- An ***epidemic*** is when there is an increased and unusually widespread infection in the community, causing waves of infection. Epidemics spread through communities and may affect those who have no active immunity to the infection.
- A ***pandemic*** is an epidemic, as defined earlier, occurring worldwide or over a very wide geographical area and crossing international boundaries, usually affecting a large number of people.

Some infectious diseases should be notified to the public health authorities so that they are aware of cases and outbreaks and can respond accordingly. Diseases that are notifiable in England and Wales are listed in **Box 42.16** (see also Further reading for the *List of Notifiable Organisms [Causative Agents]* in the UK).

The prevention of and response to epidemics and pandemics are complex and depend on the underlying cause but are multifactorial in approach and may include:

- Routine and enhanced disease surveillance
- Identification and tracing of contacts, including use of rapid diagnostics, through epidemiological investigation increasingly informed by pathogen genome sequencing
- Public awareness campaigns (e.g., focused on hand and personal hygiene or other control measures; to provide reassurance)
- Enhanced infection prevention and control procedures, including wearing of face masks, personal distancing, and prevention of mass gatherings
- Isolation or cohort nursing in healthcare facilities admitting patients and other institutions (e.g., nursing homes)
- Use of mass antimicrobials (e.g., antiviral agents in influenza epidemics or the universal test and treat approach to HIV prevention) and immunization as appropriate and available.

Efficient coordination of such responses by local, regional, national (e.g., UK Health Security Agency), or international organizations (e.g., WHO) is essential. The SARS-CoV-2 pandemic in 2020 is a good example of how a slow initial response can lead to grave consequences.

Infection prevention and control

Relatively few patients with infections present a serious risk to healthcare workers and other contacts or patients. However, the potential remains for unexpected outbreaks of life-threatening or almost untreatable infections, including within healthcare settings, associated with the appearance of diseases such as SARS-CoV-2 infection or strains of influenza with high mortality, occasional cases of Lassa fever or Ebola, bioterrorist use of microbes, and the emergence of extensively drug-resistant bacterial infections (e.g., TB, *Mycobacterium abscessus* and Gram-negative bacteria, especially

i Box 42.16 Notifiable diseases in England and Wales[a]

- *Bacillus anthracis*
- *Bacillus cereus* (only if associated with food poisoning)
- *Bordetella pertussis*
- *Borrelia* spp.
- *Brucella* spp.
- *Burkholderia mallei*
- *Burkholderia pseudomallei*
- *Campylobacter* spp.
- Carbapenemase-producing Gram-negative bacteria
- Chikungunya virus
- *Chlamydophila psittaci*
- *Clostridium botulinum*
- *Clostridium perfringens* (only if associated with food poisoning)
- *Clostridium tetani*
- *Corynebacterium diphtheriae*
- *Corynebacterium ulcerans*
- *Coxiella burnetii*
- Crimean-Congo haemorrhagic fever virus
- *Cryptosporidium* spp.
- Dengue virus
- Ebola virus
- *Entamoeba histolytica*
- *Francisella tularensis*
- *Giardia lamblia*
- Guanarito virus
- *Haemophilus influenzae* (invasive)
- Hanta virus
- Hepatitis A, B, C, delta, and E viruses
- Influenza virus
- Junin virus
- Kyasanur Forest disease virus
- Lassa virus
- *Legionella* spp.
- *Leptospira interrogans*
- *Listeria monocytogenes*
- Machupo virus
- Marburg virus
- Measles virus
- Mpox virus
- Mumps virus
- *Mycobacterium tuberculosis* complex
- *Neisseria meningitidis*
- Omsk haemorrhagic fever virus
- *Plasmodium falciparum, vivax, ovale, malariae, knowlesi*
- Polio virus (wild or vaccine types)
- Rabies virus (classical rabies and rabies-related lyssaviruses)
- *Rickettsia* spp.
- Rift Valley fever virus
- Rubella virus
- Sabia virus
- *Salmonella* spp.
- SARS-CoV-2
- *Shigella* spp.
- *Streptococcus pneumoniae* (invasive)
- *Streptococcus pyogenes* (invasive)
- Varicella zoster virus
- Variola virus
- Verocytotoxigenic *Escherichia coli* (including *E. coli* O157)
- *Vibrio cholerae*
- West Nile Virus
- Yellow fever virus
- Yersinia pestis

[a]Diseases are notifiable to Local Authority Proper Officers under the Health Protection (Notification) Regulations 2010. Report other diseases that may present significant risk to human health under the category 'other significant disease'.

SARS-CoV-2, Severe acute respiratory syndrome coronavirus 2.

carbapenemase-producing or carbapenemase-resistant *E. coli* and *K. pneumoniae*).

During the worldwide outbreak of SARS in 2003, scrupulous infection control procedures reduced the spread of infection. However, in the 'inter-epidemic' period, it is difficult to maintain a high level of alert, and indeed, many countries were not properly prepared for the SARS-CoV-2 pandemic of 2020. Healthcare workers should remain vigilant because the early symptoms of many of these diseases are non-specific. Methods of preventing infection depend on the source and route of transmission.

Healthcare-associated infections

The burden of morbidity, mortality, and costs attributed to healthcare-associated infection (HAI) has been highlighted in many HICs. Although data from LMICs are sparse, the impact of HAI is likely to be even greater. *C. difficile*, *S. aureus* (especially MRSA), vancomycin-resistant enterococci, and the increasingly difficult-to-treat multi- and extensively drug-resistant Gram-negative infections are all strongly associated with healthcare contact and remain a problem in hospitals worldwide. In 2016, the proportion of inpatients with HAIs in acute hospitals in England was 6.4%. The most frequent HAIs affect the respiratory and urinary tracts, as well as surgical sites. There has been a notable decline, however, in the incidences of and death caused by MRSA bloodstream and *C. difficile* infections in the UK. This has been attributed to higher standards of basic infection prevention and control, including rapid isolation of patients, hospital 'deep cleaning', universal MRSA screening (and decolonization therapy, if positive) on entry to acute hospitals and before surgery, and improved care of intravenous access devices.

Healthcare-associated infection prevention and control measures

Poor infection control practice in hospitals and other healthcare environments can cause the transfer of infection from person to person. Healthcare workers are strongly encouraged to wash or clean their hands before and after patient contact; whenever necessary, they should wear gloves, aprons, and other protective equipment. This is particularly important when they are performing invasive procedures, dealing with infected wounds, and manipulating indwelling devices. Additional control measures are recommended for contact with patients known to be colonized or infected with a resistant organism (e.g., carbapenemase-producing *K. pneumoniae*), those infected by communicable pathogens with high mortality (e.g., Lassa or Ebola, for which high-security isolation and extreme care are mandatory), and those with diarrhoea (e.g., *C. difficile* infection).

Interventions that health workers should focus on to prevent and control HAIs are listed next.

Care bundle approach

This is a set of three to five simple, evidence-based actions that healthcare professionals should always implement in certain clinical situations to minimize the risk of infection. For example, care bundles have been developed for the insertion and ongoing care of urinary catheters, insertion and use of central lines (to prevent central line-associated bloodstream infections [**Box 42.17**]), and to prevent ventilator-associated pneumonia or surgical site infections.

Hand hygiene

The WHO's 'Five Moments for Hand Hygiene' (see Further reading) are applied before and after all patient contacts, including before and after aseptic procedures and after contact with body fluids and the patient environment. Soap and water (rather than an alcohol-based rub) should be used after visible contamination of hands and when patients are vomiting or have diarrhoea even if gloves are used. In low-income settings, in the absence of soap and water, alcohol-based hand rub should be used if available.

Box 42.17 Central line–associated bloodstream infection care bundles

Central line insertion bundle

Hand hygiene

Wash hands or use an alcohol-based, waterless hand cleaner when caring for central lines:

- Before and after palpating the catheter insertion site
- Before and after inserting, replacing, adjusting, or dressing the site
- During palpation of the insertion site and after application of antiseptic and only if full asepsis is maintained.

Complete maximal barrier

- The operator inserting the central venous catheter should adhere to strict aseptic techniques and wear sterile gloves, a gown, a surgical cap. and a surgical mask.

Chlorhexidine 2% skin antisepsis

Prepare skin with 2% chlorhexidine in 70% alcohol using swabs and a friction scrub for at least 30 seconds. Do not wipe or blow dry and allow to dry completely before skin puncture:

- 30 seconds for a dry site
- 2 minutes for a moist site (especially femoral)

Optimal catheter site selection

- In adult patients, there is some evidence that the subclavian site has a lower risk of catheter-related bloodstream infections. However, there is usually more experience with the internal jugular site. The subclavian and internal jugulars are the preferred sites for infection-control purposes.
- Use sterile, single-use jelly
- Use sterile, single-use jelly for ultrasound-guided insertions

Central line maintenance bundle

Practice hand hygiene at five moments:

- Before touching a patient
- Before clean or aseptic procedures
- After body-fluid exposure or risk
- After touching a patient
- After touching patient surroundings

Aseptic technique for accessing and changing needleless connectors

- Scrub the access port or hub immediately before each use with an appropriate antiseptic.

Standardised tubing change

- Intravenous medication administration tubing should be changed as per the recommendation in the local organisations policy.

Daily review of catheter necessity

- Daily review of line necessity during rounds so the necessity of the lines can be determined and unnecessary lines removed.

Personal protective equipment

Gloves should be worn for all invasive procedures, including venesection, and contact with sterile sites, mucous membranes, bodily fluids, or non-intact skin. Disposable plastic aprons should be used when there is a risk of contamination of clothing or skin; if there is a risk of extensive contamination, surgical gowns should be worn and, in some situations, plastic visors should be worn. Respiratory protection should be worn for patients with respiratory-spread pathogens, such as measles, influenza, SARS-CoV-2, and TB.

Urinary catheters

These should be inserted only when it is essential to do so and then removed as soon as possible; they must be inserted using an aseptic technique. A clear care plan should be in place for longer-term

catheters. Overuse or poor care of urinary catheters increases the risk of *E. coli* bloodstream infection, a current target for HAI reduction in the UK.

Vascular access devices

A vascular catheter should only be inserted when there is a clear plan to use it, it is checked daily, and it is removed as soon as possible. Two percent chlorhexidine gluconate in 70% isopropyl alcohol is used for skin decontamination before insertion and to clean the access port or hub before use. A clear care plan should be instituted for medium- or long-term devices.

Adequate hydration

Dehydrated patients are at higher risk of UTIs. Oral hydration is preferred to intravenous hydration if possible (there is a lower risk of infection than with an intravenous device), and patients should be encouraged to drink an adequate quantity of fluids.

Communication

This is essential when transferring patients with transmissible infections (e.g., chickenpox, SARS-CoV-2, and infections caused by multi- and extensively drug-resistant bacteria) between admission areas, wards, hospitals, and institutions, including residential and nursing homes, so that appropriate infection control measures can be put in place for patients with known infections.

Isolation

The appropriateness of isolation, if available, should be considered. Infections with a high risk to public health (e.g., Lassa and Middle East respiratory syndrome [MERS]) are isolated within high-security, hospital-based, negative-pressure isolation facilities using enhanced respiratory, personal protective equipment (PPE), and decontamination procedures. In the UK, there are a small number of sites for the management of confirmed high-consequence infections transmitted by air-borne or contact spread, but many hospitals (particularly those with specialist infectious disease units) have facilities to manage patients who are being tested for these infections or to act as surge centres in the event of a large epidemic. Patients hospitalized with highly infectious respiratory-spread pathogens, such as chickenpox and measles, are isolated in negative-pressure isolation facilities. Patients with pathogens spread by direct or indirect person-to-person contact (colonization or infection), such as *C. difficile* and multidrug-resistant Gram-negative bacteria, are, ideally, housed in side-rooms with *en suite* toileting facilities and appropriate contact precautions (hand hygiene and PPE). If such facilities are limited, then basic precautions (hand hygiene, PPE) are still applied, and cohort isolation is considered (i.e., patients with the same microbiologically confirmed infection are housed on the same ward).

Other infection prevention and control measures (non–hospital-acquired infections)

Although effective antimicrobial chemotherapy is available for many diseases, the ultimate aim of any infectious disease control programme is to prevent infection occurring. This is achieved either by eliminating the source or mode of transmission of an infection *or* reducing host susceptibility to environmental pathogens.

Eradication of reservoir

In a few diseases for which humans are the only natural reservoir of infection, it may be possible to eliminate disease by an intensive programme of case finding, isolation, treatment, and immunization (as in the case of smallpox). If there is an animal or environmental reservoir, complete eradication is unlikely, but local control methods may decrease the risk of human infection (e.g., killing of rodents to control plague, management of poultry and other birds to prevent avian influenza).

For arthropod- or vector-borne infections

It may be possible to destroy the vector species or to take measures to avoid being bitten (e.g., insect repellent sprays, insecticide-impregnated bed nets, indoor residual spraying). Release of genetically modified sterile mosquitoes has potential to reduce transmission of dengue and malaria.

For food-borne infections

Improvements in food handling and preparation result in less contamination during processing, transport, or preparation. Organisms that are intrinsically present in food can be killed by appropriate preparation and cooking. Improved surveillance and regulation of the food industry, as well as better health education for the public, reduce risk.

For faecal–oral infections

Improvements in water supply and sanitation (following Water, Sanitation & Hygiene [WASH] strategies; see Further reading) may dramatically decrease the prevalence of faecal-oral infections.

For blood-borne infections

Blood transfer may be prevented by optimal use of medical equipment. Donated blood is routinely tested for infection in most HICs.

For infections spread by air-borne and direct contact

Some air-borne respiratory infections and some infections spread by direct contact can be controlled by isolating patients, although this is often difficult. Positive-pressure isolation (in which the flow of air is from the patient's room or ventilated lobby into the ward) is useful in uninfected patients with severe immunodeficiency to protect them from infection. Neuraminidase inhibitors may reduce the transmission of influenza viruses to at-risk persons.

Immunization, immunoprophylaxis, and immunotherapy

Immunization has changed the course and natural history of many infectious diseases. Passive immunization by administering preformed antibody, in the form of either immune serum or purified normal immunoglobulin, provides short-term immunity and has been effective in both the prevention (immunoprophylaxis) and the treatment (immunotherapy) of patients with a number of bacterial and viral diseases (**Box 42.18**).

The active childhood immunization schedule currently recommended by the WHO is summarized in **Box 42.19**. Long-lasting immunity is achieved by active immunization with a live attenuated or an inactivated organism, a subunit of an infectious agent, or sometimes a microbial toxin (either native or modified), that is, a toxoid (**Box 42.20**). Immunization should be kept up to date, with booster doses being given throughout life according to the vaccines booster schedule. Travellers to LMICs, especially if visiting rural areas for prolonged periods, may warrant specific immunizations (with yellow fever vaccination being mandatory for entry to some countries).

In 1974, the WHO introduced the Expanded Programme on Immunization (EPI). By 1994, more than 80% of the world's children had been immunized against TB, diphtheria, tetanus, pertussis, poliomyelitis, and measles. It is hoped that poliomyelitis will be eradicated worldwide in the near future, matching the past success of global smallpox eradication. The rapid development and testing of novel vaccines against SARS-CoV-2 saved millions of lives during the pandemic. Introduction of conjugate vaccines against

Box 42.18 Examples of passive immunization available

Infection	Antibody	Indication
Bacterial		
Tetanus	Human tetanus immune globulin	Prevention and treatment
Diphtheria	Horse serum	Prevention and treatment
Botulism	Horse serum	Treatment
Clostridioides difficile diarrhoea	Bezlotoxumab	Prevention of recurrent infection
Viral		
Hepatitis A Measles	Human normal immune globulin	Prevention (rarely required since vaccine introduced)
Hepatitis B	Human hepatitis B immune globulin	Prevention
Rabies	Human rabies immunoglobulin	Prevention
Varicella zoster	Human varicella zoster immune globulin	Prevention
SARS-CoV-2	Human neutralizing monoclonal antibody (e.g., sotrovimab)	Treatment

Box 42.19 Recommended childhood immunization schedules[a]

Time of immunization	Vaccine
As soon as possible after birth	BCG + HBV
6 weeks old	DPT + PV + HBV + Hib + PCV + rotavirus
10 weeks old	DPT + PV + HBV + Hib + PCV + rotavirus
14 weeks old	DPT + PV + Hib + PCV + rotavirus
9 or 12 months old	Measles (second dose at a minimum of 4 weeks later) + rubella
As soon as possible after 9 years old	HPV (two doses at least 6 months apart)

[a]WHO vaccination schedule for all children; different schedules or additional vaccines are recommended in some locations: https://www.who.int/immunization/policy/Immunization_routine_table2.pdf?ua=1. Mumps vaccine is also given in many low- and middle-income countries.
BCG, Bacille Calmette–Guérin; *DPT,* diphtheria, pertussis, tetanus triple vaccine; *HBV,* hepatitis B vaccine; *Hib, Haemophilus influenzae* type b; *HPV,* human papillomavirus vaccine; *PCV,* pneumococcal conjugate vaccine; *PV,* polio vaccine.

Haemophilus influenzae type b (Hib) and *S. pneumoniae* has proved highly effective in controlling invasive *H. influenzae* infection, notably meningitis, and reducing invasive disease, and can have a positive effect on antibiotic resistance (**Fig. 42.5**). A safe and immunogenic vaccine against *Neisseria meningitidis* serogroup B has now been included in the childhood vaccination schedules of some countries. Waning immune responses years after childhood immunization (e.g., tetanus and pertussis) may warrant additional or maintained vaccination boosting into adulthood.

Protection for travellers to developing and tropical countries

There has been a huge increase in the number of people travelling to low- and middle-income countries, mainly for recreation, leisure, and family visits. The risk of infection depends on the area to be visited, the type of activity, and the underlying health of the traveller. Advice should always be based on an individual assessment.

Box 42.20 Examples of active immunization available

Live attenuated vaccines
- Oral polio (Sabin) (not recommended, used only for outbreaks)
- MMR[a]
- Rotavirus
- Varicella zoster
- Yellow fever
- BCG
- Typhoid

Toxoids
- Diphtheria
- Tetanus

Recombinant vaccines
- Hepatitis B
- Varicella zoster (for shingles)
- Astra-Zeneca and Janssen/Johnson & Johnson SARS-CoV-2 vaccine (adenovirus vector)
- Respiratory syncytial virus

In activated and/or conjugate vaccines
- Hepatitis A
- Pertussis
- Typhoid: whole-cell and Vi antigen
- Polio (Salk) for routine use
- Influenza
- Human papillomavirus
- Cholera (oral): includes toxoid
- Meningococcus groups A, B, C, W135, and Y
- Rabies
- Anthrax
- Tick-borne encephalitis
- Japanese encephalitis
- Pneumococcal
- *Haemophilus influenzae* type b

Messenger RNA vaccines
- Moderna and Pfizer-BioNTech SARS-CoV-2 vaccine

Viral subunit vaccine
- Novavax SARS-CoV-2 vaccine (viral subunit)

[a]Individual vaccines for measles, mumps, and rubella (MMR) are not licensed in the UK.
BCG, Bacille Calmette–Guérin; *SARS-CoV-2,* severe acute respiratory syndrome coronavirus 2.

Protection for travellers against infection can be divided into three categories (**Box 42.21**). Because situations and risks can change rapidly, websites such as Travel Health Pro (which are regularly updated) are often the best source of advice on travel health (see Further reading).

Further reading

Llewelyn MJ, Budgell EP, Laskawiec-Szkonter M et al. Antibiotic review kit for hospitals (ARK-Hospital): a stepped-wedge cluster-randomised controlled trial. *Lancet Infect Dis* 2023; 23:207–221.
Pronovost P, Needham D, Berenholtz S et al. An intervention to decrease catheter-related bloodstream infections in the ICU. *N Engl J Med* 2006; 355:2725–2732.
Requejo J, Strong K, Agweyu A et al. Measuring and monitoring child health and wellbeing: recommendations for tracking progress with a core set of indicators in the Sustainable Development Goals era. *Lancet Child Adolesc Health* 2022; 6(5):345–352.
World Health Organization. *Five moments for hand hygiene*; https://www.who.int/gpsc/tools/Five_moments/en/.
World Health Organization. *WASH strategies*. https://www.who.int/teams/control-of-neglected-tropical-diseases/interventions/strategies/wash#:~:text=In%20March%202021%2C%20WHO%20launched,to%20drinking%20water%20and%20sanitation.
https://travelhealthpro.org.uk *Travel Health Pro*.
https://www.gov.uk/government/collections/immunisation-against-infectious-disease-the-green-book *The Green Book—Immunisation Against Infectious Disease*.
https://www.gov.uk/guidance/notifiable-diseases-and-causative-organisms-how-to-report *List of Notifiable Organisms (Causative Agents)* for the UK.

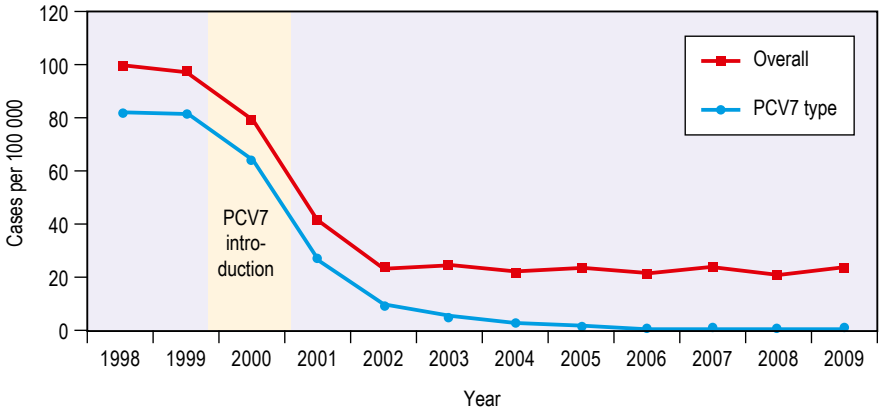


Fig. 42.5 Decline in invasive pneumococcal disease in young children after introduction of a pneumococcal conjugate vaccine (PCV) in the USA. *(From the Centers for Disease Control and Prevention, http://www.cdc.gov/pneumococcal/surveillance.html, with permission.)*

Box 42.21 Protection for travellers against infection

Personal protection, e.g.,
- Insect repellents
- Impregnated bed netting
- Avoidance of animals
- Care with food and drink
- Condoms
- Care with travel (i.e., road traffic accidents)

Chemoprophylaxis, e.g., antimalarials

Immunization, e.g.,
- Yellow fever
- Hepatitis A and B
- Typhoid
- Cholera
- Rotavirus
- Rabies

PRINCIPLES AND MECHANISMS OF INFECTION

Host–organism interactions, including the human microbiome

Although many infectious diseases are caused by exposure to exogenous pathogenic organisms (e.g., *Legionella* spp., malaria, leptospirosis, scabies), a significant proportion are caused by microorganisms that are part of the human microbiome.

Each of us is colonized by huge numbers of microorganisms (10^{14} bacteria, plus viruses, fungi, protozoa, and worms), with which we coexist. Many mucosal surfaces and organs (e.g., skin, lungs, and GI and genitourinary tracts) in humans have their own unique community of microbes, often called the ***microbiome***. Within the GI tract, for example, there are approximately 10^{11} to 10^{12} microbes per gram of content with each person carrying at least 160 species but often many more. A very high proportion of the GI tract microbiome bacteria are anaerobic, which means that antibiotics with anti-anaerobic activity (e.g., co-amoxiclav and metronidazole) can profoundly disrupt the nature of this microbial community. The human microbiome also acts as a reservoir for antimicrobial resistance, with many antimicrobials promoting the emergence and transfer of resistance elements within this habitat.

The development of the human microbiome is complex; normal vaginal delivery (as opposed to caesarean section), breastfeeding, diet, and the living environment, for example, exert major and lasting impacts. The effect of change in the human microbiome, particularly of the GI tract, on host metabolism and immunity is increasingly associated with health states, including, for example, cancer (in which it has been shown to change the patient's response to chemotherapy), diabetes and other metabolic diseases, host metabolism of medications, inflammatory bowel disease, mental health, and rheumatological diseases. Early use of antibiotics and disruption of the microbiome by other means have a prolonged impact on microbiome community structure and, it seems increasingly apparent, on health and disease states later in life. On the other hand, it is possible to protect and manipulate the microbiome therapeutically; existing examples include the use of faecal transplantation to treat recurrent *C. difficile* infection and selective decontamination of the digestive tract to prevent infections in critically ill patients. Wider therapeutic benefits of faecal transplantation and dietary manoeuvres to increase microbiome diversity are increasingly gaining an evidence base for treatment and prevention, revolutionizing our understanding of the pathogenesis of non-communicable diseases.

Development of infection by commensal organisms

Most microbiome microorganisms are commensal or symbiotic, living on or within humans without causing harm, but some are potentially pathogenic in the right circumstances. Infection and illness may be caused by usually harmless commensal microbes evading the body's defences and infecting abnormal sites. Examples include skin or mucosal commensals latching onto abnormal internal surfaces (e.g., *Staphylococcus epidermidis* infecting a prosthetic joint, an oral streptococcus causing endocarditis on an abnormal heart valve). UTI most commonly occurs, however, in the context of normal bladder epithelium, simply by physical transfer from the bowel to the bladder via the urethra. Bacterial liver abscess can occur by ascending infection of the biliary tract or via the portal venous route from a normal gut mucosa.

In other circumstances, the host immune response may be reduced, and normally commensal organisms cause 'opportunistic infection' (e.g., oral candidiasis in an immunocompromised host and following exposure to antibiotics). Common latent infections,

such as cytomegalovirus (CMV) and toxoplasma, exist in stable controlled equilibrium in host immune cells or other tissues of the immunocompetent host after initial infection but reactivate and multiply as 'opportunistic infections' after immunosuppression.

Symptoms of infection and the host response

The symptoms and signs of infection are a result of the interaction between host and pathogen. In some cases, such as the early stages of influenza, symptoms are almost entirely caused by killing of host cells by the invading organism. Usually, however, the harmful effects of infection are caused by a combination of direct microbial pathogenicity and the body's immune response to infection. In Gram-negative bloodstream infection, for example, much of the tissue damage is caused by cytokines released in an attempt to fight bacteria.

Specificity of microorganisms

Microbes are highly specific with respect to the organ or tissue they infect. For example, a number of viruses are hepatotropic, such as those responsible for hepatitis A, B, C, D, and E and for yellow fever. This predilection for specific sites in the body relates partly to the presence of appropriate receptors on different cell types, partly to the immediate environment in which the organism finds itself, and partly to the nature of the organism. *S. epidermidis*, for example, is a facultative anaerobe (i.e., can live in oxygenated and non-oxygenated environments) that usually lives on the human skin, an aerobic environment, but that has a predilection to forming biofilms on medical devices (e.g., of central venous lines and joint replacements) in less aerobic environments. In contrast, *Mycobacterium tuberculosis* is an obligate aerobe (i.e., it needs oxygen to replicate), and *Bacteroides* spp. and *Clostridium* spp. (found in the human GI tract) are obligate anaerobes (i.e., cannot live in atmospheric conditions), which partly explains the nature of the infections these bacteria cause. *Clostridium* spp. form spores, however, that enable survival in atmospheric conditions and thereby their spread to other human hosts.

Even within a species of bacterium, there are sometimes clear differences between strains with regard to their ability to cause certain clinical presentations. There are many different strains of *S. pneumoniae*, for example, with some asymptomatically colonizing the nasopharyngeal mucosa, some more likely to invade and cause pneumococcal bloodstream infection, and certain different invasive strains more likely to cause empyema or meningitis; other strains are more likely to cause pneumonia without bloodstream infection. Likewise, whereas some strains of *E. coli* (enterotoxin-producing strains) are more likely to cause GI disease, 'uropathogenic' strains are more likely to cause UTIs.

Pathogenesis

Fig. 42.6 summarizes some of the steps that are followed during the pathogenesis of infection. Pathogens have developed a variety of mechanisms to evade host defences. For example, some produce toxins directed at phagocytes—*S. aureus* (α-toxin), *S. pyogenes* (streptolysin), and *Clostridium perfringens* (α-toxin)—but others can survive within macrophages. Several pathogens possess a capsule that protects against complement activation (e.g., *S. pneumoniae*). Antigenic variation can also help evasion of host defences (e.g., antigenic shift and drift in influenza), which is the reason why influenza vaccines are modified from year to year.

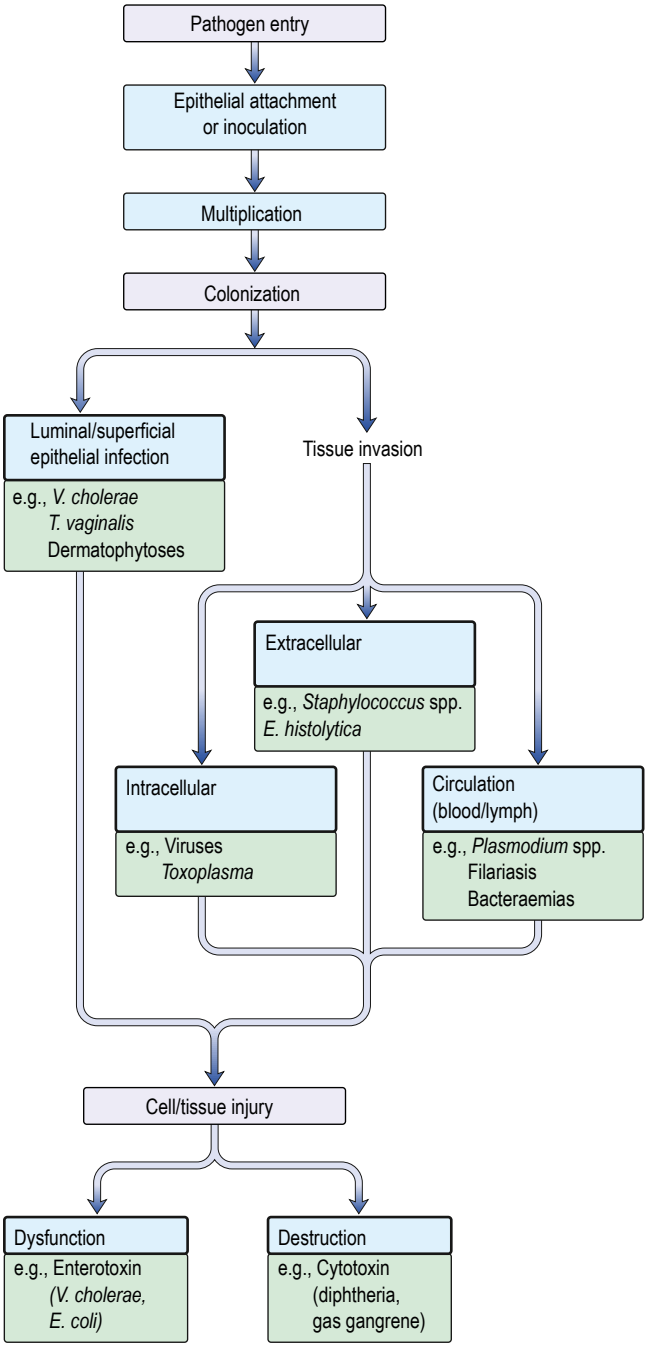


Fig. 42.6 The pathogenesis of infection.

Epithelial attachment

Many bacteria attach to epithelium cells using the adhesion molecules (called adhesins) of specific organelles called pili (or fimbriae). Enterotoxigenic and uropathogenic *E. coli* use fimbriae, for example, to attach to enterocytes and urothelial cells, respectively. After attachment, some bacteria, such as *S. epidermidis* and *Pseudomonas aeruginosa*, may produce an extracellular 'slime' layer and recruit additional bacteria to form a 'protective' biofilm. Biofilms are difficult to eradicate and are frequently associated with medical devices and persistent infections. Many viruses and protozoa (e.g., *Plasmodium* spp., *Entamoeba histolytica*) attach to

specific target-cell receptors, but others have attachment organs (e.g., the buccal plates of hookworms) that firmly grip the intestinal epithelium.

Colonization and invasion

After epithelial attachment, pathogens may remain on the surface epithelium or tissue invasion may follow. Invasion may result in:

- An intracellular location for the pathogen (e.g., viruses, *Mycobacterium* spp., *T. gondii*, *Plasmodium* spp.)
- An extracellular location for the pathogen (e.g., pneumococci, *E. coli*, *E. histolytica*)
- Invasion directly into the blood or lymph circulation (e.g., schistosome schistosomula and lymphatic filariasis).

Tissue dysfunction or damage

Microorganisms produce disease by a number of mechanisms, as described next.

Cell lysis

The presence of replicating viruses may interfere with host-cell metabolism, resulting in cell death.

Exotoxins and endotoxins

- ***Exotoxins*** have many diverse actions, including inhibition of protein synthesis (diphtheria toxin), neurotoxicity (*Clostridium tetani* and *Clostridium botulinum*) and enterotoxicity, which results in intestinal secretion of water and electrolytes *(E. coli, Vibrio cholerae).*
- ***Endotoxin*** is a lipopolysaccharide (LPS) in the cell wall of Gram-negative bacteria. It stimulates a variety of host responses characterized by release of pro-inflammatory cytokines and is responsible for many of the features of sepsis (see Chapter 12).

The clinical expression of disease caused by *S. aureus* varies according to site, invasion, and toxin production (**Box 42.22**). Furthermore, host susceptibility to infection may be linked to genetic or acquired defects in host immunity that may complicate intercurrent infection, injury, ageing, and metabolic disturbances (**Box 42.23**).

Host response to infection

Natural defences

The natural host defences to infection are those of an intact surface epithelium with local production of secretions; antimicrobial enzymes (e.g., lysozyme in the eye); and in the stomach, gastric acidity. The latter is reduced by proton pump inhibitors, increasing the risk of some GI infections, including *C. difficile* infection. The mucociliary escalator of the large airways is destroyed by smoking.

Immunological defences

The immunological response to infection is described in Chapter 6.

Metabolic and immunological consequences of infection

Fever

Body temperature is controlled by the thermoregulatory centre in the anterior hypothalamus and is usually maintained at approximately

Box 42.22 Clinical conditions produced by *Staphylococcus aureus*

Caused by invasion

Skin
- Furuncles
- Cellulitis
- Impetigo
- Carbuncles

Lungs
- Pneumonia
- Lung abscesses

Heart
- Endocarditis
- Pericarditis

Central nervous system
- Meningitis
- Brain abscesses

Bones and joints
- Osteomyelitis, arthritis

Miscellaneous
- Parotitis
- Pyomyositis
- Septicaemia
- Enterocolitis

Caused by toxin
- Staphylococcal food poisoning
- Scalded skin syndrome
- Bullous impetigo
- Staphylococcal scarlet fever
- Toxic shock syndrome

Box 42.23 Examples of host factors that increase susceptibility to staphylococcal[a] infections

Injury to skin or mucous membranes
- Abrasions
- Trauma (accidental or surgical)
- Burns
- Insect bites

Metabolic abnormalities
- Diabetes mellitus
- Uraemia

Foreign bodies[b]
- Intravenous and other indwelling catheters
- Cardiac and orthopaedic prostheses
- Tracheostomies

Abnormal leucocyte function
- Job's syndrome
- Chédiak–Higashi syndrome
- Steroid therapy
- Drug-induced leucopenia

Postviral infections
- Influenza

Miscellaneous conditions
- Excess alcohol consumption
- Malnutrition
- Malignancies
- Old age
- Recent antibiotic therapy

[a]Predominantly *Staphylococcus aureus.*
[b]Often *Staphylococcus epidermidis.*

36.8°C in health, with diurnal and inter-individual variation of about ±1.0°C. Gram-negative bacteria contain LPS and peptidoglycan; the latter is also a component of Gram-positive bacterial cell walls. Monocytes and dendritic cells recognize these, leading to the formation of inflammatory cytokines, such as interleukin (IL)-1, IL-6, IL-12, tumour necrosis factor alpha (TNF-α) and many others. These act on the thermoregulatory centre by increasing prostaglandin synthesis. The antipyretic effect of salicylates is brought about, at least in part, through their inhibitory effects on prostaglandin synthase.

Fever production has a positive effect on the course of infection. However, for every 1°C rise in temperature, there is a 13% increase in resting metabolic rate and oxygen consumption. Therefore, fever leads to increased energy requirements at a time when anorexia leads to decreased food intake, resulting in an increase in skeletal muscle breakdown to provide energy.

The inflammatory response

The inflammatory response is a fundamental biological response to a variety of stimuli, including microorganisms or their products, such as endotoxin, which acts on monocytes and macrophages. Non-phagocytic cells (lymphocytes, natural killer cells) are also involved. The release of cytokines—notably, TNF-α, IL-1, IL-6, and IL-17—leads to the release of a cascade of other mediators involved in direct antiviral responses, inflammation, and tissue remodelling, such as IFNs, prostaglandins, leukotrienes, and corticotropin. See Further reading for more information about the pathogenesis of infection by some key human bacterial, viral, and protozoan pathogens.

Further reading

Caldara M, Belgiovine C, Secchi E et al. Environmental, microbiological, and immunoloigcal features of bacterial biofilms associated with implanted medical devices. *Clin Microbiol Rev* 2022; 35(2):e0022120.
Chaguza C, Yang M, Jacques et al. Serotype 1 pneumococcus: epidemiology, genomics and disease mechanisms. *Trends Microbiol* 2022; 30(6):581–592.
Howden B, Giuleri S, Wong Fok Lung T et al. Staphylococcus aureus host interactions and adaptation. *Nat Rev Microbiol* 2023; 21(6):380–395.
Walter EJ, Hanna-Jumma S, Carraretto M et al. The pathophysiological basis and consequences of fever. *Crit Care* 2016; 20:200.
Whelan S, Lucey B, Finn K. Uropathogenic *Escherichia coli* (UPEC)-associated urinary tract infections: the molecular basis for challenges to effective treatment. *Microorganisms* 2023; 11(9):2169.
Wu Y, Zhang G, Wang Y et al. A review on maternal and infant microbiota and their implications for the prevention and treatment of allergic diseases. *Nutrients* 2023; 15(11):2483.

VIRAL INFECTIONS

Viruses are much smaller than other infectious agents and carry their genome as either DNA or RNA (but never both). Details of the structure, size, and classification of human DNA and RNA viruses are shown in **Boxes 42.24 and 42.25**, respectively. Because viruses are metabolically inert, they must live intracellularly, using the host cell for synthesis of viral proteins and nucleic acid. Viruses have a central nucleic acid (genome) surrounded by a protective protein coat (capsid) that is antigenically unique for a particular virus. The capsid imparts a helical or icosahedral structure to the virus. Some viruses also possess an outer envelope consisting of lipid and protein.

Outcomes of virus infection of a cell

Replication of viruses within a cell may result in sufficient distortion of normal cell function to result in cell death: a ***cytocidal*** or ***cytolytic*** infection. However, acute cell death is not an inevitable consequence of virus infection of a cell. In a ***chronic***, or ***persistent***, infection, virus replication continues throughout the lifespan of the cell but does not interfere with the normal cellular processes necessary for cell survival. Hepatitis B and C viruses may interact with cells in this way. Some viruses, such as those of the herpesvirus family, are able to go ***latent*** within a cell; in such a state, the virus genome is present within the cell, but there is very little, if any, production of viral proteins and no production of mature virus particles. Finally, some viruses are able to ***transform*** cells, leading to uncontrolled cell division; an example is Epstein–Barr virus (EBV) infection of B lymphocytes, resulting in the generation of an immortal lymphoblastoid cell line.

VIRUS INFECTIONS OF THE SKIN AND MUCOUS MEMBRANES

Many virus infections are associated with skin rashes (exanthems) or eruptions on the mucous membranes (enanthems) (see **Box 42.8**), so clinical syndromes based on the nature of the rash (which may be ***vesicular*** [i.e., consisting of fluid-filled vesicles] or ***maculopapular***) are often highly specific to the causative agent.

Box 42.24 Human DNA viruses

Structure		Approximate size	Family	Viruses
Symmetry	Envelope			
Icosahedral	–	80 nm	Adenovirus	**Adenoviruses** (>80 serotypes)
Icosahedral	+	100 nm (160 nm with envelope)	Herpesvirus	**Herpes simplex virus types 1 and 2** **Varicella zoster virus** **Cytomegalovirus** **Epstein–Barr virus** **Human herpesvirus type 6** **Human herpesvirus type 7** **Human herpesvirus type 8**
Icosahedral	+	42 nm	Hepadnavirus	**Hepatitis B virus**
Icosahedral	–	50 nm	Papillomavirus	**Human papillomaviruses** (>100 types)
Icosahedral	–	40–45 nm	Polyomavirus	**Polyomaviruses JC, BK, KI, WU**
Icosahedral	–	23 nm	Parvovirus	**Erythrovirus B19, bocavirus**
Complex	+	300 nm × 200 nm	Poxvirus	**Variola virus** **Vaccinia virus** **Mpox** **Cowpox** **Orf** **Molluscum contagiosum**

Box 42.25 Human RNA viruses

Structure		Approximate size	Family	Viruses
Symmetry	Envelope			
Icosahedral	–	30 nm	Picornaviruses	**Enteroviruses (including poliovirus, coxsackievirus, echovirus)** **Hepatitis A virus** **Rhinovirus**
Icosahedral	–	80 nm	Reoviruses	**Reovirus** **Rotavirus**
Icosahedral	+	50–80 nm	Togaviruses	**Alphaviruses, e.g., chikungunya**
Icosahedral	+	50–80 nm	Matonaviruses	**Rubella virus**
Icosahedral	+	50–80 nm	Flaviviruses	**Yellow fever virus** **Dengue virus** **Zika virus** **Japanese encephalitis virus** **West Nile virus** **Hepatitis C virus** **Tick-borne encephalitis virus**
Spherical	+	80–100 nm	Nairoviridae	**Congo–Crimean haemorrhagic fever**
Spherical	+	80–100 nm	Hantaviridae	**Hantavirus**
Spherical	+	80–100 nm	Phenuiviridae	**Rift Valley fever virus**
Spherical	–	35–40 nm	Caliciviruses	**Norovirus** **Sapovirus**
Spherical	–	28–30 nm	Astroviruses	**Astrovirus**
Spherical	–	30–34 nm	Hepeviruses	**Hepatitis E virus**
Helical	+	80–120 nm	Orthomyxoviruses	**Influenza viruses A, B, and C**
Helical	+	100–300 nm	Paramyxoviruses	**Parainfluenza viruses 1–4** **Measles virus** **Mumps virus** **Respiratory syncytial virus** **Human metapneumovirus** **Nipah virus** **Hendra virus**
Helical	+	80–220 nm	Coronaviruses	**229E, OC43, NL63, HK1, SARS, MERS, SARS-CoV-2**
Helical	+	60–175 nm	Rhabdoviruses	**Lyssavirus: rabies**
Helical	+	100 nm	Retroviruses	**Human immunodeficiency viruses 1 and 2** **Human T-cell leukaemia/lymphotropic viruses 1 and 2**
Helical	+	100–300 nm	Arenaviruses	**Lassa virus** **Lymphocytic choriomeningitis virus**
Pleomorphic	+	Filaments or circular forms; 100 × 130–2600 nm	Filoviruses	**Marburg virus** **Ebola virus**

MERS, Middle East respiratory syndrome; *SARS-CoV-2,* severe acute respiratory syndrome coronavirus 2.

Vesicular viral rashes

Herpes simplex virus infection

Members of the *Herpesviridae* family of viruses cause a wide range of human diseases (Box 42.26). The hallmark of all herpesvirus infections is the ability of the viruses to establish latent infections that persist for the life of the individual.

Two types of HSV (Fig. 42.7) have been identified: HSV-1 is the major cause of herpetic stomatitis, herpes labialis ('cold sore'), keratoconjunctivitis, and encephalitis, and HSV-2 causes genital herpes and may also be responsible for systemic infection in immunocompromised hosts. These divisions are not rigid, however, because HSV-1 can give rise to genital herpes and HSV-2 can cause infections in the mouth. The site of latency for both HSV-1 and 2 is the nerve cell body.

Herpes simplex virus type 1

The portal of entry of HSV-1 infection is usually via the mouth or, occasionally, the skin. The primary infection may go unnoticed or may produce a severe inflammatory reaction with vesicle formation leading to painful ulcers (gingivostomatitis; Fig. 42.8). Latent virus may be reactivated from the trigeminal ganglion by stress, trauma, febrile illnesses, and ultraviolet radiation, producing the recurrent form of the disease known as herpes labialis ('cold sore'). Approximately 70% of the population is infected with HSV-1, and recurrent infections occur in one-third of individuals. Reactivation often produces localized paraesthesiae in the lip before the appearance of a cold sore.

Complications of HSV-1 infection include transfer to the eye (dendritic ulceration, keratitis), acute encephalitis (see Chapter 11, 'Viral encephalitis' section), nail-bed infections (herpetic whitlow), and erythema multiforme (see Chapter 45, 'Erythema multiforme' section).

Box 42.26 Major diseases caused by human herpesviruses

Subfamily	Virus	Children	Adults	Immunocompromised
Alphaherpesviruses	Herpes simplex type 1	Stomatitis[a] Herpes simplex encephalitis	Cold sores Herpes simplex encephalitis Keratitis Erythema multiforme	Dissemination
	Herpes simplex type 2	Neonatal herpes[a]	Primary genital herpes[a] Recurrent genital herpes	Dissemination
	Varicella zoster virus	Chickenpox[a]	Shingles	Dissemination
Betaherpesviruses	Cytomegalovirus	Congenital[a]	Infectious mononucleosis[a] Hepatitis	Pneumonitis Retinitis Hepatitis Gastrointestinal
	Human herpesvirus type 6	Roseola infantum[a]		Pneumonitis Encephalitis
	Human herpesvirus type 7	Roseola infantum[a]		?Encephalitis
Gammaherpesviruses	Epstein–Barr virus		Infectious mononucleosis[a] Burkitt's lymphoma Nasopharyngeal carcinoma	Lymphoma
	Human herpesvirus type 8		Kaposi's sarcoma	Kaposi's sarcoma

[a]Signifies primary infection.

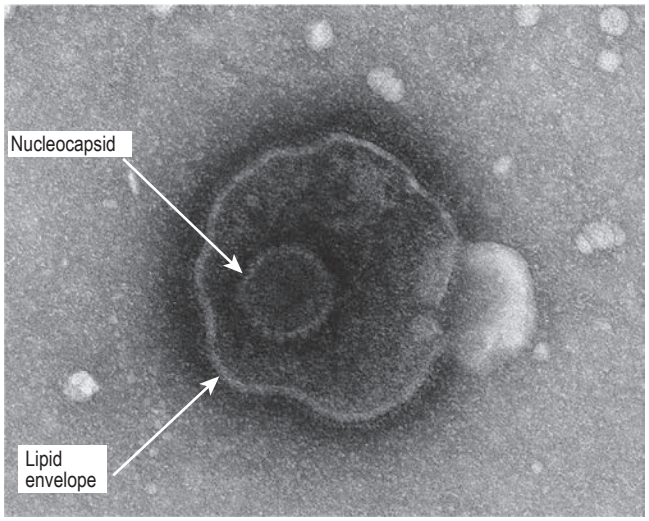


Fig. 42.7 Electron micrograph of herpes simplex virus.

Fig. 42.8 Primary herpes simplex type 1 (gingivostomatitis).

Herpes simplex virus type 2

The clinical features, diagnosis, and management of genital herpes are described in Chapter 43 ('Herpes simplex' section). The virus remains latent in the sacral ganglia and, during recurrence, can produce a radiculomyelopathy, with pain in the groin, buttocks, and upper thighs. Primary anorectal herpes infection is common in males who have sex with males.

Neonates may develop primary HSV infection after vaginal delivery in the presence of active genital HSV infection in the mother, particularly if the maternal disease is a primary, rather than a recurrent, infection. The disease in the baby varies from localized skin lesions (~10%–15%) to widespread visceral disease most often affecting the lungs, liver, and brain, with a poor prognosis. Caesarean section should therefore be performed if there is evidence of primary maternal HSV infection in the last trimester or if active genital HSV infection is present during labour.

In pregnancy and in immunocompromised patients, such as those receiving intensive cancer chemotherapy or those with AIDS, HSV infection may become disseminated, involving many of the viscera. In severe cases, death may result from hepatitis and encephalitis. Eczema herpeticum is a serious complication in individuals with eczema, in whom the non-intact skin allows spread of lesions across large areas and bloodstream access, which may lead to herpetic involvement of internal organs.

Diagnosis and management

Confirmation of clinical diagnosis is most commonly obtained by detection of HSV DNA in vesicle fluid or a peripheral blood sample

by genome amplification. Treatment of patients with HSV-associated disease is with aciclovir and derivatives.

Varicella (chickenpox) and herpes zoster (shingles)

Infection with varicella zoster virus (VZV), another herpesvirus, produces two distinct diseases: varicella (chickenpox) and herpes zoster (shingles). The ***primary infection***, chickenpox, usually occurs in childhood, the virus entering through the mucosa of the upper respiratory tract. In some countries (e.g., the Indian subcontinent), a different epidemiological pattern exists, with most infections occurring in adulthood. Chickenpox is rarely contracted twice by the same individual. Infectious virus is spread from the throat and from fresh skin lesions by air-borne transmission or direct contact. The period of infectivity in chickenpox extends from 2 days before the appearance of the rash until the skin lesions are all at the crusting stage. After recovery from chickenpox, the virus remains latent in dorsal root and cranial nerve ganglia. ***Reactivation*** of infection results in shingles.

Clinical features of chickenpox

About 14 to 21 days after exposure to VZV, a brief prodromal illness of fever, headache, and malaise heralds the eruption of chickenpox, characterized by the rapid progression of macules to papules to vesicles to pustules in a matter of hours (Fig. 42.9). In young children, the prodromal illness may be very mild or absent. The illness tends to be more severe in older children and can be debilitating in adults. The lesions occur on the face; scalp; trunk; and, to a lesser extent, the extremities. It is characteristic to see lesions at all stages of development on the same area of skin. Fever subsides as soon as new lesions cease to appear. Eventually, the pustules crust and heal without scarring.

Complications include pneumonia, which generally begins 1 to 6 days after the skin eruption, and bacterial superinfection of skin lesions. Pneumonia is more common in adults than in children, and cigarette smokers are at particular risk. The pulmonary symptoms are usually more striking than the physical findings, although a chest X-ray usually shows diffuse changes throughout both lung fields. Central nervous system (CNS) involvement occurs in about 1 per 1000 patients and most commonly presents as an acute truncal cerebellar ataxia. Immunocompromised people are susceptible to disseminated infection, with multiorgan involvement. Pregnancy increases the risk of severe chickenpox and, in addition, there is a risk of intrauterine infection with structural damage to the fetus. (If maternal infection is within the first 20 weeks of pregnancy, the risk of varicella embryopathy is 1%–2%.)

Clinical features of shingles

Shingles (see Chapter 45, 'Herpes zoster [shingles]' section) arises from the reactivation of VZV latent within the dorsal root or cranial nerve ganglia. It may occur at any age but is most common in older adults, producing skin lesions similar to those of chickenpox, although classically, they are unilateral and restricted to a sensory nerve (i.e., dermatomal) distribution (Fig. 42.10). The onset of a shingles rash is usually preceded by severe dermatomal pain, indicating the involvement of sensory nerves in its pathogenesis. The virus is disseminated from freshly formed vesicles and may cause chickenpox in susceptible contacts.

The most common complication of shingles is postherpetic neuralgia (PHN; see Chapter 28, 'Neuralgia' section). The occurrence of PHN is related to age (more likely in older adults) and the intensity of the original shingles rash.

Fig. 42.9 Chickenpox in an adult. Generalized varicella zoster virus infection.

Fig. 42.10 Shingles. Varicella zoster virus affecting a dermatome. *(Reproduced with kind permission of the Imperial College School of Medicine.)*

Diagnosis

The diseases are usually recognized clinically but can be confirmed by detection of VZV DNA within vesicular fluid, electron microscopy, immunofluorescence or culture of vesicular fluid, and serology.

Prevention and management

Chickenpox usually requires no treatment in healthy children, and infection results in life-long immunity. Aciclovir and derivatives are licensed for this indication in the USA. However, the disease may be fatal in immunocompromised people, who should therefore be offered protection after exposure to the virus, with zoster-immune globulin (ZIG) and high-dose aciclovir at the first sign of development of the disease.

Antiviral therapy with aciclovir or a similar drug may be offered to patients older than the age of 16 years if they present within 72 hours of onset. Prophylactic ZIG is recommended for susceptible pregnancies exposed to VZV, and if chickenpox develops, aciclovir treatment should be given if there are concerns around varicella pneumonia, but the drug is *not* licensed for use in pregnancy. Babies should be protected by ZIG if delivery occurs within 7 days of the onset of the mother's rash. An effective live attenuated varicella vaccine is licensed as a routine vaccination of childhood in the USA; it is available on a named patient basis in the UK and for susceptible healthcare workers.

Box 42.27 Enteroviruses

	Enterovirus A species	Enterovirus B species	Enterovirus C species	Enterovirus D species
Old nomenclature	Coxsackie A16, EV-A71	Coxsackie A9, coxsackie B viruses 1–6 All echoviruses	Coxsackie A1 Polioviruses 1–3 EV-C95 et seq	EV-D68
Disease				
Herpangina	+++	+		
Hand, foot, and mouth	+++	+		
Erythematous rashes	+	+++		
Conjunctivitis	+			
Paralysis	+	+	+++	+
Meningitis	++	+++	+	+
Encephalitis	++	++		+
Myocarditis, pericarditis	+	+++		
URTI				++
Myositis (Bornholm disease)	+	+++		

+++, Often causes; *++*, sometimes causes; *+*, rarely causes; *±*, possibly causes; *EV*, enterovirus; *URTI*, upper respiratory tract infection.

Patients with ***shingles*** involving motor nerves (e.g., the VIIth cranial nerve, leading to facial palsy) are also treated with aciclovir (or derivatives). Antiviral therapy has been shown to reduce the burden of PHN when treatment is given in the acute phase and is therefore recommended for shingles patients older than 65 years. Shingles involving the ophthalmic division of the trigeminal nerve has an associated 50% incidence of acute and chronic ophthalmic complications. Early treatment with aciclovir reduces this to 20% or less. As for chickenpox, all immunocompromised individuals should be given aciclovir at the onset of shingles, no matter how mild the attack appears when it first presents.

Both recombinant and live attenuated vaccines have been shown to reduce shingles-related morbidity and PHN. Vaccination is recommended for all adults older than the age of 60 years (USA). In the UK, the age limit is currently older than 70 years, but this may be reduced to 65 years.

Picornavirus infections

The picornaviruses (pico = small) are a large family of RNA viruses, which includes the enteroviruses (EVs), rhinoviruses, parechoviruses, and hepatitis A virus (see Chapter 23, 'Hepatitis A' section). The term 'enterovirus' refers to their enteric means of spread, that is, via the faecal–oral route. The EVs include poliovirus types 1 to 3, coxsackie A and B viruses, echoviruses, and EVs 68 to 71. There are several newly described EVs yet to be officially classified. All EVs are now grouped into four species, A to D, with the species letter being incorporated in the virus type name (e.g., EV D68). They are responsible for a broad spectrum of disease involving the skin and mucous membranes; muscles; nerves; heart (**Box 42.27**); and, rarely, other organs, such as the liver and pancreas. They are frequently associated with pyrexial illnesses and are the most common causes of aseptic meningitis (see later).

Herpangina

This is mainly caused by coxsackie A viruses. It presents with a vesicular eruption on the fauces, palate, and uvula, and the lesions evolve into ulcers. The illness is usually associated with fever and headache but is short-lived, with recovery taking place within a few days.

Hand, foot, and mouth disease

This is mainly caused by coxsackievirus A16 or A10 (EV A16, EV A10, respectively). It is also the main feature of infection with EV A71. Oral lesions are similar to those seen in herpangina but may be more extensive in the oropharynx. Vesicles and a maculopapular eruption also appear, typically on the palms of the hands and the soles of the feet but also on other parts of the body. This infection commonly affects children. Recovery occurs within 1 week.

Poxvirus infections

Smallpox (variola)

This disease was eradicated in 1977 after an aggressive vaccination policy. Its possible use in bioterrorism has resulted in the reintroduction of smallpox vaccination in some countries.

Mpox (formerly known as monkeypox)

This was, until 2022, a rare zoonosis arising from infection with monkeypox virus in small villages in tropical rainforests in several countries of West and Central Africa. Its clinical effects, including a generalized vesicular rash, were indistinguishable from those of smallpox, but person-to-person transmission was unusual. Most infections occurred in children who were not vaccinated against smallpox. Disease could be severe, with mortality rates of 10% to 15% in unvaccinated individuals. Serological surveys indicated that several species of squirrel were the likely animal reservoir. The disease was renamed mpox in 2022 because there is very little evidence of monkeys being involved in its natural history. The virus was introduced into the USA in 2003 via small West African mammals illegally imported as pets. Widespread infection of prairie dogs resulted, and there were 37 laboratory-confirmed cases in humans.

However, in 2022, the WHO declared a Public Health Emergency of International Concern because of an unprecedented surge in reported cases in more than 70 countries around the world, with many previously unrecognized clinical features. In the UK, cases were overwhelmingly in males who have sex with males, presenting with penile, perianal, and rectal symptoms, with or without skin lesions (see Chapter 43, 'Mpox' section and Chapter 45, 'Mpox [formerly monkeypox]' section). The lesion count was often low at presentation, with some atypical single lesions mimicking abscesses e.g., in the

throat. Prodromal illness was often absent. Only a minority of patients required hospital admission and then mostly for pain management.

Diagnosis is by demonstration of mpox virus DNA in a lesion swab, the most common differential being HSV infection. Treatment is symptomatic, although tecovirimat, an antiviral developed for the treatment of smallpox, has also been recommended for severe cases. Prevention is by active immunization with a modified vaccinia Ankara vaccine, although there have been problems in generating adequate supplies.

Cowpox

Cowpox produces large vesicles that are classically found on the hands of individuals in contact with infected cows. The lesions are associated with regional lymphadenitis and fever. Cowpox virus has been found in a range of species, including domestic and wild cats, and the reservoir is thought to exist in a range of rodents.

Vaccinia virus

This laboratory virus does not occur in nature in either humans or animals. Its origins are uncertain, but it has been invaluable in its use as the vaccine to prevent smallpox.

Orf

This poxvirus causes contagious pustular dermatitis in sheep and hand lesions in humans.

Molluscum contagiosum

This is discussed in Chapter 43 ('Molluscum contagiosum' section).

Human papillomavirus infections

These cause warts (see Chapter 43, 'Human papillomavirus [anogenital warts]' section) and cervical cancer (see Chapter 47, 'Gynaecological malignancy' section).

Maculopapular viral rashes

Measles (rubeola)

Measles virus is a paramyxovirus (see **Box 42.25**). Measles is a highly communicable disease that occurs worldwide. With the introduction of aggressive immunization policies, the global incidence of and mortality rate from measles fell by 80% between 2000 and 2017, but there were still approximately 128 000 measles-related deaths in 2021, mostly in children younger than the age of 5 years and largely in Africa and South-east Asia. It is spread by droplet infection, and the period of infectivity is from 4 days before until 2 days after the onset of the rash.

Clinical features

The incubation period is 8 to 14 days. Two distinct phases of the disease can be recognized.

- The ***pre-eruptive and catarrhal stage***, the stage of viraemia and viral dissemination. Malaise, fever, rhinorrhoea, cough, conjunctival suffusion, and the pathognomonic Koplik's spots (small, greyish, irregular lesions surrounded by an erythematous base, found in greatest numbers on the buccal mucous membrane opposite the second molar tooth) are present during this stage.
- The ***eruptive or exanthematous stage***, characterized by the presence of a maculopapular rash that initially occurs on the face, chiefly the forehead, and then spreads rapidly to involve the rest of the body (**Fig. 42.11**). At first, the rash is discrete, but later it may become confluent and patchy, especially on the face and neck. It fades in about 1 week and leaves behind a brownish discoloration.

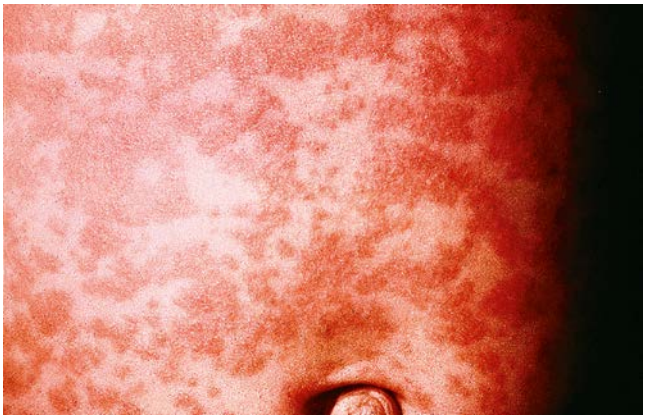


Fig. 42.11 Measles. *(Courtesy of Dr MW McKendrick, Royal Hallamshire Hospital, Sheffield.)*

The most feared complication in an immunocompetent child is post-measles encephalomyelitis, occurring within 2 weeks of the rash in 1 in 1000 to 1 in 5000 cases of measles. This is most likely an autoimmune disorder triggered by the measles infection because virus is not present in the brain substance. The prognosis is poor, with a high mortality rate (30%) and severe residual damage in survivors.

Measles carries a high mortality rate in malnourished people and those who have other diseases. Complications are common in such individuals and include bacterial pneumonia, bronchitis, otitis media, and severe diarrhoea. Less commonly, myocarditis, hepatitis, and encephalomyelitis occur. In malnourished people and those with defective cell-mediated immunity (CMI), the classical maculopapular rash may not develop, and there may be widespread desquamation. The virus also causes the rare condition subacute sclerosing panencephalitis, which may follow measles infection occurring early in life (before 18 months of age). Persistence of the virus with reactivation before puberty results in accumulation of virus in the brain, progressive mental deterioration, and a fatal outcome.

Maternal measles, unlike rubella, does not cause fetal abnormalities. It is, however, associated with spontaneous abortions and premature delivery.

Diagnosis and management

Most cases of measles are diagnosed clinically, but detection of measles-specific IgM in blood or oral fluid or genome or antigen detection from nasopharyngeal aspirates or throat swabs should be used to confirm the diagnosis.

Management is supportive. Antibiotics are indicated only if secondary bacterial infection occurs. Vitamin A supplements lower the number of deaths from measles by 50% and reduce eye damage and blindness; therefore, two doses of vitamin A supplements given 24 hours apart is recommended for all children with measles.

Prevention

A previous attack of measles confers a high degree of immunity, and second attacks are uncommon. Normal human immunoglobulin given within 5 days of exposure effectively aborts an attack of measles. It is indicated after exposure for previously unimmunized children younger than 3 years of age, in pregnancy, and for those with debilitating disease.

Active immunization

Children in the UK are immunized with the combined mumps, measles, and rubella (MMR) vaccine. In low- and middle-income countries, the first measles vaccination is given at 9 months.

Rubella

Rubella ('German measles') is caused by a spherical, enveloped RNA virus of the *Rubivirus* genus belonging to the Matonaviridae family. Although the disease can occur sporadically, epidemics are not uncommon. It has a worldwide distribution. Spread of the virus is via droplets; maximum infectivity occurs before and during the time the rash is present.

Clinical features

The incubation period is 14 to 21 days, averaging 18 days. The clinical features are largely determined by age, symptoms being mild or absent in children younger than 5 years.

During the prodrome, the patient may develop malaise, fever, mild conjunctivitis, and lymphadenopathy involving particularly the suboccipital, post-auricular, and posterior cervical groups of lymph nodes. Small petechial lesions on the soft palate (Forchheimer spots) are suggestive but not diagnostic. Splenomegaly may be present.

The eruptive or exanthematous phase usually occurs within 7 days of the initial symptoms. The rash first appears on the forehead and then spreads to involve the trunk and limbs. It is pinkish red, macular, and discrete, though some of these lesions may coalesce (Fig. 42.12). It usually fades by the second day and rarely persists beyond 3 days after its appearance.

Complications

Complications are rare. They include superadded pulmonary bacterial infection, arthralgia (more common in females), haemorrhagic manifestations caused by thrombocytopenia, and encephalitis. Maternal infection in pregnancy may result in the congenital rubella syndrome. Rubella affects the fetuses of up to 90% of all infections during the first trimester of pregnancy. The incidence of congenital abnormalities diminishes in the second trimester, and no ill-effects result from infection in the third trimester.

Congenital rubella syndrome is characterized by the presence of fetal cardiac malformations (especially patent ductus arteriosus and ventricular septal defect), eye lesions (especially cataracts), microcephaly, mental retardation, and deafness. Hepatosplenomegaly, myocarditis, interstitial pneumonia, and metaphyseal bone lesions also occur.

Diagnosis and management

Laboratory diagnosis is essential, especially in pregnancy, to distinguish the illness from other virus infections (e.g., erythrovirus B19, echovirus) and drug rashes. This is achieved by detection of rubella-specific IgM by enzyme-linked immunosorbent assay (ELISA) in an acute serum sample, preferably confirmed by demonstration of IgG seroconversion (or a rising titre of IgG) in a subsequent sample taken 14 days later. Viral genome can be detected in throat swabs (or oral fluid samples); urine; and, in the case of intrauterine infection, the products of conception.

Management is supportive.

Prevention

Several live attenuated rubella vaccines have been used and have had great success in preventing this illness. They have been successfully combined with measles and mumps vaccines in the MMR vaccine. Use of the vaccine is contraindicated during pregnancy or if there is a likelihood of pregnancy within 3 months of immunization. Inadvertent use of the vaccine during pregnancy has not, however, revealed a risk of teratogenicity.

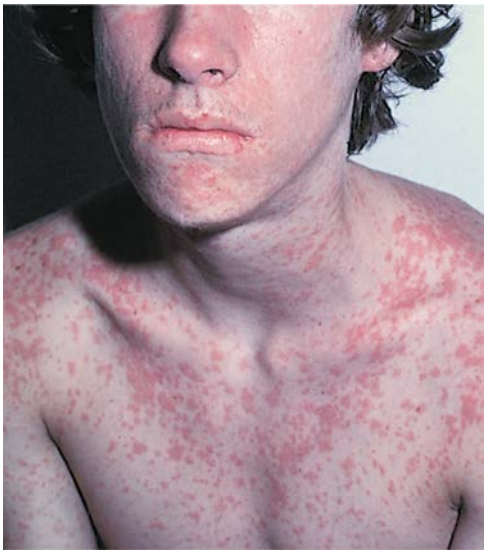


Fig. 42.12 **Rubella rash.**

Erythrovirus infections

Human erythrovirus (also known as parvovirus) B19 causes erythema infectiosum (fifth disease), a common infection in schoolchildren. The rash is typically on the face (the ***'slapped cheek'*** appearance). The patient is well, and the rash can recur over weeks or months. Asymptomatic infection occurs in 20% of children. Non-specific respiratory tract illness may also occur. In adults, the rash may be clinically indistinguishable from that of rubella. Moderately severe self-limiting polyarthropathy is common if infection occurs in adulthood, especially in females (this is also true of rubella). The virus infects bone marrow cells, and aplastic crisis may occur in patients with chronic haemolysis (e.g., sickle cell disease). Chronic infection with anaemia occurs in immunocompromised subjects. Hydrops fetalis (3% risk) and spontaneous abortion (9% risk) may result from maternal infection in the first and second trimesters of pregnancy.

Diagnosis of acute infection is by detection of parvovirus-specific IgM or DNA.

Human herpesvirus types 6 and 7 infection

These human herpesviruses (HHVs) occur worldwide and exist as latent infections in more than 85% of the adult population. They are spread by contact with oral secretions. HHV type 6 (HHV-6) causes roseola infantum (exanthem subitum), which presents as a high fever followed by a generalized macular rash in infants. HHV-6 is a common cause of febrile convulsions, and aseptic meningitis or encephalitis is a rare complication.

The full spectrum of disease caused by HHV-7 has not yet been fully characterized, but like HHV-6, it may cause roseola infantum in infants. Both viruses have been associated with encephalitis in heavily immunosuppressed patients, such as bone marrow transplant recipients.

Management

Supportive management is recommended for patients with the common infantile disease. Ganciclovir can be used in immunocompromised patients.

Further reading

Gessain A, Nakoune E, Yazdanpanah Y. Monkeypox. *N Engl J Med* 2022; 387:1783–1793.
Hubschen JM, Gouandjika-Vasilache I, Dina J. Measles. *Lancet* 2022; 399:678–690.

VIRUS INFECTIONS OF THE RESPIRATORY TRACT

Many viruses can cause URTIs, for example, in the nose (rhinitis), throat (tonsillitis, pharyngitis), sinuses (sinusitis), ear (otitis media), eye (conjunctivitis), or larynx (laryngitis) (**Box 42.28**). Infections below the larynx are referred to as LRTIs. URTIs are common but relatively trivial; LRTIs are less common but may result in hospitalization and even death.

Upper respiratory tract infections

The common cold: coryza and rhinitis

Rhinovirus infection is the most common cause of the common cold (see Chapter 20, 'Viral rhinitis' section). Peak incidence rates occur in the colder months, especially spring and autumn. There are multiple rhinovirus serotypes (>100), which explains why infection occurs throughout life and makes vaccine control impracticable. In contrast to EVs, which replicate at 37°C, rhinoviruses grow best at 33°C (the temperature of the upper respiratory tract), which explains the localized disease characteristic of common colds.

Human coronaviruses (CoVs) were first isolated in the mid-1960s, and the majority of isolates (related to the reference strains 229E and OC43) have been associated with common colds. In 2004 and 2005, two new CoV infections of humans were described, NL63 and HKU1. These are also associated with coryzal symptomatology. CoVs have recently become of interest owing to the discovery of three new viruses causing life-threatening LRTI (see later).

Parainfluenza

Parainfluenza is caused by the parainfluenza viruses types I to IV; these have a worldwide distribution and cause acute respiratory disease.

Parainfluenza is essentially a disease of children and presents with features similar to those of the common cold. When infection is severe, a brassy cough with inspiratory stridor and features of laryngotracheitis (croup) are present. Fever usually lasts for 2 to 3 days and may be more prolonged if pneumonia develops. The development of croup is caused by submucosal oedema and consequent airway obstruction in the subglottic region. This may lead to cyanosis, subcostal and intercostal recession, and progressive airway obstruction. Infection in immunocompromised patients is usually prolonged and may be severe. Management is supportive with oxygen, humidification, and sedation when required. The role of steroids and the antiviral agent ribavirin is controversial.

Adenovirus infection

More than 80 adenovirus serotypes split into 7 species labelled A to G have been identified as human pathogens, infecting a number of different cell types and therefore resulting in different clinical syndromes. Adenovirus infection commonly presents as an acute pharyngitis, and extension of infection to the larynx and trachea in infants may lead to croup. By school age, the majority of children show serological evidence of previous infection. Certain subtypes produce an acute conjunctivitis associated with pharyngitis. In adults, adenovirus causes acute follicular conjunctivitis and, rarely, pneumonia that is clinically similar to that produced by *M. pneumoniae*. Adenoviruses F40 and F41 cause gastroenteritis without respiratory disease, and adenovirus infection may be responsible for acute mesenteric lymphadenitis in children and young adults, which may lead to intussusception in infants. Infection in an immunocompromised host, such as a bone marrow transplant recipient, may result in multisystem failure and fatal disease. In 2022, a number of countries experienced a dramatic increase in cases of severe hepatitis in children mostly younger than the age of 10 years. There is considerable evidence that infection with Ad-F41, together with adeno-associated virus type 2, in individuals carrying particular HLA-D alleles, was a major contributor to this syndrome.

Box 42.28 Virus infections of the respiratory tract

Virus	Disease
Upper respiratory tract infections	
Rhinoviruses (>100 serotypes)	Common cold, rhinitis
Parainfluenza viruses 1–4	Croup
Coronaviruses OC43, 229E	Common cold
Respiratory syncytial virus	Cough, sore throat
Human metapneumovirus	Cough, sore throat
Adenoviruses (>80 serotypes)	Pharyngitis, conjunctivitis Common cold (see Chapter 20, 'Viral rhinitis' section)
Enteroviruses (coxsackie, echo, EV D68)	Bronchiolitis in young babies
Lower respiratory tract infections	
Influenza A, B	Epidemics and pandemics of influenza
Respiratory syncytial virus	Bronchiolitis, pneumonia in young babies
SARS-CoV-2	COVID-19
Rarely	
Cytomegalovirus	Pneumonitis in immunocompromised
Varicella zoster virus	Pneumonia in adults with primary infection
Measles	Giant-cell pneumonia in immunocompromised
Coronaviruses	Pneumonia, regional or widespread outbreaks

COVID-19, Coronavirus disease 2019; *EV,* enterovirus; *SARS-CoV-2,* severe acute respiratory syndrome coronavirus 2.

Other viral causes of upper respiratory tract infection

Human erythrovirus B19 infection may present as a URTI. Bocavirus is a recently identified parvovirus, which accounts for around 3% to 5% of respiratory tract infections in young children. EVs, such as coxsackie and echoviruses, are occasionally found in respiratory tract secretions of young babies, and there has been much interest recently in EV D68, which appears to infect the respiratory tract predominantly. WU and KI polyomaviruses may be associated with respiratory tract infections in young children.

Lower respiratory tract infections

Viral infections resulting in life-threatening LRTIs are SARS-CoV-2, the MERS-CoV and the original SARS-CoV-2, influenza, and respiratory syncytial virus (RSV). However, viral pneumonia may occasionally arise from infection with a number of other viruses (see **Box 20.29**), and as evidenced by the emergence of SARS-CoV-2, other, as yet unknown, viruses may yet appear.

Box 42.29 Emergent human coronaviruses in the 21st century

	SARS-CoV (2002–2003)	MERS-CoV (2012–)[a]	SARS-CoV-2 (2019–)[b]
Cases *(n)*	8096	2604	676 609 951
R_0 (uncontrolled)	2–5	<1–8	3.1
Deaths *(n)*	774	936	6 881 955
Case fatality rate (%)	9.6	36	1–2

[a]Data derived from World Health Organization, May 2023
[b]Data derived from Johns Hopkins COVID tracker March 2023.
MERS-CoV, Middle East respiratory syndrome coronavirus; *SARS-CoV-2,* severe acute respiratory syndrome coronavirus 2.

Coronavirus infections

CoVs are a large family of positive single-stranded RNA viruses that infect a wide range of species, causing a variety of clinical syndromes. Human CoV OC43, 229E, NL63, and HKU1 have been recognized for many years as the second commonest cause (behind rhinoviruses) of a common-cold-like illness. Interest in HCoV dramatically increased in November 2002 when an apparently new viral LRTI occurred in China and spread rapidly in other parts of the Far East and, from there, across the world. This disease, of which bronchopneumonia was a major feature, became known as 'severe acute respiratory syndrome' (SARS); it was caused by a previously unknown CoV (SARS-CoV-2). Similarity of this virus to CoVs isolated from civet cats, raccoons, and ferret badgers indicates that SARS is a zoonotic disease. Bats are the likely host species. The epidemic was finally brought under control in the summer of 2003 with implementation of strict quarantining of infected individuals and contacts of infected individuals to prevent further spread, by which time there had been more than 8000 cases with approximately 800 deaths (**Box 42.29**). Naturally acquired infection has not been reported since.

In 2012, a new CoV, MERS-CoV, was identified in a patient who died of acute respiratory failure in Saudi Arabia. There have been more than 2600 cases since, across 27 countries, although mostly in the Middle East; the mortality rate is around 36%. The most likely source is from dromedary camels. Person-to-person spread within healthcare settings is another major source, with the largest outbreaks in Saudi Arabia, United Arab Emirates, and the Republic of Korea. New cases continue to occur sporadically.

In December 2019, yet another new CoV was identified in an outbreak of severe respiratory disease in Wuhan, China, which spread rapidly to cause a major pandemic. This virus is now known as SARS-CoV-2, and the associated disease is coronavirus disease 2019 (COVID-19). Although the case-fatality rate is not as high as with SARS-CoV or MERS-CoV, this virus is clearly more transmissible and has spread explosively around the world. The impact of this pandemic has been incalculable, with global infection rates of almost 1 billion people and almost 7 million recorded deaths. As with all RNA viruses, there is a relatively high rate of spontaneous mutation during replication, which results in genetic drift. Thus, the SARS-CoV-2 pandemic has seen global waves of infection associated with the evolution of novel variants. Major waves have involved alpha, beta, delta, and omicron variants.

The clinical presentation of SARS-CoV-2 infection is as an acute illness with high temperature, cough, and evidence of impaired lung function. The pathogenesis of disease is thought to involve initial viral replication followed a few days later by an exaggerated inflammatory response ('cytokine storm'), which is responsible for the severe pulmonary complications. An unusual manifestation is loss of taste, smell, or both. Many surviving patients have long-term clinical consequences, in a syndrome referred to as 'long COVID'. The mortality rate is age related, with much higher fatality rates in those older than 65 years old and in patients with underlying co-morbidities. The management and complications of COVID-19 illness is discussed in detail in Chapter 20 ('COVID-19 pneumonitis' section).

Diagnosis

Diagnosis of CoV infection is by detection of virus in a respiratory sample, with nose or throat swab being the most useful. The most sensitive approach is that of genome detection, usually by reverse transcription polymerase chain reaction (RT-PCR). This is mostly done in laboratories, but there are point-of-care PCR kits that can be used for less centralized diagnosis. Antigen detection assays, usually using lateral flow technology, are also available for SARS-CoV-2. These allow for self-sampling and rapid diagnostic testing in the home but are intrinsically less sensitive than genome detection assays. Both assays may generate false-negative results if there is sufficient genomic or antigenic change arising from virus evolution.

Serological assays for detection of antibodies have been developed, but they are mainly used for surveillance purposes. It is possible to differentiate antibody positivity arising from immunization (will comprise anti-spike antibodies only) as opposed to those arising from natural infection, in which a range of antibodies against different viral proteins are present.

Management

The sheer scale of the COVID-19 pandemic enabled very large clinical trials to be conducted very rapidly to test the efficacy of a variety of therapeutic approaches. Many existing drugs were repurposed for use in SARS-CoV-2 infection, but none showed significant effects. Three novel antiviral agents, remdesivir, molnupiravir, and nirmatrelvir, have been shown to be effective in reducing the risks of serious and life-threatening complications of infection, with the latter being the most effective. A number of monoclonal anti-spike antibodies (e.g., sotrovimab) have been developed, but these lose potency and clinical usefulness as the virus evolves in different waves of infection. Anti-inflammatory approaches, including dexamethasone and monoclonal antibodies, which block IL-6 receptors or Janus kinase, are of benefit in seriously ill hospitalized patients.

Prevention

Infection control approaches to reduce the spread and burden of SARS-CoV-2 infection have included relatively straightforward measures such as isolation of infected patients admitted to hospital; wearing of PPE, including face masks, by healthcare workers and by the public in general; social distancing; and encouraging adequate hand hygiene through to drastic ones such as school closures and societal lockdowns, which had major health, economic, and psychological consequences.

There has been an extremely rapid development of vaccines, most of which have focused on inducing an immune response to the virus spike (S) protein. There are a number of strategies in use, including mRNA, DNA vector (e.g., adenovirus), whole inactivated virus, and S protein subunit vaccines. Large-scale clinical trials have generated impressive protection data, and the emergence of this phalanx of vaccines has been critical in controlling the pandemic.

Influenza

Influenza viruses belong to the family of orthomyxoviruses, having a segmented negative-strand RNA genome. The influenza virus is a spherical or filamentous enveloped virus. Haemagglutinin (H), a surface glycoprotein, aids attachment of the virus to the surface of susceptible host cells at specific sialic acid receptor sites. Release of progeny viruses from the cell surface, effected by budding through the cell membrane, is facilitated by the enzyme neuraminidase (N), which is also present on the viral envelope. Three types of influenza virus are recognized—A, B, and C—distinguishable by the nature of their internal proteins. In addition, 18 H subtypes (H1–H18) and 11 N subtypes (N1–N11) have been identified for influenza A viruses but only H1, H2, H3, N1, and N2 have established stable lineages in the human population since 1918.

- ***Influenza A*** is generally responsible for pandemics and epidemics.
- ***Influenza B*** often causes smaller or localized, milder outbreaks but can also result in severe disease and death. There are no subtypes of influenza B.
- ***Influenza C*** rarely produces disease in humans.

Antigenic shift generates new influenza A subtypes carrying different H and/or N proteins on their surface, which emerge at irregular intervals and give rise to influenza pandemics. Possible mechanisms include:

- Genetic reassortment of the eight RNA segments of the virus with that of an avian influenza virus. This requires co-infection of a host with both human and avian viruses. Pigs are thought to be the likely mixing vessel, although this could theoretically also occur in humans.
- Trans-species transmission of an avian influenza virus to humans. Viruses transmitted in this way are usually not well adapted to growth in their new host, but adaptation may occur as a result of spontaneous mutations, leading to the emergence of a pandemic strain.

Antigenic drift (minor changes in influenza A and B viruses) results from Darwinian evolution, with randomly occurring point mutations leading to amino acid changes in the two surface glycoproteins, haemagglutinin and neuraminidase, against the selection pressure of previously existing humoral immunity. This enables the virus to evade previously induced immune responses and is the process whereby annual influenza epidemics arise.

Thus, changes caused by antigenic shift or drift render the individual's immune response less able to combat the new variant.

Incidence increases during the winter months. Spread is mainly by droplet infection, but fomites and direct contact have also been implicated. Influenza pandemics of the 20th century are listed in Box 42.30.

In 1997, avian influenza A/H5N1 viruses were first isolated from humans in Hong Kong, raising the spectre of another pandemic. Since then, the virus has spread across the world. There have been several hundred human cases, with a mortality rate approaching 50%. Although this virus is highly pathogenic to humans, because of the induction of a cytokine storm within the lungs, it still has not evolved to achieve efficient human-to-human spread. However, anxieties remain that either genetic reassortment will occur in a human co-infected with human A(H1N1)pdm09 or A(H3N2) virus or adaptive mutations will occur within infected human hosts, such that a truly pandemic strain will emerge.

In April 2009, a novel influenza A virus, now referred to as A(H1N1)pdm09, was identified in patients with severe respiratory illness in Mexico and North America. The virus quickly spread across the world, with the WHO declaring an official pandemic in June 2009. The virus was the end-product of several reassortments between pre-existing swine, avian, and human virus lineages, with the swine H1 protein showing around 20% amino acid sequence divergence from previously circulating human seasonal H1N1 influenza viruses. Although unquestionably highly transmissible (with estimates of millions of infections worldwide within 1 year), this pandemic virus was, fortunately, not especially virulent. Most infections occurred in children; adults older 50 years of age had evidence of pre-existing protective immunity. A minority of infections resulted in serious disease, with an estimated 200 000 deaths worldwide. Risk factors for serious disease included pre-existing underlying medical conditions, age younger than 5 years, obesity, and pregnancy. The pandemic was declared officially over by the WHO in August 2010, the virus now behaving as a normal seasonal influenza virus and replacing the previously circulating A(H1N1) virus.

Box 42.30 Influenza pandemics of the 20th century

Year	Virus subtype	Comments
1918–2019	H1N1	'Spanish' flu; >40 million deaths
1957	H2N2	'Asian' influenza
1968	H3N2	
1976	H1N1	'Russian' influenza; not thought to be a natural occurrence
2009	H1N1pdm09	Derived from swine influenza

Novel influenza viruses may continue to infect humans sporadically, raising the possibility of a new pandemic. In 2013, more than 100 cases of infection with an H7N9 virus were reported from China. As of December 2018, there were 1567 confirmed cases, with 615 deaths. Most patients gave a history of exposure to live animals, including chickens.

Purified haemagglutinin and neuraminidase from recently circulating strains of influenza A and B viruses are incorporated in current vaccines.

The clinical features, diagnosis, treatment, and prophylaxis of influenza are discussed in Chapter 20 ('Influenza' section).

Respiratory syncytial virus infection

RSV is a paramyxovirus that causes many respiratory infections in epidemics each winter. It is a common cause of bronchiolitis in infants, complicated by pneumonia in approximately 10% of cases. The infection normally starts with upper respiratory symptoms. After an interval of 1 to 3 days, a cough and low-grade fever may develop. The onset of bronchiolitis is characterized by dyspnoea and hyperexpansion of the chest with subcostal and intercostal recession. The disease may be severe and potentially fatal in babies with underlying cardiac, respiratory (including prematurity), or immunodeficiency disease. The virus undergoes antigenic drift; consequently, re-infection occurs throughout life. RSV is occasionally the cause of outbreaks of influenza-like illness or pneumonia in older adults (in residential homes) and immunocompromised individuals.

Transfer of infection between children in hospital (hospital-acquired infection) commonly occurs unless infected patients are isolated or cohorted. Meticulous attention to hand-washing and other infection control measures reduces the risk of transmission by staff members.

Diagnosis and management

Genome detection or immunofluorescence on nasopharyngeal aspirates is the usual way of confirming the diagnosis.

Management is generally supportive, but aerosolized ribavirin has been given to patients with severe disease, particularly those with underlying cardiac or respiratory disease.

Prevention

Recent clinical trials have shown the efficacy of both passive and active immunization in preventing the severe consequences of RSV infection in neonates, infants, and in older adults (older than 75 years of age). Passive protection is with nirsevimab, a long-acting monoclonal antibody directed at the viral F protein; active vaccines use a pre-fusion version of the F protein as antigen. National immunization campaigns targeted at the elderly and also pregnant women have now been initiated in several countries to reduce the burden of RSV infection.

Metapneumovirus

Human metapneumovirus belongs to the same virus family as RSV and causes approximately 10% of LRTIs in infants and young children. Infection is clinically indistinguishable from that caused by RSV.

Further reading

Barouch DH. COVID-19 vaccines—immunity, variants, boosters. *N Engl J Med* 2022; 387:1011–1020.
Dalziel SR, Haskell L, O'Brien S et al. Bronchiolitis. *Lancet* 2022; 400:392–406.
Graham BS. The journey to RSV vaccines—heralding an era of structure-based design. *N Engl J Med* 2023; 388:579–581.
Uyeki TM, Hui DS, Zambon M et al. Influenza. *Lancet* 2022; 400:693–706.

SYSTEMIC VIRAL INFECTIONS

Dengue

This is the most common arthropod-borne viral infection in humans, and its global incidence has increased dramatically in recent years; an estimated 100 to 400 million cases occur every year in the tropics, with more than 10000 deaths from severe dengue (previously known as haemorrhagic fever) (Fig. 42.13). Dengue is caused by a flavivirus. The disease is endemic in all tropical regions, including northern Australia, most South-east Asian countries, tropical Africa and the Middle East, and Caribbean countries. Cases of dengue are also imported into the continental USA and Europe (e.g., Italy) via tourists returning from endemic countries.

Viruses spread by insects are collectively referred to as arthropod-borne viruses, or arboviruses (Boxes 42.31 and 42.32). Four different antigenic serotypes of dengue virus are recognized, and all are transmitted by the daytime-biting mosquito *Aedes aegypti*, which breeds in standing water in refuse dumps in inner cities. *Aedes albopictus* is a less common transmitter. Humans are infective during the first 3 days of the illness (viraemic stage; Fig. 42.14). Mosquitoes become infective about 2 weeks after feeding on an infected individual and remain so for life. The disease is usually endemic. Heterotypic immunity between serotypes after the illness is partial and lasts only a few months, although homotype immunity is life-long.

Clinical features

The incubation period is 5 to 6 days after the mosquito bite. Asymptomatic or mild infections are common. Two clinical forms are recognized (see Fig. 42.15).

Classic dengue fever

Classic dengue fever is characterized by the abrupt onset of fever, malaise, headache, facial flushing, retrobulbar pain that worsens on eye movements, conjunctival suffusion, and severe backache, which is a prominent symptom. Lymphadenopathy, petechiae on the soft palate, and transient morbilliform skin rashes may also

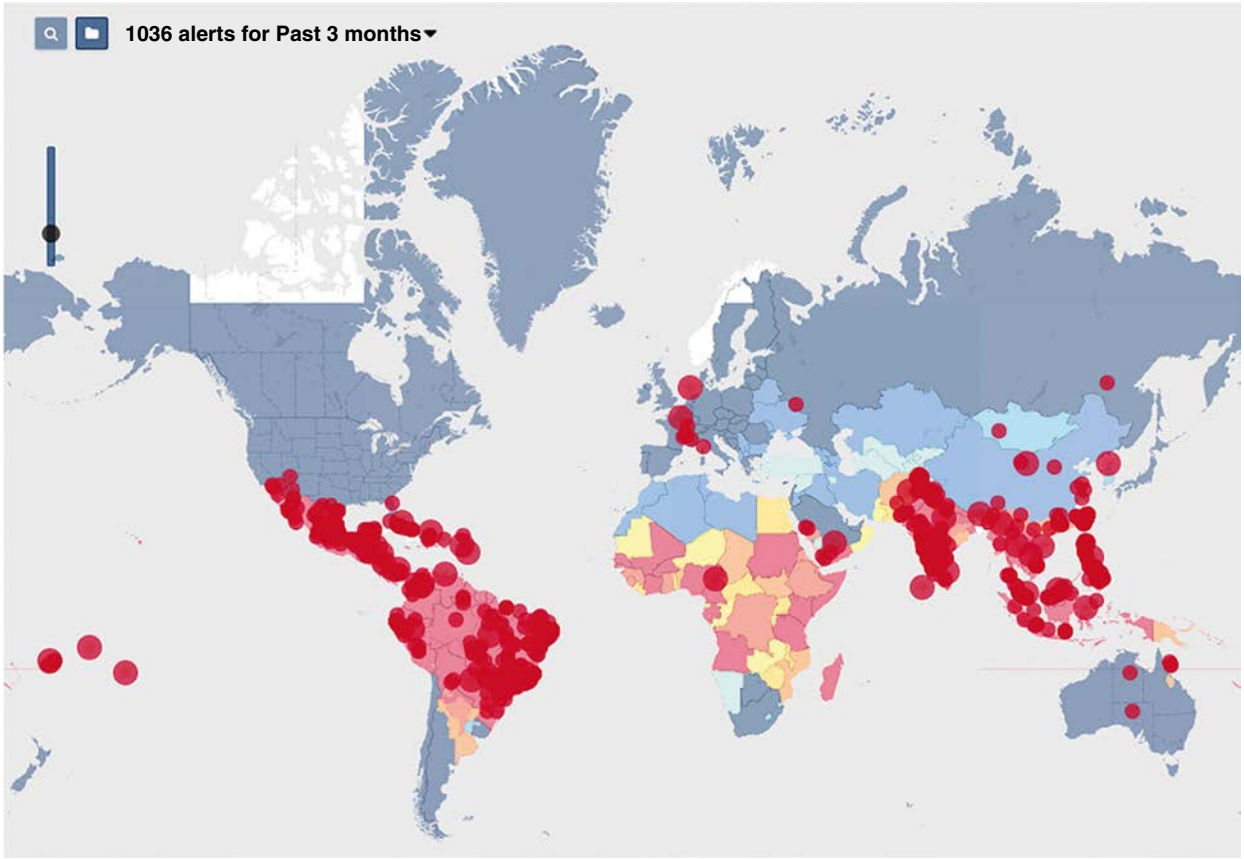


Fig. 42.13 Global dengue burden. Map lines delineate study areas and do not necessarily depict accepted national boundaries.

Box 42.31 Arbovirus (arthropod-borne) infection

Defining features of arboviruses

- Arboviruses are zoonotic viruses, transmitted through the bites of insects, especially mosquitoes and ticks.
- >385 viruses are classified as arboviruses.
- Most are members of the families of togaviruses, flaviviruses, and bunyaviruses (see **Box 42.32**).
- *Culex*, *Aedes*, and *Anopheles* mosquitoes account for transmission of the majority of these viruses.

Clinical features of arbovirus infection

- Most arbovirus diseases are generally mild; epidemics are frequent, and when they occur, the mortality rate is high.
- In general, the incubation period is short (<10 days). Common features include a biphasic illness, pyrexia, conjunctival suffusion, a rash, retro-orbital pain, myalgia, and arthralgia. Lymphadenopathy is seen in patients with dengue.
- ***Haemorrhage*** (from increased vascular permeability, capillary fragility, consumptive coagulopathy) is a feature of some arbovirus infections (see **Box 42.37**).
- ***Encephalitis*** resulting from cerebral invasion may be prominent in some arbovirus fevers.

Box 42.32 Some arboviruses

Family	Genus	Viruses	Transmission	Predominant clinical syndrome
Togaviridae	Alphaviruses	Eastern, Western, Venezuelan equine encephalitis	Mosquito borne	E
		Chikungunya		F
		Ross River		F
Flaviviridae	Flaviviruses	St. Louis encephalitis	Mosquito borne	E
		Japanese encephalitis		E
		Murray Valley encephalitis		E
		Yellow fever (see 'Yellow fever' section)		H
		Dengue (see 'Dengue' section)		F
		Zika		E
		West Nile		
		Louping ill	Tick-borne	E
		Tick-tome encephalitis		E
		Kyasanur Forest virus		H
		Omsk haemorrhagic fever		H
Peribunyaviridae	Orthobunyaviruses	La Crosse	Mosquito borne	E
Phenuiviridae	Phleboviruses	Rift Valley fever	Mosquito borne	F
Nairoviridae	Orthonairoviruses	Congo-Crimean haemorrhagic fever	Tick borne	H

E, Encephalitis; *F*, febrile illness; *H*, haemorrhagic disease.

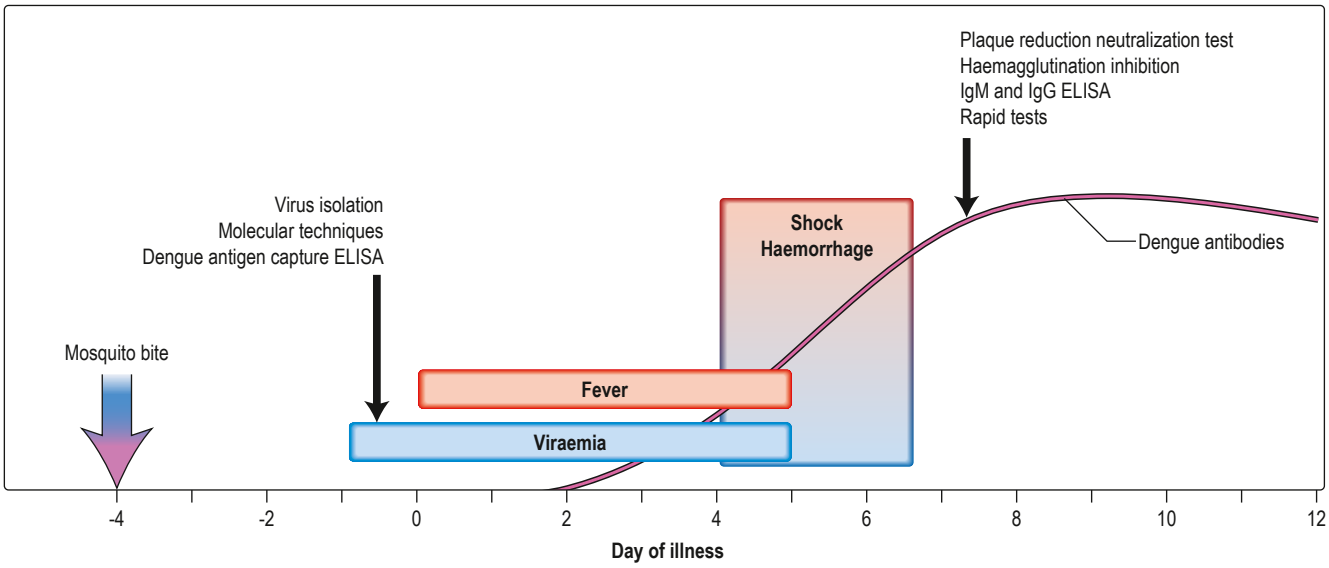


Fig. 42.14 Dengue infection: course and timing of diagnosis. *ELISA*, Enzyme-linked immunosorbent assay; *Ig*, immunoglobulin. *(From Halstead SB. Dengue.* Lancet *2007; 370:1644–1652, with permission.)*

appear on the limbs with subsequent spread to involve the trunk. Desquamation occurs subsequently. Cough is uncommon. The fever subsides after 3 to 4 days, the temperature returning to normal for a couple of days, after which the fever returns, together with the features already mentioned but milder. This biphasic or 'saddleback' pattern is considered characteristic. Severe fatigue, a feeling of being unwell, and depression are common for several weeks after the fever has subsided.

Severe dengue

Severe dengue is potentially deadly and is believed to arise from two or more sequential infections with different dengue serotypes. It is characterized by the capillary leak syndrome, thrombocytopenia, haemorrhage, hypotension, and shock. It is characteristically a disease of children, occurring most commonly in South-east Asia. The disease has a mild start, often with symptoms of an URTI. This is followed by the abrupt onset of shock and haemorrhage into the skin and ear, epistaxis, haematemesis, and melaena, known as the ***dengue shock syndrome***, which has a mortality rate of up to 44%. Serum complement levels are depressed, and there is laboratory evidence of DIC (see Chapter 38, 'Disseminated intravascular coagulation' section).

Diagnosis and management

- Isolation of dengue virus by tissue culture or detection of viral RNA by genome amplification in sera obtained during the first few days of illness is diagnostic.
- Detection of virus-specific IgM antibodies or of rising IgG titres in sequential serum samples confirms the diagnosis.
- Blood tests show leucopenia and thrombocytopenia.

Management is supportive, with analgesics and adequate fluid replacement. Corticosteroids are of no benefit, and convalescence can be slow. In dengue haemorrhagic fever, blood transfusion may be necessary, as well as intensive care support.

Prevention

Travellers should be advised to sleep under impregnated nets, but this is not very effective because the mosquito bites in the daytime. Topical insect repellents should be used. Adult mosquitoes should be destroyed by sprays, and breeding sites, such as small pools of stagnant water, should be eradicated. A live recombinant tetravalent dengue vaccine has been licensed, although it should only be administered to individuals who have evidence of previous dengue infection because in naïve individuals, immunization may increase the risk of subsequent dengue shock syndrome.

Zika virus infection

Zika virus, a flavivirus closely related to dengue viruses, came to prominence in 2016 after an explosive pandemic throughout South and Central America and the Caribbean. Like dengue, Zika is an arbovirus, transmitted to humans by *Aedes* mosquitoes. Illness, when it arises, is usually unremarkable, with fever, myalgia, eye pain, prostration, and maculopapular rash, with spontaneous resolution, although infection may also be associated with Guillain–Barré syndrome, neuropathy and myelitis. However, the epidemiological association of Zika virus outbreaks with subsequent dramatic increases in the number of cases of microcephaly in infants (e.g., Brazil reported a 20-fold increase in incidence from 2014 to 2015) led the WHO to declare Zika a public health emergency of international concern. Accurate diagnosis of infection is complicated by serological cross-reactivity with dengue viruses.

Chikungunya

Chikungunya virus is an alphavirus, a genus within the family of togaviruses. There are eight alphaviruses that result in human disease, and all are arboviruses, transmitted by mosquitoes. They are globally distributed and tend to acquire their names from the location where they were first isolated (e.g., Ross River, Eastern, Venezuelan, and Western equine encephalitis viruses) or from the local expression for a major symptom caused by the virus (e.g., chikungunya, meaning 'doubled up'). Chikungunya infection is characterized by fever, headache, maculopapular skin rash, arthralgia, myalgia, and sometimes encephalitis. After 1 year, at least 20% of patients still have recurrent joint pains. The mortality rate is approximately 0.1%, mostly in very old and very young patients.

Major epidemics of chikungunya, spread via *A. aegypti* or *A. albopictus*, were reported in India, Sri Lanka, and islands in the Indian Ocean (including Reunion, Mauritius, and the Seychelles) in 2005 and 2006, with at least 1 million cases and several hundred deaths. The severity of these epidemics is possibly due to a viral strain, with mutations resulting in a higher neurovirulence. Several European countries reported cases of chikungunya in returning travellers, including 133 cases in the UK in 2006. The increased numbers of infected returning travellers may result in the local *Aedes* mosquito population becoming infected, giving rise to outbreaks in countries where it has not previously been described, for example, in Italy (2007) and the French mainland (2010). In 2013, an outbreak in the Caribbean islands spread to contiguous Central and South American countries, and more than 200 cases were reported in 2014 in the USA. Clinical diagnosis is complicated by the similarities with dengue and zika virus infections. Detection of viral RNA in blood samples taken early in the illness is the gold standard for laboratory confirmation.

A number of chikungunya vaccines are currently undergoing clinical trials.

Infectious mononucleosis: Epstein–Barr virus infection

Globally, most individuals are infected with the herpesvirus EBV at an early age (0–5 years), at which time clinical symptoms are unusual. Infection at an older age is associated with an acute febrile illness known as infectious mononucleosis (glandular fever), which occurs worldwide in adolescents and young adults. EBV is probably transmitted in saliva and by aerosol.

Clinical features

The predominant symptoms of infectious mononucleosis are fever, headache, malaise, and sore throat. Palatal petechiae and a transient macular rash are common, the latter occurring in 90% of patients who have received ampicillin (inappropriately) for the sore throat. Cervical lymphadenopathy, particularly of the posterior cervical nodes, and splenomegaly are characteristic. Mild hepatitis is common, resulting in clinical jaundice in 10% of young adults, but other complications such as myocarditis, meningitis, encephalitis, mononeuritis multiplex, cerebellar ataxia, and mesenteric adenitis are rare. Splenic rupture may occur in the first 3 weeks of illness, and contact sport should be avoided during this period. Some young adults may remain debilitated and depressed for some months after infection.

The site of EBV latency is resting memory B lymphocytes. Evidence for reactivation of latent virus in healthy individuals is controversial, although this does occur in immunocompromised patients. Severe, often fatal, infectious mononucleosis may result from a rare

X-linked lymphoproliferative syndrome affecting young boys. Those who survive have an increased risk of hypogammaglobulinaemia and lymphoma. EBV is the cause of oral hairy leucoplakia in AIDS patients and is intimately linked to the generation of a number of malignancies (see later). EBV is a cause of haemophagocytic lymphohistiocytosis (HLH), a rare condition presenting with fever, rash, jaundice, hepatosplenomegaly, and enlarged lymph nodes. Blood tests show cytopenia and a high ferritin level, and the bone marrow shows haemophagocytosis. HLH can be primary (inherited) or secondary (e.g., to EBV) and has a high mortality rate.

Diagnosis

EBV infection should be strongly suspected if atypical mononuclear cells (activated CD8-positive T lymphocytes) are found in the peripheral blood in large numbers. It can be confirmed during the second week of infection by a positive Paul–Bunnell reaction, which detects heterophile antibodies (IgM) that agglutinate sheep erythrocytes, in around 90% of cases. False-positives can occur in other conditions, such as viral hepatitis, Hodgkin lymphoma, and acute leukaemia. The Monospot test is a sensitive and easily performed screening test for heterophile antibodies. Specific EBV IgM antibodies indicate recent infection by the virus. Clinically similar illnesses are produced by CMV, toxoplasmosis, and acute HIV infection (the so-called seroconversion illness), but these can be distinguished serologically.

Management

The majority of cases require no specific treatment, and recovery is rapid. Corticosteroid therapy is advised when there is neurological involvement (e.g., encephalitis, meningitis, Guillain–Barré syndrome), marked thrombocytopenia or haemolysis, or tonsillar enlargement that is so marked as to cause respiratory obstruction.

Cytomegalovirus infection

Clinically significant CMV infection arises in two patient groups in particular:

- Fetuses who acquire the infection transplacentally and are born congenitally infected
- Patients who are immunocompromised, such as transplant recipients and people with HIV infection.

As with all herpesviruses, CMV persists for life, usually as a latent infection in cells of the monocyte/macrophage lineage, so it may become reactivated, especially in immunocompromised patients. More than 50% of the adult population has serological evidence of latent CMV infection.

Clinical features

In healthy children and adults, CMV infection is usually asymptomatic but occasionally may result in a glandular fever–like illness, with fever, lymphocytosis with atypical lymphocytes, and hepatitis with or without jaundice. The result of the Paul–Bunnell test for heterophile antibody is negative. Infection may be spread in saliva (accounting for extensive person-to-person spread in childcare units), sexual intercourse, or blood transfusion and transplacentally to fetuses. In an immunocompromised patient, CMV infection may involve multiple organ systems, common manifestations being retinitis, pneumonitis, diffuse involvement of the GI tract, and encephalitis.

Intrauterine infection may arise from either primary or reactivated maternal infection. CMV is, by far, the most common congenital infection; in high-income countries, 0.3% to 1% of all babies are born congenitally infected with CMV. Around 5% to 10% of such babies have severe disease evident at birth, with a poor prognosis (cytomegalic inclusion disease). CNS involvement may cause microcephaly and motor disorders. Jaundice and hepatosplenomegaly

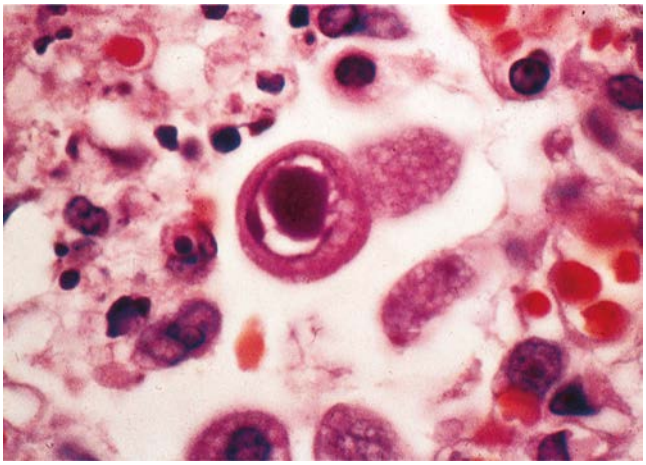


Fig. 42.15 Typical 'owl's eye' inclusion-bearing cell infected with cytomegalovirus.

are common, and thrombocytopenia and haemolytic anaemia also occur. Periventricular calcification is seen on skull X-ray. A further 5% to 10% of infected babies are normal at birth, but developmental abnormalities become apparent later (e.g., sensorineural nerve deafness). The remaining 80% to 85% of infected babies are normal at birth and develop normally.

Diagnosis

Serological tests can identify latent (IgG) or primary (IgM) infection. However, most infections are now diagnosed by detection and quantification of CMV DNA or RNA using molecular amplification techniques in blood or other body fluid samples. The virus can also be identified in tissues by the presence of characteristic intranuclear 'owl's eye' inclusions (see Fig. 42.15) on histological staining and by direct immunofluorescence.

Management

In immunocompetent patients, infection is usually self-limiting, and no specific treatment is required. In immunosuppressed patients, ganciclovir reduces retinitis and GI damage and can eliminate CMV from blood, urine, and respiratory secretions. It is less effective against pneumonitis. In patients who are continually immunocompromised, maintenance therapy may be necessary. Drug resistance may arise during long-term therapy, such as in transplant recipients. Bone marrow toxicity is common. Valganciclovir, foscarnet, cidofovir (and its derivative brincidofovir), and maribavir are also available for treatment and letermovir for prophylaxis (see 'Antiviral drugs' section). Treatment of CMV in neonates is difficult, but therapy of infected babies with evident CNS involvement has been shown to improve long-term hearing outcome, and 6 months of valganciclovir therapy is now recommended for symptomatic neonates.

Mumps

Mumps virus is a paramyxovirus. It is spread by droplet infection, by direct contact, or through fomites. Humans are the only known natural hosts. The peak period of infectivity is 2 to 3 days before the onset of the parotitis and for 3 days afterward.

Clinical features

The incubation period averages 18 days. Although no age is exempt, this is primarily a disease of school-aged children and young adults; it is uncommon before the age of 2 years. The prodromal symptoms are non-specific and include fever, malaise, headache, and

anorexia. This is usually followed by severe pain over the parotid glands, with either unilateral or bilateral parotid swelling (Fig. 42.16). The enlarged glands obscure the angle of the mandible and may elevate the ear lobe, which does not occur in cervical lymph node enlargement. Trismus because of pain is common at this stage. Submandibular gland involvement occurs less frequently.

Complications

CNS involvement is the most common extra-salivary gland manifestation of mumps. Clinical meningitis occurs in 5% of all infected patients, and 30% of patients with CNS involvement have no evidence of parotid gland involvement.

Epididymo-orchitis develops in about one-third of patients who develop mumps after puberty. Bilateral testicular involvement results in sterility in only a small percentage of these patients. Pancreatitis, oophoritis, myocarditis, mastitis, hepatitis, and polyarthritis may also occur.

Diagnosis and management

The diagnosis of mumps is on the basis of the clinical features. Viral RNA can be demonstrated by genome detection assays in saliva, throat swabs, urine, and CSF. Serological demonstration of a mumps-specific IgM response in an acute blood or oral fluid sample is also diagnostic.

Treatment is supportive. Attention should be paid to adequate nutrition and mouth care. Analgesics should be used to relieve pain.

Prevention

Children in the UK are immunized with the live attenuated MMR vaccine (see Box 42.19), and the mumps vaccine is given in most low- and middle-income countries. Vaccination is contraindicated in immunosuppressed individuals, pregnant females, and in those with severe febrile illnesses.

HIV infections

The HIVs belong to the family of retroviruses (Box 42.33) and are characterized by their ability to replicate through a DNA intermediate using the enzyme reverse transcriptase.

HIV-1 and the related virus HIV-2 are further classified as lentiviruses ('slow' viruses) because of their slowly progressive clinical effects. They give rise to the AIDS and are discussed in Chapter 44.

Human T-cell leukaemia/lymphotropic virus infection

Human T-cell leukaemia/lymphotropic virus type 1 (HTLV-1) causes adult T-cell leukaemia/lymphoma and tropical spastic paraparesis.

Myocarditis and skeletal muscle infection

EV infection is a cause of acute myocarditis and pericarditis (see Box 42.27), from which, in general, there is complete recovery. However, these viruses can also cause chronic congestive cardiomyopathy and, rarely, constrictive pericarditis. Skeletal muscle involvement, particularly of the intercostal muscles, is a feature of Bornholm disease, a febrile illness usually caused by coxsackie B viruses. The pain may be of such intensity as to mimic pleurisy or an acute abdomen. The infection affects both children and adults and may be complicated by meningitis or cardiac involvement. Myocarditis and pericarditis are also discussed in Chapter 17 ('Myocarditis' and 'Pericardial disease' sections).

Postviral or chronic fatigue syndrome

Many viral infections, including EBV, coxsackie B viruses, echoviruses, CMV, and hepatitis A virus, are associated with a postviral chronic fatigue syndrome (see Chapter 30, 'Medically unexplained symptoms and functional neurological disorders' section).

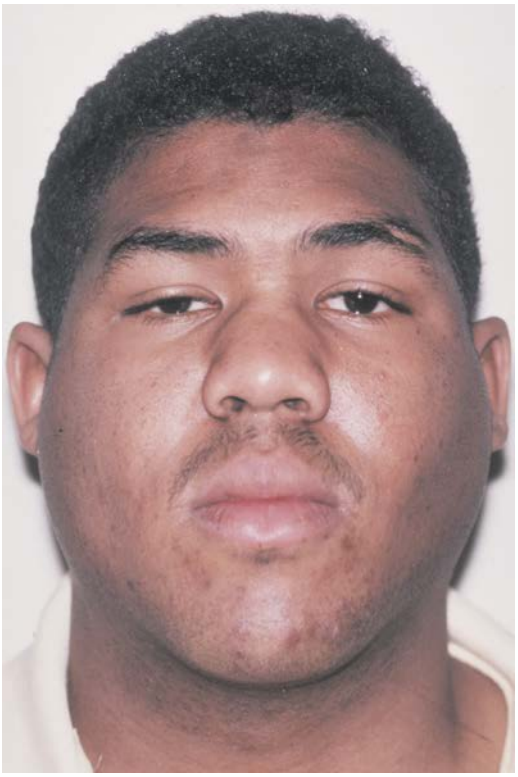


Fig. 42.16 Mumps: a child with a swollen face. *(From Chaudhry B, Harvey DR.* Mosby's Color Atlas and Text of Paediatrics and Child Health. *Edinburgh: Mosby; 2001. © Elsevier 2001.)*

Box 42.33 Human lymphotropic retroviruses

Subfamily	Virus	Disease
Lentiviruses	HIV-1	AIDS
	HIV-2	AIDS
Oncoviruses	HTLV-1	Adult T-cell leukaemia/lymphoma Tropical spastic paraparesis
	HTLV-2	Myelopathy

AIDS, Acquired immunodeficiency syndrome; *HIV,* human immunodeficiency virus; *HTLV,* human T-cell leukaemia/lymphotropic virus.

Further reading

De Amorin Vilharba BL, Yamamura M, de Azevedo MV et al. Disease burden of congenital Zika virus syndrome in Brazil and its association with socioeconomic data. *Sci Rep* 2023; 13(1):11882.
Hamer DH. Dengue—perils and prevention. *N Engl J Med* 2021; 384:2252–2253.
Hviid A, Rubin S, Muhlemann K. Mumps. *Lancet* 2008; 371:932–944.
Khawaja F, Spallone A, Kotton CN, Chemaly RF. Cytomegalovirus infection in transplant recipients: newly approved additions to our armamentarium. *Clin Microbiol Inf* 2023; 29: 44–50.
Weaver SC, Lecuit M. Chikungunya virus and the global spread of a mosquito-borne disease. *N Engl J Med* 2015; 372:1231–1239.
Xie C, Zhong LY, Bu GL et al. Anti-EBV antibodies: roles in diagnosis, pathogenesis and antiviral therapy. *J Med Virol* 2023; 95(5):e28793.

VIRUS INFECTIONS OF THE NERVOUS SYSTEM

Acute flaccid paralysis, including poliomyelitis

A number of EVs may infect the nervous system (see Box 42.24), with different clinical outcomes. Infection with polioviruses 1, 2, and 3, classified within EV species C, may result in acute flaccid paralysis (AFP). Polioviruses are excreted in faeces and spread via the

faecal–oral route. They have a propensity for the nervous system, especially the anterior horn cells of the spinal cord and cranial motor neurones.

Clinical features

The incubation period is 7 to 14 days. Although polio is essentially a disease of childhood, no age is exempt. The clinical manifestations vary considerably. The most common outcome (95% of individuals) is asymptomatic seroconversion. AFP, also known as ***paralytic poliomyelitis***, arises in only approximately 0.1% and 1% of infected children and adults, respectively. It follows about 4 to 5 days after an initial illness of fever, sore throat, and myalgia. Meningeal irritation and muscle pain occur and are followed by the onset of asymmetric flaccid paralysis without sensory involvement. Factors predisposing to the development of paralysis include male sex, exercise early in the illness, trauma, surgery, or intramuscular injection, which localize the paralysis, and recent tonsillectomy (bulbar poliomyelitis).

Aspiration pneumonia, myocarditis, paralytic ileus, and urinary calculi are late complications of poliomyelitis.

AFP may also arise from infection with a number of other EVs, and indeed, AFP is now more commonly caused by these viruses than by polioviruses. EV A71, more usually associated with hand, foot, and mouth disease, has a particular predilection for neuroinvasion. Epidemics of EV A71 are not infrequent; an estimated 6 million cases have occurred globally since 2009, with more than 2000 deaths from a variety of serious neuromotor syndromes. Inactivated EV A71 vaccines have been approved for clinical use in China. Acute flaccid myelitis has also been associated recently with outbreaks of EV D68 infection in the USA and UK.

Diagnosis

All cases of AFP must be investigated to exclude poliovirus infection. Diagnosis is by detection of EV RNA in a throat swab, faecal sample, or CSF followed by sequencing to identify precisely which EV is present. Sequencing also enables a distinction to be made between wild poliovirus and vaccine strains.

Management

Management is supportive. Bed rest is essential during the early course of the illness. Respiratory support with intermittent positive-pressure respiration is required if the muscles of respiration are involved.

Prevention and control of poliomyelitis

In 1988, the World Health Assembly adopted the goal of eliminating poliomyelitis worldwide. Since then, the estimated 350 000 cases arising annually in 125 countries have been reduced by more than 99%, with only Pakistan and Afghanistan never having terminated indigenous transmission. Only 30 cases of wild-type poliomyelitis were documented in 2022. The eradication campaign has relied on improvements in sanitation, hygiene, and the widespread use of polio vaccines. However, spread of infection via travel from endemic countries is an ongoing threat. The creation of a poliomyelitis-free world remains the goal, and the WHO Polio Eradication Strategy 2022–2026 describes a pathway to a world free of this infection. However, a temporary halt to immunization campaigns during the COVID pandemic put more than 80 million children at risk, with outbreaks of circulating vaccine-derived poliovirus (cVDPV, arising from mutations in the oral polio vaccine strains after delivery) tripling from 2019 to 2020, and more than 1100 children paralyzed. Vaccination remains the bedrock of the campaign. New vaccine preparations include an oral vaccine containing a more genetically stable polio type 2 virus and inactivated intramuscular poliovirus vaccines (IPV) with greater potency than the original Salk IPV. The greater reliability of IPV in hot climates and the scientific and ethical problems of continuing to use oral polio vaccine (OPV) in countries free from poliomyelitis mean that IPV has replaced OPV in the routine immunization schedules in many countries.

Meningitis

EVs are the most common viral cause of aseptic meningitis, although a number of other viruses may also result in this syndrome (**Box 42.34**).

Encephalitis

Encephalitis may arise by direct virus invasion of nervous tissue or by induction of an aberrant cross-reactive immune response (post-infectious encephalitis). Although HSV is the most common cause of sporadic encephalitis worldwide, a number of other virus infections may also present with an encephalitic illness.

Japanese encephalitis

Japanese encephalitis is a mosquito-borne encephalitis caused by a flavivirus. It has been reported most frequently in the rice-growing countries of South-east Asia and the Far East. *Culex tritaeniorhynchus* is the usual vector and feeds mainly on pigs, as well as birds such as herons and sparrows. Humans are accidental hosts.

The incubation period is 5 to 15 days. Most infections are asymptomatic. When disease arises, the onset is heralded by severe rigors. Fever, headache, and malaise last 1 to 6 days. Weight loss is prominent. In the acute encephalitic stage, the fever is high (38°–41°C); neck rigidity occurs; and neurological signs, such as altered consciousness, hemiparesis, and convulsions, develop. Mental deterioration occurs over a period of 3 to 4 days and culminates in coma. The mortality rate varies from 7% to 40% and is higher in children. Residual neurological defects, such as deafness, emotional lability, and hemiparesis, occur in about 70% of patients who have had CNS involvement. Convalescence is prolonged. Antibody detection in serum and CSF by IgM capture ELISA is a useful rapid diagnostic test. Vaccines containing formalin-inactivated viruses derived from mouse brain are available and effective. Management is supportive.

West Nile encephalitis

West Nile virus was first recognized in the Western hemisphere (New York, USA) in 1999, having been previously reported in Africa, Asia, and parts of Europe. By the end of 2009, the US outbreak had resulted in more than 25 000 human cases and more than 1100 deaths. The vast majority of infections are asymptomatic. In a minority of cases, infection presents as a febrile illness with a maculopapular rash, 1% resulting in severe encephalitis. Disease severity and mortality are age related, being greatest in older adults. The primary hosts of infection are birds. It is spread by mosquitoes (*Culex* spp.) and may also infect humans and horses. It can also be transmitted by blood transfusions, breast-feeding, and organ donation from an infected individual. Diagnosis is by genome detection in appropriate samples or specialized serology for the detection of IgM virus-specific antibodies.

Tick-borne encephalitis

This arises from infection with a flavivirus (actually a series of closely related viruses) transmitted by *Ixodes* spp. ticks. It occurs in an area extending from Western Europe to Japan. The tick is the main reservoir for the virus, which is transmitted when it feeds on mice and other rodents.

Box 42.34 Viral causes of meningitis and encephalitis

Virus	Comments	Outcome
Meningitis		
Enteroviruses, especially enterovirus B species (coxsackie B and echoviruses)	Commonest cause of viral meningitis	Relatively benign prognosis
Mumps virus	Rare in countries with MMR vaccine policy. May occur in absence of parotitis	Relatively benign prognosis
Herpes simplex viruses	More common in women and may recur—Mollaret's meningitis	Benign prognosis in absence of signs of encephalitis
Encephalitis, direct invasion		
Herpes simplex viruses	Most common cause of sporadic encephalitis; any age	High morbidity and mortality
Mumps virus	Spread from meninges, resulting in meningo-encephalitis	Most patients recover
Measles virus	Subacute sclerosing encephalitis	Rare, high mortality rate
Varicella zoster virus	May cause vasculitis	May arise with reactivation (shingles)
Cytomegalovirus	Only in immuno-compromised	Poor prognosis
Enteroviruses	EV 71; otherwise, usually in immuno-compromised only	
HIV	Dementia is the main feature	
Japanese encephalitis virus	Flavivirus, mosquito borne	
West Nile encephalitis virus	Flavivirus, mosquito borne	Birds are the primary hosts; higher mortality rate in older adults
Tick-borne encephalitis viruses	Flaviviruses, tick borne	
Lymphocytic choriomeningitis virus	Arenavirus	House mouse is primary host
Hendra, Nipah viruses	Paramyxoviruses	Zoonotic infections from horses, pigs
Rabies virus	Acquired from animal or bat bite	Almost always fatal
JC virus	Only in immuno-compromised	Results in progressive multifocal leucoencepha-lopathy
Eastern, Venezuelan encephalitis viruses	Alphaviruses, mosquito borne	Geographically restricted to the Americas
Encephalitis, post-infectious		
May arise from a number of different virus infections, including measles (1 in 1000–5000 cases of measles), influenza, varicella zoster		

EV, Enterovirus; *HIV,* human immunodeficiency virus.

The disease starts 4 to 28 days after a bite from an infected tick and is biphasic in 80% of patients. After a symptom-free period of about 7 days, fever, malaise, headache, and fatigue are followed by encephalitis. There may be associated limb paralysis, which is caused by anterior horn cell involvement, mainly of the cervical region. Cranial nerve involvement also occurs. Tick-borne encephalitis virus IgM and IgG antibodies are present, and the virus can be detected in blood by genome detection assays. The overall mortality rate is about 1% (but can be considerably higher for certain strains); 30% of patients have impairment in neurological function with persistent paralysis in 6%. A preventive vaccine is available.

Lymphocytic choriomeningitis

This infection is a zoonosis, the natural reservoir of lymphocytic choriomeningitis virus (an arenavirus; see **Box 42.25**) being the house mouse. Infection is by inhalation from urine-contaminated rubbish and is characterized by:

- Non–nervous system illness, with fever, malaise, myalgia, headache, arthralgia, and vomiting
- ***Meningitis*** in addition to the above symptoms.

Occasionally, a more severe form occurs, with encephalitis leading to disturbance of consciousness.

This illness is generally self-limiting and requires no specific treatment.

Hendra and Nipah virus infection

Hendra and Nipah viruses cause zoonotic infections in humans who have been in contact with infected animals (horses and pigs, respectively). The viruses are named after the locations where they were first isolated: Hendra in Australia and Nipah in Malaysia. Both are classified as paramyxoviruses, and fruit bats are their natural hosts. Hendra virus has caused severe respiratory distress in horses and humans, and Nipah virus caused a major outbreak of viral encephalitis (265 cases and 105 deaths) in Malaysia between September 1998 and April 1999. Subsequent outbreaks have been reported in India and Bangladesh. Management of these conditions is supportive, and no vaccines are currently available.

Rabies

Rabies is a major problem in some countries, with an estimated tens of thousands of deaths per year worldwide. Established infection is almost invariably fatal; there are only a handful of recorded cases of survival from clinical rabies. It is caused by the rabies virus, a single-stranded RNA virus of the lyssavirus genus within the family of rhabdoviruses. The virus is bullet shaped and has spike-like structures arising from its surface containing glycoproteins that cause the host to produce neutralizing, haemagglutination-inhibiting antibodies. The virus has a marked affinity for nervous tissue and the salivary glands. It exists in two major epidemiological settings:

- ***Urban rabies*** is most frequently transmitted to humans through rabid dogs and, less frequently, cats.
- ***Sylvan (wild) rabies*** is maintained in the wild by a host of animal reservoirs such as foxes, skunks, jackals, mongooses, and bats.

With the exceptions of Australia, New Zealand, and the Antarctic, human rabies has been reported on all continents. Transmission is usually through the bite of an infected animal, most commonly a dog. However, the percentage of rabid bites leading to clinical disease ranges from 10% (on the legs) to 80% (on the head). Other forms of transmission, if they occur, are rare.

The virus replicates in the muscle cells near the entry wound. It penetrates the nerve endings and travels in the axoplasm to the

spinal cord and brain. In the CNS, the virus again proliferates before spreading to the salivary glands, lungs, kidneys, and other organs via the autonomic nerves.

Clinical features

The incubation period is variable, ranging from a few weeks to several years; on average, it is 1 to 3 months. In general, bites on the head, face, and neck have a shorter incubation period than those elsewhere. In humans, two distinct clinical varieties of rabies are recognized:

- ***Furious rabies***, the classic variety. The only characteristic feature in the prodromal period is the presence of pain and tingling at the site of the initial wound. Fever, malaise, and headache are also present. About 10 days later, marked anxiety and agitation or depressive features develop. Hallucinations, bizarre behaviour, and paralysis may also occur. Hyperexcitability, the hallmark of this form of rabies, is precipitated by auditory or visual stimuli. Hydrophobia (fear of water) is present in 50% of patients and is caused by severe pharyngeal spasms on attempting to eat or drink. Aerophobia (fear of air) is considered pathognomonic of rabies. Examination reveals hyper-reflexia, spasticity, and evidence of sympathetic overactivity indicated by pupillary dilation and diaphoresis. The patient goes on to develop convulsions, respiratory paralysis, and cardiac arrhythmias. Death usually occurs in 10 to 14 days.
- ***Dumb rabies***, the paralytic variety. This presents with a symmetrical ascending paralysis that resembles Guillain–Barré syndrome.

Diagnosis

The diagnosis of rabies is generally made clinically. Skin-punch biopsies are used to detect antigen with an immunofluorescent antibody test on frozen section. Viral RNA can be detected by genome amplification. Isolation of virus from saliva or the presence of antibodies in blood or CSF may establish the diagnosis. Negri bodies are detected at postmortem in 90% of all patients with rabies; these are eosinophilic, cytoplasmic, ovoid bodies, 2 to 10 nm in diameter, seen in greatest numbers in the neurones of the hippocampus and the cerebellum. The diagnosis should be made pathologically on the biting animal using genome amplification, immunofluorescence on brain tissue using fluorescently labelled anti-rabies antibody, or tissue culture of the brain.

Management

When CNS disease is established, therapy is symptomatic because death is virtually inevitable. The patient should be nursed in a quiet, darkened room. Nutritional, respiratory, and cardiovascular support may be necessary.

Drugs such as morphine, diazepam, and chlorpromazine should be used liberally in patients who are excitable.

Prevention and control

Pre-exposure prophylaxis

This is given to individuals with a high risk of contracting rabies, such as laboratory workers, animal handlers and veterinarians. Three doses of human diploid (HDCV) or purified chick embryo cell vaccine, given by deep subcutaneous or intramuscular injection on days 0, 7, and 28, provide effective immunity. A reinforcing dose is given after 12 months, and additional reinforcing doses are given every 3 to 5 years, depending on the risk of exposure. Vaccines of nervous tissue origin are still used in some parts of the world. These, however, are associated with significant side-effects and should not be used if HDCV is available.

Post-exposure prophylaxis

The wound should be cleaned carefully and thoroughly with soap and water, and left open. Human rabies immunoglobulin should be given immediately (20 IU/kg); half should be injected around the area of the wound, and the other half should be given intramuscularly. Four 1.0-mL doses of HDCV should be given intramuscularly: the first dose is given on day 0 and is followed by injections on days 3, 7, and 21. Reaction to the vaccine is uncommon. Protection mediated by this protocol is almost 100%.

Control

Domestic animals should be vaccinated if there is any risk of rabies in the country. In the UK, control has been by quarantine of imported animals for 6 months, and no indigenous case of rabies has been reported for many years. The Pet Travel Scheme (PETS) enables certain pet animals to enter or re-enter Great Britain without quarantine if they come from qualifying countries via designated routes, are carried by authorized transport companies, and meet the conditions of the scheme. Wild animals in 'at-risk' countries must be handled with great care.

Progressive multifocal leucoencephalopathy

JC virus, a polyomavirus, is the cause of progressive multifocal leucoencephalopathy (PML), which presents as dementia in the immunocompromised and is caused by multiple foci of demyelination in cerebral white matter. Virus replicates in the nucleus of oligodendrocytes, resulting in cell lysis and breakdown of the myelin sheath. The virus is usually acquired in childhood and may reactivate if the host immune system becomes compromised in later life. Reactivation of JC virus and development of PML have recently been described in patients with multiple sclerosis treated with the monoclonal antibody natalizumab directed against the cell adhesion molecule, α4-integrin.

Further reading

Gundamraj V, Hasbun R. Viral meningitis and encephalitis: an update. *Curr Opin Infect Dis* 2023; 36(3):177–185.
Hemachuda T, Ugolini G, Wacharapluesadee S. Human rabies: neuropathogenesis, diagnosis and treatment. *Lancet Neurol* 2013; 12:498–513.
Hill M, Bandyopadhyay AS, Pollard AJ. Emergence of vaccine-derived poliovirus in high-income settings in the absence of oral polio vaccine use. *Lancet* 2022; 400:713–715.

VIRUS INFECTIONS OF THE GASTROINTESTINAL TRACT

Virus infections that commonly result in acute gastroenteritis (diarrhoea and vomiting) are listed in **Box 42.35**.

Rotavirus infection

Rotaviruses are a major cause of infantile gastroenteritis worldwide. The virus (Latin *rota*, meaning wheel) is so named because of its electron microscopic appearance with a characteristic circular outline and radiating spokes (**Fig. 42.17**). It belongs to the reoviruses family, which have double-stranded RNA segmented genomes. More than 200 000 infected children younger than the age of 5 years are estimated to die annually in resource-deprived countries. The prevalence is higher during the winter months in non-tropical areas. Asymptomatic infections are common and bottle-fed babies are more likely to be symptomatic than those that are breast-fed.

Adults may become infected with rotavirus, but the symptoms are usually mild or absent. The virus may, however, cause diarrhoea

Box 42.35 Viruses associated with gastroenteritis

- Rotaviruses (groups A, B, and C)
- Enteric adenoviruses (types 40 and 41)
- Caliciviruses (includes noroviruses and sapoviruses)
- Astroviruses

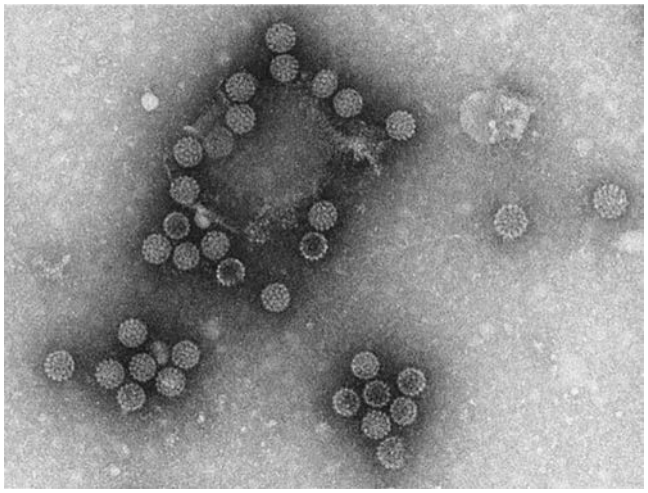


Fig. 42.17 Electron micrograph of human rotavirus.

in immunosuppressed adults or outbreaks in patients on care in older adult wards.

Clinical features

The illness is characterized by vomiting, fever, diarrhoea, and the metabolic consequences of water and electrolyte loss.

Diagnosis and differential diagnosis

The diagnosis can be established by genome amplification, or ELISA for the detection of rotavirus antigen in faeces and by electron microscopy of faeces. Histology of the jejunal mucosa in children shows shortening of the villi, with crypt hyperplasia, and mononuclear cell infiltration of the lamina propria.

Management and prevention

Management is directed at overcoming the effects of water and electrolyte imbalance with adequate oral rehydration therapy and, when indicated, intravenous fluids (**Box 42.36**). Antibiotics should not be prescribed.

Rotavirus vaccines

Despite a major setback when the first licensed rotavirus vaccine was rapidly withdrawn from the market in 1999 after reports of increased rates of intussusception, two live vaccines have been developed and are now used in many countries. One (Rotarix) contains an attenuated human strain, the relevant antigens being P[8] and G1; the other (Rotateq) is based on a bovine parent strain and comprises five single-gene reassortants, each containing a human-strain outer capsid gene encoding the most common human antigenic types (P[8] and G1–4). Routine rotavirus vaccination (two doses at 8 and 12 weeks of age) was introduced in the UK in 2012. Vaccine efficacy against severe rotavirus gastroenteritis is around 90%.

Calicivirus infection

The caliciviruses are an extensive virus family, named after the cup-shaped (Latin *calyx*, meaning cup) indentations on their viral surface seen by electron microscopy. The family contains four genera, two of which, the noroviruses and sapoviruses, infect humans and cause gastroenteritis.

Norovirus is the major cause of acute non-bacterial gastroenteritis, causing outbreaks in nursing homes, hospitals, schools, leisure centres, and restaurants and on cruise ships. In countries that have adopted routine rotavirus vaccination of infants, norovirus has become the leading cause of clinically significant acute gastroenteritis in children. Transmission is mostly faecal–oral, with outbreaks suggesting a common source, such as food and water, and fomites. Aerosol transmission also occurs, and noroviruses can be detected in vomit. Illness is usually self-limiting (12–48 hours) and mild, consisting of nausea, headache, and abdominal cramps followed by diarrhoea and vomiting, which may be the only feature (winter vomiting). Diagnosis is by demonstration of viral nucleic acid or antigen in diarrhoeal faeces. Treatment is with oral rehydration solutions. Prevention can be difficult, but hand-washing and good hygienic food preparation are required.

Sapovirus causes gastroenteritis, mainly in children.

Further reading

Jagirdhar GSK, Pulakirthi YS, Chigurupati HD, Surani S. Gastrointestinal tract and viral pathogens. *World J Virol* 2023; 12:136–150.
Omatola CA, Olaniran AO. Rotaviruses: from pathogenesis to disease control—a critical review. *Viruses* 2022; 14(5):875.

VIRAL HEPATITIS

A number of distinct virus infections may result in inflammation of the liver (hepatitis). The viruses belong to different virus families, are spread by different routes and may have differing clinical consequences (see **Box 34.7**). Viral hepatitis is discussed further in Chapter 23 ('Viral hepatitis' section).

VIRUSES AND MALIGNANT DISEASE

Around 15% to 20% of all human malignant disease arises from infection. Recognition of the infectious cause of a particular cancer raises the possibility of developing an anticancer vaccine, as has now happened for both HBV and human papillomavirus (HPV).

Hepatitis viruses and primary hepatocellular carcinoma

Globally, hepatocellular carcinoma is the second most common cause of death from malignant disease. A high percentage arises in patients infected with either HBV or HCV (see Chapter 23).

Human papillomaviruses and cancer of the uterine cervix

Papillomaviruses tend to produce chronic infections (see Chapter 45, 'Human papillomavirus' section). HPVs, of which there are at least 100 types, are responsible for the common skin and genital warts, and certain types (mainly 16 and 18) are the cause of carcinoma of the cervix, cancers in other genital areas (vagina, vulva, penis, anus), and some oral cancers (type 16). Multivalent vaccines against a number of HPV types (but always including 16 and 18) are available. The current recommendations in many countries are for vaccination of all girls at age 9 to 14 years, although some countries (including the UK) are increasingly opting to vaccinate both boys and girls. (For ***anogenital warts***, see Chapter 43, 'Human papillomavirus [anogenital warts]' section.)

Box 42.36 Oral rehydration solutions and intravenous solutions used in moderate and severe diarrhoea[a]

Solution	Salts (mmol/L)				Substance added (per L of water)	
	Na^+	K^+	Cl^-	Glucose	Substance	Quantity (g)
Oral[b]						
WHO (hypo-osmolar formulation)	75	20	65	75	NaCl	2.6
					Na citrate	2.9
					KCl	1.5
					Glucose	13.5
Cereal-based	85	–	80	–	Cooked rice	80
					Salt (NaCl)	5
Household	85	–	80	111	Glucose	20
					Salt (NaCl)	5
UK and Europe	35–60	20	37	90–200	Pre-prepared solutions	
Intravenous						
Ringer's lactate[c]	131	5	111	0	Pre-prepared	

[a]In children younger than 5 years of age, zinc supplements (10 mg/day in those younger than 6 months; 20 mg/day in those older than 6 months) should ideally be given along with oral rehydration solution.
[b]Oral rehydration solution should be administered by encouraging the patient to take small, regular sips of fluid. If vomiting occurs, administration should be paused for 10 minutes and then cautiously restarted. Oral rehydration solution can be administered via a nasogastric tube if swallowing is impaired. If patients are severely dehydrated, intravenous fluids are preferred.
[c]Contains Ca^{2+} 2 mmol/L and HCO_3^- (as lactate).
WHO, World Health Organization.

Epstein–Barr virus and malignant disease

EBV infection has been associated with a number of malignant diseases, including Burkitt lymphoma, undifferentiated nasopharyngeal carcinoma, post-transplant lymphoproliferative disease, the immunoblastic lymphoma of patients with AIDS, some forms of Hodgkin lymphoma, and gastric cancer. Different levels of expression of EBV latency genes occur in patients with these proliferative conditions, and various co-factors are also involved in their pathogenesis; for example, in Burkitt lymphoma, the most common tumour of childhood in sub-Saharan Africa, epidemiological evidence points to an interplay between EBV infection and the presence of hyperendemic (i.e., present all year round) malaria.

Kaposi's sarcoma and human herpesvirus type 8

HHV-8 is strongly associated with the aetiology of all forms of Kaposi's sarcoma. Antibody prevalence is high in those with tumours but relatively low in the general population of most industrialized countries. High rates of infection (>50% population) have been described in Central and Southern Africa, and this matches the geographical distribution of classical Kaposi's sarcoma before the era of AIDS. HHV-8 is transmitted sexually and through exposure to blood from needle sharing. It is thought that salivary transmission may be the predominant route in Africa. HHV-8 RNA transcripts have been detected in Kaposi's sarcoma cells and in circulating mononuclear cells from patients with the tumour. This virus also has an aetiological role in two rare lymphoproliferative diseases: multicentric ***Castleman's disease*** (a disorder of the plasma cell type) and ***primary effusion lymphoma*** (body cavity–based lymphoma), which is characterized by pleural, pericardial, or peritoneal lymphomatous effusions in the absence of a solid tumour mass.

Other rare malignancies caused by virus infections

HTLV-1, a lentivirus belonging to the same virus family as HIV, is the aetiological agent of adult T-cell leukaemia/lymphoma. Merkel cell polyomavirus has been identified in the malignant tissue of the rare Merkel cell carcinoma of the skin.

Further reading

Markowitz LE, Unger ER. Human papillomavirus vaccination. *N Engl J Med* 2023; 388:1790–1798.

VIRAL HAEMORRHAGIC FEVERS

These comprise infection with a range of different viruses (**Box 42.37**), which have in common the clinical feature of haemorrhagic manifestations.

Yellow fever

Yellow fever, caused by a flavivirus, is an illness of widely varying severity. It is confined to Africa (90% of cases) and South America between latitudes 15°N and 15°S. For poorly understood reasons, yellow fever has not been reported from Asia even though climatic conditions are suitable and the vector, *A. aegypti*, is common. The infection is transmitted in the wild by *Aedes africanus* in Africa and *Haemagogus* spp. in South and Central America. Extension of infection to humans (via the mosquito from monkeys) leads to the occurrence of 'jungle' yellow fever. *A. aegypti*, a domestic mosquito that lives in close relationship to humans, is responsible for human-to-human transmission in urban areas (urban yellow fever). After being infected, a mosquito remains so for life.

Clinical features

The incubation period is 3 to 6 days. Mild infection is indistinguishable from other viral fevers such as influenza and dengue.

Three phases are recognized in the severe (classical) illness. Initially, the patient presents with a high fever of acute onset, usually 39° to 40°C, which then returns to normal in 4 to 5 days. During this time, headache is prominent. Retrobulbar pain, myalgia, arthralgia, a flushed face, and suffused conjunctivae are common. Epigastric discomfort and vomiting are present when the illness is severe.

Box 42.37 Viral infections associated with haemorrhagic manifestations[a]

Flavivirus
- Yellow fever (urban and sylvan)
- Dengue haemorrhagic fever
- Kyasanur Forest disease
- Omsk haemorrhagic fever

Phenuivirus
- Rift Valley fever

Orthonairovirus
- Congo–Crimean haemorrhagic fever

Hantavirus
- Hantavirus infections

Arenavirus
- Argentinian haemorrhagic fever
- Bolivian haemorrhagic fever
- Lassa fever
- Epidemic haemorrhagic fever

Filovirus
- Marburg
- Ebola

[a]Most of these are arboviruses. Some (e.g., hantavirus, Lassa fever) have a rodent vector. The source and transmission route of filoviruses are not known.

Relative bradycardia (Faget's sign) is present from the second day of illness. The patient then makes an apparent recovery and feels well for several days. After this 'phase of calm', the patient again develops increasing fever, deepening jaundice, and hepatomegaly. Ecchymosis, bleeding from the gums, haematemesis, and melaena may occur. Coma, which is usually a result of uraemia or haemorrhagic shock, occurs for a few hours preceding death. The mortality rate is up to 40% in severe cases. Liver pathology shows mid-zone necrosis and eosinophilic degeneration of hepatocytes (Councilman bodies).

Diagnosis and management

The diagnosis is established by a history of travel and vaccination status and genome detection or by isolation of the virus (when possible) from blood during the first 3 days of illness. Serodiagnosis is possible, but in endemic areas, cross-reactivity with other flaviviruses is a problem.

Management is supportive. Bed rest (under mosquito nets), analgesics, and maintenance of fluid and electrolyte balance are required.

Prevention and control

Yellow fever is an internationally notifiable disease. It is easily prevented by using the attenuated 17D chick embryo vaccine. Vaccination is not recommended for children younger than 9 months of age or in pregnant or immunosuppressed patients unless there are compelling reasons. For the purposes of international certification, immunization is valid for 10 years, but protection lasts much longer than this and probably for life. The WHO EPI includes yellow fever vaccination in endemic areas.

Congo–Crimean haemorrhagic fever

Congo–Crimean haemorrhagic fever (CCHF) virus belongs to the *Orthonairovirus* genus of the family Nairoviridae. This family contains more than 200 viruses, grouped into a number of genera, most of which are arthropod borne.

CCHF disease is found mainly in Asia and Africa. The primary hosts are cattle and hares, and the vectors are *Hyalomma* ticks. After an incubation period of 3 to 6 days, there is an influenza-like illness with fever and haemorrhagic manifestations. The mortality rate is 10% to 50%.

Hantavirus infection

Hantaviruses belong to the *Orthohantavirus* genus of the Hantaviridae family and are enzootic viruses of wild rodents, which are spread by aerosolized excreta and not by insect vectors. The most severe form of this infection is Korean haemorrhagic fever (also known as haemorrhagic fever with renal syndrome). This condition has a mortality rate of 5% to 10% and is characterized by fever, shock, and haemorrhage followed by an oliguric phase. Milder forms of the disease are associated with related viruses (e.g., Puumala virus) and may present as nephropathia epidemica, an acute fever with renal involvement, seen in Scandinavia and other European countries in people who have been in contact with bank voles. In the USA, a hantavirus (transmitted by the deer mouse) termed 'Sin Nombre' was identified as the cause of outbreaks of acute respiratory disease (hantavirus pulmonary syndrome) in adults. Other hantavirus types and rodent vector systems have been associated with this syndrome.

Diagnosis of hantavirus infection is made by ELISA detection of specific antibodies.

Rift Valley fever

Rift Valley fever, caused by a virus from the *Phlebovirus* genus of the Phenuiviridae family, is primarily an acute febrile illness of livestock: sheep, goats, and camels. It is found in southern and eastern Africa. The vectors are *Culex pipiens* in East Africa and *Aedes caballus* in southern Africa, but disease can be transmitted by the bite of an infected animal. After an incubation period of 3 to 6 days, the patient has an acute febrile illness that is difficult to distinguish clinically from other viral fevers. The temperature pattern is usually biphasic. The initial febrile illness lasts 2 to 4 days and is followed by a remission and a second febrile episode. Complications are indicative of severe infection and include retinopathy, meningoencephalitis, haemorrhagic manifestations, and hepatic necrosis. The mortality rate approaches 50% in severe forms. Management is supportive. Animals can be vaccinated.

Lassa fever

Lassa fever virus belongs to the family *Arenaviridae*. Arenaviruses are pleomorphic, round, or oval viruses with diameters ranging from 50 to 300 nm. The virion surface has club-shaped projections, and the virus itself contains a variable number of characteristic electron-dense granules that represent residual, non-functional host ribosomes. Lassa fever illness was first documented in the town of Lassa, Nigeria, in 1969 and is confined to sub-Saharan West Africa (Nigeria, Liberia, and Sierra Leone). The multimammate rat *Mastomys natalensis* is known to be the reservoir. Humans are infected by ingesting foods contaminated by rat urine or saliva containing the virus. Person-to-person spread by body fluids also occurs. Only 10% to 30% of infections are symptomatic. Related arenaviruses are the causative agents of Argentinian and Bolivian haemorrhagic fevers.

Clinical features

The incubation period is 7 to 18 days. The disease is insidious in onset and characterized by fever, myalgia, severe backache, malaise, and headache. A transient maculopapular rash may be present. A sore throat, pharyngitis, and lymphadenopathy occur in more than 50% of patients. In severe cases, there may be epistaxis and GI bleeding, hence the classification of Lassa fever as a viral haemorrhagic fever. The fever usually lasts 1 to 3 weeks, and recovery within 1 month of the onset of illness is usual. However, death occurs in 15% to 20% of hospitalized patients, usually from irreversible hypovolaemic shock.

Diagnosis

The diagnosis is established by serial serological tests (including the Lassa virus-specific IgM titre) or by genome detection in throat swab, serum, or urine.

Management

Management is supportive. In addition, clinical benefit and a reduction in the mortality rate can be achieved with ribavirin therapy if given in the first week.

In non-endemic countries, strict isolation procedures should be used, the patient ideally being nursed in a flexible-film isolator. Specialized units for the management of Lassa fever and other haemorrhagic fevers have been established in the UK. Because Lassa fever virus and other causes of haemorrhagic fever (Marburg/Ebola and Congo–Crimean haemorrhagic fever viruses; see **Box 42.37**) have been transmitted from patients to staff in healthcare situations, great care should be taken in handling specimens and clinical material from these patients.

Marburg virus disease and Ebola virus disease

These severe, haemorrhagic, febrile illnesses are discussed together because their clinical manifestations are similar, and the causative negative-strand RNA viruses belong to the same virus family, the *Filoviridae*, so called because the viruses have a filamentous appearance on electron microscopy. The diseases are named after Marburg in Germany and the Ebola river region in the Sudan and Zaire, where these viruses first appeared. Filoviruses are zoonotic pathogens, with fruit bats thought to be the natural host. The mode of spread from one individual to another involves contact with infected animals or the body fluids (or objects contaminated by the body fluids) of an infected patient. The genus *Ebolavirus* contains six defined species, which vary in terms of case fatality rates: less than 40% for Bundibugyo, 50% for Sudan, and 40% to 90% for Ebola viruses. There has been only one (non-fatal) case of Tai Forest virus infection. Reston virus, the only Asian species, and Bombali virus have not been associated with death in humans.

Epidemics have occurred periodically in recent years, mainly in sub-Saharan Africa. The illness is characterized by the acute onset of severe headache, myalgia, vomiting, and high fever followed by prostration. On about the fifth day of illness, a non-pruritic maculopapular rash develops on the face and then spreads to the rest of the body. Diarrhoea is profuse and associated with abdominal cramps and vomiting. Haematemesis, melaena, or haemoptysis may occur between days 7 and 16. Hepatosplenomegaly and facial oedema are usually present. In Ebola virus disease (EBOV), chest pain and a dry cough are prominent symptoms.

The largest outbreak of EBOV began in West Africa in August 2014 and was declared over by the WHO in June 2016. More than 28 000 individuals were infected, and more than 11 000 people died. The most recent outbreak occurred in Uganda from September 2022 to January 2023 with 164 cases and 77 deaths; that of Marburg virus disease is ongoing in Equatorial Guinea.

Treatment is symptomatic, although a number of experimental therapies, including monoclonal antibodies (e.g., Mab114 and REGN-EB3) and the antivirals remdesivir and favipiravir, are being tried. Ebola vaccines, including a live-attenuated recombinant vesicular stomatitis virus vaccine containing the Ebola virus glycoprotein, have been developed and are recommended by the WHO for use in outbreak situations. In the absence of effective therapy, accurate diagnosis and infection control precautions are the only intervention strategies. Lack of infrastructure to support rapid molecular diagnostic techniques and to allow institution of the meticulous patient isolation necessary to interrupt human-to-human transmission of infection has contributed to generating the recent large outbreaks. Asymptomatic people may be infected but have a low viral count, making transmission unlikely.

Further reading

Okokhere P. Finding a safe and effective vaccine for the Lassa virus. *Lancet* 2023; 401:1242–1243.
PREVAC study team. Randomised trial of vaccines for Zaire Ebola virus disease. *N Engl J Med* 2022; 387:2411–2424.

ANTIVIRAL DRUGS

Drugs for HIV infection are discussed in Chapter 44 ('Anti-retroviral therapy' section).

Anti-herpesvirus drugs

Nucleoside analogues

Aciclovir

Aciclovir (**Fig. 42.18A**; also known as acycloguanosine) is an acyclic nucleoside analogue that acts as a chain terminator of herpesvirus DNA synthesis. This drug is converted to aciclovir monophosphate by a virus-encoded thymidine kinase produced by the alpha herpesviruses, HSV types 1 and 2, and VZV (**Box 42.38**). Conversion to the triphosphate is then achieved by cellular enzymes. Aciclovir triphosphate acts as a potent inhibitor of viral (but not cellular) DNA polymerase and competes with deoxyguanine triphosphate for incorporation into the growing chains of herpesvirus DNA, thereby resulting in chain synthesis termination because of its acyclic structure. This highly specific mode of activity, targeted only to virus-infected cells, means that aciclovir has very low toxicity. Crystallization in the renal tubules is a well-recognized adverse effect, and patients should be well hydrated, particularly when high-dose intravenous therapy is being used. Intravenous, oral, and topical preparations are available for the treatment of patients with HSV and VZV infections. Treatment does not eliminate latent virus, so relapses do occur.

Valaciclovir

Valaciclovir is an oral prodrug of aciclovir. Coupling of valine to the acyclic side chain of aciclovir allows better intestinal absorption. The valine is removed by enzymic action, and aciclovir is released into the circulation. A similar prodrug of a related nucleoside analogue (penciclovir) is the antiherpes drug famciclovir. The mode of action and efficacy of famciclovir are similar to those of aciclovir.

Ganciclovir

The guanine analogue ***ganciclovir*** is structurally similar to aciclovir, with extension of the acyclic side chain by a carboxymethyl group.

It is active against HSVs and VZV by the same mechanism as aciclovir. In addition, phosphorylation by a protein kinase, encoded by the *UL97* gene of CMV, renders it potently active against this virus.

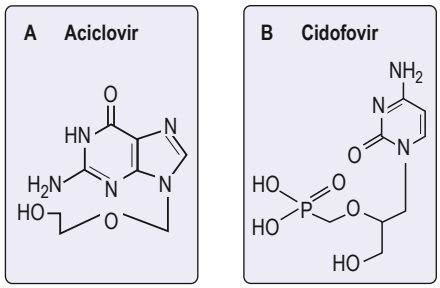


Fig. 42.18 The structures of antiviral drugs. (A) Aciclovir. (B) Cidofovir.

Thus, ganciclovir is currently the first-line treatment for patients with CMV disease. Intravenous and oral preparations are available, as is an oral prodrug, valganciclovir. Unlike aciclovir, ganciclovir has a significant toxicity profile, including neutropenia; thrombocytopenia; and, rarely, sterilization by inhibition of spermatogenesis. It is therefore usually reserved for the treatment or prevention of life- or sight-threatening CMV infection in immunocompromised patients, but it is also recommended for the treatment of symptomatic babies with congenital CMV infection.

Nucleotide analogues

Cidofovir

Cidofovir is a phosphonate derivative of an acyclic nucleoside that acts as a viral DNA polymerase inhibitor (Fig. 42.18B). Phosphorylation to the active triphosphate moiety is entirely by host cell enzymes. It is administered intravenously for the treatment

Box 42.38 Antiviral agents[a]

Virus	Structure or mode of action	Drug	Route of administration	Notes
Anti-herpesvirus	Nucleoside analogue	Aciclovir	Oral, i.v., topical	HSV, VZV
		Valaciclovir	Oral	HSV, VZV
		Famciclovir	Oral	HSV, VZV
		Ganciclovir	i.v.	CMV
		Valganciclovir	Oral	CMV
	Nucleotide analogue	Cidofovir	i.v.	CMV
		Brincidofovir	Oral	CMV
	Pyrophosphate analogue	Foscarnet	i.v.	CMV
	Helicase primase inhibitor	Pritelivir	Oral	HSV
	UL97 inhibitor	Maribavir	Oral	CMV
	Terminase inhibitor	Letermovir	Oral	CMV
Anti-SARS-CoV-2	Inhibitor of viral RNA replication	Remdesivir	i.v.	
	Induces copying errors in viral RNA replication	Molnupiravir	Oral	
	Protease inhibitor	Nirmatrelvir	Oral	Co-administered with ritonavir
Anti-influenza	Neuraminidase inhibitor	Zanamivir	Inhaled	Influenza A, B
		Oseltamivir	Oral	Influenza A, B
		Peramivir	i.v.	Influenza A, B
	Endonuclease inhibitor	Baloxavir	Oral	Influenza A, B
Anti-hepatitis B virus	Nucleoside analogue	Lamivudine	Oral	Rapid resistance
		Telbivudine	Oral	Rapid resistance
		Entecavir	Oral	High barrier to resistance
	Nucleotide analogue	Adefovir	Oral	
		Tenofovir	Oral	High barrier to resistance
Anti-hepatitis C virus (directly acting antivirals, DAA)	NS3 protease inhibitors	Suffix -previr, e.g., grazoprevir	Oral	Usually co-formulated with an NS5a inhibitor
	NS5a inhibitors	Suffix -asvir, e.g., velpatasvir	Oral	Usually co-formulated with an NS3 protease inhibitor or an NS5b polymerase inhibitor
	NS5b polymerase inhibitors	Suffix -buvir, e.g., sofosbuvir	Oral	Usually co-formulated with an NS5a inhibitor
Other	Nucleoside analogue	Ribavirin	Oral, i.v.	Broad spectrum; RSV, HCV
	Interferons		i.m.	Pegylated
	Monoclonal antibodies	Palivizumab Nirsevimab	i.m.	RSV

[a]For drugs used in human immunodeficiency virus, see Chapter 44; for drugs used in the treatment of hepatitis C virus infection, see Chapter 23.
CMV, Cytomegalovirus; *HCV,* hepatitis C virus; *HSV,* herpes simplex virus; *i.m.,* intramuscular; *i.v.,* intravenous; *RSV,* respiratory syncytial virus; *SARS-CoV-2,* severe acute respiratory syndrome coronavirus 2; *VZV,* varicella zoster virus.

of severe CMV infections when other drugs are inappropriate. It is given with probenecid, and because it is nephrotoxic, particular attention should be paid to hydration and monitoring of renal function.

Brincidofovir

Brincidofovir consists of cidofovir bound to a lipid moiety, enabling an oral route of administration and avoiding renal toxicity; it has shown promise in clinical trials.

Pyrophosphate analogue

Foscarnet (sodium phosphonoformate) is a simple pyrophosphate analogue that inhibits viral DNA polymerases. It is active against herpesviruses, and its main roles are as a second-line treatment for severe CMV disease and in the treatment of patients with aciclovir-resistant HSV infection. It is given intravenously, but the potential for severe side-effects, particularly nephrotoxicity, limits its use.

Other anti-herpesvirus agents

Maribavir, an inhibitor of the protein kinase product of the *UL97* gene of CMV, which plays a key role in allowing viral egress from an infected cell, is licensed for use as treatment of CMV-associated disease in transplant recipients. Letermovir, an inhibitor of the CMV terminase enzyme complex, which cuts the string of CMV genomes synthesized in an infected cell into unit-length genomes for packaging in the capsid, is recommended for prevention of CMV reactivation and disease in haemopoietic stem cell transplant recipients. New drugs undergoing clinical trials include pritelivir, an inhibitor of the herpes simplex viral helicase–primase enzyme complex.

Anti-SARS-CoV-2 drugs

Remdesivir

Remdesivir, broad-spectrum adenosine nucleotide analogue, was first developed for the treatment of hepatitis C virus (HCV) infection and was subsequently trialled in the West Africa Ebola virus epidemic. It is administered by intravenous injection and therefore is limited to use in hospital or similar settings. After diffusion into cells and phosphorylation, it interferes with the function of the viral RNA-dependent RNA polymerase enzyme, causing termination of viral RNA synthesis.

Molnupiravir

Molnupiravir is a cytosine nucleoside analogue that introduces copying errors, leading to hypermutation of the genome as the virus replicates. It is administered orally.

Nirmatrelvir

Nirmatrelvir, administered orally, is an inhibitor of the viral protease enzyme responsible for digesting the newly synthesized polyprotein into its component individual proteins. It is co-administered with ritonavir, which was originally developed as an HIV protease inhibitor but more usefully acts as an inhibitor of the liver cytochrome P450 system, which would otherwise metabolise nirmatrelvir.

Anti-influenza drugs

Adamantanes

The use of ***amantadine*** and derivatives in the treatment and prophylaxis of influenza has largely been superseded by the neuraminidase inhibitors. Amantadine is active prophylactically and therapeutically against influenza A (but not B) virus. It is an ion-channel blocker that inhibits viral uncoating after cell entry. CNS side-effects, such as insomnia, dizziness, and headache, may occur, and the drug is poorly tolerated, especially in older adults.

Neuraminidase inhibitors

Zanamivir (administered by inhalation) and ***oseltamivir*** (an oral preparation). These drugs inhibit the action of the neuraminidase of influenza A and B. Both have been shown to be effective in reducing the duration of illness in influenza if given within the first 48 hours. Both may also be used as prophylaxis of influenza among contacts of an index case. Intravenous ***peramivir*** and zanamivir are available for complicated and potentially life-threatening infection resistant to other treatments.

Novel anti-influenza drugs

Baloxavir inhibits the cap-dependent endonuclease activity of the influenza polymerase. Its efficacy is similar to that of oseltamivir, but resistance arises rapidly even after a single dose caused by a mutation in the polymerase gene.

Anti–hepatitis B and anti–hepatitis C drugs

A number of nucleoside and nucleotide analogues have been shown to inhibit the reverse transcriptase function of HBV and thereby suppress viral replication (see Chapter 23). The first such agents, ***lamivudine*** and ***adefovir***, have largely been superseded by the more potent drugs ***entecavir*** and ***tenofovir***, both of which also have a considerably higher barrier to resistance. ***Telbivudine*** is also a nucleoside analogue inhibitor of HBV DNA polymerase activity.

There are three classes of directly acting antiviral drugs (DAAs) used in the treatment of patients with chronic HCV infection. Inhibitors of the viral NS3 protease have the suffix -previr (e.g., grazoprevir), inhibitors of the NS5a multi-functional protein have the suffix -asvir (e.g., pibrentasvir), and inhibitors of the NS5b polymerase enzyme have the suffix -buvir (e.g., sofosbuvir).

The treatment of patients with hepatitis C and B virus infections is discussed in Chapter 23.

Other antiviral drugs

Ribavirin is a synthetic purine nucleoside derivative that interferes with 5′ capping of messenger RNA. It is active against several RNA and DNA viruses, at least *in vitro*. It is occasionally used in the treatment of patients with chronic hepatitis C infection in combination with other agents and for those with chronic hepatitis E virus infection. It is administered orally. Haemolytic anaemia is the most frequent adverse reaction. It is also administered by a small-particle aerosol generator to infants with acute RSV infection. Another indication is in the treatment of patients with Lassa fever virus infection.

Tecovirimat was developed and approved as a treatment of patients with smallpox, with a mechanism of action of inhibiting the orthopoxvirus VP37 envelope wrapping protein. It has subsequently been recommended as a possible therapy for mpox virus infection.

Palivizumab. This monoclonal antibody is specifically indicated to prevent seasonal RSV infection in infants at high risk of severe complications of this infection (including those with bronchopulmonary dysplasia and congenital heart disease). However, this is likely to be replaced by use of nirsevimab, which has an extended half-life and has been trialled successfully for the prevention of RSV disease in neonates and infants.

Interferons

These are naturally occurring proteins with a multiplicity of actions, including antiviral, immunomodulatory, and antiproliferative effects. IFNs are produced by virus-infected cells, macrophages, and lymphocytes. They induce an antiviral state in uninfected cells through activation of a complex set of biochemical pathways. They have been synthesized commercially and are licensed for therapeutic use.

The potency of IFN-α has been enhanced by coupling the protein with polyethylene glycol. The resulting pegylated IFN, given once weekly, has been shown to improve the response to and reduce the side-effects from treatment for hepatitis B and C, although its use in the latter condition has been replaced by the advent of directly acting antiviral agents.

Further reading

Burdet C, Ader F. Real-world effectiveness of oral antivirals for COVID-19. *Lancet* 2022; 400:1175–1176.

TRANSMISSIBLE SPONGIFORM ENCEPHALOPATHIES (PRION DISEASES)

Transmissible spongiform encephalopathies are caused by the accumulation in the nervous system of a protein, termed a 'prion', which is an abnormal isoform (PrP^{Sc}) of a normal host protein (PrP^{c}).

Although familial forms of prion disease are known to exist, these conditions can be transmissible, particularly if brain tissue enters another host. There is no convincing evidence for the presence of nucleic acid in association with prions, and it is the abnormal prion protein itself that is infectious and can trigger a conversion of the normal protein into the atypical isoform. After infection, a long incubation period is followed by CNS degeneration associated with dementia or ataxia, which invariably leads to death. Histology of the brain reveals spongiform change with an accumulation of the abnormal prion protein in the form of amyloid plaques.

The human prion diseases are CJD, including the sporadic, familial, iatrogenic, and variant forms of the disease; Gerstmann–Straussler–Scheinker syndrome; fatal familial insomnia; and kuru.

- ***CJD.*** CJD occurs sporadically worldwide with an annual incidence of 1 per million of the population. Although in most cases, the epidemiology remains obscure, iatrogenic transmission to others has occurred as a result of administration of human cadaveric growth hormone or gonadotrophin, from dura mater and corneal grafting, and in neurosurgery from reuse of contaminated instruments and electrodes (iatrogenic CJD).
- ***Variant CJD (vCJD).*** In the UK, knowledge that large numbers of cattle with the prion disease, bovine spongiform encephalopathy (BSE), had gone into the human food chain led to enhanced surveillance for emergence of the disease in humans. Based on transmission studies in mice and on glycosylation patterns of prion proteins, the evidence is convincing that this has occurred, and to date, there have been approximately 180 confirmed and suspected cases of vCJD (human BSE) in the UK and 50 in the rest of the world. In contrast to sporadic CJD, which presents with dementia at a mean age of onset of 60 years, vCJD presents with ataxia, dementia, myoclonus, and chorea at a mean age of onset of 29 years. Although the epidemic in the 1990s has now passed and new cases are extremely rare, there remains a concern about longer incubation periods, and there are ongoing international restrictions on blood transfusion from those who were resident in or received transfusions from the UK at that time.
- ***Gerstmann–Straussler–Scheinker syndrome and fatal familial insomnia.*** These are rare prion diseases, usually occurring in families with a positive history. The pattern of inheritance is autosomal dominant with some degree of variable penetrance. The gene encoding PrP^{C} in these families often contains mutations.
- ***Kuru.*** This disease was described and characterized in the Fore highlanders in northeast New Guinea. Transmission was associated with ritualistic cannibalism of deceased relatives. With the cessation of cannibalism by 1960, the disease has gradually diminished, and recent cases were all exposed to the agent before 1960.

The infectious agents of prion disease have remarkable characteristics. In an infected host, there is no evidence of inflammatory, cytokine, or immune reactions. The agent is highly resistant to decontamination, and infectivity is not reliably destroyed by autoclaving or by treatment with formaldehyde and most other gas or liquid disinfectants. It is very resistant to γ-irradiation. Autoclaving at a high temperature (134°–137°C for 18min) is used for decontamination of instruments and hypochlorite (20 000 ppm available chlorine) or 1-molar sodium hydroxide is used for liquid disinfection. Uncertainty about the reliability of any methods for safe decontamination of surgical instruments has necessitated the introduction of guidelines for patient management.

Further reading

Zerr I. Laboratory diagnosis of Creutzfeldt-Jakob disease. *N Engl J Med* 2022; 386:1345–1350.

BACTERIAL INFECTIONS

Bacteria are single-celled organisms that form the major part of the normal human microbiome. Bacterial infection is caused by commensal bacteria invading normally sterile parts of the body or by pathogenic bacteria that are not normal commensal organisms. Infection can be the result of translocation of bacteria from non-sterile to sterile areas (e.g., from the gut to the urinary tract), acquisition from other people or animals, or transmission on fomites or in the environment.

In some cases, bacterial infection may be localized to a specific site, without systemic symptoms. In other circumstances, a localized infection may be accompanied by systemic illness or sepsis (see Chapter 12). Sometimes there is no obvious focus of infection, and the presentation is with undifferentiated fever. In a few cases, bacteria cause illness through release of a specific toxin, which produces symptoms distant from the actual site of infection (**Box 42.39**).

Some bacterial infections are found worldwide, but others are geographically restricted. Often these relate to particular climates and habitats that support the transmitting vector or disease or certain communities, such as those that have limited access to fresh water, good hygiene, and medical care. These factors are often interdependent, and many of the conditions that are traditionally labelled as 'tropical' are, in fact, related to socioeconomic conditions rather than climate.

Many of the common organ-specific bacterial conditions are covered in the relevant system chapters and sepsis is described in Chapter 12. This section deals with those bacterial infections that

Box 42.39 Classification of more common bacteria affecting humans

Bacteria	Cocci	Bacilli	Others
Aerobic			
Gram positive	*Staphylococcus aureus* *Staphylococcus epidermidis* *Streptococcus pneumoniae* *Streptococcus pyogenes* (group A) *Streptococcus agalactiae* (group B) Enterococci *Viridans* streptococci	*Listeria monocytogenes* *Corynebacterium diphtheriae* *Bacillus anthracis* *Bacillus cereus*	
Gram negative	*Neisseria gonorrhoeae* *Neisseria meningitidis* *Moraxella catarrhalis* *Bordetella pertussis*	*Escherichia coli* *Klebsiella* spp. *Proteus* spp. *Haemophilus influenzae* *Legionella* spp. *Salmonella* spp. *Shigella* spp. *Campylobacter jejuni* *Helicobacter pylori* *Pseudomonas* spp. *Brucella* spp. *Acinetobacter* spp. *Burkholderia* spp. *Vibrio cholerae* *Yersinia pestis*	
Anaerobic			
Gram positive	Peptococci Peptostreptococci	*Actinomyces* spp. *Clostridium perfringens* *Clostridioides difficile* *Clostridium botulinum* *Clostridium tetani*	
Gram negative		*Bacteroides fragilis* group *Fusobacterium* spp.	
Spirochaetes			*Treponema pallidum* *Leptospira* spp. *Borrelia* spp.
Others			*Mycobacterium* spp. *Mycoplasma pneumoniae* *Ureaplasma* spp. *Chlamydia* spp. *Rickettsia* spp.

do not fit readily into an organ-specific model and those that are more defined by geographical and economic constraints than by body site.

BACTERIAL INFECTIONS OF THE SKIN AND SOFT TISSUES

The most common bacterial infections of skin and soft tissue are covered in Chapter 11, although some of the more unusual causes are described here and listed in **Box 42.40**. Although the commonest cause of cellulitis is beta-haemolytic streptococci, the majority of infections localized to skin and soft tissue structures (folliculitis, boils, infected insect bites, wounds, and ulcers) are caused by *S. aureus*. Staphylococci are part of the normal microflora of the human skin and nasopharynx and up to 25% of people are carriers of *S. aureus*. Although most *S. aureus* infections remain superficial, this organism can spread from the skin to cause a variety of more serious infections (see **Box 42.22**). Other species of staphylococci are only rarely pathogenic.

Invasive staphylococcal infection

Invasive staphylococcal infection is often associated with breaches in the skin, for example, those caused by injecting drug use, iatrogenic cannulation, surgery, or trauma, although an entry point is not always identified. Other predisposing factors are listed in **Box 42.23**. In clinical situations, scrupulous attention to disinfection and hygiene when performing invasive procedures can minimize the risk of infection. Although *S. aureus* is the most common cause of intravenous catheter–related infections, other normally non-pathogenic species, such as *S. epidermidis*, may be found. The distinction is important: *S. epidermidis*, sometimes known as coagulase-negative *Staphylococci*, are ubiquitous on the skin and may 'contaminate' blood culture samples, leading to false-positive results. When identified as a genuine cause of bloodstream infection, *S. epidermidis* usually colonizes indwelling catheters or prosthetic heart valves, and although generally less aggressive than *S. aureus*, it is increasingly resistant to beta-lactam antibiotics.

Flucloxacillin is a beta-lactam antibiotic and remains the first choice for staphylococcal infection when the organism is known to be sensitive

Box 42.40 Bacterial causes of superficial skin and soft tissue infections

Specific risk factors	Likely organisms
None	*Streptococcus pyogenes* *Staphylococcus aureus*
Diabetes, peripheral vascular disease	Group B streptococci
Animal bite	*Pasteurella multocida*, *Capnocytophaga canimorsus*
Freshwater exposure	*Aeromonas hydrophila*
Seawater exposure	*Vibrio vulnificus*
Lymphoedema, stasis dermatitis	Groups A, C, and G streptococci
Hot tub exposure	*Pseudomonas aeruginosa*
Malignant otitis externa	*Pseudomonas aeruginosa*
Human bite	*Fusobacterium* spp.

(see Chapter 12 , 'Treatment of infection caused by multidrug-resistant organisms' section and Fig. 12.2), but when MRSA is confirmed or suspected, alternative anti-staphylococcal agents are required.

Staphylococcal virulence factors

S. aureus can produce a variety of toxins and virulence factors that affect the type and severity of infection. These include staphylococcal enterotoxin A, superantigenic staphylococcal exotoxins, toxic shock toxin 1, and PVL. The last of these has been found mainly in community-associated strains of *S. aureus* (both MRSA and meticillin-sensitive *S. aureus* [MSSA]) rather than in hospital-acquired or epidemic strains. PVL-producing MRSA and MSSA are common causes of invasive soft tissue infections in some countries such as the USA, although it is unclear whether PVL itself is directly responsible for the increased virulence. Severe lung infections with PVL-associated strains also occur.

Meticillin-resistant *Staphylococcus aureus*

Penicillin was introduced for public use in 1946 and transformed the treatment of patients with skin and soft tissue infections. However, by the late 1960s, more than 80% of *S. aureus* infections had become resistant to penicillin, and the semi-synthetic meticillin (now rarely used) and flucloxacillin became the drugs of choice. However, since their introduction, strains of MRSA with resistance to a much wider range of antibiotics have emerged and are now a major cause of concern.

MRSA is usually found as a skin commensal, especially in hospitalized patients and nursing home residents in the UK. Often the organisms found in community settings are the typical 'hospital' strains of MRSA and have been acquired directly or indirectly from a healthcare setting (e.g., in workers in care homes). However, there is an increasing prevalence in some countries of true community-associated MRSA (CA-MRSA), with no discernible links to hospital or residential care. These CA-MRSA strains have different resistance profiles to those of typical hospital strains (often retaining sensitivity to tetracyclines, clindamycin, and co-trimoxazole) and are more likely to produce PVL. They are particularly associated with surgical wound infections. Eradication of the organisms is difficult, and patients who are known to be colonized should be isolated or barrier nursed. Topical decolonization is used to decrease bacterial load before invasive procedures but is of limited efficacy in completely clearing carriage. Careful attention to hand-washing and hygiene is the main method of controlling spread.

Pasteurellosis

Pasteurella multocida is found in the oropharynx of up to 90% of cats and 70% of dogs. It can cause soft tissue infections after animal bites. Although the infection initially resembles other forms of cellulitis, there is a greater tendency to spread to deeper tissues, resulting in osteomyelitis, tenosynovitis, or septic arthritis. It can also spread haematogenously to other organs, including the lungs, brain, and heart. The organism is sensitive to penicillin, but because infections are often polymicrobial, co-amoxiclav is recommended after animal bites.

Cat-scratch disease

Cat-scratch disease is a zoonosis caused by *Bartonella henselae*. Up to one-third of cats, and in particular kittens, carry *B. henselae* in their blood but are generally asymptomatic. Human infection is uncommon and is probably caused by direct inoculation from the claws or via cat flea bites. Regional lymphadenopathy appears 1 to 2 weeks after infection; the nodes become tender and may suppurate. Histology of the nodes shows granuloma formation, and the illness may be mistaken for mycobacterial infection or lymphoma. There are usually few systemic symptoms in immunocompetent patients, although more severe disease may be seen in immunocompromised patients. In these patients, tender cutaneous or subcutaneous nodules are seen ***(bacillary angiomatosis)***, which may ulcerate. The lymphadenopathy resolves spontaneously over weeks or months, although surgical drainage of very large suppurating nodes may be necessary. Rarely, *B. henselae* may cause 'culture-negative' endocarditis, retinitis, or meningoencephalitis. It is sensitive to azithromycin, doxycycline, and ciprofloxacin, but drug selection and clinical benefit of treatment vary according to the primary site of the infection and the immune status of the patient.

Toxin-mediated skin disease

A number of skin conditions, although caused by bacteria, are mediated by exotoxins rather than direct local tissue damage.

Staphylococcal scalded skin syndrome

The scalded skin syndrome is caused by a toxin-secreting strain of *S. aureus*. It principally affects children younger than the age of 5 years. The toxin, exfoliatin, causes intra-epidermal cleavage at the level of the stratum corneum, leading to the formation of large, flaccid blisters that shear readily. It is a relatively benign condition and responds to treatment with flucloxacillin.

Toxic shock syndrome

TSS is usually caused by toxin-secreting staphylococci, but toxin-secreting streptococci have also been implicated. Although historically associated with vaginal colonization and tampon use, changes in tampon constituents have made this a rare portal of entry now. The exotoxin (normally toxic shock syndrome toxin 1) induces cytokine release, causing abrupt onset of fever and shock, with a diffuse macular rash and desquamation of the palms and soles. Many patients are severely ill, and the mortality rate is about 5%. It is important to look for and remove the source. Management is mainly supportive, although the organism should be eradicated, and there is evidence to support use of intravenous immunoglobulin.

BACTERIAL INFECTIONS OF THE RESPIRATORY TRACT

Infections of the respiratory tract are divided into those affecting the upper respiratory tract and those affecting the lower respiratory tract, which are separated by the larynx. The majority of common respiratory infections are covered elsewhere.

Upper respiratory tract infections

Scarlet fever

Scarlet fever occurs when the infectious organism (usually GAS but occasionally group C or G streptococci) produces erythrogenic toxin in an individual who does not possess neutralizing antitoxin antibodies. Infections may be sporadic or epidemic, occurring in residential institutions such as schools, prisons, and military establishments. It was an important cause of mortality in children in the early part of the 20th century, but improvements in living conditions and sanitation between the wars led to a decline in the mortality rate even before the introduction of antibiotics.

Clinical features

The onset of this relatively mild disease, which mainly affects children, is 2 to 4 days after a streptococcal infection (usually in the pharynx). It is often sore to swallow. Common features include regional lymphadenopathy, fever, rigors, headache, and vomiting. The rash usually appears on the second day of illness and is generalized but typically absent from the face, palms, and soles. It is described as having a sandpaper texture and blanches on pressure. It usually lasts for about 5 days and is followed by extensive desquamation of the skin (Fig. 42.19). The face is flushed, with characteristic circumoral pallor. Early in the disease, the tongue has a white coating, through which prominent bright red papillae can be seen ('strawberry tongue'). Later the white coating disappears, leaving a raw-looking, bright red colour ('raspberry tongue'). The patient is infectious for 10 to 21 days after the onset of the rash unless treated with antibiotics. A runny nose, hoarse voice, and conjunctivitis are usually absent, helping to differentiate scarlet fever from a viral infection.

Scarlet fever may be complicated by peritonsillar or retropharyngeal abscesses and otitis media.

The diagnosis is established by a combination of the typical clinical features with culture of a throat swab. Patients with scarlet fever should be treated with penicillin, to which GAS are invariably sensitive.

Diphtheria

Diphtheria occurs worldwide and is caused by *Corynebacterium diphtheriae*. It is a Gram-positive bacillus and is mainly transmitted through air-borne droplet spread. Only strains that carry the *tox+* gene are capable of toxin production. *C. diphtheriae* is closely related to *Corynebacterium ulcerans*, which is primarily a disease of livestock, although toxigenic strains have been reported to cause human disease. Once common, diphtheria is now seldom seen in countries with effective vaccination programmes and is most likely to occur in migrants who have not been vaccinated.

Clinical features

Diphtheria was formerly a disease of childhood but may affect adults in countries where childhood immunization has been interrupted or uptake is poor. The incubation period is 2 to 7 days. The manifestations are local (caused by the membrane) or systemic (caused by exotoxin). The presence of a membrane, however, is not essential to the diagnosis.

Fig. 42.19 **Scarlet fever rash, showing desquamation.**

Nasal diphtheria is characterized by the presence of a unilateral, serosanguineous nasal discharge that crusts around the external nares.

Pharyngeal diphtheria is associated with the greatest toxicity and is characterized by marked tonsillar and pharyngeal inflammation and usually the presence of a membrane. The tough, greyish-yellow membrane is formed by fibrin, bacteria, epithelial cells, mononuclear cells, and polymorphs and is firmly adherent to the underlying tissue. Regional lymphadenopathy, often tender, is prominent and produces the so-called 'bull-neck'.

Laryngeal diphtheria is usually a result of extension of the membrane from the pharynx. A husky voice, brassy cough, and later dyspnoea and cyanosis caused by respiratory obstruction are common features.

Diphtheria toxin principally affects the heart and the nervous system. Clinically evident myocarditis occurs in a proportion of patients with pharyngeal or laryngeal diphtheria, sometimes weeks after initial infection. Acute circulatory failure caused by myocarditis may occur in convalescent individuals around the 10th day of illness and can be fatal. Neurological manifestations occur either early in the disease (palatal and pharyngeal wall paralysis) or several weeks after its onset (cranial nerve palsies; paraesthesiae; polyneuropathy; or, rarely, encephalitis).

Cutaneous diphtheria, caused by *C. ulcerans*, is uncommon. It is mostly seen in association with residence in or travel to endemic countries, with farm or companion animal contact, and in individuals with poor personal hygiene. Typically, the ulcer is punched out with undermined edges and is covered with a greyish white to brownish adherent membrane. Constitutional symptoms are uncommon.

Diagnosis

This must be made on clinical grounds because therapy is usually urgent; the untreated mortality rate is about 10%. It is confirmed by bacterial culture from a pharyngeal swab and toxin studies.

Management

Antitoxin therapy is the only specific treatment. One trial showed that it reduced the mortality rate from 12.2% to 3.3%. It must be given promptly to prevent further fixation of toxin to tissue receptors because fixed toxin is not neutralized by antitoxin. Depending on the site and severity, 40 000 to 100 000 units of horse-serum antitoxin should be administered intramuscularly after an initial test dose to exclude any allergic reaction. Intravenous therapy may be required in very severe cases. There is a small risk of anaphylaxis after antitoxin administration and of serum sickness 2 to 3 weeks later (Box 42.41). However, the risk of death outweighs the problems of anaphylaxis. Antibiotics (penicillin, amoxicillin, or a macrolide) should be administered concurrently to eliminate the organisms and thereby remove the source of toxin production.

Box 42.41 Antitoxin administration

- Many antitoxins are heterologous and therefore dangerous.
- Hypersensitivity reactions are common.

Before treatment

- Question the patient about:
  - Allergic conditions (e.g., asthma, hay fever)
    - Previous antitoxin administration
- Read the instructions on the antitoxin package carefully.
- Always give a subcutaneous test dose.
- Have adrenaline (epinephrine) available.

The cardiac and neurological complications need intensive supportive therapy. Recovery and rehabilitation may take many weeks.

Prevention

Patients with suspected diphtheria must be kept in respiratory isolation, and the local public health authorities should be informed. Staff caring for the patient should have documented immunization status. All contacts of the patient should have throat swabs sent for culture; those with positive results should be treated with antibiotics and given active immunization or a booster dose of diphtheria toxoid. Diphtheria is prevented by immunization in childhood (see Chapter 9, 'Vaccination' section). Travellers to high-risk countries may require a booster dose, usually in combination with tetanus and pertussis, because monovalent diphtheria vaccine is no longer available.

Acute epiglottitis

Acute epiglottitis has been virtually eliminated among children in countries that have introduced *H. influenzae* vaccination. Infections are occasionally seen in adults. The clinical features are described in Chapter 21 ('The throat' section).

Lower respiratory tract infections

Pneumonia, lung abscess, and tuberculosis

For community-acquired pneumonia, see Chapter 11 ('Community-acquired pneumonia' section); hospital-acquired pneumonia, Chapter 11 ('Hospital-acquired pneumonia' section); and pneumonia in immunocompromised persons, Chapter 11 ('Pneumonia in immunocompromised patients' section.

Lung abscess is described in Chapter 11 ('Lung abscess' section).

TB is described in Chapter 20 ('Tuberculosis' section).

Further reading

Truelove SA, Keegan LT, Moss WJ et al. Clinical and epidemiological aspects of diphtheria: a systematic review and pooled analysis. *Clin Infect Dis* 2020; 71(1):89–97. Published online 2019 Aug 19.

Turner NA, Sharma-Kuinkel BK, Maskarinec SA et al. Methicillin-resistant *Staphylococcus aureus*: an overview of basic and clinical research. *Nat Rev Microbiol* 2019; 17:203–218.

https://cks.nice.org.uk/topics/bites-human-animal/management/managing-a-cat-or-dog-bite/.

BACTERIAL INFECTIONS OF THE GASTROINTESTINAL TRACT

Gastroenteritis

The most common form of acute GI infection is gastroenteritis, causing diarrhoea with or without vomiting. Children in LMICs can expect, on average, 3 to 6 bouts of severe diarrhoea annually. Although oral rehydration programmes have cut the death toll significantly, more than half a million children die every year as a direct result of diarrhoeal disease. In high-income settings, diarrhoea is both less common and less likely to cause death but remains a major cause of morbidity, especially in older adults. Other groups at increased risk of infectious diarrhoea include travellers to LMICs, people in residential care, and infants in day-care facilities.

Viral gastroenteritis (see Chapter 11, 'Acute gastrointestinal infections' section) is the commonest cause of diarrhoea and vomiting in young children but is less commonly seen in adults, other than in the context of common source outbreaks (usually caused by norovirus). Protozoal and helminthic gut infections (see 'Protozoal infections' section) are rare in HICs but relatively common in LICs. The most common cause of significant adult gastroenteritis worldwide is bacterial infection.

Bacteria can cause diarrhoea in three different ways (**Box 42.42**). Some species may use more than one of these methods. In addition to these direct mechanisms, a proportion of people develop post-infectious irritable bowel syndrome, a functional bowel disorder triggered by infection but persisting after resolution of the inflammation (see Chapter 22, 'Eosinophilic gastroenteritis' section).

Mucosal adherence

Most bacteria causing diarrhoea must first adhere to specific receptors on the gut mucosa; a number of different molecular adhesion mechanisms have been elaborated. For some pathogens, this is merely the prelude to invasion or toxin production, but others, such as enteropathogenic *E. coli*, cause mucosal lesions and produce a secretory diarrhoea as a direct result of adherence. Enteroaggregative *E. coli* adhere in an aggregative pattern, with the bacteria clumping on the cell surface, and their toxin causes persistent diarrhoea. Diffusely adhering *E. coli* adheres in a uniform manner and

Box 42.42 Pathogenic mechanisms of bacterial gastroenteritis

Pathogenesis	Mode of action	Clinical presentation	Examples
Mucosal adherence	Effacement of intestinal mucosa	Moderate watery diarrhoea	Enteropathogenic *Escherichia coli* Enteroaggregative *E. coli* *E. coli* O14:H4 Diffusely adhering *E. coli*
Mucosal invasion	Penetration and destruction of mucosa	Dysentery	*Shigella* spp. *Campylobacter* spp. Enteroinvasive *E. coli*
Toxin production			
Enterotoxin	Fluid secretion without mucosal damage	Profuse watery diarrhoea	*Vibrio cholerae* *Salmonella* spp. *Campylobacter* spp. Enterotoxigenic *E. coli* *Bacillus cereus* *Staphylococcus aureus* producing enterotoxin B *Clostridium perfringens* type A
Cytotoxin	Damage to mucosa	Dysentery	*Salmonella* spp. *Campylobacter* spp. Enterohaemorrhagic *E. coli* O157 *E. coli* O104:H4

may also cause diarrhoea. Both are seen mainly in LICs with poor hygiene and inadequate clean water supply.

E. coli O104:H4, which has caused huge outbreaks of gastroenteritis in Germany and other high-income countries, has two different diarrhoea-causing *E. coli* pathotypes: typical enteroaggregative *E. coli* and Shiga-toxin-producing *E. coli*.

Mucosal invasion

Invasive pathogens, such as *Shigella* spp., enteroinvasive *E. coli* (EIEC) and *Campylobacter* spp., penetrate the intestinal mucosa. Initial entry into the mucosal cells is facilitated by the production of 'invasins', which disrupt the host-cell cytoskeleton. Subsequent destruction of the epithelial cells allows further bacterial entry and associated colitis, with the typical symptoms of dysentery: low-volume bloody diarrhoea with abdominal pain.

Toxin production

Gastroenteritis can be caused by different types of bacterial toxin. ***Enterotoxins*** produced by bacteria adhering to the intestinal epithelium induce excessive fluid secretion into the bowel lumen, leading to watery diarrhoea without physically damaging the mucosa (e.g., cholera, enterotoxigenic *E. coli* [ETEC]). Some enterotoxins that are pre-formed in food primarily cause vomiting (e.g., enterotoxigenic *S. aureus* and *Bacillus cereus*). ***Cytotoxins*** damage the intestinal mucosa and, in some cases, vascular endothelium as well (e.g., *E. coli* O157).

Clinical syndromes

Bacterial gastroenteritis can be divided on clinical grounds into two broad syndromes: ***watery diarrhoea*** (usually due to enterotoxins or adherence) and ***dysentery*** (characterized by bloody diarrhoea; usually caused by mucosal invasion and damage in the colon) (**Box 42.43**). With some pathogens, such as *Campylobacter jejuni*, there may be overlap between the two syndromes. A third clinical type of infective gastroenteritis, ***chronic diarrhoea***, is rarely bacterial in origin.

Management of infective diarrhoea is usually syndromic (see later) because except during outbreaks, the causative organism is rarely known at presentation. Bacterial gastroenteritis is usually self-limiting and in otherwise healthy individuals does not require antibacterials unless the infection is particularly severe or prolonged. Patients with dysentery are generally treated with antibiotics in all age groups; in children, this is proven to reduce mortality rates (**Box 42.44**).

Box 42.43 Bacterial causes of watery diarrhoea and dysentery

Watery diarrhoea	Dysentery
• *Bacillus cereus* • *Staphylococcus aureus* } plus profuse vomiting	• *Shigella* spp. • Salmonella spp.
• *Vibrio cholerae*	• *Campylobacter* spp.
• *Enterotoxigenic Escherichia coli*	• *Enteroinvasive E. coli*
• Enteropathogenic *E. coli*	• Enterohaemorrhagic *E. coli*
• *Salmonella* spp.	• *Yersinia enterocolitica*
• *Campylobacter jejuni*	• *Vibrio parahaemolyticus*
• *Clostridium perfringens*	• *Clostridioides difficile*
• *Clostridioides difficile*	

Salmonella spp.

Gastroenteritis can be caused by many of the numerous serotypes of non-typhoidal *Salmonella enterica*, but the most commonly implicated are *S. enterica enteritidis* and *S. enterica typhimurium* (*S. enterica typhi* and *S. enterica paratyphi*, the causes of enteric fever, which are covered in the 'Enteric fever' section). These organisms, which are found all over the world, are commensals in the bowels of livestock (especially poultry) and oviducts of chicken (where the eggs can become infected). They are usually transmitted to humans in contaminated foodstuffs and water.

Salmonellae can affect both the large and the small bowel and induce diarrhoea both by production of enterotoxins and by invasion. The typical symptoms commence abruptly, 12 to 48 hours after infection, and consist of nausea, cramping abdominal pain, diarrhoea, and sometimes fever. The diarrhoea can vary from profuse and watery to a bloody dysentery syndrome. Spontaneous resolution usually occurs in 3 to 6 days, although the organism may persist in the faeces for several weeks. Bloodstream infection occurs in 1% to 4% of cases. Some serotypes are more prone to cause bacteraemia than others, but it is also more common in older adults and immunosuppressed people; in settings of high HIV prevalence, bacteraemia without gastroenteritis is common. Occasionally, bacteraemia is complicated by metastatic infection, especially of atheroma on vascular endothelium, with potentially devastating consequences. Salmonella osteomyelitis is seen in people with sickle cell disease. In healthy adults, *Salmonella* gastroenteritis is usually a relatively minor illness, but young children and older adults are at risk of significant dehydration.

Specific diagnosis is made by isolating the organism in blood or faeces, but management is usually empirical and includes oral rehydration. Antibiotic resistance in salmonellae is increasing, with more than 10% of European human isolates resistant to ciprofloxacin and in some countries more than 50%. Resistance third-generation cephalosporins and azithromycin remains rare (<1%).

Campylobacter jejuni

The Gram-negative bacillus *C. jejuni* is a zoonotic organism, existing as a bowel commensal in many species of livestock, including poultry and cattle. It is found worldwide and is an important cause of childhood gastroenteritis in low- and middle-income countries. Adults in these countries may be tolerant of the organism, excreting it asymptomatically. In the West, it is a common cause of sporadic food-borne outbreaks of gastroenteritis. About 50 000 cases are confirmed annually in England and Wales, but the actual incidence is much higher. The most common sources are undercooked meat (especially beefburgers and chicken) and contaminated milk products and water. In many high-income countries, there is significant seasonal variation, with peaks associated with warm weather and barbecues.

Like *Salmonella*, *Campylobacter* spp. can affect the large and small bowel and can cause a wide variety of symptoms. The incubation period is usually 2 to 4 days, after which there is an abrupt onset of nausea, diarrhoea, and severe abdominal cramps. The diarrhoea is usually profuse and watery, but an invasive haemorrhagic colitis is sometimes seen. Bacteraemia is very rare, and infection is usually self-limiting in 3 to 5 days. Diagnosis is made from culture or, increasingly, molecular tests such as multiplex PCR, on stool. Quinolone resistance is now widespread in Europe (>60%); this is due in part to the use of antibiotics in intensive livestock farming. Antibiotics are not usually indicated, but if symptoms are severe, azithromycin is the drug of choice.

Box 42.44 Antibiotics in adult acute bacterial gastroenteritis

Condition	Indications	Drug of choice	Other drugs	Benefits
Cholera	Severely dehydrated patients only (World Health Organization guidance)	Ciprofloxacin 500 mg twice daily	Tetracycline 250 mg four times daily Azithromycin 1 g single dose Doxycycline 300 mg single dose	Relieve symptoms Shorten illness
Treatment of confirmed *Salmonella* spp. infection[a] (depending on resistance profile)	Symptoms not improving (rarely needed)	Azithromycin 500 mg once daily	Ciprofloxacin 500 mg twice daily for 7–10 days; ceftriaxone 2 g i.v. daily	May shorten illness
Treatment of confirmed *Campylobacter* spp. infection[a]	Symptoms not improving (rarely needed)	Azithromycin 500 mg once daily for 3 days	Erythromycin 500 mg four times a day for 3 days	May shorten illness
Treatment of confirmed shigellosis (depending on resistance profile)	Most patients (unless symptoms resolved)	Ciprofloxacin 500 mg twice daily	Azithromycin 500 mg once daily Ceftriaxone 2 g daily (i.v. or i.m.)	Relieve symptoms Shorten illness Reduce mortality rate in children Decrease transmission
Clostridioides difficile	Most cases (unless symptoms resolved)	Vancomycin 125–250 mg four times daily	Metronidazole 400 mg three times daily Fidaxomicin 200 mg twice daily	Relieve symptoms Shorten illness May reduce relapse rate
Empirical therapy of watery diarrhoea[a]	Severe symptoms Prolonged illness Older adult patients Immunosuppressed	Azithromycin 500 mg once daily	Ceftriaxone 2 g/day i.v.	Relieve symptoms Shorten illness May decrease complications
Travellers' diarrhoea[a]	Rarely used	Azithromycin 500 mg/day for three days	Ciprofloxacin 500 mg twice daily for 3 days	Relieve symptoms Shorten illness
Empirical therapy for dysentery (unless E, histolytica suspected)	Most patients	Azithromycin 500 mg/day for 3 days	Ciprofloxacin 500 mg twice daily for 3 days	Relieve symptoms Shorten illness

[a]Antibiotics are not needed for the majority of adult cases in high-income countries.
i.m., Intramuscular; *i.v.*, intravenous.

Shigella spp.

Shigellae are enteroinvasive Gram-negative bacteria, which cause classical bacillary dysentery. The principal species causing gastroenteritis are *Shigella dysenteriae*, *Shigella Flexneri*, and *Shigella sonnei*, which are found with varying prevalence in different parts of the world. All cause a similar syndrome as a result of damage to the intestinal mucosa. Some strains of *S. dysenteriae* also secrete a cytotoxin affecting vascular endothelium. Although shigellae are found worldwide, transmission is strongly associated with poor hygiene, and of the 165 million episodes of shigella dysentery estimated by the WHO to occur every year, 99% are in economically low- and middle-income countries. The main burden falls on children: *Shigella* dysentery kills nearly 50 000 children younger than 5 years old every year. There have been recent major European outbreaks of *Shigella* among males who have sex with males, often associated with significant antimicrobial resistance.

The organism is spread from person to person (either directly or via contaminated food or water), and only a small number of bacteria need to be ingested to cause illness (<200 compared with 10^4 for *Campylobacter* spp. and >10^5 for *Salmonella* spp.). Symptoms start 24 to 48 hours after ingestion and typically consist of frequent, small-volume stools containing blood and mucus. Dehydration is not as significant as in the secretory diarrhoeas, but systemic symptoms and intestinal complications are worse. The illness is usually self-limiting in 7 to 10 days, but in children in low-income settings, it can carry a significant mortality rate.

Antibiotic treatment decreases the severity and duration of diarrhoea, reduces the mortality rate in children, and possibly lowers the risk of further transmission. Resistance to antibiotics is widespread and, wherever possible, treatment should be based on known local sensitivity patterns. Antibiotics such as co-trimoxazole, which were widely used in the past, are no longer recommended by the WHO because of extensive resistance. This is increasingly the case for fluoroquinolones also, and most guidelines increasingly recommend azithromycin as the first-choice antibiotic, with ceftriaxone for severe disease. Resistance to both these antibiotics has been recognized in a recent outbreak in males who have sex with males.

Enteroinvasive *Escherichia coli*

EIEC cause an illness that is indistinguishable from shigellosis. Definitive diagnosis is made by isolation in stool, but most patients are probably treated empirically as having shigellosis.

Enterohaemorrhagic *Escherichia coli*

Enterohaemorrhagic *E. coli* (EHEC; usually serotype O157:H7, also known as verotoxin-producing *E. coli*) is a well-recognized cause of gastroenteritis in humans. It is a zoonosis that is usually associated with cattle, and there have been a number of major outbreaks linked to contaminated food. A variety of modes of transmission have been reported and the epidemiology of EHEC infection is a paradigm for all enteric livestock-associated zoonoses (Fig. 42.20). EHEC secretes a toxin (Shiga-like toxin 1) that affects vascular endothelial cells in the gut and in the kidney. After an incubation period of 12 to 48 hours, it causes diarrhoea (frequently bloody), associated with abdominal pain and nausea. Some days after the onset of symptoms, the patient may develop thrombotic thrombocytopenic purpura (see Chapter 38, 'Thrombotic

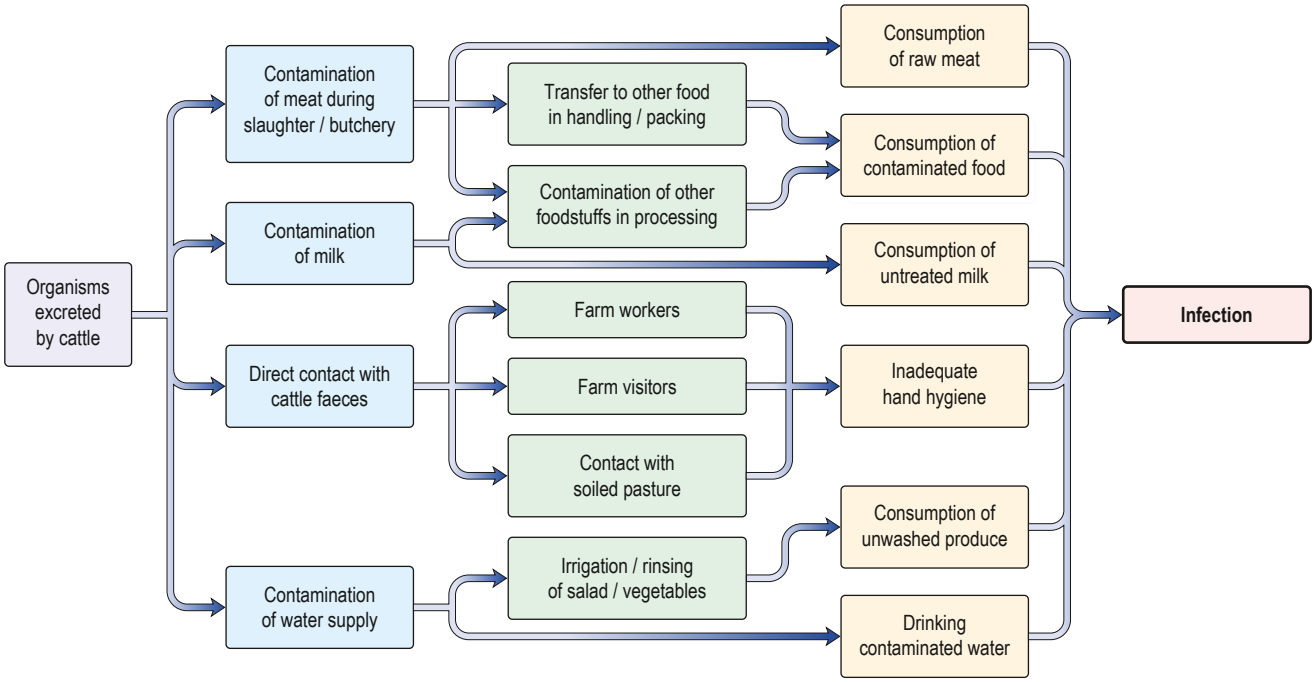


Fig. 42.20 **Routes of human infection with *Escherichia coli* O157.**

thrombocytopenic purpura' section) or haemolytic uraemic syndrome (HUS; see Chapter 26, 'Haemolytic uraemic syndrome' section). These are more common in children and may lead to permanent renal damage or death. Non-O157 serotypes are of increasing concern.

Between May and June 2011, the largest ever recorded outbreak of *Shiga* toxin–producing *E. coli* causing HUS was recorded in Germany. The outbreak was caused by the O104 serotype, and more than 2000 people were affected. High rates of HUS were observed in adults not in the typical 'at-risk' age range. The increased virulence of this strain is possibly due to its having two different pathotypes.

Treatment is mainly supportive; there is evidence that antibiotic therapy might precipitate HUS by causing increased toxin release, although this remains controversial.

Enterotoxigenic *Escherichia coli*

ETEC produce both heat-labile and heat-stable enterotoxins, which stimulate secretion of fluid into the intestinal lumen. The result is watery diarrhoea of varying intensity, which usually resolves within a few days. Transmission is normally from person to person via contaminated food and water. The organism is common in low- and middle-income countries and is a major cause of travellers' diarrhoea (see later).

Vibrio spp.

Cholera, caused by *V. cholerae*, is the prototypic pure enterotoxigenic diarrhoea; it is described in the 'Cholera' section.

Vibrio parahaemolyticus can cause acute watery diarrhoea after consumption of contaminated raw fish or shellfish. Explosive diarrhoea, abdominal cramps, and vomiting occur, with fever in 50%. The infection is self-limiting but can last for up to 10 days.

Yersiniosis

Yersinia enterocolitica infection is a zoonosis of a variety of domestic and wild mammals. Human disease can arise via either contaminated food products, such as pork, or direct animal contact. *Y. enterocolitica* can cause a range of gastroenteric symptoms, including watery diarrhoea, dysentery, and mesenteric adenitis. The illness is usually self-limiting, but ciprofloxacin may shorten the duration. *Y. pseudotuberculosis* is a much less common human pathogen; it causes mesenteric adenitis and terminal ileitis.

Enterotoxin-producing *Staphylococcus aureus*

Some strains of *S. aureus* can produce a heat-stable toxin (enterotoxin B), which causes massive secretion of fluid into the intestinal lumen. It is a common cause of food-borne gastroenteritis in Europe and the USA, where outbreaks usually occurring as a result of poor food hygiene. Because the toxin is pre-formed in the contaminated food, onset of symptoms is rapid, often within 2 to 4 hours of consumption. There is violent vomiting followed within hours by profuse watery diarrhoea. Symptoms usually settle within 24 hours.

Bacillus cereus

Bacillus cereus produces two toxins. The first causes vomiting 2 to 4 hours after ingestion of pre-formed heat-stable exotoxins. The second causes watery diarrhoea up to 12 hours after ingestion of uncooked food that contains spores or viable bacteria that multiply and produce a toxin within the bowel.

Clostridioides difficile

C. difficile can cause watery diarrhoea, colitis, and pseudomembranous colitis. It is a Gram-positive, anaerobic, spore-forming bacillus that is found in the bowel flora of 3% to 5% of the asymptomatic population (increasing to up to 20% in hospitalized people).

Pathogenesis

C. difficile produces two toxins; whereas toxin A is an enterotoxin, toxin B is cytotoxic and causes bloody diarrhoea. It usually causes illness either in patients who have been given antibiotic therapy that has eliminated other bowel commensals or in those who are debilitated for other reasons. Almost all antibiotics have been implicated, but an increase in cases has been attributed in part to greater use of quinolones. Hospital-acquired infection has been a major cause

of morbidity in recent decades, partly because of increased spread from person to person and via fomites and partly because of high rates of antibiotic use in hospitalized people. Stewardship interventions have reduced disease prevalence over the past decade, but an increasing proportion of cases appears to have no association with healthcare settings.

Clinical features

C. difficile–associated diarrhoea can begin any time from 2 days to some months after taking antibiotics. Older hospitalized patients are most frequently affected. It is unclear why some carriers remain asymptomatic. Symptoms can range from mild diarrhoea to profuse, watery, haemorrhagic colitis, along with lower abdominal pain. The colonic mucosa is inflamed and ulcerated and can be covered by an adherent membrane-like material (pseudomembranous colitis). The disease is usually more severe in older adults and can cause intractable diarrhoea; occasionally, it can cause toxic megacolon. Markers of severity include:

- Temperature greater than 38.5°C
- WCC greater than 15×10^9/L
- Serum creatinine greater than 50% above baseline
- Raised serum lactate
- Severe abdominal pain.

Diagnosis

Diagnosis is made by detecting A or B toxins in the stools using ELISA or PCR for *C. difficile* nucleic acid.

Management

Treatment is with oral vancomycin 125 mg four times daily, which has been shown to be more effective than metronidazole, the previously recommended first-line antibiotic. Fidaxomicin, higher vancomycin doses, and intravenous metronidazole are indicated in more severe or recurrent cases. Other antibiotics should be discontinued if possible. In refractory or relapsing cases, instilling faeces from the bowel of a healthy donor (faecal transplant) can restore normal bowel flora and eradicate the *C. difficile* infection. This therapy is increasingly becoming the standard for refractory *C. difficile* infections; bezlotoxumab, a monoclonal antibody targeting toxin B, has demonstrated efficacy in reducing recurrence.

Prevention

Infection control relies on:

- Responsible use of antibiotics (see Chapter 12, 'Antimicrobial stewardship' section)
- Hygiene, which should involve all health workers, as well as patients and relatives. Washing hands thoroughly using soap and water is essential because alcohol disinfectants do not kill spores.
- Isolation of patients with symptomatic *C. difficile* in hospitals or residential care.

Clostridium perfringens

C. perfringens infection is caused by inadequately cooked food, usually meat or poultry allowed to cool for a long time, during which period the spores germinate. The ingested organism produces an enterotoxin that causes watery diarrhoea with severe abdominal pain, usually without vomiting.

Travellers' diarrhoea

Travellers' diarrhoea is defined as the passage of three or more unformed stools per day in a traveller to an LMIC. Infection is usually food- or water-borne and commonly occurs in the first week of travel. Reported attack rates vary from country to country but approach 50% for a 2-week stay in some nations. The disease is usually benign and self-limiting; treatment with ciprofloxacin or azithromycin may hasten recovery but is not normally necessary. The common causative organisms are listed in Box 42.45.

? Box 42.45 Commonly identified causes of travellers' diarrhoea[a]

Organism	Frequency (%) (varies from country to country)
Enterotoxigenic *Escherichia coli*	30–70
Shigella spp.	0–15
Campylobacter spp.	0–15
Salmonella spp.	0–10
Viral pathogens	0–10
Giardia intestinalis	0–3

[a]In most cases, no microbiological diagnosis is made.

Management of acute gastroenteritis

In children in LICs, untreated diarrhoea has a high mortality rate caused by dehydration, often on a background of malnutrition and other chronic infection. Death and serious morbidity are less common in adults but still occur, particularly in low- and middle-income countries and in older adults.

The mainstay of treatment for all types of gastroenteritis is oral rehydration solutions (Fig. 42.21; see Box 42.36). The use and formulation of oral rehydration solutions are discussed earlier in this chapter. Antibiotics may sometimes have a role, especially in children in low- and middle-income countries. It is also important to remember that diarrhoea can be a 'false localizing symptom' in sepsis, and a resulting delay in diagnosis of, for example, malaria, can have major consequences. Most cases of acute gastroenteritis resolve within 10 days; if symptoms persist, protozoal causes such as *Giardia* spp. in the context of a relevant exposure history or inflammatory bowel disease and other non-infectious causes of colitis become more likely.

Food poisoning

Food poisoning is a legally notifiable condition in England and Wales. There is some overlap between food poisoning (defined as 'any disease of an infective or toxic nature caused by or thought to be caused by the consumption of food and water') and gastroenteritis. However, not all cases of gastroenteritis are caused by food poisoning because the pathogens are not always food or water borne. Conversely, some types of food poisoning (e.g., botulism) do not primarily cause gastroenteritis. Common bacterial causes of food poisoning are listed in Box 42.46. Food poisoning may also be caused by a number of non-infectious organic and inorganic toxins (Box 42.47).

Food poisoning is associated with modern intensive farming practices in the production and distribution of food. Livestock is frequently contaminated with *Salmonella* or *Campylobacter* spp., and flaws in the processing, storage, and distribution of food products allow massive amplification of the infection, resulting in extensive contamination. The internationalization of the food supply encourages widespread and distant transmission of the resulting infections.

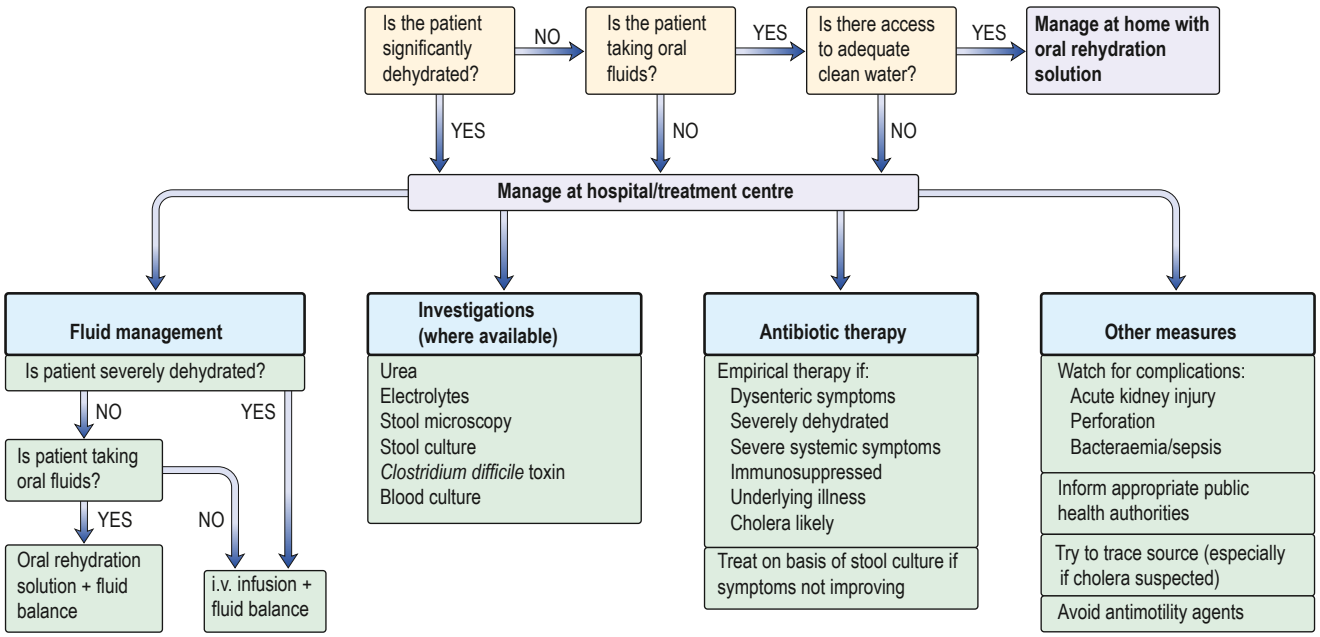


Fig. 42.21 Gastroenteritis: management plan. *i.v.,* Intravenous.

Box 42.46 Bacterial causes of food poisoning

Organism	Source/vehicles	Incubation period	Symptoms	Diagnosis	Recovery
Staphylococcus aureus	Humans: contaminated food and water	2–4 h	Diarrhoea, vomiting, dehydration	Culture of organism in vomitus or remaining food	<24 h
***Escherichia coli* (ETEC)**	Salads, water, ice	24 h	Watery diarrhoea	Stool PCR	1–4 days
***E. coli* O157:H7**	Cattle: contaminated foodstuffs	12–48 h	Watery diarrhoea ± haemorrhagic colitis, HUS	Stool culture or PCR	10–12 days
Yersinia enterocolitica	Milk, pork	2–14 h	Abdominal pain, vomiting, diarrhoea	Stool culture with specific media	2–30 days
Bacillus cereus	Environment: rice, ice cream, chicken	1–6 h 6–14 h	Vomiting Diarrhoea	Culture of organism in faeces and food	Rapid
Clostridium perfringens	Environment: contaminated food	8–22 h	Watery diarrhoea, cramping pain	Culture of organism in faeces and food	2–3 days
Listeria monocytogenes	Environment: milk, raw vegetables, dairy products, unpasteurized cheese	Variable	Colic, diarrhoea, vomiting	Stool culture	Variable
Vibrio parahaemolyticus	Seafood	2–48 h	Diarrhoea, vomiting	Stool, food	2–10 days
Clostridium botulinum	Environment: bottled or canned food	18–24 h	Brief diarrhoea and paralysis caused by neuromuscular blockade	Demonstration of toxin in food or faeces	10–14 days
***Salmonella* spp.**	Cattle and poultry: eggs, meat	12–48 h	Abrupt diarrhoea, fever, vomiting	Stool culture or PCR	Usually 3–6 days but may be up to 2 weeks
Campylobacter jejuni	Cattle and poultry: meat, milk	48–96 h	Diarrhoea ± blood, fever, malaise, abdominal pain	Stool culture or PCR	3–5 days
***Shigella* spp.**	Humans: contaminated food and water	24–48 h	Acute watery, bloody diarrhoea	Stool culture or PCR	7–10 days

ETEC, Enterotoxigenic *Escherichia coli*; *HUS,* haemolytic uraemic syndrome; *PCR,* polymerase chain reaction.

Box 42.47 Organic toxins causing food poisoning[a]

Toxin	Source	Illness
Scombrotoxin	Tuna, mackerel	Histamine fish poisoning
Ciguatoxin	Barracuda, snapper	Ciguatera (diarrhoea and paraesthesia)
Dinoflagellate plankton toxin	Shellfish	Neurotoxic shellfish poisoning
Haemagglutinins	Inadequately prepared dried kidney beans	Diarrhoea and vomiting
Unknown	Buffalo fish	Haff disease (toxic rhabdomyolysis)
Phallotoxins, amatoxins	Mushrooms	Various
Alpha-galactosidase	Lone star tick	Mammalian meat allergy

[a]See online topics.

Enteric fever, *Helicobacter pylori*, Whipple's disease, and bacterial peritonitis

For ***enteric fever***, see 'Enteric fever' section. For ***Helicobacter pylori***, see Chapter 22 ('*Helicobacter pylori* infection' section). For ***Whipple's disease***, see Chapter 28 ('Whipple's disease' section). For ***bacterial peritonitis***, see Chapter 23 ('Spontaneous bacterial peritonitis' section).

Further reading

Kim J-S, Lee M-S, Kim J. Recent updates on outbreaks of Shiga-toxin producing Escherichia coli and its potential reservoirs. *Front Cell Infect Microbiol* 2020; 10:273.

Muzembo A, Kitahara K, Ohgno A et al. Colonization with extended-spectrum beta-lactamase-producing Escherichia coli and traveler's diarrhea attack rates among travelers to India: a systematic review and meta-analysis. *Trop Dis Travel Med Vaccines* 2022; 8(1):22.

Ressler A, Wang J, Rao K. Defining the black box: a narrative review of factors associated with adverse outcomes from severe *Clostridioides difficile* infection. *Ther Adv Gastroenterol* 2021; 14:17562848211048127.

Siddiq M, O'Flanagan H, Richarson D, Llewellyn C. Factors associated with sexually transmitted shigella in men who have sex with men: a systematic review. *Sex Transm Infect* 2023; 99(1):58–63.

BACTERIAL INFECTIONS OF THE CARDIOVASCULAR SYSTEM

Infective endocarditis is discussed in Chapter 17 ('Infective endocarditis' section).

Rheumatic fever

Rheumatic fever is a multisystem disorder but is included here because the cardiac effects are usually the most significant. (About 35 million people worldwide have rheumatic heart disease.) It is an immune-mediated inflammatory disease that occurs in children and young adults as a result of infection with GAS. It affects the heart, skin, joints and CNS. It is relatively common in LMICs but very rare in Western Europe and North America. The decline in the incidence of rheumatic fever in these settings (from 10% of children in the 1920s to 0.01% today) parallels the reduction in all streptococcal infections and is largely attributable to improved hygiene and the use of antibiotics.

Pharyngeal infection with GAS is followed, in a small proportion of cases, by the clinical syndrome of rheumatic fever. This is thought to develop because of an autoimmune reaction triggered by molecular mimicry between the cell-wall M proteins of the infecting *S. pyogenes* and cardiac myosin and laminin. The condition is not caused by direct infection of the heart or the production of a toxin.

Clinical features

The disease presents 2 to 3 weeks after the initial GAS infection with fever, joint pains, and malaise. Diagnosis relies on the presence of two or more major clinical manifestations or one major manifestation plus two or more minor features in addition to evidence of current or recent streptococcal infection. These are known as the modified Jones criteria (**Box 42.48**).

Carditis manifests as new or changed heart murmurs, development of cardiac enlargement or cardiac failure, appearance of a pericardial effusion, or electrocardiographic changes of pericarditis, myocarditis, or cardiac arrhythmias.

Non-cardiac features include fever, arthritis, Sydenham's chorea (see Chapter 28, 'Chorea' section), and skin lesions. The arthritis is typically a fleeting migratory polyarthritis affecting large joints such as the knees, elbows, ankles, and wrists. When the acute inflammation disappears, the rheumatic process leaves the joints normal. Skin manifestations include erythema marginatum, a transient pink rash with slightly raised edges, which occurs in 20% of cases, and small painless subcutaneous nodules.

Management

The streptococcal infection should be eradicated with penicillin. If the diagnosis is suspected clinically, antibiotics should be administered even if nasal or pharyngeal swabs do not culture the streptococci.

The arthritis of rheumatic fever usually responds to non-steroidal anti-inflammatory drugs, although these have no impact on long-term cardiac sequelae. There is no good evidence that steroids are of benefit, although some experts give high-dose prednisolone if there is severe carditis. In areas of high prevalence, recurrence is common, and persistent cardiac damage is more likely. Further infection can largely be prevented by prophylaxis with daily oral or monthly intramuscular benzathine penicillin until the age of 20 years or for 10 years after the latest attack. Erythromycin or clarithromycin is used if the patient is allergic to penicillin. Any streptococcal infection that does develop should be treated promptly.

Prognosis

More than 50% of those who suffer acute rheumatic fever with carditis later (after 10–20 years) develop chronic rheumatic valvular disease, predominantly affecting the mitral and aortic valves.

BACTERIAL INFECTIONS OF THE NERVOUS SYSTEM

The central and peripheral nervous systems can be affected by a variety of microorganisms, including bacteria, viruses, and protozoa, which cause disease by direct invasion or via toxins. The nervous system is also vulnerable to prion disease (see 'Transmissible spongiform encephalopathies [prion diseases]' section).

Bacterial meningitis

The most common bacterial disease affecting the CNS is acute meningitis (see Chapter 28, 'Meningitis' section), which causes around 200 000 deaths per year, predominantly in LICs. *S. pneumoniae* is now the commonest cause of community-acquired bacterial meningitis followed by *N. meningitidis* (or meningococcus), with

Box 42.48 Modified Jones criteria for the diagnosis of rheumatic fever

Evidence of antecedent streptococcal infection
- Positive throat culture for group A streptococcus
- Good clinical history (e.g., of scarlet fever)
- Elevated antistreptolysin O titre (or other serological assay for streptococci)

Major criteria
- Carditis
- Polyarthritis
- Chorea
- Erythema marginatum
- Subcutaneous nodules

Minor criteria
- Fever
- Arthralgia (unless arthritis counted as major criterion)
- Previous rheumatic fever
- Raised erythrocyte sedimentation rate or C-reactive protein
- Leucocytosis
- Prolonged PR interval on electrocardiogram (unless carditis counted as major criterion)

tuberculous meningitis (see Chapter 20, 'Tuberculosis' section) also common in sub-Saharan Africa and parts of Asia.

Hib was once a common cause of meningitis in children, but since an effective vaccine has been available, serious *H. influenzae* infections have become rare in countries that have also instituted immunization programmes; however, invasive *H. influenzae* infection remains common in some parts of the world.

Other less common causes of meningitis in adults include group B streptococci, *L. monocytogenes*, *S. aureus*, and Gram-negative bacilli. These organisms are usually associated with an underlying illness or immunocompromising condition or with a CSF leak.

N. meningitidis is also an important cause of bloodstream infection and sepsis. There is a spectrum of presentation ranging from pure meningitis with few systemic features to a typical sepsis syndrome without clinical evidence of meningitis.

Meningococcal sepsis

N. meningitidis is found worldwide. Twelve serogroups have been identified, of which six (A, B, C, W, X, and Y) cause significant outbreaks of infection. The main burden falls on the so-called 'meningitis belt', which includes 25 countries across Africa stretching from Senegal in the west to Ethiopia in the east. Traditionally, massive seasonal outbreaks of serotype A occurred in the dry season every 5 to 12 years, with attack rates up to 1000 per 100 000 children per year, associated with disruption to host pharyngeal defences. The Menafrivac conjugate vaccine roll-out, focused on serotype A, has been highly successful but at the expense of replacement with other serotypes, including C and W, which are now also causing epidemics. In Europe and North America, cases associated with all serotypes have declined enormously over the past decades, with introduction of serotype-specific vaccines into the routine schedule, but sporadic cases do occur.

Humans are the only known reservoir for the organism, which is carried asymptomatically in the nasopharynxes of 5% to 20% of the general population. Meningococcal disease occurs when the bacteria invade the nasal mucosa and enter the bloodstream; this happens only in a small percentage of those colonized. Invasion depends on both host and bacterial factors. It is more likely to take place soon after colonization has occurred and after viral upper respiratory infections.

Clinical features

Invasive meningococcal infection may cause meningitis, the meningococcal sepsis syndrome, or both. Meningitic disease (see Chapter 28, 'Meningitis' section) usually presents with the classical triad of headache, fever, and neck stiffness. Vomiting, diminished consciousness, and focal neurological signs sometimes occur, although many patients, especially in the early stages, have only mild symptoms. Meningococcal bloodstream infection causes the typical features of septic shock, such as fever, myalgia, and hypotension (see Chapter 12), and may be accompanied by a petechial or haemorrhagic rash (see Fig. 42.4). False localizing symptoms such as diarrhoea can cause a fatal delay in diagnosis. Patients can deteriorate rapidly, with shock, DIC, and multiorgan failure.

Diagnosis

The presence of meningitis and sepsis syndrome with a typical rash is strongly suggestive of meningococcal disease. The rash is not always present, however, and its absence does not rule out the diagnosis. Gram-negative diplococci may be seen on Gram-stain of CSF or of aspirate from petechiae; meningococci can also be cultured from CSF or blood or detected by PCR. PCR can also be used to identify the serogroup, which may be important for public health management.

Management

N. meningitidis is sensitive to benzylpenicillin (in most cases), third-generation cephalosporins, and chloramphenicol; cefotaxime or ceftriaxone is the recommended treatment in the UK (see Box 26.57). Meningococcal sepsis should be managed in the same way as any other form of bacterial sepsis (see Chapter 12). Whereas the mortality rate from meningococcal sepsis in high-income countries is currently approximately 10%, that from meningococcal meningitis alone is less than 5% (see later). Mild neurological sequelae (especially vestibular nerve damage) are common, but serious brain damage is relatively unusual.

The meningococcal C conjugate vaccine was introduced to routine childhood immunization schedules in the UK in 1999, group B protein-based vaccine in in 2017, and the MenACYW conjugate vaccine for adolescents in 2015.

In an outbreak, close contacts of a case of meningococcal disease should be given prophylaxis with oral rifampicin or ciprofloxacin to eradicate the bacteria from the nasopharynx and reduce the risk of onward spread. (This has little impact in epidemic areas.) In some outbreak situations, contacts are offered immunization.

Cerebral abscess

This is covered in Chapter 28 ('Brain abscess' section).

Toxin-mediated infections

Botulism

C. botulinum is a common environmental organism that produces spores that can survive heating to 100°C. It can cause illness through a number of routes:
- Contamination of canned or bottled foodstuffs, in which the anaerobic organism can multiply and elaborate a neurotoxin
- Infection of wounds
- Contamination of injectable drugs (especially heroin).

In botulism (food poisoning), the toxin is ingested and causes profound neuromuscular blockade, leading to autonomic and motor paralysis. The first symptoms, occurring 18 to 24 hours after ingestion, are nausea and diarrhoea. These are followed by cranial nerve palsies and then progressive symmetrical paralysis, leading to respiratory failure.

The diagnosis is usually clinical and is confirmed by detection of toxin in faeces or in contaminated food. Management is mainly supportive, with mechanical ventilation if necessary. Antitoxin is available in some countries (including the UK); the risk of anaphylaxis is relatively high, and antitoxin should be used only in severe cases. A subcutaneous test dose should be given before intravenous or intramuscular injection. Antibiotics have no proven role. The overall mortality rate from botulism is high, but patients who survive the acute paralysis can make a full recovery.

A similar clinical picture can follow the contamination of wounds or injection of street heroin contaminated with *C. botulinum*; in infants, botulism may be related to bowel colonization by the organism.

Tetanus

Tetanus is also caused by a toxin-secreting clostridium: *C. tetani*. The organism is found in soil, and illness usually results from a contaminated wound. The injury itself may be trivial and disregarded by the individual. Tetanus can also complicate injecting drug use if the drug (usually heroin) is contaminated with *C. tetani*. In low- and middle-income countries, neonatal tetanus follows contamination of the umbilical stump, often after the area is dressed with dung. The overall incidence of tetanus across the world has fallen significantly in the past 30 years mainly because of active immunization.

The organism is not invasive and clinical manifestations of the disease are caused by the potent neurotoxin tetanospasmin. Tetanospasmin causes disinhibition at synapses, neuromuscular blockade, and skeletal muscle spasm and acts on the sympathetic nervous system. The end results are marked flexor muscle spasm and autonomic dysfunction.

Clinical features

The incubation period varies from a few days to several weeks. The most common form of the disease is generalized tetanus. General malaise is rapidly followed by trismus (lockjaw) caused by masseter muscle spasm. Spasm of the facial muscles produces the characteristic grinning expression known as *risus sardonicus*. If the disease is severe, painful reflex spasms develop, usually within 24 to 72 hours of the initial symptoms. The interval between the first symptom and the first spasm is referred to as the 'onset time'. The spasms may occur spontaneously but are easily precipitated by noise, handling of the patient, or light. Respiration may be impaired because of laryngeal spasm; oesophageal and urethral spasm lead to dysphagia and urinary retention, respectively, and there is arching of the neck and back muscles (opisthotonus). Autonomic dysfunction produces tachycardia, a labile blood pressure, sweating, and cardiac arrhythmias. Patients with tetanus are mentally alert.

Death results from aspiration, hypoxia, respiratory failure, cardiac arrest, or exhaustion. Mild cases with rigidity usually recover. Poor prognostic indicators include a short incubation period, short onset time, and extremes of age.

Localized tetanus is a milder form of the disease. Pain and stiffness are confined to the site of the wound, with increased tone in the surrounding muscles. Recovery is usual.

Cephalic tetanus is uncommon but invariably fatal. It usually occurs when the portal of entry of *C. tetani* is the middle ear. Cranial nerve abnormalities, particularly of the VIIth nerve, are usual. Generalized tetanus may or may not develop.

Neonatal tetanus is characterized by failure to thrive, poor sucking, grimacing, and irritability followed by the rapid development of intense rigidity and spasms. The mortality rate approaches 100%. Aims of the WHO's EPI are to eliminate this condition by immunizing all females of child-bearing age, providing clean delivery facilities, and strengthening surveillance in high-risk areas.

Diagnosis

Few diseases resemble tetanus in its fully developed form, and the diagnosis is therefore usually clinical. Rarely, *C. tetani* is isolated from wounds. Phenothiazine overdosage, strychnine poisoning, meningitis, and tetany can occasionally mimic tetanus.

Management

If tetanus is suspected, any wound must be cleaned and debrided to remove the source of toxin. Human tetanus immunoglobulin 250 units should be given along with an intramuscular injection of tetanus toxoid. If the patient is already protected, a single booster dose of the toxoid is given; otherwise, the full three-dose course of adsorbed vaccine is required (see later).

In established disease, management consists of supportive medical and nursing care. Improvement in this area has contributed, more than any other single measure, to the decrease in the mortality rate from 60% to nearer 20%. Patients are nursed in a quiet, isolated, well-ventilated, darkened room. Benzodiazepines are used to control spasms and sedate the patient; if the airway is compromised, intubation and mechanical ventilation may be necessary. Magnesium sulphate infusion may decrease the need for antispasmodics.

Antibiotics and antitoxin should be administered, even in the absence of an obvious wound. Intravenous metronidazole for 7 to 10 days is the drug of choice, although penicillin is also effective. Human tetanus immunoglobulin (HTIG) should be given as soon as possible, by either intravenous infusion or intramuscular injection, to neutralize any circulating toxin. If HTIG is not available, human normal immunoglobulin (HNIG) can be used. If the patient recovers, active immunization should be instituted because immunity after tetanus is incomplete.

Prevention

An effective, safe and relatively cheap vaccine means that tetanus is an eminently preventable disease. After initial active immunization at 2, 3, and 4 months of age, followed by pre-school and teenage doses, further boosters are no longer recommended at 10-year intervals for all adults; however, if there is any doubt regarding receipt of all five doses, then further boosters are recommended, typically in the context of potential contaminated injuries. Infant immunization schedules in all countries include tetanus. Protection by passive immunization with HTIG or HNIG is short-lived, lasting only about 2 weeks.

Further reading

Auala T, Zavale B, Mbakwem A, Mocumbi A. Acute rheumatic fever and rheumatic heart disease: highlighting the role of group a streptococcus in the global burden of cardiovascular disease. *Pathogens* 2022; 11(5):496.
Mercer A. Protection against severe infectious disease in the past. *Pathog Glob Health* 2021; 115(3):151–167.
Wall EC, Chan J, Gil E, Heyderman R. Acute bacterial meningitis. *Curr Opin Neurol* 2021; 34(3):386–395.

BACTERIAL BONE AND JOINT INFECTIONS

See Chapters 32 ('Infection of joints' section) and 33 ('Bone infections' section).

BACTERIAL INFECTIONS OF THE URINARY TRACT

See Chapter 26 ('Urinary tract infection' section).

SYSTEMIC AND MULTISYSTEM BACTERIAL INFECTIONS

Many infections are confined to a particular body organ or system, owing to the metabolic requirements of the organism, the route of infection, or the response of host defences. Other infections can potentially affect several systems or the entire body. Under unusual circumstances, such as altered host immunity, infections that are normally circumscribed may become systemic. The general principles of systemic infections, bacteraemia, and sepsis are described in Chapter 12, as are their common causes. This section describes some specific and less common bacterial infections that cause multisystem disease in immunocompetent hosts.

Leptospirosis

Leptospirosis is a zoonosis caused by the spirochaete *Leptospira interrogans*. There are more than 200 serotypes. The main types affecting humans are:

- *Leptospira i. icterohaemorrhagiae* (normal host rodents)
- *Leptospira i. canicola* (dogs and pigs)
- *Leptospira i. hardjo* (cattle)
- *Leptospira i. pomona* (pigs and cattle).

Leptospires are excreted in animal urine and enter the host through a skin abrasion or intact mucous membranes. They can also be transmitted by ingestion of contaminated water. The organism can survive for many days in warm freshwater and for up to 24 hours in seawater.

In England and Wales, between 100 and 150 cases of leptospirosis are reported every year, although many mild infections probably go undiagnosed. (About 90%–95% of infections are subclinical or cause only a mild fever.) It is traditionally an occupational disease of farmers, veterinarians, and others who work with animals but is increasingly associated with recreational activities that bring people into closer contact with rodents or contaminated water. Outbreaks of leptospirosis have also been associated with flooding, particularly in urban areas of LMICs.

Clinical features

In 1886, Adolf Weil described a severe illness consisting of jaundice, haemorrhage, and renal impairment caused by *L. i. icterohaemorrhagiae*, but the disease has long been recognized around the world in communities at risk such as 'autumn fever' in Chinese rice paddy workers and 'dairy farm fever' in New Zealand cattle farmers.

The incubation period of leptospirosis is usually 7 to 14 days, and the illness typically has two phases: a leptospiraemic phase, which lasts for up to 1 week, followed after a couple of days' interval by an immunological phase. The first phase is characterized by severe headache, malaise, fever, anorexia, and myalgia. Most patients have conjunctival suffusion. Hepatosplenomegaly, lymphadenopathy, and various skin rashes are sometimes seen. The second phase is usually mild. Meningism is common; if lumbar puncture is performed, there is often CSF lymphocytosis. The majority of patients recover uneventfully at this stage.

In severe disease, there may not be a clear distinction between phases. After the initial symptoms, patients progressively develop liver and kidney injury, haemolytic anaemia, and circulatory collapse. Cardiac failure and pulmonary haemorrhage may also occur. Even with full supportive care, the mortality rate is around 10%, rising to 15% to 20% in older adults.

Diagnosis

The diagnosis is usually a clinical one, in combination with epidemiological risk and laboratory data. Leptospires can be cultured from blood or CSF during the first week of illness, but culture requires special media and may take several weeks. Leptospiral DNA can be detected by PCR in the blood and urine. A minority of patients may also excrete the organism in their urine from the second to the fourth week. Confirmation is usually serological. Specific IgM antibodies start to appear from the end of the first week and the diagnosis is often made retrospectively with a microscopic agglutination test showing a fourfold rise. There is typically a leucocytosis and transaminitis and, in severe infection, thrombocytopenia and an elevated creatine kinase.

Management

Early antibiotic therapy may limit progress of the disease, and penicillin or ceftriaxone should be given in severe cases. The efficacy of antibiotics in mild disease is debated; if they are used, oral doxycycline is first-line therapy. Patients who develop hepatorenal syndrome may need intensive care.

Brucellosis

Brucellosis (Malta fever, undulant fever) is caused by various species of *Brucella*, a Gram-negative intracellular coccobacillus. It is a zoonosis, especially of goats, sheep, and cattle, and has a worldwide distribution. Some countries, such as the UK, have been declared 'brucellosis free', although a handful of imported human cases are diagnosed each year. There are around half a million new cases a year with the highest incidence occurring in the Mediterranean, Middle East, and tropical belt.

The bacteria usually gain entry to the human body via the mouth; less frequently, they may enter via the respiratory tract, genital tract, or broken skin. The bacilli travel in the lymphatics and infect lymph nodes. This is followed by haematogenous spread with ultimate localization in the reticulo-endothelial system. Acquisition is usually by the ingestion of raw milk from infected cattle, goats, or camels, although occupational exposure is also common. Person-to-person transmission is rare.

Clinical features

The incubation period of acute brucellosis is 1 to 3 weeks. The onset is insidious, with malaise, headache, weakness, generalized myalgia, and night sweats. The fever pattern is classically waxing and waning (undulant), although continuous and intermittent patterns are also seen. Lymphadenopathy and hepatosplenomegaly are common; sacroiliitis, arthritis, osteomyelitis, epididymo-orchitis, meningoencephalitis, and endocarditis have all been described.

Untreated brucellosis can give rise to chronic infection, lasting 1 year or more. This is characterized by easy fatigability, myalgia, and occasional bouts of fever and depression. Splenomegaly is usually

present. Occasionally, brucellosis can cause a localized infection in almost any organ that is not associated with systemic features.

Diagnosis

Blood and bone marrow cultures are positive in more than half of patients with acute *Brucella* (higher in *Brucella melitensis*) but prolonged culture is needed and the success rate falls with time. In chronic disease, serological tests are of greater value. The *Brucella* agglutination test, which demonstrates a fourfold or greater rise in titre (>1 in 160) over a 4-week period, is highly suggestive of brucellosis. An elevated serum IgG level is evidence of current or recent infection; a negative test result excludes chronic (non-localized) brucellosis. In localized brucellosis, antibody titres are low, and the diagnosis is usually established by culturing the organisms from the involved site. Species-specific PCR tests are also available.

Management and prevention

Brucellosis should be treated with a combination of doxycycline, rifampicin, and an aminoglycoside (usually gentamicin) for 6 weeks. Prevention and control involve careful attention to hygiene when handling infected animals, eradication of infection in animals through vaccination, and the pasteurization of milk. No vaccine is available for use in humans.

Listeriosis

L. monocytogenes is an environmental organism that is widely disseminated in soil and decayed matter. It affects both animals and humans; the most common route of human infection is in contaminated foodstuffs. The organism can grow at temperatures as low as 4°C, and the most commonly implicated foods are unpasteurized soft cheeses, raw vegetables, and chicken pâtés. Listeriosis is a rare but serious infection affecting mainly neonates, pregnant females, older adults, and immunocompromised people. *L. monocytogenes* has also been recognized as a cause of self-limiting, food-borne gastroenteritis in healthy adults, but the incidence of this is unknown.

In pregnancy, *Listeria* causes an influenza-like illness, but infection of the fetus can lead to septic abortion, premature labour, and stillbirth in up to half of cases when treatment is delayed. In older adults and immunocompromised patients, *Listeria* can cause bacteraemia, meningoencephalitis, and various focal infections. The diagnosis is established by culture of blood, CSF, or other body fluids. The treatment of choice for adult listeriosis is ampicillin plus gentamicin. Co-trimoxazole and meropenem are also effective, but the organism is resistant to cephalosporins.

Q fever

In 1937 Edward Derrick documented a new disease in Australian slaughterhouse workers and called it ‘Q fever’ for ‘query cause’. We now know that Q fever is a zoonosis caused by *C. burnetii*, a Gram-negative intracellular bacterium. The infection is widespread in domestic, farm, and other animals; birds; and arthropods, and it is transmitted mainly by ticks. Animal-to-human spread is through contaminated dust, aerosol, and unpasteurized milk from infected cows. The formation of spores means that *C. burnetii* can survive in extreme conditions for long periods; moreover, the infective dose is very small, so minimal animal contact is required. An outbreak affecting more than 4000 people in the Netherlands linked to goat farming had the main risk factor as being resident within 5 km of an affected farm. Diagnosis is difficult, so the disease is underdiagnosed, particularly in resource-limited settings. One prospective study in northern Tanzania found that almost 10% of cases of severe non-malarial fever were caused by *C. burnetii*.

Clinical features

Symptoms begin insidiously 2 to 4 weeks after infection. Fever is accompanied by influenza-like symptoms with myalgia and headache. The acute illness usually resolves spontaneously, but pneumonia or hepatitis may develop. In pregnancy, Q fever can cause severe fetal damage. Occasionally, infection can become chronic, resulting in endocarditis, myocarditis, uveitis, osteomyelitis, and other focal infections.

C. burnetii is an obligate intracellular organism and does not grow on standard culture media. Diagnosis is made serologically using an immunofluorescent assay. Antibody tests for two different bacterial antigens allow distinction between acute and chronic infection. A PCR assay is available, but the sensitivity is low.

Management

Treatment with doxycycline or azithromycin shortens the duration of the acute illness, and there is emerging evidence that it may reduce the incidence of chronic sequelae. For chronic Q fever, including endocarditis, doxycycline is often combined with rifampicin or hydroxychloroquine, but even prolonged courses of treatment may not clear the infection. A vaccine is available for those at high risk.

Lyme disease

Lyme disease is caused by spirochaetes of the genus *Borrelia*. *Borrelia burgdorferi* is the sole cause of Lyme disease in the USA. In Europe, *Borrelia afzelii* and *Borrelia garinii* are also implicated. Lyme disease is a zoonosis of deer and other wild mammals. It has increased in both incidence and detection, and it is now known to be widespread in the USA, Europe, Russia, and the Far East. The vector is not present in most of the Southern Hemisphere, and case reports from South America, sub-Saharan Africa, and South Asia are rare.

More than 1000 laboratory-confirmed infections are reported in England and Wales each year, with two to three times more diagnosed clinically. By comparison, around 30 000 cases per year are reported to the US Centers for Disease Control and Prevention. Infection is transmitted from animal to human by *Ixodid* ticks and is most likely to occur in wooded areas, including urban gardens and parks, in spring and early summer when the ticks are feeding. In places where other tick-borne infections such as ehrlichia and babesia are endemic, simultaneous co-infection may occur.

Clinical features

Early localized disease presents about a week after the tick bite with erythema migrans: a macular ‘bull’s eye’ rash usually occurring around the site of the tick bite (**Fig. 42.22**). This progresses to regional lymphadenopathy and associated fever and headache. Most people recover spontaneously with no further symptoms or signs. A small number may develop early disseminated disease, which occurs days to weeks after the appearance of erythema migrans. The rash may spread, and a small proportion of untreated patients may develop neurological complications such as meningitis, encephalitis, cranial or peripheral neuritis, or radiculopathies. Cardiac involvement is seen in the USA but is rare elsewhere. Myalgia and migratory arthritis may also occur at this stage. Late Lyme disease, which is rare, can cause chronic arthritis, encephalomyelitis, and other neurological disorders, along with acrodermatitis chronica atrophicans. Much of the damage in late Lyme disease is likely to be immune-mediated, and although antibiotics should be given to eradicate any persistent infection, there is no evidence to support a role for extended treatment.

Fig. 42.22 **The skin lesion erythema migrans at the site of a tick bite.**

Diagnosis

The clinical features and epidemiological considerations are usually strongly suggestive. The presence of erythema migrans after a risk exposure requires antibiotic treatment without further diagnostic tests (which may well be negative at this stage anyway). In patients with a high clinical suspicion of Lyme disease but no sign of erythema migrans, it is reasonable to start treatment while waiting for a confirmatory test using a combined IgG and IgM ELISA. If the ELISA result is equivocal or positive, then follow-up with a Lyme immunoblot test because of the high rate of false-positive results. When the ELISA is negative and Lyme disease is still suspected, the ELISA is repeated 4 to 6 weeks after the onset of symptoms, or if the symptoms have been present for more than 3 months, an immunoblot test should be performed. The diagnosis can be confirmed only rarely by isolation of the organisms from blood, skin lesions, or CSF, and although it is occasionally possible to find patients who have PCR evidence of Lyme disease and negative serology, this is the exception; an alternative diagnosis should be sought in patients with persistently negative Lyme serology. As in all cases, but particularly regarding assays for Lyme disease, samples should only be sent to accredited laboratories that take part in external quality assurance programmes.

Management

Antibiotics given early in the course of disease shorten the duration of illness and significantly reduce the risk of complications. Doxycycline given for 3 weeks is first-line treatment for erythema migrans and most cases of early disseminated disease. Amoxicillin is an alternative. Ceftriaxone is the treatment of choice for CNS disease. It is reasonable to offer a second course of a different class of antibiotics for patients with persisting symptoms. However, after this, when the diagnosis is secure, care is principally supportive. Patients may have chronic pain, fatigue, depression, and loss of function over a prolonged period, requiring specialist support and information.

Repellents and protective clothing should be worn in areas with high tick burden. Checking for ticks at the end of the day and prompt removal are essential because infection is unlikely to take place unless the tick has been attached for more than 48 hours. Ticks should be grasped with forceps near to the point of attachment to the skin and then withdrawn by gentle traction. Antibiotic prophylaxis after a tick bite is not usually justified even in areas where Lyme disease is common. There is no effective vaccine.

Tularaemia

Tularaemia is caused by infection with *Francisella tularensis*, a Gram-negative coccobacillus. It is primarily a zoonosis, usually acquired from rodents. Infection can be transmitted by arthropod vectors or by handling infected animals, when the microorganisms enter through minor abrasions or intact mucous membranes. Occasionally, infection occurs from contaminated water or from uncooked meat. The organism is widely distributed in North America, Northern Europe, and Asia, but the particularly virulent type A subspecies is seen only in the USA. It is relatively rare, occurring mainly in hunters, trappers, and others in close contact with animals; about 800 cases per year are recorded in Europe, mainly in Scandinavia.

The incubation period of 2 to 7 days is followed by a systemic illness. The most common presentation is ulceroglandular tularaemia. A papule occurs at the site of inoculation; this ulcerates and is followed by tender, suppurative lymphadenopathy. Rarely, this can be followed by septicaemia, pneumonia, or meningitis. These forms of the disease carry a high mortality rate if left untreated.

Routine culture results are often negative, and diagnosis has traditionally been made serologically. PCR is now becoming the standard diagnostic test.

Doxycycline can be used in mild disease, but the most effective antibiotics are aminoglycosides and fluoroquinolones.

Further reading

Eldin C, Mélenotte C, Mediannikov O et al. From Q Fever to Coxiella burnetii Infection: a paradigm change. *Clin Microbiol Rev* 2017; 30(1):115–190.
Herrick JA, Lederman RJ, Sullivan B, Powers JH, Palmore TN. Brucella arteritis: clinical manifestations, treatment, and prognosis. *Lancet Infect Dis* 2014; 14(6):520–526.
Maurin M, Gyuranecz M. Tularaemia: clinical aspects in Europe. *Lancet Infect Dis* 2016; 16:113–124.
Pam H, Tran M-H. One Health: An effective and ethical approach to leptospirosis control in Australia. *Trop Med Infect Dis* 2022; 7(11):389.
Verschoor Y, Vrijlandt A, Spijker R. Persistent Borrelia burgdorferi *Sensu Lato* Infection after antibiotic treatment: systematic overview and appraisal of the current evidence from experimental animal models. *Clin Microbiol Rev* 2022; 35(4):e0007422.
https://www.nice.org.uk/guidance/ng95.

BACTERIAL INFECTIONS SEEN IN LOW- AND MIDDLE-INCOME COUNTRIES

Skin, soft tissue, and eye diseases

Leprosy (Hansen's disease)

Leprosy is caused by the acid-fast bacillus *Mycobacterium leprae*. Unlike other mycobacteria, this does not grow in artificial media or even in tissue culture, although it can be grown in the footpads of mice for research purposes. Apart from nine-banded armadillos, humans are the only natural host of *M. leprae*.

Leprosy has a global distribution, and new cases are currently reported in 127 countries. The WHO has provided free drug treatment since 1995, and the campaign to control the disease has been hugely successful. In 2016 the WHO published a new Global Leprosy Strategy, with the aim of reducing the number of children diagnosed and presenting with physical deformities to zero. However, by the end of 2019, 202 256 patients worldwide were still undergoing treatment for leprosy with a prevalence of 22.9 cases per million. Several countries, principally India, Brazil, and Indonesia, continue to report high numbers of new cases with more than half of them being in children, indicating ongoing early exposure to infection.

The precise mode of transmission of leprosy is still uncertain, but it is likely that nasal secretions play a role. Infection is related

to poverty and overcrowding. After an individual has been infected, subsequent progression to clinical disease appears to be dependent on several factors. Males are more susceptible than females, and there is evidence from twin studies of a genetic susceptibility. The main factor, however, is the response of the host's cell-mediated immune system.

Two ***polar types*** of leprosy are recognized, with most patients falling on a spectrum between the two (Fig. 42.23):

- ***Tuberculoid leprosy*** is a localized disease that occurs in individuals with a high degree of CMI. The T-cell response to the antigen releases IFN-γ, which activates macrophages to destroy the bacilli (Th1 response) but with an associated destruction of tissue.
- ***Lepromatous leprosy*** is a generalized disease that occurs in individuals with impaired CMI. Here the tissue macrophages fail to be activated, and the bacilli multiply intracellularly. Th2 cytokines are produced, and patients develop a fever and systemic illness.

In 1966 Ridley and Jopling divided leprosy into five clinicopathological groups (Fig. 42.24). This remains the 'gold standard' classification, but for the purposes of treatment, the WHO has defined two main patterns of disease:

- ***Paucibacillary leprosy***, with five or fewer skin lesions and no bacilli on microscopy
- ***Multibacillary leprosy***, with six or more lesions that may have bacilli on microscopy.

Slit-skin smears are not always available in practice, and the distinction must be made on clinical grounds alone. Concern has been raised that this may lead to undertreatment of some people incorrectly identified as having paucibacillary leprosy, but overall, the WHO regimen has been remarkably successful.

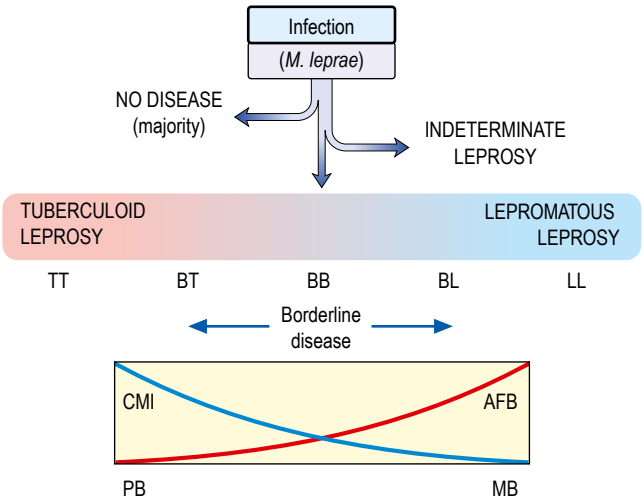


Fig. 42.23 Clinical spectrum of leprosy with the combined Ridley–Jopling and World Health Organization classification. *AFB,* Acid-fast bacilli; *BB,* borderline; *BL,* borderline lepromatous; *BT,* borderline tuberculoid; *CMI,* cell-mediated immunity; *LL,* lepromatous leprosy; *MB,* multibacillary; *PB,* paucibacillary; *TT,* tuberculoid leprosy.

Clinical features

The incubation period is 5 years on average, although it may be as short as 1 year or as long as 20 years and differs between tuberculoid (shorter) and lepromatous (longer). The onset is generally insidious, although acute onset is known to occur. Patients may present with a transient rash, features of an acute febrile illness, evidence of nerve involvement, or any combination of these.

Diagnosis

The diagnosis of leprosy is essentially clinical, with anaesthetic hypopigmented or reddish skin patches and thickened peripheral nerves. Acid-fast bacilli (AFB) may be seen on slit-skin smears. These are 3- to 5-mm long and 2- to 3-mm deep incisions taken

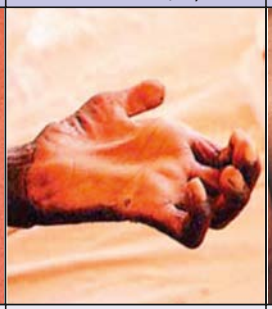
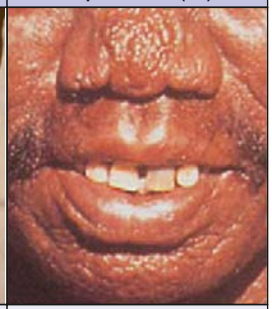
A Tuberculoid (TT)	B Borderline tuberculoid (BT)	C Borderline (BB)	D Borderline lepromatous (BL)	E Lepromatous (LL)
				
Skin lesion (usually single) - hypopigmented - anaesthetic - clearly defined edges - thickened local nerve(s) May heal spontaneously	Several skin lesions - similar to TT Thickened peripheral nerves - deformities of hands/feet	Numerous skin lesions - varying size Widespread nerve involvement and deformity	Numerous florid skin lesions - variable size and form - contain many AFB Variable nerve involvement Neurotrophic atrophy of digits	Extensive skin infiltration Mucous membrane involvement Nasal septal perforation - 'Leonine facies' Glove and stocking anaesthesia Multiple nerve palsies Neurotrophic atrophy of digits

Fig. 42.24 The spectrum of disease in leprosy. *AFB,* Acid-fast bacilli. *(C, From Cohen J, Powderly JG.* Infectious Diseases, *3rd edn. London: Mosby; 2010, with permission. E, From Goering R, et al.* Mims' Medical Microbiology, *5th edn. London: Mosby; 2013, courtesy of DA Lewis, with permission.)*

from the ear lobes, elbows, and knees of the patient with a scalpel and examined under the microscopy for AFB using a modified Ziel-Neelsen stain. Results are expressed on a semi-logarithmic scale as the bacillary index. Occasionally, nerve biopsies are helpful. Detection of *M. leprae* DNA is possible in all forms of leprosy and can be used to assess the efficacy of treatment. The Lepromin test and leprosy serology are poorly sensitive and specific and are not routinely used for diagnosis in clinical practice.

Management

In some areas up to 40% of bacilli are resistant to dapsone and so multidrug therapy (MDT) is essential. Much shorter courses of treatment are now being used; the current WHO-recommended drug regimens for leprosy are shown in **Box 42.49** but longer therapy is required in severe cases. Follow-up, including skin smears where possible, is obligatory. Immunological reactions ('lepra reactions') can occasionally occur after treatment is started, especially in borderline and lepromatous disease (**Box 42.50**).

Patients with leprosy should be treated in specialist centres with adequate physiotherapy and occupational therapy support. Surgery and physiotherapy also play a role in the management of trophic ulcers and deformities of the hands, feet, and face. Patients should be taught self-care of their anaesthetic hands and feet to prevent ulcers. Cheap canvas shoes with cushioned insoles are protective. Psychosocial and financial support are as important as drug therapy in addressing the societal impact of the disease on patients and their families.

The prognosis in tuberculoid leprosy is good, even if untreated; lepromatous leprosy is progressive if untreated.

Prevention

The prevention and control of leprosy depend on the rapid treatment of infected patients, particularly those with lepromatous and borderline lepromatous types, to decrease the bacterial reservoir. Bacille Calmette–Guérin vaccination has a modest effect in protection against leprosy, but the effect tends to wane, and its use remains controversial.

Anthrax

Anthrax is caused by *Bacillus anthracis*. The spores of these Gram-positive bacilli are extremely hardy and withstand extremes of temperature and humidity. The organism is capable of toxin production, and this property correlates most closely with its virulence. The disease occurs worldwide but is most common in Africa and Southern Asia. Surveillance data are limited, but the likely global annual incidence is in the tens of thousands. In Europe and the USA, there are only very occasional cases. Transmission is through direct contact with an infected animal or its carcass; infection is most frequently seen in farmers, butchers, and dealers in wool and animal hides. Spores can also be ingested or inhaled. Cases in the USA have been caused by the deliberate release of anthrax spores as a bioterrorist weapon.

Box 42.49 Recommended treatment regimens for leprosy in adults (modified World Health Organization guidelines)

Multibacillary leprosy (LL, BL, BB)	Paucibacillary leprosy (BT, TT)
• Rifampicin 600 mg once monthly, supervised	• Rifampicin 600 mg once monthly, supervised
• Clofazimine 300 mg once monthly, supervised	• Dapsone 100 mg/day, self-administered • Clofazimine 300 mg once monthly plus 50 mg daily
	• Treatment continued for 6 months
• Clofazimine 50 mg/day, self-administered	**Single-lesion paucibacillary leprosy**
• Dapsone 100 mg/day, self-administered • Treatment continued for12 months	• Rifampicin 600 mg • Ofloxacin 400 mg • Minocycline 100 mg (as a single dose)

Clinical features

The incubation period is 1 to 10 days.

Cutaneous anthrax is the most common form. The small, erythematous, maculopapular lesion is initially painless. It may subsequently vesiculate and ulcerate, with formation of a central black eschar. The illness is self-limiting in most patients, but occasionally, perivesicular oedema and regional lymphadenopathy may be marked, and toxaemia can occur.

Inhalational anthrax (wool-sorter's disease) follows inhalation of spores. A febrile illness is accompanied by a non-productive cough and retrosternal discomfort; pleural effusions are common. If the disease is untreated, the mortality rate is about 90%; in bioterrorism cases in the USA, it was 45% despite treatment.

GI anthrax is caused by consumption of undercooked, contaminated meat. It presents as severe gastroenteritis, sometimes with haematemesis and bloody diarrhoea. In some cases, systemic toxin spread may lead to shock and death.

Diagnosis

The diagnosis is established by demonstrating the organism in smears from cutaneous lesions or by culture of blood and other body fluids. Serological confirmation can be obtained using an ELISA that detect antibodies to both the organism and the toxin. PCR tests are available.

Management

Ciprofloxacin is considered the best treatment. In mild cutaneous infections, oral therapy for 2 weeks is adequate. In more severe infections, high doses of intravenous antibiotics are needed, along with appropriate supportive care. The monoclonal antibody raxibacumab has been shown in animal studies to improve survival in those with inhalational anthrax. All suspected cases should be reported to the relevant authority.

Control

All infected animals that die should be burned and the area where it was housed disinfected. Where animal husbandry is poor, mass vaccination of animals may prevent widespread contamination but needs to be repeated annually. A human vaccine is available for those at high risk, and prophylactic ciprofloxacin is used after exposure. Some countries have established public health policies to deal with the deliberate release of anthrax spores.

Mycobacterial ulcer (buruli ulcer)

Buruli ulcer, caused by *Mycobacterium ulcerans*, is occasionally reported in humid parts of the tropics near slow-moving or stagnant water, but it is particularly prevalent in rural west Africa. It is found in water, soil, and fish and invertebrates that live in or near water. The mode of transmission is unknown, but it does not seem to be spread by human-to-human contact. About 3000 cases per year are reported to the WHO, but the true prevalence is much higher.

Box 42.50 Leprosy reactions after commencement of anti-mycobacterial treatment

Type 1
- Occurs in borderline disease
- Type IV hypersensitivity reaction
- Lasts for many months
- Upgrading (→TT) or downgrading (→BB)
- Abrupt onset of nerve palsies in upgrading
- Treatment: prednisolone

Type 2 (erythema nodosum leprosum)
- Occurs in lepromatous disease
- Type III (immune complex) hypersensitivity reaction
- Lasts days to weeks
  - Fever
  - Arthralgia
  - Subcutaneous nodules
  - Iridocyclitis
- Rare since clofazimine included in treatment
- Treatment[a]
  - Analgesia
  - Clofazimine (if not already started)
  - Prednisolone if eye involvement
  - Continuation of anti-mycobacterial treatment

[a]Thalidomide no longer has any role in the management of lepra reactions.
BB, Borderline; *TT*, tuberculoid.

A small subcutaneous nodule at the site of infection gradually ulcerates, involving subcutaneous tissue, muscle and fascial planes, and occasionally bone. The ulcers are usually large with undermined edges and markedly necrotic bases due to mycolactone (a toxin produced by the mycobacterium). Smears taken from necrotic tissue generally reveal numerous AFB; a PCR assay is widely available. Combination therapy with rifampicin plus one of streptomycin, clarithromycin, or moxifloxacin, given for 8 weeks, heals ulcers in many cases. However, in more severe cases, surgical debridement and, where possible, skin grafting may be needed as well.

Tropical ulcer

Tropical ulcer is a non-specific infection, usually polymicrobial, seen in damp, hot climates. A wide variety of organisms have been implicated, but there is no characteristic bacterium associated with the condition. Good wound hygiene is the mainstay of management, with antibiotic therapy used (either empirically or based on culture) as needed.

Endemic treponematoses

Yaws

Yaws is caused by *Treponema pallidum* subspecies *pertenue*. It is the most widespread and common of the endemic treponemal diseases and primarily occurs in children younger than 15 years. Most cases are now reported from West Africa and the South Pacific.

After an incubation period of weeks or months, a primary inflammatory reaction occurs at the inoculation site, from which organisms can be isolated. Dissemination of the organism leads to multiple papular lesions containing treponemes; these skin lesions usually involve the palms and soles. Left untreated, there may be bone involvement with a painful periostitis, commonly affecting the long bones and those of the hand.

Approximately 10% of those infected go on to develop late yaws. Bony gummatous lesions may progress to cause gross destruction and disfigurement, particularly of the skull and facial bones, interphalangeal joints, and long bones. Plantar hyperkeratosis is characteristic. As in syphilis, there may be a latent period between the early and late phases of the disease, but visceral, neurological, and cardiovascular problems do not occur.

The diagnosis is clinical. The causative organism can be identified from the exudative lesions under dark-ground microscopy. Serological tests for syphilis are positive and do not differentiate between the conditions. Rapid point-of-care, species-specific tests are used but do not distinguish between past and current infection. PCR diagnosis is not widely available but may become important in the yaws eradication campaign because it is necessary to distinguish active from treated infection.

Single-dose oral azithromycin has now replaced intramuscular benzathine penicillin as the treatment of choice. This drug has now been placed on the WHO list of essential medicines as part of the international strategy to eradicate yaws through mass treatment.

Bejel (endemic syphilis)

Bejel is seen in the arid parts of Africa, the Middle East, and Central Asia. The causative organism *(Treponema endemicum)* enters through abrasions in the skin directly or by mouth-to-mouth contact. Bejel differs from venereal syphilis in that the primary lesion is small and often in the mouth. The late stages resemble syphilis, but cardiological and neurological manifestations are rare.

Pinta

Pinta (caused by *Treponema carateum*) is restricted mainly to Central and South America. There is no effective surveillance, and the current prevalence is unknown. It is milder than the other treponematoses and confined to the skin. The primary lesion is a pruritic red papule, usually on the hand or foot. It may become scaly but never ulcerates. Lesions can occur for up to 1 year after onset and are often associated with generalized lymphadenopathy. Eventually, the lesions heal, leaving hyperpigmented or depigmented patches.

The diagnosis of bejel and pinta is largely clinical, but spirochaetes are visible on dark-ground microscopy of early lesions. Serological tests for syphilis cross-react and are usually positive. Treatment is with benzathine penicillin, although it is likely that azithromycin will prove to be effective. Surveillance, diagnosis, and treatment are less well coordinated than for yaws.

Trachoma

Trachoma is caused by the intracellular bacterium *Chlamydia trachomatis*, which can also cause a number of STIs (see Chapter 43). It is the most common infectious cause of blindness worldwide. The WHO launched a global elimination programme in 1996, and as a result, the number of people at risk has fallen from 1.5 billion to 115 million with 17 countries having eliminated the disease.

Trachoma is a disease of poverty. It was once found throughout the world but is now mainly limited to the tropics and the Middle East and is entirely preventable. It commonly occurs in children and is spread by direct contact, on fomites such as dirty washcloths, or through flies landing on the eyes. Isolated infection is self-limiting, but repeated infection leads to chronic inflammation and scarring.

Clinical features

Infection is bilateral and begins in the conjunctiva, with marked follicular inflammation and subsequent scarring. Damage to the upper

eyelid causes entropion, leaving the cornea exposed to further abrasion from the eyelashes turning in and rubbing against it (trichiasis) and eventually to corneal opacification and blindness.

Trachoma may also occur as an acute ophthalmic infection in neonates.

Diagnosis and management

The diagnosis is generally made clinically.

Systemic therapy with a single dose of azithromycin is the treatment of choice, but re-infection is common in endemic areas. Annual mass drug administration is recommended for communities with a prevalence of active disease greater than 10%. After the infection has been controlled, surgery may be required for patients with trichiasis to direct the eyelashes away from the cornea.

Prevention

Community health education, improvements in water supply and sanitation (pit latrines), and earlier case reporting have a substantial impact on disease prevalence. This is reflected in the 'SAFE' approach to trachoma:

- Surgery
- Antibiotics
- Facial cleanliness
- Environmental improvement

Further reading

Burton MJ, Rajak SN, Hu VH et al. Pathogenesis of progressive scarring trachoma in Ethiopia and Tanzania and its implications for disease control: two cohort studies. *PLoS Negl Trop Dis* 2015; 9(5):e0003763.
Hendricks K, Person MK, Bradley JS et al. Clinical features of patients hospitalized for all routes of anthrax, 1880-2018: a systematic review. *Clin Infect Dis* 2022; 75(suppl 3):S341–S353.
Holmes A, Tildesley MJ, Solomon AW et al. Modeling treatment strategies to inform yaws eradication. *Emerg Infect Dis* 2020; 26(11):2685–2693.
Lockwood DN, McIntosh A, Armstrong M et al. Diagnosing and treating leprosy in a non-endemic setting in a national centre, London, United Kingdom 1995-2018. *PLoS Negl Trop Dis* 2022; 16(10):e0010799.
Mitjà O, Marks M, Bertran L et al. Integrated control and management of neglected tropical skin diseases. *PLoS Negl Trop Dis* 2017; 11(1):e0005136.

Gastrointestinal infections

Cholera

Cholera is caused by the curved, flagellated, Gram-negative bacillus *V. cholerae*. The organism is killed by temperatures of 100°C in a few seconds but can survive in ice for up to 6 weeks.

The fertile, humid plains of the River Ganges in West Bengal have traditionally been regarded as the home of cholera. However, a series of pandemics have spread the disease across the world, usually following trade routes. Until recently, all outbreaks of cholera were caused by *V. cholerae* O1, which is responsible for the recent pandemic (which spread internationally between the 1960s and 1990s). More recently, a new serogroup (O139 Bengal) caused outbreaks of an illness indistinguishable from O1 cholera, but it has not been reported outside of Bangladesh, India, and South-east Asia. Outbreaks of cholera are common when there is disruption of water supplies and sanitation, for example, with natural disasters or conflict. It was John Snow's observations linking cases of cholera to the use of a water pump in Broad Street, London, in 1854, which is seen as the origin of the science of epidemiology.

Transmission is by the faecal–oral route, and although contaminated water plays the major role in its spread, contaminated foodstuffs and contact carriers may contribute in cholera epidemics. Achlorhydria or hypochlorhydria facilitates passage of the cholera bacilli into the small intestine. There they attach to the wall and proliferate and produce an exotoxin that causes a massive secretion of isotonic fluid into the lumen through increasing the secretion of chloride and blocking the absorption of sodium. Cholera toxin also releases serotonin from enterochromaffin cells in the gut, which activates a neural secretory reflex in the enteric nervous system. This may account for at least 50% of cholera toxin's secretory activity. *V. cholerae* also produces other toxins, contributing to its pathogenic effect.

Clinical features

The incubation period varies from a few hours to 6 days. The vast majority of patients with cholera have a mild illness that is clinically indistinguishable from diarrhoea from other infective causes. In severe cases, there is abrupt onset of profuse painless diarrhoea followed by vomiting. As the illness progresses, the typical 'rice water' stool, flecked with mucus, may be seen. There is massive fluid loss, and if this is not replaced, the features of hypovolaemic shock (cold, clammy skin; tachycardia, hypotension; and peripheral cyanosis) and dehydration (sunken eyes, hollow cheeks, and a diminished urine output) can develop within hours. Muscle cramps may be severe. Children may present with convulsions owing to electrolyte disturbance and hypoglycaemia.

Severe cholera has an untreated mortality rate of up to 50% but as long as the patient can be adequately rehydrated, the prognosis is good, with a gradual return to normal clinical and biochemical parameters in 1 to 3 days.

Diagnosis

This is largely clinical, particularly in epidemic settings. Microscopy of freshly passed stools may demonstrate rapidly motile organisms, although this is not diagnostic because *C. jejuni* may give a similar appearance. A rapid dipstick test is also available in the field. Stool and rectal swabs should be taken for culture to confirm the diagnosis and establish antibiotic sensitivity. Cholera should always be reported to the appropriate public health authority.

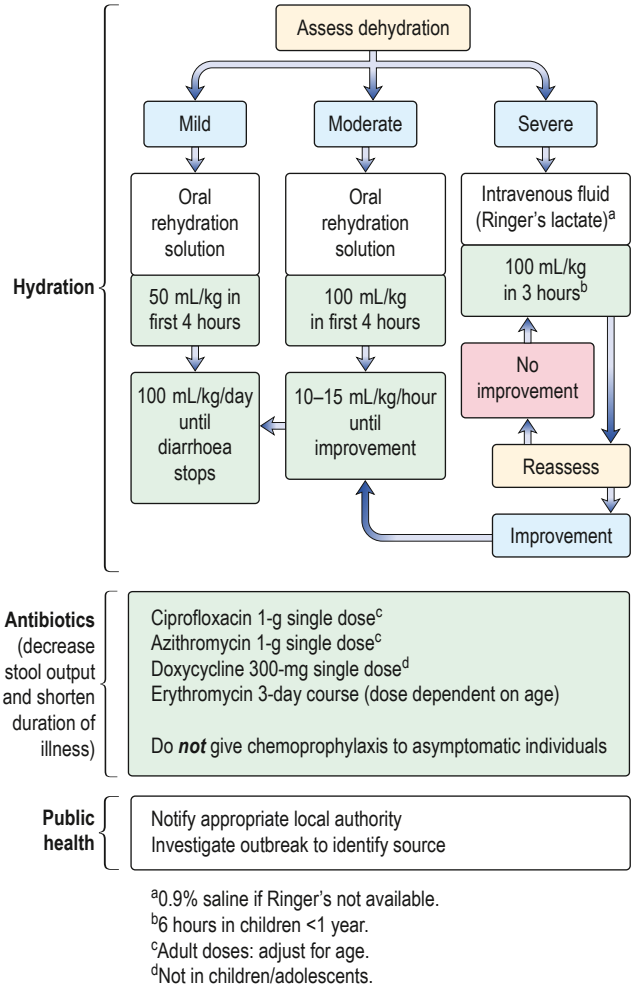
Management

The mainstay of treatment is rehydration. If done effectively the mortality decreases to less than 1%. Oral rehydration is usually adequate, but intravenous therapy is sometimes required (**Fig. 42.25**).

Oral rehydration solutions are based on the observation that glucose and other carbohydrates enhance sodium and water absorption in the small intestine even in the presence of secretory loss caused by toxins. Additions such as amylase-resistant starch to glucose-based oral rehydration solutions have been shown to increase the absorption of fluid. Cereal-based electrolyte solutions have been found to be as effective as sugar and salt oral rehydration solutions and reduce stool volume, as well as rehydrating. The WHO and UNICEF recommend the use of reduced-osmolarity oral rehydration solution with zinc supplementation for all types of diarrhoea; the small risk of hyponatraemia with this formulation is far outweighed by the improved efficacy. Suitable solutions for rehydration are listed in **Box 42.36**.

Infection gives protection for at least 3 years. Immunization is recommended in potential outbreak situations. The WHO recommends any of three oral, killed, whole-cell vaccines with efficacies of around 50% to 60% in outbreak settings. There is also an oral live-attenuated vaccine for travellers to endemic areas.

Antibiotics may shorten the duration and severity of diarrhoea but should be given only to patients with severe dehydration. Tetracyclines are the first-line treatment, although resistance is developing. Other options include co-trimoxazole, macrolides, and fluoroquinolones.



[a]0.9% saline if Ringer's not available.
[b]6 hours in children <1 year.
[c]Adult doses: adjust for age.
[d]Not in children/adolescents.

Fig. 42.25 The management of cholera.

The best preventive measures, however, are good hygiene and improved sanitation. Studies in Bangladesh have shown that using a folded sari cloth to filter river drinking water decreases the incidence of cholera by up to 50%. The cloth should be rinsed in clean water between use and dried in the sun to kill any bacteria.

In 2017 the WHO announced a strategy to end cholera by 2030. Although eradication is not possible because the organism persists in the environment, it is theoretically possible to eliminate human-to-human transmission. The strategy has three elements:

- Early detection and quick response to contain outbreaks
- Prevention of cholera recurrence in cholera 'hotspots'
- Coordination of technical support, advocacy, resource mobilization, and partnership at local and global levels.

Enteric fever

Enteric fever is an acute systemic illness characterized by fever, headache and abdominal discomfort. ***Typhoid***, the typical form of enteric fever, is caused by *Salmonella typhi*. A similar illness known as ***paratyphoid*** is caused by infection with *S. paratyphi* A, B, or C. Humans are the only natural host for *S. typhi*, which is transmitted in contaminated food or water. The incubation period is 10 to 14 days.

In 2017 typhoid fever was estimated to cause 10.9 million cases and 116 800 deaths worldwide with an overall case-fatality rate of 1%. South Asia accounted for three-quarters of all cases.

Clinical features

After ingestion, the bacteria invade the small bowel wall via Peyer's patches, from where they spread to the regional lymph nodes and then to the blood and reticulo-endothelial system. The onset of illness is insidious and non-specific, with intermittent fever, headache, and abdominal pain. Physical findings include abdominal tenderness and hepatosplenomegaly. Historically, a scanty, blanching maculopapular rash called 'rose spots' was observed, but this is less common in the antibiotic era. Without treatment and occasionally even after treatment, serious complications can arise, usually in the third week of illness. These include shock, jaundice, encephalopathy, intestinal perforation, and intestinal haemorrhage. The fourth week is characterized by gradual improvement, but in areas with limited healthcare, up to one-third of those infected die, and 10% of untreated survivors relapse.

After clinical recovery, 2% to 5% of patients continue to excrete *S. typhi* for several months and are termed ***convalescent carriers***. A smaller proportion continue to shed the organism in stool for more than 1 year and become ***chronic carriers***. The usual site of carriage is the gall bladder; it is associated with gallstones, which is perhaps one reason why chronic carriers are more likely to be older and female. In parts of the Middle East and Africa where urinary schistosomiasis is prevalent, chronic carriage of *S. typhi* in the urinary bladder is also common.

Paratyphoid fever is associated with a milder and shorter illness, and complications are uncommon.

Diagnosis

Diagnosing enteric fever in resource-limited settings can be challenging because the clinical syndrome is similar to other non-specific febrile illnesses. Patients are often treated empirically, given the risk of severe disease. The Widal test is commonly used but has moderate sensitivity and specificity because of patients having cross-reactive antibodies; its use is not recommended. The definitive diagnosis requires significant laboratory infrastructure to perform cultures for *S. typhi* or *S. paratyphi*. Blood culture is positive in 50% to 70% of cases if taken before antibiotics. Culture of intestinal secretions, faeces, and urine is also used, although care must be taken to distinguish acute infection from chronic carriage. Bone marrow culture is the most sensitive test but is rarely required except in patients who have already received antibiotics or when the diagnosis is in doubt. Leucopenia and transaminitis are common but non-specific features.

Management

Most patients can be managed in an outpatient setting, but it is often necessary to start treatment before making a microbiological diagnosis. The choice of empirical treatment changes frequently over time and geography. Increasing antibiotic resistance in *S. typhi* to the first-line drugs of chloramphenicol, co-trimoxazole, and amoxicillin was initially seen in South-east Asia and the Indian subcontinent in the 1970s. This became known as multidrug-resistant (MDR) typhoid because resistance occurs to all three antibiotics and is mediated by a single transmissible plasmid, which can be exchanged between bacteria. By the 1990s fluoroquinolones had become the first-line treatment in most of Asia. Unfortunately, resistance to fluoroquinolones followed shortly after because of the accumulation of mutations in the *S. typhi* topoisomerase gene. Empirical treatment shifted to azithromycin for mild cases and ceftriaxone for severe cases.

Efforts to control the widespread use of antibiotics have resulted in a gradual reduction in rates of MDR typhoid more recently.

However, a new strain of typhoid, resistant to all three first-line drugs, plus fluoroquinolones and third-generation cephalosporins, was reported in Pakistan in 2016. By 2019 this extensively drug-resistant strain had become the commonest cause of typhoid fever in Pakistan. Sporadic reports exist from other countries, but for now it is generally possible to identify a link with travel to Pakistan. At the time of writing in 2024, azithromycin remains effective for most strains of *S. typhi* around the world.

The patient's temperature may remain elevated for several days after antibiotics are started, and this alone is not a sign of treatment failure. Prolonged antibiotic therapy may eliminate the carrier state, but in the presence of gall bladder disease, it is rarely effective. Cholecystectomy is not usually justified on clinical or public health grounds.

Prevention

This is mainly through improved sanitation and clean water, which is easier said than done. It often requires government investment in infrastructure to provide uncontaminated drinking water, community campaigns to change people's behaviour, and even laws to ban industrial practices such as irrigating crops with untreated sewage.

Vaccination with either injectable inactivated or oral live attenuated vaccines gives partial protection and is recommended for travellers to high-risk areas. A new conjugate vaccine that protects up to 80% of children and has longer-lasting immunity has recently been developed. A single dose is now recommended by the WHO as part of the standard childhood vaccination programme in high-prevalence countries.

Further reading

Malembaka EB, Bugeme PM, Hutchins C et al. Effectiveness of one dose of killed oral cholera vaccine in an endemic community in the Democratic Republic of the Congo: a matched case-control study. *Lancet Infect Dis* 2024; 24(5):514-522. Erratum in: *Lancet Infect Dis* 2024; 24(3):e156.
Meiring JE, Khanam F, Basnyat B et al. Typhoid fever. *Nat Rev Dis Primers* 2023; 9:71.

Systemic infections

Tuberculosis

See Chapter 20 ('Tuberculosis' section).

Non-tuberculous mycobacterial infections

Environmental mycobacteria can cause a wide variety of infections in susceptible hosts (**Box 42.51**). These are discussed in more detail in Chapter 44 ('*Mycobacterium avium-intracellulare* infection' section).

? Box 42.51 Examples of non-tuberculous mycobacteria causing disease in humans

Clinical	Common cause	Rare cause
Chronic lung disease	*Mycobacterium avium-intracellulare* *Mycobacterium kansasii*	*Mycobacterium malmoense* *Mycobacterium xenopi*
Local lymphadenitis	*M. avium-intracellulare* *Mycobacterium scrofulaceum*	*Mycobacterium malmoense* *Mycobacterium fortuitum*
Skin and soft tissue infection		
Fish tank granuloma	*Mycobacterium marinum*	
Abscesses, ulcers, sinuses	*Mycobacterium fortuitum* *Mycobacterium chelonae*	*Mycobacterium haemophilum*
Bone and joint infection	*Mycobacterium kansasii* *M. avium-intracellulare*	*M. scrofulaceum*
Disseminated infection (in HIV)	*M. avium-intracellulare*	

HIV, Human immunodeficiency virus.

Plague

Plague is caused by *Yersinia pestis*, a Gram-negative bacillus. Sporadic cases of plague, as well as occasional epidemics, occur worldwide, but it is relatively uncommon: about 1000 cases per year are reported to the WHO, with a 10% mortality rate. The majority are in Madagascar and Central Africa, although the disease is occasionally seen in high-income countries, usually associated with hunting. The main reservoirs are woodland rodents, which transmit infection to domestic rats *(Rattus rattus)*. The usual vector is the rat flea, *Xenopsylla cheopis*. These fleas bite humans when there is a sudden decline in the rat population. Occasionally, spread of the organisms may be through infected faeces being rubbed into skin wounds or through inhalation of droplets. A modern re-analysis of the 1665 outbreak of plague in the isolated English village of Eyam found that three-quarters of cases were transmitted by human-to-human contact within families.

Clinical features

Four clinical forms are recognized: bubonic, pneumonic, septicaemic, and cutaneous.

- ***Bubonic plague*** is the most common form, accounting for about 90% of cases. The incubation period is 2 to 7 days. The onset of illness is acute, with high fever; chills; headache; myalgia; nausea; vomiting; and, when severe, prostration. This is rapidly followed by the development of lymphadenopathy (buboes), most commonly involving the inguinal region. Characteristically, these are matted and tender and suppurate in 1 to 2 weeks.
- ***Pneumonic plague*** is characterized by the abrupt onset of features of a fulminant pneumonia with bloody sputum, marked respiratory distress, cyanosis, and death in almost all affected patients.
- ***Septicaemic plague*** presents as an acute fulminant infection with evidence of shock and DIC. If it is left untreated, death usually occurs in 2 to 5 days.
- ***Cutaneous plague*** can present as a pustule, eschar, or papule or as extensive purpura, which can become necrotic and gangrenous.

Diagnosis

This is based on clinical, epidemiological, and laboratory findings. Microscopy (on blood or lymph node aspirate) or a rapid dipstick antigen detection test can provide a presumptive diagnosis in an appropriate clinical setting. Confirmation is by blood or lymph node culture or paired serological tests.

Management

Treatment is urgent and should be instituted before the results of culture studies are available. *Y. pestis* remains sensitive to most antibiotics for Gram-negative bacteria. Gentamicin is the most widely used agent; doxycycline, ciprofloxacin, and chloramphenicol are also effective.

Prevention

Prevention of plague is largely dependent on control of the flea population. Outhouses or huts should be sprayed with insecticides

that are effective against the local flea. During epidemics, rodents should not be killed until the fleas are under control because the fleas will leave dead rodents to bite humans. Tetracycline is an effective chemoprophylactic agent for those at risk during an outbreak. Both killed and attenuated vaccines have been available for many years but have limited efficacy and are not widely used.

Relapsing fevers

These conditions are so named because after apparent recovery from the initial infection, recurrences may occur after 1 week or more without fever. They are caused by spirochaetes of the genus *Borrelia*.

Louse-borne relapsing fever

This condition (caused by *Borrelia recurrentis*) is spread by body lice, and only humans are affected. Classically, it is an epidemic disease of armies and refugees, although it is also endemic in the highlands of Ethiopia and Yemen and in the Andes. Lice are spread from person to person when humans live in close contact in impoverished conditions. The spirochaete penetrates through the skin or mucosa when infected lice are crushed by scratching or by biting, a traditional means of killing the insect, hence the adage 'man bites louse is more dangerous than louse bites man'. Symptoms begin 3–10 days after infection and consist of an abrupt high fever, myalgia, and headache. A petechial or ecchymotic rash may be seen. In severe cases, the condition then deteriorates, with delirium, hepatosplenomegaly, jaundice, bleeding, and circulatory collapse. A proportion of those who recover have no further symptoms, but the majority experience one or more relapses of diminishing intensity over the weeks after the initial illness. The severity varies enormously, with many people experiencing only mild symptoms and some developing antibodies with no history of clinical illness. In some epidemics, however, the mortality rate in untreated people has exceeded 50%.

Tick-borne relapsing fever

This is caused by *Borrelia duttoni* and other *Borrelia* species, spread by soft (argasid) ticks of the genus *Ornithodoros*. Rodents are the main reservoir of infection, and humans are incidental hosts, acquiring the spirochaete from the saliva of the infected tick. The disease is found worldwide and is endemic in parts of North America and Europe, but the clinical burden falls mainly in Africa. It is particularly common where traditional mud huts are the principal form of human shelter because the ticks live in the walls and feed on the inhabitants at night. The illness is generally similar to the louse-borne disease, with the same variability in presentation and severity.

Diagnosis and management

Spirochaetes can be demonstrated microscopically in the blood during febrile episodes; organisms are more numerous in louse-borne relapsing fever. PCR is more sensitive and is replacing traditional microscopy where available. Treatment is usually with tetracycline. A severe drug reaction (Jarish-Herxheimer reaction) occurs in many patients, especially those who have the louse-borne form.

Prevention

Control of infection relies on eliminating the vector. Ticks live for years and remain infected, passing the infection to their progeny. These reservoirs of infection may be controlled by spraying houses with insecticides and reducing the number of rodents. Post-exposure prophylaxis has been shown to be effective in areas where tick-borne disease is highly endemic. Infested patients should be deloused by washing with a suitable insecticide. All clothes must be thoroughly disinfected.

Rickettsiae and rickettsia-like infections

Rickettsiae, which include the genera *Ehrlichia*, *Orientia*, and *Coxiella*, are small intracellular bacteria that are spread to humans by arthropod vectors, including body lice, fleas, hard ticks, and larval mites. Rickettsiae inhabit the alimentary tracts of these arthropods, and infection is spread to the human host by inoculation of their faeces through broken human skin, generally produced by scratching. Rickettsiae multiply intracellularly and can enter most mammalian cells, although the main lesion produced is a vasculitis caused by invasion of endothelial cells of small blood vessels. Multisystem involvement is usual.

Typhus is the collective name given to a group of diseases caused by *Rickettsia* species. **Box 42.52** provides a list of some of the more important rickettsial diseases; with the advent of molecular techniques, more than 30 pathogenic species of *Rickettsia* are now recognized. Various classifications have been proposed, based on genetic typing, clinical presentation, and geography. None of these is ideal because species that are genetically very similar can occur in different parts of the world, and there is considerable overlap between the various rickettsial syndromes. A useful clinical approach is to divide rickettsial infections into three groups: typhus, spotted fever, and scrub typhus, although the organisms causing the last of these are now strictly classified as orientiae and not rickettsiae.

Typhus group

Epidemic typhus

The vector of epidemic typhus is the human body louse and, as with louse-borne relapsing fever, epidemics are associated with war and refugees.

The incubation period of 1 to 3 weeks is followed by an abrupt febrile illness associated with profound malaise and generalized myalgia. Headache is severe, and there may be conjunctivitis with orbital pain. A measles-like eruption appears around the fifth day. At the end of the first week, signs of meningoencephalitis appear, and CNS involvement may progress to coma. At the height of the illness, splenomegaly, pneumonia, myocarditis, and gangrene at the peripheries

? Box 42.52 Infections caused by rickettsiae

Disease	Organism	Reservoir	Vector
Typhus fever group			
Epidemic typhus	*Rickettsia prowazekii*	Human	Human body louse
Endemic (murine) typhus	*Rickettsia typhi*	Rodent	Rat flea
Spotted fever group			
African tick typhus	*Rickettsia africae*	Various mammals	Hard tick
Mediterranean spotted fever	*Rickettsia conorii*	Rodent, dog	Hard tick
Rocky Mountain spotted fever	*Rickettsia rickettsii*	Rodent, dog	Hard tick
Rickettsial pox	*Rickettsia akari*	Rodent	Mite
Flea-borne spotted fever	*Rickettsia felis*	Various mammals	Flea
Scrub typhus	*Orientia tsutsugamushi*	Trombiculid mite	Trombiculid mite

may be evident. Oliguric acute kidney injury occurs in fulminating disease, which is usually fatal. Recovery begins in the third week but is generally slow. If untreated, the disease may recur many years after the initial attack, owing to rickettsiae that lie dormant in lymph nodes. The recrudescence, which is now rarely seen because of appropriate use of antibiotics in the primary phase, is known as Brill–Zinsser disease.

Endemic (murine) typhus

This is an infection of rodents that is inadvertently spread to humans by rat fleas. The disease closely resembles epidemic typhus but is much milder and rarely fatal.

Spotted fever group

A variety of *Rickettsia* species, collectively known as the spotted fever group rickettsiae, cause the illnesses known as spotted fevers. In most cases, the vector is a hard tick. Although the causative organism and the name of the illness vary from place to place, the clinical course is common to all.

The typical feature of the spotted fevers is a widespread petechial rash (Fig. 42.26), although a variety of other types of skin lesion are seen, especially in Rocky Mountain spotted fever. An exception is the form of African tick typhus caused by *Rickettsia africae*, which often presents without a rash.

After an incubation period of 4 to 10 days, an eschar (a black, crusted, necrotic papule) may develop at the site of the bite in association with regional lymphadenopathy (Fig. 42.27). There is abrupt onset of fever, myalgia, and headache accompanied by a maculopapular rash that may become petechial. Neurological, haematological, and cardiovascular complications can occur, as in epidemic typhus, although these are uncommon.

Scrub typhus

Found throughout Asia and the Western Pacific (and also reported occasionally from Africa), this disease is spread by larval trombiculid mites (chiggers). An eschar can often be found at the site of the bite. The clinical illness is very variable, ranging from a mild illness to fulminant and potentially fatal disease. The more severe cases resemble epidemic typhus. Unlike in other types of typhus, the organism is passed on to subsequent generations of mites, which consequently act as both reservoir and vector.

Diagnosis

The diagnosis of all types of rickettsial infection is generally made on the basis of the history and clinical course. It can be confirmed serologically or by PCR.

Management and prevention

Doxycycline given for 5 to 7 days is the treatment of choice for uncomplicated disease. Ciprofloxacin is also effective. Seriously ill patients may need intensive care, and the addition of azithromycin to doxycycline improves outcome. Prophylactic doxycycline 200 mg weekly protects against scrub typhus; it is reserved for highly endemic areas. Rifampicin is also used when resistance to tetracycline is present. Control of typhus is achieved by eradication of the arthropod vectors: lice and fleas can be eradicated from clothing with insecticides (0.5% malathion or DDT). Control of rodents is necessary in endemic typhus and some of the spotted fevers. Areas of vegetation infested with trombiculid mites can be cleared by chemical spraying from the air. Bites from ticks and mites should be avoided by wearing protective clothing on exposed areas of the body. The likelihood of infection from ticks is related to the duration of feeding; in high-risk areas, the body should be inspected twice a day because the bites are painless, and any ticks should be removed.

Bartonellosis

Bartonella spp. are intracellular bacteria closely related to the rickettsiae. A number of human diseases can be caused by these organisms, including cat-scratch disease and bacillary angiomatosis *(Bartonella henselae)*. As in rickettsial disease, infection is usually spread from animals via an arthropod vector (Box 42.53).

Carrion's disease *(Bartonella bacilliformis)*

This disease is largely restricted to the habitat of its main vector, the sandfly, in the river valleys of the Andes mountains at an altitude of 500 to 3000 m. Two clinical presentations are seen, which may occur alone or consecutively. ***Oroya fever*** is an acute febrile illness causing myalgia, arthralgia, severe headache, and confusion followed by a haemolytic anaemia. ***Verruga peruana*** consists of eruptions of reddish-purple haemangiomatous nodules, resembling bacillary angiomatosis. It may follow 4 to 6 weeks after Oroya fever or be the presenting feature of infection. Spontaneous resolution may occur over a period of months or years. Carrion's disease is frequently complicated by superinfection, especially with *Salmonella* spp.

The diagnosis is made by culturing bacilli from blood or peripheral lesions. Serological tests have been developed but are not widely available.

Fig. 42.26 **The petechial rash of rickettsial spotted fever.**

Fig. 42.27 **An eschar developing at the site of the bite in rickettsial disease.**

Box 42.53 Human infections caused by *Bartonella* spp. and *Ehrlichia* spp.

Disease	Organism	Reservoir	Vector
***Bartonella* spp.**			
Carrion's disease	*Bartonella bacilliformis*	Unknown	Sandfly
Cat-scratch disease	*Bartonella henselae*	Cat	Cat flea
Bacillary angiomatosis	*Bartonella henselae*	Cat	Cat flea
Trench fever	*Bartonella quintana*	Human	Body louse
***Ehrlichia* spp.**			
Human monocytic ehrlichiosis (HME)	*Ehrlichia chaffeensis*	Deer	Hard tick
Human granulocytic ehrlichiosis (HGE)[a]	*Ehrlichia ewingii* *Anaplasma phagocytophilum*[b]	Small mammals and deer	Hard tick

[a]Also known as human granulocytic anaplasmosis.
[b]Formerly known as Ehrlichia phagocytophilia.

Treatment with chloramphenicol or tetracycline is effective in acute disease but less so in verruga peruana. It has no known hosts other than humans.

The disease is named after Daniel Alcides Carrion Garcia, a Peruvian medical student who died from the disease in 1885. Carrion was trying to demonstrate the link between the acute Oroya fever and chronic verruga peruana when he inoculated himself with blood from the wart of a 14-year-old patient. It has the potential for eradication, but there is currently no surveillance programme, and it is a paradigm for a neglected disease of poverty.

Trench fever

Trench fever is caused by *Bartonella quintana* and transmitted by human body lice. It is mainly seen in refugees and homeless people. It is characterized by cyclical fever, chills, and headaches accompanied by myalgia and pretibial pain. The disease is usually self-limiting, but it can be treated with erythromycin or doxycycline if symptoms are severe.

Ehrlichiosis

Ehrlichiosis and anaplasmosis are infections caused by tick-borne, rickettsia-like bacteria. At least three species have been implicated: *Ehrlichia chaffeensis*, which causes human monocytic ehrlichiosis, and *Ehrlichia ewingii* and *Anaplasma phagocytophilum* (formerly known as *Ehrlichia phagocytophilum*), which cause human granulocytic ehrlichiosis (also known as human granulocytic anaplasmosis). All cause a rather non-specific febrile illness with fever, myalgia, and headache. Treatment is with doxycycline. The vectors are hard ticks, and the main reservoir hosts are deer. As with most tick-borne zoonoses, the avoidance of tick bites and the prompt removal of feeding ticks are the best forms of prevention.

OTHER BACTERIAL INFECTIONS

Melioidosis

This term refers to infections caused by the Gram-negative bacterium *Burkholderia pseudomallei*. This environmental organism, which is found in soil and surface water, is distributed widely in the tropics and subtropics. Most clinical cases of melioidosis occur in South-east Asia and the Northern Territories of Australia with a peak presentation during the rainy season. Infection follows inhalation or direct inoculation. More than half of all patients with melioidosis have predisposing underlying disease; it is particularly common in people with diabetes. The organism may remain latent for many years and may reactivate when patients become immunosuppressed from conditions such as renal failure, diabetes, and alcoholism.

B. pseudomallei causes a wide spectrum of disease, and the majority of infections are probably subclinical. Illness may be acute or chronic and localized or disseminated, but one form may progress to another, and individual patients may be difficult to categorize. The most serious form is septicaemic melioidosis, which is often complicated by multiple metastatic abscesses; it is frequently fatal. Serological tests are available, but definitive diagnosis depends on isolating the organism from blood or appropriate tissue. *B. pseudomallei* has extensive intrinsic antibiotic resistance. The most effective agent is ceftazidime, given intravenously for 2 to 4 weeks; this should be followed by several months of co-amoxiclav or co-trimoxazole to prevent relapses.

Actinomycosis

Actinomyces spp. are branching, Gram-positive higher bacteria, which are normal mouth and intestine commensals; they are particularly associated with poor mouth hygiene. *Actinomyces* spp. have a worldwide distribution but are a rare cause of disease in the West.

Cervicofacial actinomycosis, the most common form, usually occurs after dental infection or extraction. It is often indolent and slowly progressive, is associated with little pain, and results in induration and localized swelling of the lower part of the mandible. Sinuses and tracts develop, with discharge of 'sulphur' granules. ***Thoracic actinomycosis*** follows inhalation of organisms, usually into a previously damaged lung. The clinical picture is not distinctive and is often mistaken for malignancy or TB. Symptoms such as fever, malaise, chest pain, and haemoptysis are present. Empyema occurs in 25% of patients, and local extension produces chest-wall sinuses with discharge of 'sulphur' granules.

Abdominal and pelvic actinomycosis most frequently affects the caecum. Characteristically, an indurated mass is felt in the right iliac fossa. Later, sinuses develop. The differential diagnosis includes malignancy, TB, Crohn's disease, and amoeboma. The incidence of pelvic actinomycosis appears to be increasing with wider use of intrauterine contraceptive devices. Occasionally, actinomycosis becomes disseminated to involve any site.

Diagnosis is by microscopy and culture of the organism. The presence of firm 'sulphur granules' containing collections of filamentous rods is virtually pathognomonic. Treatment often involves surgery, as well as very long courses of antibiotics. Penicillin is the drug of choice and is usually given until 3 to 4 months after clinical resolution. Tetracyclines and sulphonamides are also effective.

Nocardia infections

Nocardia spp. are Gram-positive branching bacteria found in soil and decomposing organic matter. *Nocardia asteroides* and, less often, *Nocardia brasiliensis* are the main human pathogens.

Mycetoma is the most common illness. It is a result of local invasion by *Nocardia* spp. and presents as a painless swelling, usually on the sole of the foot (Madura foot). The swelling of the affected part of the body continues inexorably. Nodules gradually appear, which eventually rupture and discharge characteristic 'grains', which are colonies of organisms. Systemic symptoms

and regional lymphadenopathy are rare. Sinuses may occur several years after the onset of the first symptom. A similar syndrome may be produced by other branching bacteria and also by species of eumycete fungi, such as *Madurella mycetomi*. Most cases of Madura foot occur in rural parts of India, sub-Saharan Africa, and Mexico.

Pulmonary disease, which follows inhalation of the organism, presents with cough, fever, and haemoptysis; it is usually seen in immunocompromised people. Pleural involvement and empyema occur. In severely immunosuppressed patients, initial pulmonary infection may be followed by disseminated disease.

The diagnosis is often delayed, partly because of inadequate specimens and partly because *Nocardia* spp. require prolonged culture on standard media. Severe pulmonary or disseminated infection may require parenteral treatment. Co-trimoxazole, linezolid, ceftriaxone, and amikacin have all been used successfully, but *in vitro* sensitivities are variable, and there is no consensus on the best treatment.

Antibiotic therapies

Drugs used to combat bacterial infections are described in Chapter 12 ('Antibiotic therapies' section).

Further reading

Abdad MY, Abou Abdallah R, Fournier PE, Stenos J, Vasoo S. A concise review of the epidemiology and diagnostics of rickettsioses: Rickettsia and Orientia spp. *J Clin Microbiol* 2018; 56(8):e01728-17.
Esquivel Gomez LR, Savin C, Andrianaivoarimanana V et al. Phylogenetic analysis of the origin and spread of plague in Madagascar. *PLoS Negl Trop Dis* 2023; 17(5):e0010362.
Fahal AH, Bakhiet SM. Mycetoma and the environment. *PLoS Negl Trop Dis* 2023; 17(11):e0011736.
Garcia-Quintanilla M, Dichter AA, Guerra H, Kempf VAJ. Carrion's disease: more than a neglected disease. *Parasit Vectors* 2019; 12(1):141.
Stewart JD, Smith S, Binotto E et al. The epidemiology and clinical features of melioidosis in Far North Queensland: Implications for patient management. *PLoS Negl Trop Dis* 2017; 11(3):e0005411.
Varghese GM, Dayanand D, Gunasekaran K et al. INTREST Trial Investigators. Intravenous Doxycycline, Azithromycin, or both for severe scrub typhus. *N Engl J Med* 2023; 388(9):792–803.

FUNGAL INFECTIONS

Morphologically, fungi can be grouped into three major categories:

- ***Yeasts and yeast-like fungi***, which reproduce by budding
- ***Moulds***, which grow by branching and longitudinal extension of hyphae
- ***Dimorphic fungi***, which behave as yeasts in the host but as moulds *in vitro*.

Even though fungi are ubiquitous, systemic fungal infections are relatively rare, in contrast to superficial fungal infections of the skin, nails, and orogenital mucosae. Systemic mycoses are usually seen in immunocompromised patients and in critical care settings and are becoming more prevalent as this population increases.

Fungal infections are transmitted by inhalation of spores, contact with the skin, or direct inoculation. This last can occur through penetrating injuries, injecting drug use, or iatrogenic procedures. Fungi may also produce allergic pulmonary disease. Some fungi, such as *Candida albicans*, are human commensals. Fungal infections are usually divided into systemic, subcutaneous, or superficial (**Box 42.54**). However, these divisions can be misleading because many fungi that cause superficial infection in healthy hosts can cause invasive or disseminated infection in immunocompromised patients, but 'deep' subcutaneous infections can invade other organs to cause systemic disease.

Systemic fungal infections

Candidiasis

Candidiasis is the most common fungal infection in humans and is predominantly caused by *Candida albicans*, although other species of *Candida* are increasingly seen. *Candida* are small asexual yeasts. They are found worldwide, and most species that cause serious invasive disease are also normal oropharyngeal and GI commensals.

Clinical features

Most human candidiasis is superficial. Vaginal thrush and oral thrush are the most common forms, typically seen in very young patients, older adults, patients on antibiotic therapy, and patients who are immunosuppressed. Cutaneous candidiasis typically occurs in intertriginous areas. It is also a cause of paronychia.

Invasive candidiasis is usually associated with intravascular devices (especially in intensive care units) and with profound immunosuppression (e.g., caused by treatment of haematological malignancy). Dissemination can lead to meningitis, visceral and pulmonary abscesses, endophthalmitis, and osteomyelitis. *Candida* oesophagitis presents with painful dysphagia and is generally seen in patients with cancer, diabetes, and cellular immunosuppression (e.g., from long-term steroid use and uncontrolled HIV infection).

Diagnosis and management

The fungi can be demonstrated in scrapings from infected lesions; from tissue secretions; or in invasive disease, from standard blood cultures. Serological tests for beta-D-glucan, a fungal cell wall component, can be used to look for systemic fungal infection but are common to many pathogenic fungi, and the tests do not distinguish different infections.

Treatment varies, depending on the site and severity of infection and on the sensitivity of the organism. Superficial lesions may respond to topical nystatin or amphotericin B, although systemic agents may be needed for more extensive infection. For invasive

Box 42.54 Important fungal infections

Systemic
- Histoplasmosis
- Cryptococcosis
- Coccidioidomycosis
- Blastomycosis
- Zygomycosis (mucormycosis)
- Candidiasis
- Aspergillosis
- *Pneumocystis jirovecii* (formerly *Pneumocystis carinii*)

Subcutaneous
- Sporotrichosis
- Subcutaneous zygomycosis
- Chromoblastomycosis
- Mycetoma

Superficial
- Dermatophytosis
- Superficial candidiasis
- *Malassezia* infections

infection, parenteral therapy with amphotericin B, azoles, or echinocandins is necessary. Many non-*albicans* species are intrinsically resistant to many antifungals, and even *C. albicans* can develop extensive resistance (especially to azoles).

Candida auris

Since 2010 there has been an increase in human colonization and infection with *C. auris*, a species first identified in Japan in 2009 and that has no known environmental niche. *C. auris* is particularly worrying because it is intrinsically resistant to many antifungals and sometimes to all three major classes of drug. It is also spread readily from person to person (especially in healthcare settings) and can survive for extended periods on hard surfaces despite normal cleaning and disinfection processes. Laboratory diagnosis is difficult and it can easily be mistaken for other species. Healthcare-associated outbreaks have been reported from a number of countries across all continents, and it is common to screen patients who have recently been in hospital in countries with high rates of nosocomial *C. auris*.

Histoplasmosis

Histoplasmosis is caused by *Histoplasma capsulatum*. Spores can survive in moist soil for several years, particularly when the soil is enriched by bird and bat droppings. Histoplasmosis occurs worldwide but is most commonly seen in the Ohio and Mississippi river valleys, where more than 80% of the population have been subclinically exposed. An increasing number of cases are being reported from parts of Asia (especially China). Transmission is mainly by inhalation of the spores, particularly when clearing out attics, barns, or bird roosts or exploring caves.

Clinical features

Primary pulmonary histoplasmosis is usually asymptomatic. Calcification in the lungs, spleen, and liver occurs in patients from areas of high endemicity. When symptomatic, primary pulmonary histoplasmosis generally presents as a mild influenza-like illness, although symptoms can be more severe. Occasionally, there can be complications such as secondary bacterial pneumonia, pleural effusions, erythema nodosum, and erythema multiforme.

Chronic pulmonary histoplasmosis is clinically similar to pulmonary TB. It is usually seen in White American males older than the age of 50 years. Disseminated histoplasmosis resembles disseminated TB, with fever, lymphadenopathy, hepatosplenomegaly, weight loss, leucopenia, and thrombocytopenia. Rarely, features of meningitis, hepatitis, hypoadrenalism, endocarditis, and peritonitis may dominate the clinical picture. Histoplasmosis sometimes remains latent for years and reactivates when the patient becomes immunosuppressed. This long interval between exposure and disease may result in delayed recognition.

Diagnosis

Definitive diagnosis is possible by culturing the fungi (e.g., from sputum) or by demonstrating them on histological sections. *H. capsulatum* glycoprotein can be detected in the urine and serum of those with acute pulmonary and disseminated infections. Diagnostic tests using urinary antigens are becoming more widely available. Antibodies usually develop within 3 weeks of the onset of illness and are best detected by complement fixation or immunodiffusion (sensitivities of 95% and 90%, respectively).

Management

Only symptomatic acute pulmonary histoplasmosis, chronic histoplasmosis, and acute disseminated histoplasmosis require therapy. Itraconazole is effective in mild to moderate disease. Severe infection is treated with intravenous amphotericin B followed by itraconazole for a total of 3 months or with voriconazole. Surgical excision of histoplasmomas (granulomatous pulmonary nodules caused by *H. capsulatum*) or chronic cavitatory lung lesions and release of adhesions after mediastinitis may be required.

African histoplasmosis

This is caused by a specific subspecies, *Histoplasma capsulatum* var. *duboisii*, the spores of which are larger than those of *H. capsulatum*. Skin lesions (e.g., abscesses and nodules), lymph node involvement, and lytic bone lesions are prominent. Pulmonary lesions do not occur. Treatment is similar to that for *H. capsulatum* infection.

Aspergillosis

Aspergillosis is caused by one of several species of dimorphic fungi of the genus *Aspergillus*. Of these, *Aspergillus fumigatus* is the most common, although *Aspergillus flavus* and *Aspergillus niger* are also recognized. These fungi are ubiquitous in the environment and are commonly found on decaying leaves and trees. Humans are infected by inhalation of the spores. Disease manifestation depends on the dose of the spores inhaled, as well as the immune response of the host. Three major forms of the disease are recognized: bronchopulmonary allergic aspergillosis, aspergilloma, and invasive aspergillosis.

Diagnosis and treatment are described in Chapter 20 (‘Aspergilloma and invasive aspergillosis’ section).

Cryptococcosis

Cryptococcosis is caused by the yeast-like fungus *Cryptococcus neoformans*. It has a worldwide distribution, with bird droppings being an important mode of spread. The spores enter the body through the respiratory tract, where they elicit a granulomatous reaction. Pulmonary symptoms are, however, uncommon, although lung cavitation, hilar lymphadenopathy, pleural effusions, and pulmonary fibrosis occur. Skin and bone involvement are rare. Cryptococcal meningitis is primarily a disease of advanced HIV infection and often presents subacutely with fever and headache caused by raised intracranial pressure. It sometimes occurs in patients with lymphoma or iatrogenic suppression of cellular immunity.

Diagnosis and management

Diagnosis is established by demonstrating the organisms in appropriately stained tissue sections. A positive latex cryptococcal agglutinin test performed on the CSF is diagnostic of cryptococcal meningitis. Patients need indefinite treatment until their cellular immunity recovers.

Guidelines for the treatment of HIV-related cryptococcal meningitis, which carries a high mortality rate, have recently changed based on the Ambition study. This study showed that a single dose of liposomal amphotericin B, at 10 mg/kg given in combination with flucytosine and fluconazole is as effective as a standard 2-week course of amphotericin B plus flucytosine, better tolerated, and far cheaper. Therapeutic lumbar puncture has an important role in acute illness.

Coccidioidomycosis

Coccidioidomycosis is caused by the non-budding spherical form (spherule) of *Coccidioides immitis*. This is a soil saprophyte and is found in the southern USA, Central America, and parts of South America. Humans are infected by inhalation of the thick-walled,

barrel-shaped spores called arthrospores. Occasionally, epidemics of coccidioidomycosis have been documented after dust storms.

Clinical features

The majority of patients are asymptomatic, and the infection is detected only by the conversion of a skin test using coccidioidin (extract from a culture of mycelial growth of *C. immitis*) from negative to positive. Acute pulmonary coccidioidomycosis presents after an incubation period of about 10 days with fever, malaise, and cough. Erythema nodosum; erythema multiforme; phlyctenular conjunctivitis; and less commonly, pleural effusions may occur. Complete recovery is usual.

Pulmonary cavitation with haemoptysis, pulmonary fibrosis, meningitis, lytic bone lesions, hepatosplenomegaly, skin ulcers, and abscesses may occur in severe disease.

Diagnosis

The organism can be identified in respiratory secretions and cultured in specialist laboratories. Serological tests are also widely used for diagnosis. These include the highly specific latex agglutination and precipitin tests (IgM); the test results are positive within 2 weeks of infection and decline thereafter. Other tests include complement fixation, ELISA, and radioimmunoassay.

A complement fixation test (IgG) performed on the CSF is diagnostic of coccidioidomycosis meningitis; it becomes positive within 4 to 6 weeks and remains so for many years.

Management

Mild pulmonary infections are self-limiting and require no treatment, but progressive and disseminated disease mandates urgent therapy. Itraconazole or fluconazole for 6 months is the treatment of choice for patients with primary pulmonary disease, with more prolonged courses for cavitating or fibronodular disease. Fluconazole in high dose (600–1000 mg/day) is given for meningitis. Voriconazole or posaconazole is used for poor responders. Surgical excision of cavitatory pulmonary lesions or localized bone lesions may be necessary.

Blastomycosis

Blastomycosis is a systemic infection caused by the biphasic fungus *Blastomyces dermatitidis*. Although previously believed to be confined to certain parts of North America, it has now been reported in South America, India, and the Middle East.

Clinical features

Blastomycosis primarily involves the skin, causing non-itchy papular lesions that later develop into ulcers with red verrucous margins. The ulcers are initially confined to the exposed parts of the body but later involve the unexposed parts as well. Atrophy and scarring may occur. Pulmonary involvement presents as a solitary lesion resembling a malignancy or gives rise to radiological features similar to the primary complex of TB. Systemic symptoms, such as fever, malaise, cough, and weight loss, are usually present. Bone lesions are common and present as painful swellings.

Diagnosis and management

The diagnosis is confirmed by demonstrating the organism in histological sections or by culture, although results can be negative in 30% to 50% of cases. Enzyme immunoassay may be helpful, although there is some cross-reactivity of antibodies to *Blastomyces* spp. with *Histoplasma* spp.

Itraconazole is preferred for treating mild to moderate disease in immunocompetent patients for periods of up to 6 months. In severe or unresponsive disease and in immunocompromised patients, amphotericin B is usually used.

Paracoccidioidomycosis

Sometimes known as 'South American blastomycosis', paracoccidioidomycosis is caused by dimorphic fungi of the genus *Paracoccioides*. It is confined to localized areas of Brazil and neighbouring countries. Like the other pathogenic dimorphic fungi, it is usually acquired by inhalation. It can cause a wide variety of systemic and local manifestations, although skin lesions are particularly common. Itraconazole is generally the best treatment, although amphotericin B can be used in severe cases.

Mucormycosis

The terminology relating to mucormycosis and zygomycosis is confusing. ***Invasive zygomycosis*** (also known as ***mucormycosis***) is rare and is caused by several fungi, including *Mucor*, *Rhizopus*, and *Absidia* spp. It usually occurs in immunocompromised or severely ill patients and has a high mortality rate. The hallmark of the disease is vascular invasion with marked haemorrhagic necrosis.

Rhinocerebral mucormycosis is a specific, locally invasive form of the condition. Nasal stuffiness; facial pain and oedema; and necrotic, black nasal turbinates are characteristic. It is rare and is mainly seen in people with poorly controlled diabetes.

Other forms of invasive zygomycosis include pulmonary and disseminated infection (in immunosuppressed patients) and GI infection (in malnutrition).

Subcutaneous zygomycosis is described later. Management of locally invasive and systemic disease is with amphotericin B and surgical debridement, but even with the best care, the mortality rate exceeds 50%. The newer triazoles, posaconazole and isavuconazole, have been shown to be effective in mucormycosis, but their role in therapy remains unclear.

Further reading

Eyre DW, Sheppard AE, Madder H et al. A *Candida auris* outbreak and its control in an intensive care setting. *N Engl J Med* 2018; 379:1322–31.
Jarvis JN, Lawrence DS, Meya DB et al. Ambition Study Group. Single-dose liposomal amphotericin b treatment for cryptococcal meningitis. *N Engl J Med* 2022; 386(12):1109–1120.
Ocansey BK, Kosmidis C, Agyei M et al. Histoplasmosis in Africa: current perspectives, knowledge gaps, and research priorities. *PLoS Negl Trop Dis* 2022; 16(2):e0010111.
Özbek L, Topçu U, Manay M et al. COVID-19-associated mucormycosis: a systematic review and meta-analysis of 958 cases. *Clin Microbials Infect* 2023; 29(6):722–731.
Haydour Q, Hage CA, Carmona EM et al. Diagnosis of fungal infections. A systematic review and meta-analysis supporting American Thoracic Society Practice Guideline. *Ann Am Thorac Soc* 2019; 16(9):1179–1188.
Thompson GR 3rd, Le T, Chindamporn A et al. Global guideline for the diagnosis and management of the endemic mycoses: an initiative of the European Confederation of Medical Mycology in cooperation with the International Society for Human and Animal Mycology. *Lancet Infect Dis* 2021; 21(12):e364–e374. Erratum in: *Lancet Infect Dis* 2021; 21(11):e341.

Subcutaneous fungal infections

Sporotrichosis

Sporotrichosis is caused by the saprophytic fungus *Sporothrix schenckii*, which is found worldwide. It is common in soil and on plants, and infection classically follows cutaneous inoculation from thorns ('rose handler's disease). A reddish, non-tender, maculopapular lesion develops at the site of inoculation; this is referred to as 'plaque sporotrichosis'. Satellite lesions may develop over

time. Pulmonary involvement and disseminated disease occur very rarely.

Treatment with itraconazole for 3 to 6 months is usually curative.

Subcutaneous zygomycosis

Subcutaneous zygomycosis can be a form of mucormycosis in immunocompromised patients, as described earlier. However, the term is also used to describe a largely tropical condition caused by filamentous fungi of the *Basidiobolus* genus, typically associated with inoculation injuries and trauma. The disease usually remains confined to the subcutaneous tissues and muscle fascia. It presents as a brawny, woody infiltration involving the limbs, neck, and trunk. Amphotericin B is the drug of choice, usually in combination with surgical debridement. Even the cutaneous and subcutaneous form of the infection carries a significant mortality.

Chromoblastomycosis

Chromoblastomycosis (chromomycosis) is caused by fungi of various genera, including *Phialophora*, *Wangella*, and *Fonsecaea* spp. These are found mainly in tropical and subtropical countries. Chromoblastomycosis presents initially as a small papule, usually at the site of a previous injury, which persists for several months before ulcerating. The lesion later becomes warty and encrusted and gradually spreads. Satellite lesions may be present. Itching is common. The drug of choice is amphotericin B in combination with itraconazole or voriconazole. Cryosurgery is used to remove local lesions.

Mycetoma (Madura foot)

Mycetoma may be caused by subcutaneous infection with fungi (*Eumycetes* spp.) or bacteria. It is largely confined to the tropics and is a disease of poverty in patients with poor or no footwear. Infection results in local swelling, which may discharge through sinuses. Bone involvement may follow.

Management consists of surgical debridement combined with antimicrobials chosen according to the aetiological agent.

Pneumocystis jirovecii infection

Genetic analysis has shown *P. jirovecii* to be homologous with fungi. *P. jirovecii* disease is almost invariably associated with immunodeficiency states, particularly HIV infection (see Chapter 44).

Superficial fungal infections

Dermatophytosis

Dermatophytoses are chronic fungal infections of keratinous structures, such as the skin, hair, or nails. The main causative organisms are species of the genera *Trichophyton*, *Microsporum*, and *Epidermophyton*.

Malassezia infection

Malassezia spp. are found on the scalp and greasy skin and are responsible for seborrhoeic dermatitis, pityriasis versicolor (hypo- or hyperpigmented rash on the trunk), and *Malassezia* folliculitis (itchy rash on the back).

Treatment of any superficial fungal infection is with topical antifungals or with oral triazole drugs if infection is refractory or more extensive.

Antifungal drugs

Fungal infections range from trivial superficial skin and nail infections to life-threatening systemic conditions. The antifungal armory is relatively limited, and because fungal cells share more enzymes and other pathways with human cells than do bacteria, targeting the pathogen without harming host metabolism is more difficult. There are three main groups of parenteral drugs for treating severe fungal infections: amphotericin, triazoles, and echinocandins (**Box 42.55**). Although each has specific licenced indications, antifungal drugs are often given empirically in severely ill patients. The best choice from among the available agents depends on several factors, including local resistance patterns and availability, potential for toxicity, and cost. Many countries have specific national guidelines.

Polyenes

Polyenes react with the sterols in fungal membranes, increasing permeability and thus damaging the organism. The most potent is ***amphotericin***, which is used intravenously in severe systemic fungal infections. Nephrotoxicity is a major problem, and dosage levels must take background renal function into account. Liposomal formulations of amphotericin are less toxic and are used in preference, although they are very expensive. ***Nystatin*** is not absorbed through mucous membranes and is therefore useful for the treatment of oral and enteric candidiasis and for vaginal infection. It can only be given orally or as pessaries.

Azoles

Imidazoles

Imidazoles are broad-spectrum antifungal drugs. They act by inhibiting fungal sterol synthesis, resulting in damage to the cell wall. ***Ketoconazole*** is active orally but can produce liver damage and is now rarely used systemically. It continues to be used topically, along with ***clotrimazole*** and ***miconazole***, for the treatment of ringworm and cutaneous and genital candidiasis. ***Econazole*** is used for the topical treatment of cutaneous and vaginal candidiasis and dermatophyte infections, and ***tioconazole*** is indicated for fungal nail infections.

Box 42.55 Antifungal agents

Polyenes
- Amphotericin, nystatin

Echinocandins
- Caspofungin
- Anidulafungin
- Micafungin

Azoles
- Miconazole, ketoconazole, fluconazole, itraconazole, voriconazole, posaconazole, isavuconazole
- Topical clotrimazole, sulconazole, tolnaftate, econazole, tioconazole

Allylamines
- Terbinafine

Other antifungals
- Amorolfine (topical only)
- 5-Flucytosine
- Griseofulvin

Triazoles

These drugs also act by interfering with fungal sterol synthesis. ***Itraconazole*** was the first triazole to be used widely in clinical practice, but its relatively poor absorption and failure to penetrate CSF mean that it has largely been superseded (except for a few specific conditions such as histoplasmosis and blastomycosis). ***Fluconazole*** is well absorbed and has excellent CNS penetration; it is used for mucosal and invasive candidiasis and for the treatment of CNS infection with *C. neoformans*. Some non-*albicans Candida* species are intrinsically resistant to fluconazole, and resistance can develop in *C. albicans*. ***Voriconazole*** has broad-spectrum activity that includes most *Candida*, *Cryptococcus*, and *Aspergillus* spp. and other filamentous fungi. It is available for oral and intravenous use. Adverse effects include visual disturbance, abnormalities of liver enzymes, and a variety of dermatological problems (including a possible link with squamous cell carcinoma). It also has significant drug–drug interactions with drugs metabolized by the CYP450 enzymes, and like all the triazoles, can cause QTc prolongation. It is indicated for invasive aspergillosis and severe *Candida* infections unresponsive to amphotericin and fluconazole, respectively, and for empirical treatment of suspected or confirmed severe fungal infection. ***Posaconazole*** has a similar range of activity and is also associated with hepatotoxicity and CYP450 interactions. ***Isavuconazole*** is a new triazole: it is licensed for use in invasive mould disease and mucormycosis but (largely because of cost) is generally a second- or third-line agent.

Echinocandins

These act by inhibiting the cell-wall polysaccharide glucan. ***Caspofungin***, which is administered intravenously, is active against *Candida* spp. and *Aspergillus* spp. and is indicated for invasive candidiasis and invasive aspergillosis unresponsive to other drugs, as well as for empirical therapy in seriously ill patients. Other echinocandins include ***micafungin*** and ***anidulafungin***, approved for the treatment of patients with disseminated candidiasis. Severe hepatotoxicity has been reported with micafungin, and liver biochemistry must be monitored.

Flucytosine

The fluorinated pyridine derivative flucytosine is used in combination with amphotericin B for cryptococcal meningitis. Side-effects are uncommon, although it may cause bone marrow suppression. It is active when given orally or parenterally, but there is very limited availability in resource-poor settings.

Allylamines

Terbinafine has broad-spectrum antifungal and also has anti-inflammatory activity. It is well absorbed orally and accumulates in keratin. It is useful for the treatment of patients with superficial mycoses, such as dermatophyte infections, onychomycosis, and cutaneous candidiasis. A topical formulation is also available to treat those with fungal skin infections.

Further reading

Kruithoff C, Gamal A, McCormick TS, Ghannoum MA. Dermatophyte infections worldwide: increase in incidence and associated antifungal resistance. *Life* (Basel) 2023; 14(1):1.
Pristov KE, Ghannoum MA. Resistance of Candida to azoles and echinocandins worldwide. *Clin Microbiol Infect* 2019; 25:792–798.

PROTOZOAL INFECTIONS

Protozoa are unicellular eukaryotic organisms. They are more complex than bacteria and belong to the animal kingdom. Although many protozoa are free living in the environment, some have become parasites of vertebrates, including humans, often developing complex life cycles involving more than one host species. To be transmitted to a new host, some protozoa transform into hardy cyst forms, which can survive harsh external conditions. Others are transmitted by an arthropod vector, in which a further replication cycle takes place before infection of a new vertebrate host.

BLOOD AND TISSUE PROTOZOA

Malaria

Human malaria is usually caused by one of five species of the genus *Plasmodium: Plasmodium falciparum*, *Plasmodium vivax*, *Plasmodium ovale*, *Plasmodium malariae*, and rarely *Plasmodium knowlesi* (an occasional spillover from primates in Indonesia). Human malaria probably originated from animal malarias in Central Africa but was spread around the globe by human migration. Public health measures and changes in land use have eradicated malaria in much of the world where it was previously endemic, although the potential for malaria transmission still exists in many areas. Some 247 million people were infected in 2021, although 29 countries account for 96% cases, with more than 600 000 deaths (mainly in African children). The COVID pandemic stalled global reductions in incidence and mortality seen over prior decades. More than 30 000 infections per year occur in travellers from non-malarious countries.

Epidemiology

Malaria is transmitted by the bite of female anopheline mosquitoes. The parasite undergoes a temperature-dependent cycle of development in the gut of the insect, and its geographical range therefore depends on the presence of the appropriate mosquito species and an adequate temperature. The disease occurs in endemic or epidemic form throughout the tropics and subtropics except for some areas above 2000 m (Fig. 42.28), although an increasing number of countries have been declared malaria free in the past 10 years.

In areas of so-called ***stable transmission*** (including much of sub-Saharan Africa), transmission occurs consistently year-round. The bulk of the deaths are seen in children, and those who survive to adulthood acquire significant immunity; low-grade parasitaemia is still present but causes few symptoms. ***Unstable transmission*** occurs when there is erratic, seasonal, or low-level transmission (e.g., in the Sahel belt, where mosquitoes feed only in the rainy season). In these circumstances, little protective immunity develops, and symptomatic malaria occurs at all ages. Changes in environmental or social conditions in such areas can lead to epidemics with substantial mortality rates in all age groups.

Malaria can also be transmitted in contaminated blood transfusions. It has occasionally been seen in patients with a hospital-acquired infection related to contaminated equipment. Rare cases are acquired outside the tropics when mosquitoes are transported from endemic areas (‘airport malaria’) or when the local mosquito population becomes infected by a returning traveller.

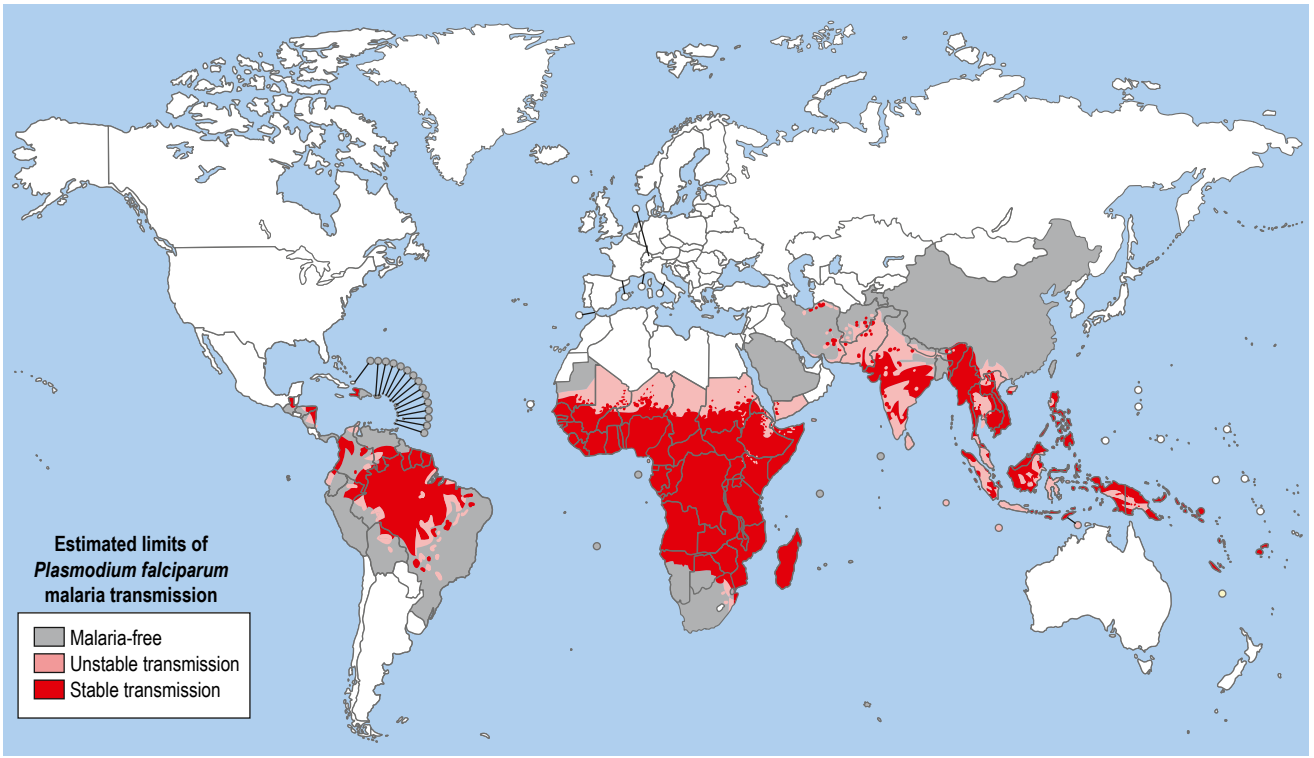


Fig. 42.28 ***Plasmodium falciparum*** **malaria transmission: geographical distribution.** Map lines delineate study areas and do not necessarily depict accepted national boundaries. *(From Malaria Atlas Project, with permission; https://data.malariaatlas.org/maps?layers=Malaria:202206_Global_Pf_Parasite_Rate.)*

Parasitology

The female mosquito becomes infected after taking a blood meal containing gametocytes, the sexual form of the malarial parasite (Fig. 42.29). The developmental cycle in the mosquito usually takes 7 to 20 days (depending on temperature), culminating in the migration of infective sporozoites to the insect's salivary glands. The sporozoites are inoculated into a new human host, and those not destroyed by the immune response are rapidly taken up by the liver. There they multiply inside hepatocytes as merozoites; this is pre-erythrocytic (or hepatic) sporogeny. After a few days, the infected hepatocytes rupture, releasing merozoites into the blood, from where they are rapidly taken up by erythrocytes. In the case of *P. vivax* and *P. ovale*, a few parasites remain dormant in the liver as hypnozoites. These may reactivate at any time subsequently, causing relapsing infection.

Inside the red blood cell, the parasites again multiply, changing from merozoite to trophozoite to schizont and finally appearing as 8 to 24 new merozoites. The erythrocyte ruptures, releasing the merozoites to infect further cells. Each cycle of this process, which is called erythrocytic schizogony, takes about 48 hours in *P. falciparum*, *P. vivax*, and *P. ovale* disease and about 72 hours in *P. malariae* disease. *P. knowlesi* has the shortest cycle at 24 hours. *P. vivax* and *P. ovale* mainly attack reticulocytes and young erythrocytes, and *P. malariae* tends to attack older cells; *P. falciparum* parasitizes any stage of erythrocyte.

A few merozoites develop not into trophozoites but into gametocytes. These are not released from the red cells until taken up by a feeding mosquito to complete the life cycle.

Pathogenesis

The pathology of malaria is related to anaemia; cytokine release; and in the case of *P. falciparum*, widespread organ damage because of impaired microcirculation. The anaemia seen in malaria is multifactorial (Box 42.56). In *P. falciparum* malaria, red blood cells containing schizonts adhere to the lining of capillaries in the brain, kidneys, gut, liver, and other organs. As well as causing mechanical obstruction, these schizonts rupture, releasing toxins and stimulating further cytokine release.

After repeated infections, partial immunity develops, allowing the host to tolerate parasitaemia with minimal ill effects. This immunity is largely lost if there is no further infection for a few years. Certain genetic traits also confer some immunity to malaria. People who lack the Duffy antigen on the red blood cell membrane (a common finding in West Africa) are not susceptible to infection with *P. vivax*. Certain haemoglobinopathies (including sickle cell trait) also give some protection against the severe effects of malaria, explaining the persistence of these otherwise harmful mutations in tropical countries. Iron deficiency may also have some protective effect. The spleen appears to play a role in controlling infection, and splenectomized people are at risk of overwhelming malaria. Pregnancy increases susceptibility to severe disease.

Clinical features

Typical malaria is seen in non-immune individuals. This includes children in any area, adults in areas of unstable transmission, and any visitors from a non-malarious region.

The normal incubation period is 10 to 21 days, but it can be longer. The most common symptom is fever, although malaria may present initially with general malaise, headache, vomiting, or diarrhoea. At first the fever may be continual or erratic; the classical tertian or quartan fever appears only after some days. The temperature often reaches 41°C and is accompanied by rigors and drenching sweats.

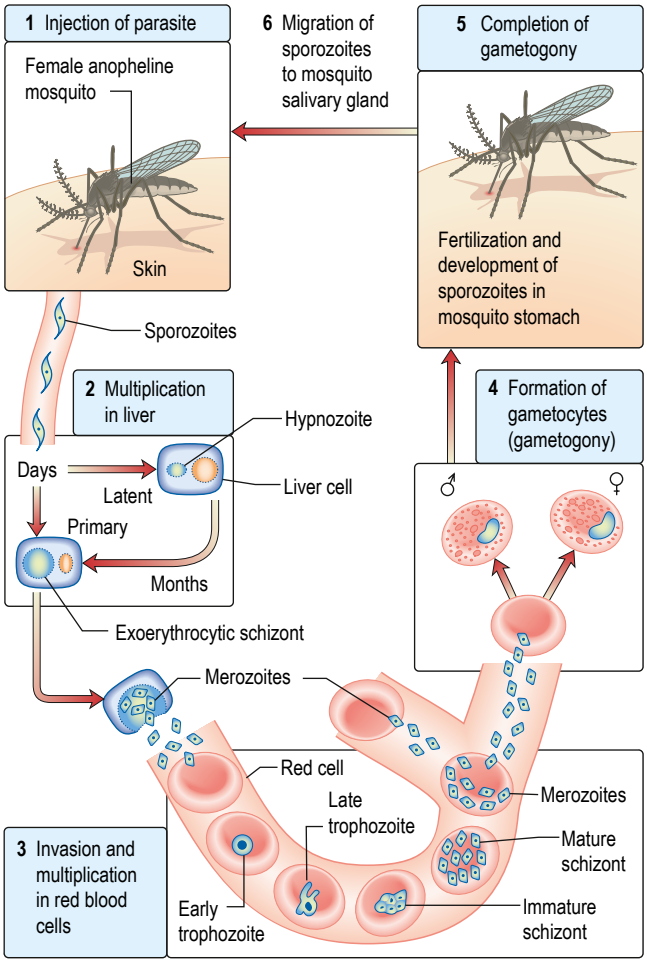


Fig. 42.29 **A schematic life cycle of *Plasmodium vivax*.**

Box 42.56 Causes of anaemia in malaria infection

- Haemolysis of infected red blood cells
- Haemolysis of non-infected red cells (blackwater fever)
- Dyserythropoiesis
- Splenomegaly and sequestration
- Folate depletion
- Haemolysis after intravenous artesunate treatment (probably caused by decreased erythrocyte lifespan)

The illness caused by ***P. vivax*** **or** ***P. ovale*** infection is usually relatively mild (although *P. vivax* can occasionally cause severe disease). Anaemia develops slowly, and there may be tender hepatosplenomegaly. Spontaneous recovery usually occurs within 2 to 6 weeks, but hypnozoites in the liver can cause relapses for many years after infection. Repeated infections often cause chronic ill health because of anaemia and hyper-reactive malarial splenomegaly.

P. malariae infection also causes a relatively mild illness but tends to run a more chronic course. Parasitaemia may persist for years, with or without symptoms. In children, *P. malariae* infection is associated with glomerulonephritis and nephrotic syndrome.

P. knowlesi is primarily a parasite of macaque monkeys but causes human malaria in certain areas of South-east Asia. The clinical picture is very variable, ranging from a chronic relatively mild illness similar to *P. malariae*, through to fulminant and fatal disease.

P. falciparum infection causes, in many cases, a self-limiting illness similar to the other types of malaria, although the paroxysms of fever are usually less marked. However, it may also cause serious complications (Fig. 42.30), and the vast majority of malaria deaths are caused by *P. falciparum*. Patients can deteriorate rapidly, and children in particular progress from reasonable health to coma and death within hours. A high parasitaemia (>10% of red blood cells infected; >2% is used as a cut-off in non-endemic settings) is an indicator of severe disease, although patients with apparently low parasite levels may also develop complications (peripheral parasite counts can be misleading because many parasitized cells are sequestered in the microcirculation). Cerebral malaria is marked by diminished consciousness, confusion, and convulsions, often progressing to coma and death. Untreated, it is universally fatal. Blackwater fever is caused by widespread intravascular haemolysis, affecting both parasitized and unparasitized red cells, giving rise to dark urine. In children, concurrent bacteraemia is common.

Hyper-reactive malarial splenomegaly (tropical splenomegaly syndrome) is seen in older children and adults in areas where malaria is hyperendemic. It is associated with an exaggerated immune response to repeated malaria infections and is characterized by anaemia, massive splenomegaly, elevated IgM levels, and anti-plasmodium antibodies. Malaria parasites are scanty or absent, but PCR increases parasitological yield. Traditionally treated with prolonged courses of prophylactic antimalarials where patients are in a setting of persistent re-exposure, a single treatment course is effective in non-endemic settings.

Diagnosis

Malaria should be in the differential diagnosis of anyone who presents with a febrile illness in, or having recently left, a malarious area. *Falciparum* malaria is unlikely to present more than 3 months after exposure even if the patient has been taking prophylaxis, but *vivax* malaria may cause symptoms for the first time up to 1 year after leaving a malarious area.

Diagnosis is usually made by identifying parasites on a Giemsa-stained thick blood film (with thin films subsequently used to speciate the parasite and determine parasitaemia). Rapid antigen detection tests (RDTs) are increasingly used for near-patient diagnosis or where laboratory staff with sufficient competence in microscopy are unavailable. Repeat films (often three are recommended) should be examined if the diagnosis remains a concern despite initial negative test results because both RDTs and thick films have insufficient sensitivity to rule out infection. In many endemic areas, malaria is over-diagnosed on clinical grounds (with the risk that the true cause of the febrile illness is missed), and a definite diagnosis should be made wherever possible. There is an increasing prevalence of parasites with a gene deletion in the histidine-rich protein 2 (pfhrp2), which leads to a false-negative result when using RDTs as a result of selection pressure through widespread use of these tests, raising concerns with regard to their sensitivity. Serological tests are of no diagnostic value, but PCR and other molecular amplification techniques are gaining traction.

Parasitaemia is common in endemic areas, and the presence of parasites does not necessarily mean that malaria is the cause of the patient's symptoms. Further investigation, including a lumbar puncture, may be needed to exclude other infections, and presumptive antibiotics are recommended in children with confirmed *P. falciparum* infection who are septic or comatose.

Management

Uncomplicated malaria

Chloroquine is still widely used to treat patients with non-*falciparum* malaria (Box 42.57), and in many areas, it remains effective. However,

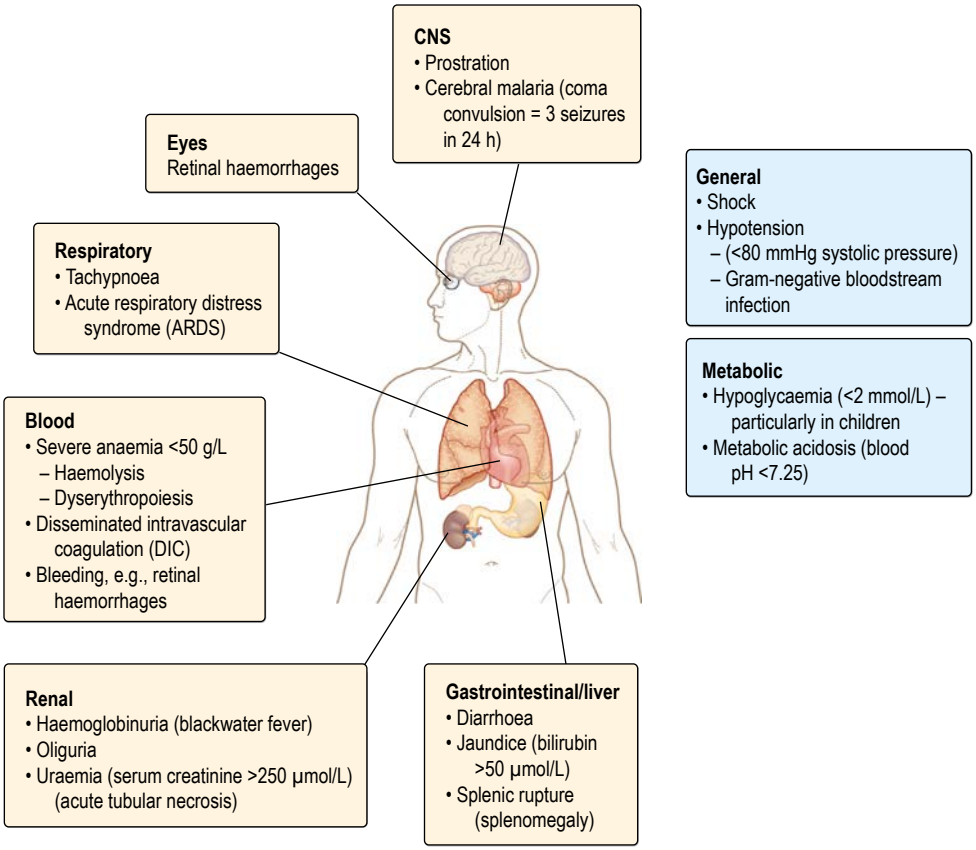


Fig. 42.30 Features of severe *falciparum* malaria. *CNS,* Central nervous system.

Box 42.57 Drug treatment of uncomplicated malaria

Parasite	Drug	Regimen	Plus	Regimen
Plasmodium vivax *Plasmodium ovale* *Plasmodium malariae*	ACT	3 days	Primaquine	0.25–0.5 mg/kg per day for 2–3 weeks
Plasmodium falciparum (adults, endemic zone)	ACT	3 days		
	or (if not available)			
	Quinine + doxycycline	7 days		
P. falciparum (pregnancy)	ACT	3 days		
P. falciparum (infants)	ACT	3 days; appropriate dose for body weight		
P. falciparum (returning travellers)	ACT or atovaquone + proguanil *or* quinine + doxycycline	3 days		

there is increasing resistance to chloroquine in some strains of *P. vivax*, and co-infection with *P. falciparum* is common in some parts of the world. It is therefore sensible to use oral artemisinin combination therapy, where available, for all cases of malaria. After successful treatment of a patient with *P. vivax* or *P. ovale* malaria, it is necessary to give a 2- to 3-week course of primaquine (0.25–0.5 mg/day) to eradicate the hepatic hypnozoites and prevent relapse ('radical cure'). This drug can precipitate haemolysis in patients with glucose-6-phosphate dehydrogenase deficiency, which should be excluded before administration (see Chapter 38, 'Glucose-6-phosphate dehydrogenase deficiency' section). Tafenoquine given as a single dose (developed for its long half-life) is also licensed in some countries.

The artemisinin-based drugs are the most effective treatment for both uncomplicated and severe infections with *P. falciparum* in adults and in children. Artemisinin-based combination therapy (ACT) is the recommended oral treatment for uncomplicated *P. falciparum* malaria worldwide, in which the artemisinin compound is co-administered with a long half-life partner drug. These drugs are now widely available and standard of care globally, partly through the efforts of the Global Fund. Several different fixed-dose combinations are recommended by the WHO (**Box 42.58**); the choice

Box 42.58 Suitable artemisinin combination therapies for malaria

Drugs	Regimen
Fixed-dose combination tablets available	
Artemether–lumefantrine	4 tablets twice daily for 3 days[a]
Artesunate–amodiaquine	4 mg/kg per day of artesunate for 3 days
Dihydroartemisinin–piperaquine	4 mg/kg per day of dihydroartemisinin for 3 days
Available as fixed-dose, co-packaged, separate tablets	
Artesunate–mefloquine[b]	4 mg/kg per day of artesunate for 3 days
Artesunate–sulfadoxine–pyrimethamine[c]	4 mg/kg per day of artesunate for 3 days
Alternatives where no combination packages are available	
Artesunate + clindamycin	2 mg/kg per day + 10 mg/kg twice daily for 7 days
Artesunate + doxycycline	2 mg/kg per day + 3.5 mg/kg per day for 7 days

[a]Adult dose: reduce dose by body weight for children.
[b]Combination tablet manufactured but not widely available.
[c]Not suitable for *Plasmodium vivax* or mixed infections. Different fixed-dose combinations are available.

Box 42.59 Drug treatment of severe *falciparum* malaria in adults and children

Drug and route	Immediate dose	Subsequent doses
Severe malaria is an emergency: after rapid assessment and confirmation of diagnosis, if possible, treatment should be started with whatever parenteral treatment is available. The options, in order of preference, are:		
1. i.v. artesunate	2.4 mg/kg	2.4 mg/kg at 12 and 24 h, then daily (up to 7 days)
2. i.v. quinine	20 mg/kg[a]	10 mg/kg 8-hourly (up to 7 days)
3. i.m. artesunate	2.4 mg/kg	2.4 mg/kg at 12 and 24 h, then daily (up to 7 days)
4. i.m. artemether	3.2 mg/kg	1.6 mg/kg daily
5. Rectal artesunate	10 mg/kg	Transfer to centre where parenteral therapy available
Continue parenteral treatment for at least 24 hours regardless of improvement in condition. After this, if the patient is improving, switch to oral therapy to complete 7 days with:		
ACT		
Or		
Quinine + doxycycline		

[a]10 mg/kg if patient has already received oral quinine or mefloquine.
ACT, Artemisinin-based combination therapy; *i.m.,* intramuscular; *i.v.,* intravenous.

should be based on local resistance to the 'partner'. To limit the development of resistance, artemisinin derivatives should never be given as monotherapy. Partial resistance to artemesinins is already widespread in parts of South-east Asia and has emerged in India, South America, and *de novo* in East Africa. As long as combination therapy is used and the parasite is sensitive to the partner drug, ACT will still be effective. Longer courses of artemether may also overcome the partial resistance, and in areas along the Mekong river in South-east Asia, triple ACT is now required.

In addition to treating the acute infection, the WHO recommends that a single dose of primaquine should be given as a gametocide to decrease onward transmission.

Severe *P. falciparum* malaria

Severe malaria, indicated by the presence of any of the complications discussed earlier or a parasite count of more than 10% (lower thresholds, e.g., 2%, are used in non-immune patients) is a medical emergency (**Box 42.59**). Anyone involved in managing patients with malaria should be familiar with the latest WHO guidelines.

- Intravenous artesunate has replaced intravenous quinine because it is more effective, and clinicians should ensure its availability in all acute hospital settings. It is now recommended in all trimesters of pregnancy.
- Intensive care facilities may be needed, including mechanical ventilation and dialysis.
- Patients with severe anaemia may require transfusion.
- Careful monitoring of fluid balance is essential; both pulmonary oedema and pre-renal failure are common.
- Hypoglycaemia can be induced both by the infection itself and by quinine treatment, when used.
- Superadded bacterial infection is common, particularly in children.

Prevention and control

Malaria is a priority for the WHO, which announced its 'Roll Back Malaria' campaign in 1998. This has had considerable success, and the global malaria-specific mortality rate decreased by nearly 40% between 2000 and 2019. However, progress has stalled, with a deteriorating picture through the COVID pandemic. Although some countries either have eradicated malaria or plan to do so soon, for much of Africa, the aim remains control rather than eradication. Recent WHO recommendations have been widely endorsed and supported by governments and non-governmental organizations (**Box 42.60**).

As with many vector-borne diseases, control of malaria relies on a combination of case treatment, vector eradication, and personal protection from vector bites, such as that provided by insecticide (permethrin)-treated nets. Mosquito eradication is usually achieved by a combination of insecticide use (e.g., house spraying with DDT) and manipulation of the habitat (e.g., marsh drainage). Alongside these elements, there has been renewed interest in seasonal and perennial malaria chemoprevention, in which children in endemic areas are given monthly doses of antimalarials during the rainy season or year-round, respectively. This has been shown to reduce the incidence of malaria but may lead to an increased risk of drug resistance. Safe and effective intermittent chemoprevention is now recommended during pregnancy and in school-age children because of the greater risk of complications. Post-discharge malaria chemoprevention, given for 6 months to children found to be severely anaemic for any reason, is recommended in high-transmission settings to reduce subsequent deaths and hospitalizations. Mass drug administration is recommended in emergency settings and in areas of moderate to high transmission, as well as in settings with very low to low transmission and to reduce *P. vivax* transmission.

Enormous effort (and resource) has been devoted to the search for a malaria vaccine, and significant progress has been made. In October 2021, the WHO recommended the RTS,S/AS01 malaria vaccine for the prevention of *P. falciparum* malaria in children living in regions with moderate to high transmission (as defined by the WHO). After further phase 3 trials demonstrating 75% efficacy, in 2022 WHO issued prequalification approval for the RTS,S vaccine, and UNICEF announced that it had secured 18 million doses over

Box 42.60 Key elements in the strategy for reducing the global burden of malaria by 90%[a]

- Vector control
  - Long-lasting insecticide-treated bed nets
  - Indoor residual spraying
- Adequate entomological surveillance and monitoring
  - Periodic assessment of vector species, abundance, seasonality, and behaviour
  - Insecticide susceptibility status and underlying resistance mechanisms
  - Routine monitoring of coverage and impact of interventions, including use and condition of bed nets
- Managing insecticide resistance and residual transmission
  - Use insecticides with different modes of action
  - Forecast vector control product requirements and support pooled procurement
- Expanding preventive treatment to prevent disease in the most vulnerable groups
  - Intermittent preventive treatment of pregnant females
  - Intermittent preventive treatment of infants and children
  - Seasonal chemoprevention for children aged younger than 5 years
  - Protecting all non-immune travellers and migrants
- Ensuring universal diagnostic testing of all suspected malaria cases
  - All suspected cases to have malaria should have the diagnosis confirmed by quality-assured microscopy or a rapid diagnostic test before treatment
  - Track and report all cases
- Providing quality-assured treatment to all patients
- Scaling up community-based diagnostic testing and treatment
- Monitoring safety and efficacy of antimalarial medicines and managing antimalarial drug resistance
- Eliminating *falciparum* malaria from the Greater Mekong subregion where artesunate resistance is endemic
- Removing all inappropriate antimalarial medicines from markets

[a]These strategies include ensuring universal access to malaria prevention, diagnosis, and treatment.

Box 42.61 Malaria prophylaxis for adult travellers

Area visited	Advice or prophylaxis
Countries or regions of low risk or certain higher risk areas out of malaria risk season	Bite avoidance measures advised
Countries or regions of significant malaria risk	Mefloquine 250 mg, one tablet weekly *or* Atovaquone (250 mg) plus Proguanil (100 mg) combination tablet, one tablet daily *or* Doxycycline 100 mg, one tablet daily
Thai–Myanmar border (high levels of mefloquine resistance)	Atovaquone (250 mg) plus Proguanil (100 mg) combination tablet, one tablet daily *or* Doxycycline 100 mg, one tablet daily

the next 3 years (2023–2025). It is now recommended in a four- or five-dose schedule to all children from 5 months of age living in regions with moderate or high transmission.

Non-immune travellers to malarious areas should take measures to avoid insect bites, such as using a repellent (diethyltoluamide [DEET] 20%–50% in lotions and sprays) and sleeping under a mosquito net. Antimalarial prophylaxis should also be taken in most cases, although this is never 100% effective (**Box 42.61**). The precise choice of prophylactic regimen depends both on the individual traveller and on the specific itinerary; further details can be found in national formularies or travel advice centres.

Trypanosomiasis

African trypanosomiasis (sleeping sickness)

Sleeping sickness is caused by trypanosomes transmitted to humans by the bite of the tsetse fly (genus *Glossina*), although mother-to-child transmission has also occasionally been reported. It is endemic in a belt across sub-Saharan Africa, extending to about 14 degrees north and 20 degrees south; this marks the natural range of the tsetse fly. Two subspecies of trypanosome cause human sleeping sickness: *Trypanosoma brucei gambiense* ('West African trypanosomiasis') and *Trypanosoma brucei rhodesiense* ('East African trypanosomiasis').

Epidemiology

Sleeping sickness caused by *T. b. gambiense* accounts for 97% of reported new cases. It is currently found in 24 countries in an area stretching from Uganda in Central Africa, west to Cameroon, and south as far as Angola; however, the Democratic Republic of Congo (with 61% cases) and a few other Central African states are now the only countries still reporting significant numbers. Humans are the major reservoir and infection is transmitted by riverine *Glossina* spp. (e.g., *Glossina palpalis*).

Sleeping sickness caused by *T. b. rhodesiense* occurs in 13 countries in East and Central Africa, but cases are rare and sporadic. It is a zoonosis of both wild and domestic animals. In endemic situations, it is maintained in game animals and transmitted by savannah flies such as *Glossina morsitans*. Previous epidemics have been related to cattle and the vectors were riverine flies.

Despite disruption to established treatment and control programmes related to repeated political upheaval and armed conflict in Central Africa, numbers have reduced greatly, with fewer than 1000 cases in 2022 compared with 500 000 in 1997. The WHO believes that it is on course to eliminate African trypanosomiasis (zero cases) by 2030.

Parasitology

Tsetse flies bite during the day, and unlike most arthropod vectors, both males and females take blood meals. An infected insect may deposit metacyclic trypomastigotes (the infective form of the parasite) into the subcutaneous tissue. These cause local inflammation ('trypanosomal chancre') and regional lymphadenopathy. Within 2 to 3 weeks, the organisms invade the bloodstream, subsequently spreading to all parts of the body, including the brain.

Clinical features

T. b. gambiense causes a chronic, slowly progressive illness. Episodes of fever and lymphadenopathy occur over months or years, and hepatosplenomegaly may develop. Eventually, infection reaches the CNS, causing headache, behavioural changes, confusion, and daytime somnolence. As the disease progresses, patients may develop tremors, ataxia, convulsions, and hemiplegias; eventually, coma and death supervene. Histologically, there is a lymphocytic meningoencephalitis, with scattered trypanosomes visible in the brain substance.

T. b. rhodesiense sleeping sickness is a much more acute disease. Early systemic features may include myocarditis, hepatitis, and serous effusions, and patients can die before the onset of CNS disease. If they survive, cerebral involvement occurs within weeks of infection and is rapidly progressive.

Diagnosis

Diagnosis involves three stages:

- Screening using serological tests (only available for *T. b. gambiense*) and clinical signs
- Confirmation of the presence of parasites in body fluids
- Assessment of stage of disease

The card agglutination test for trypanosomiasis is a robust and easy-to-use field assay for screening. If the result is positive (or there are clinical signs of disease), Giemsa-stained smears of thick or thin blood films or of lymph node aspirate should be examined for parasites. Blood films are usually positive in *T. b. rhodesiense* infection but are often negative in *T. b. gambiense* disease; concentration techniques may increase the yield. The main investigation for staging disease is lumbar puncture; examination of CSF is essential in patients with evidence of trypanosomal infection. CNS involvement causes lymphocytosis and elevated protein in the CSF, and parasites may be seen in concentrated specimens. Suramin should be given to patients with *T. b. rhodesiense* parasitaemia before lumbar puncture to avoid inoculation into the CSF.

Management

The treatment of patients with sleeping sickness is improving with better drugs now available for *T. b. gambiense* infection and ongoing clinical trials to define their use in *T. b. rhodesiense*. In both forms of disease, treatment is usually effective if given in the first stage (i.e., before the onset of CNS involvement) (**Box 42.62**). The treatment of choice for those with second-stage (CNS) disease in *T. b. gambiense* is now oral fexinidazole, although a combination of eflornithine and nifurtimox is effective and still recommended in young children or those with severe disease. Fexinidazole is also now indicated for all stages of *T.b.rhodesiense*, replacing melarsoprol for all except young children; melarsoprol is extremely toxic, causing an acute encephalopathy in up to 10% with a high mortality rate.

Control

Control programmes coordinated by the WHO have been effective in many areas. As in many vector-borne diseases, prevention depends largely on elimination, control, or avoidance of the vector.

South American trypanosomiasis (Chagas' disease)

Chagas' disease is widely distributed in rural areas of South and Central America, where up to 7 million people are infected. Since 2009, more cases have been detected in other parts of the world because of increased migration from Latin America. The disease is caused by *Trypanosoma cruzi*, which is transmitted to humans in the faeces of blood-sucking reduviid bugs (also called cone-nose or assassin bugs). Faeces infected with *T. cruzi* trypomastigotes are rubbed in through skin abrasions, mucosa, or conjunctiva. The bugs, which live in mud or thatch buildings, feed on a variety of vertebrate hosts (e.g., rats, opossums) at night, defecating as they do so.

The parasites spread in the bloodstream before entering host cells and multiplying. Cell rupture releases them back into the circulation, where they can be taken up by a feeding bug. Further multiplication takes place in the insect gut, completing the trypanosome life cycle. Human infection can also occur via contaminated blood transfusion or, occasionally, by transplacental spread.

Box 42.62 Drugs used in the treatment of African trypanosomiasis

Disease stage	*Trypanosoma brucei gambiense*	*Trypanosoma brucei rhodesiense*
1	Fexnidazole; pentamidine if <6 yrs old or <20 kg	Fexinidazole (suramin if <6 yrs old or <20 kg)[a]
2 (central nervous system)	If not severe: fexinidazole If severe: eflornithine + nifurtimox or eflornithine monotherapy	Fexinidazole (melarsoprol if <6 yrs old or <20 kg)

[a]Severe allergic reactions common: give test dose.

Clinical features

Acute infection usually occurs in children and often passes unnoticed. A firm, reddish papule is sometimes seen at the site of entry, associated with regional lymphadenopathy. In the case of conjunctival infection, there is swelling of the eyelid, which may close the eye (Romaña's sign). There may be fever; lymphadenopathy; hepatosplenomegaly; and rarely, meningoencephalitis. Acute Chagas' disease is occasionally fatal in infants, but normally there is full recovery within a few weeks or months.

Some 10% to 30% of people go on to develop chronic Chagas' disease after a latent period of many years. The pathogenesis remains controversial, although the most likely explanation is an autoimmune response to persistent parasite DNA integrated into the host genome. The heart is commonly affected, with conduction abnormalities, arrhythmias, aneurysm formation, and cardiac dilation. GI involvement leads to progressive dilation of parts of the GI tract; this commonly results in mega-oesophagus (causing dysphagia and aspiration pneumonia) and megacolon (causing severe constipation).

Diagnosis

Trypanosomes may be seen on a stained blood film during the acute illness. PCR is also sensitive in the acute phase. The chronic phase is typically 'indeterminate', with positive serology but no evidence of organ involvement, which develops in 20% to 30% cases over time. Blood PCR results may be positive in a proportion.

Management and control

Nifurtimox and benznidazole are the main drugs used. Both are highly effective in acute infection, with a cure rate of more than 90% but much less so in those with chronic disease. They are relatively toxic, with adverse reactions in up to 40% of patients, and new drugs are urgently needed. Antiarrhythmic drugs and pacemakers may be needed in patients with cardiac disease, and surgical treatment is sometimes required for GI complications.

In the long term, prevention of Chagas' disease relies on improved housing and living conditions and eradication of vectors. Several coordinated multinational vector control programmes have been implemented, with some success in reducing transmission. Screening of blood donors and pregnancies in high-risk areas can reduce transmission through non-vector routes; reactivation should be considered in the differential diagnosis of patients with suspected opportunistic infection after major iatrogenic immunosuppression such as organ transplantation.

Leishmaniasis

This group of diseases is caused by protozoa of the genus *Leishmania*, which are transmitted by the bite of the female phlebotomine sandfly (**Box 42.63**). Leishmaniasis is seen in localized areas of Africa,

Box 42.63 *Leishmania* species causing visceral and cutaneous disease

Disease type	Species complex	Species
Visceral leishmaniasis	*Leishmania donovani*	*L. donovani* *Leishmania infantum* *Leishmania chagasi*
Cutaneous leishmaniasis	*Leishmania tropica*	*L. tropica*
	Leishmania major	*L. major*
	Leishmania aethiopica	*L. aethiopica*
	Leishmania mexicana	*L. mexicana* *Leishmania amazonensis* *Leishmania garnhami* *Leishmania pifanoi* *Leishmania venezuelensis*
	Leishmania viannia complex	*Leishmania braziliensis* *Leishmania guyanensis* *Leishmania panamanensis* *Leishmania peruviana*
Mucocutaneous leishmaniasis		*Leishmania viannia*

Asia (particularly India and Bangladesh), Europe, the Middle East, and South and Central America. Certain parasite species are specific to each geographical area. The clinical picture is dependent on the species of parasite and on the host's cell-mediated immune response. Asymptomatic infection, in which the parasite is suppressed or eradicated by a strong immune response, is common in endemic areas. Symptomatic infection may be confined to the skin (cutaneous leishmaniasis, sometimes with spread to the mucous membranes) or widely disseminated throughout the body (visceral leishmaniasis); the propensity to these different forms depends predominantly on leishmania subspecies. Relapse of previously asymptomatic infection is seen in patients such as those with HIV infection or increasingly those with iatrogenic immunosuppression.

In some areas, leishmaniasis is primarily zoonotic, whereas in others, humans are the main reservoirs of infection. In the vertebrate host, the parasites are found as oval amastigotes (Leishman–Donovan bodies). These multiply inside the macrophages and cells of the reticuloendothelial system and are then released into the circulation as the cells rupture. Parasites are taken into the gut of a feeding sandfly (genus *Phlebotomus* in the Old World; genus *Lutzomyia* in the New World), where they develop into the flagellate promastigote form. These migrate to the salivary glands of the insect, where they can be inoculated into a new host.

Visceral leishmaniasis

Clinical features

Visceral leishmaniasis (kala azar) is caused by *Leishmania donovani*, *Leishmania infantum*, or *Leishmania chagasi* and is prevalent in localized areas of Asia, Africa, the Mediterranean littoral, and South America. In parts of India, where humans are the main host, the disease occurs in epidemics. In most other areas, it is endemic, and it is mainly children and visitors to the area who are at risk. The main animal reservoirs in Europe and Asia are dogs and foxes, but in Africa, it is carried by various rodents. Between 50 000 and 100 000 new cases occur each year, the vast majority in Brazil, East Africa, and the Indian subcontinent. Mortality rates have fallen in Asia because of improved treatment programmes, but not in Africa, where conflicts have disrupted management initiatives.

The incubation period is usually 1 to 2 months but may be several years. The onset of symptoms is insidious, and the patient may feel quite well despite markedly abnormal physical findings. Fever is common and may be high and intermittent, although is usually low grade. The liver and especially the spleen become enlarged. The skin becomes rough and pigmented, resulting in the common name 'kala azar', which means 'black fever' in Hindi. If the disease is not treated, profound pancytopenia develops, and the patient becomes wasted and immunosuppressed. Death usually occurs within 1 year and is normally caused by bacterial infection or uncontrolled bleeding.

Diagnosis

The specific diagnosis is made by demonstrating the parasite in stained smears of aspirates of bone marrow, lymph node, spleen, or liver. The organism can also be cultured from these specimens, but PCR is more sensitive. Specific serological tests are positive in 95% of cases. Pancytopenia, hypoalbuminaemia, and hypergammaglobulinaemia are common.

Management

Intravenous liposomal amphotericin B (which may be curative as a single-dose treatment) is effective but not universally available. The most widely used drugs for visceral leishmaniasis remain the pentavalent antimony salts (e.g., sodium stibogluconate and meglumine antimoniate) despite toxicity and increasing resistance. The oral drug miltefosine has been shown in India to be highly effective and may be used especially in combination with liposomal amphotericin or with paromomycin; combination therapies are increasingly being used to shorten treatment courses and limit resistance. The recommended treatment regimens vary from region to region (e.g., miltefosine has been less effective in Africa and the Mediterranean).

Successful treatment may be followed in a small proportion of patients by a skin eruption called ***post-kala azar dermal leishmaniasis***. It starts as a macular or maculopapular nodular rash, which spreads over the body. It is most often seen in Sudan and India and is difficult to treat, although it may improve with miltefosine. It is a major driver for onward transmission caused by high parasite numbers in the skin.

HIV co-infection

Visceral leishmaniasis is strongly associated with HIV-related immunosuppression (see Chapter 44), and the two infections may be passed on together through injecting drug use. Anti-retroviral therapy has largely controlled the problem, but in regions with suboptimal viral suppression, there are increasing case numbers.

Cutaneous leishmaniasis

Cutaneous leishmaniasis is caused by a number of geographically localized species, which may be zoonotic or anthroponotic. It is by far the most common form of the disease, with up to 1 million new cases worldwide every year. Some 95% of these occur in just seven countries: Iran, Iraq, Syria, Algeria, Afghanistan, Brazil, and Colombia.

After a sandfly bite, *Leishmania* amastigotes multiply in dermal macrophages. The local response depends on the species of *Leishmania*, the size of the inoculum, and the host immune response. Single or multiple painless nodules appear 1 week to 3 months after the bite (Fig. 42.31). These enlarge and ulcerate with a characteristic erythematous raised border. An overlying crust may develop. The

Fig. 42.31 Hyperkeratotic nodular cutaneous leishmaniasis lesions and leishmaniasis recidivans (on the right cheek) caused by *Leishmania tropica*. (*From Bailey MS, Lockwood DNJ. Cutaneous leishmaniasis.* Clin Dermatol *2007; 25:203–211.)*

lesions heal slowly over months or years, sometimes leaving a disfiguring scar.

Leishmania major and *Leishmania tropica* are found in Russia and Eastern Europe, the Middle East, Central Asia, the Mediterranean littoral, and sub-Saharan Africa. Whereas the reservoir for *L. major* is desert rodents, *L. tropica* has a mainly urban distribution with dogs and humans as reservoirs. *Leishmania aethiopica* is found in the highlands of Ethiopia and Kenya, where the animal reservoir is the hyrax. The skin lesions caused by these species usually heal spontaneously with scarring; this may take 1 year or more in the case of *L. tropica*. Leishmaniasis recidivans is a rare chronic relapsing form caused by *L. tropica*.

Leishmania mexicana is found predominantly in Mexico, Guatemala, Brazil, Venezuela, and Panama; infection usually runs a benign course with spontaneous healing within 6 months. *Leishmania braziliensis* infections (which are seen throughout tropical South America) also usually heal spontaneously but may take longer.

L. mexicana amazonensis and *L. aethiopica* may occasionally cause diffuse cutaneous leishmaniasis. This is rare and is characterized by diffuse infiltration of the skin by Leishman–Donovan bodies. Visceral lesions are absent.

Diagnosis and management

The diagnosis can often be made clinically in a patient who has been in an endemic area. Giemsa stain on a split-skin smear demonstrates *Leishmania* parasites in 80% of cases. Biopsy tissue from the edge of the lesion can be examined histologically and parasites identified by PCR; culture is less often successful. The leishmanin skin test result is positive in more than 90% of cases but does not distinguish between active and resolved infection. Serology is unhelpful.

Small lesions usually require no treatment and resolve spontaneously. Large lesions and those in cosmetically sensitive sites can sometimes be treated locally by topical antiparasitic agents (e.g., intralesional injections of meglumine antimonate, or topical paromomycin cream). Curettage and cryotherapy are also effective. In other cases, systemic treatment with pentavalent antimonials, amphotericin, or triazole antifungal drugs may be needed. Speciation by PCR of biopsy material should be performed for all lesions acquired in the 'New World' to identify species that are associated with subsequent mucosalization; systemic treatment is recommended to prevent this complication.

Mucocutaneous leishmaniasis

Mucocutaneous leishmaniasis occurs in 3% to 10% of infections with *L. b. braziliensis* and is most common in Bolivia and Peru. It can also follow infection with *L. aethiopica*. The cutaneous sores are followed months or years later by indurated or ulcerating lesions affecting mucosa or cartilage, typically on the lips or nose ('espundia'). The condition can remain static, or there may be progression over months or years affecting the nasopharynx, uvula, palate, and upper airways.

Diagnosis and management

Biopsies usually show only very scanty organisms, although parasites can be detected by PCR; results of serological tests are frequently positive.

Liposomal amphotericin B is the treatment of choice, if available, although systemic antimonial compounds are widely used; miltefosine may also be effective. Relapses are common after treatment. Patients may die because of secondary bacterial infection or, occasionally, laryngeal obstruction.

Prevention of leishmaniasis relies on control of vectors or reservoirs of infection. Insecticide spraying, control of host animals, and treatment of infected humans may all be helpful. Personal protection against sandfly bites is also important, especially in travellers visiting endemic areas. Sandflies are poor fliers and sleeping off the ground helps prevent bites.

Other protozoal diseases of the blood and tissues

Toxoplasmosis

Toxoplasmosis is caused by the intracellular protozoan parasite *T. gondii*. The sexual form of the parasite lives in the gut of the definitive host, the cat, where it produces oocysts. After a period of maturing in the environment, these oocysts become the source of infection for secondary hosts, which may ingest them. In the secondary hosts (which include humans, cattle, sheep, pigs, rodents, and birds) there is disseminated infection. After a successful immune response, the infection is controlled, but dormant parasites remain encysted in host tissue for many years. The life cycle is completed when carnivorous felines eat infected animal tissue. Humans are infected either from contaminated cat faeces or undercooked infected meat; transplacental infection may also occur.

Clinical features

Toxoplasmosis is common: seroprevalence in adults in the UK is about 25%, rising to 90% in some parts of Europe. Most infections are asymptomatic or trivial. Symptomatic patients usually present with lymphadenopathy, mainly in the head and neck. There may be fever, myalgia, and general malaise; occasionally, there are more severe manifestations, including hepatitis, pneumonia, myocarditis, and choroidoretinitis. Lymphadenopathy and fatigue can sometimes persist for months after the initial infection.

Congenital toxoplasmosis may also be asymptomatic but can produce serious disease. Clinical manifestations include microcephaly, hydrocephalus, encephalitis, convulsions, and mental retardation. Choroidoretinitis is common; occasionally, this may be the only feature.

Immunocompromised patients, especially those with HIV infection, are at risk of serious infections with *T. gondii*. In acquired

immunodeficiency states, this is usually caused by reactivation of latent disease.

Diagnosis

Toxoplasma lymphadenitis is often suspected after characteristic histological findings ('microgranulomas') identified through urgent cancer pathways. Diagnosis of this and other acute presentations is usually confirmed serologically. IgG antibodies detectable by the Sabin–Feldman dye test remain positive for years; acute infection can be confirmed by demonstrating a high titre of specific IgM. In immunosuppressed patients with suspected reactivation, diagnosis (e.g., of cerebral toxoplasmosis) is made on the basis of typical radiological findings but may be supported by PCR, whose results are often positive in tissue or fluid (e.g., CSF).

Management

Acquired toxoplasmosis in an immunocompetent host rarely requires treatment. In those with severe disease (especially eye involvement), sulfadiazine and pyrimethamine are given for 4 weeks, along with folinic acid. Azithromycin may also be effective. The management of toxoplasmosis in pregnancy aims to decrease the risk of fetal complications. However, there is only weak evidence that giving spiramycin, either alone or in combination with sulfadiazine (which is the recommended treatment), has any significant effect on the frequency or severity of fetal damage. Infected infants should be treated from birth. The treatment of toxoplasmosis in HIV-positive patients is covered in Chapter 44 ('Toxoplasmosis' section).

Babesiosis

Babesiosis is a tick-borne parasitic disease, diagnosed most commonly in North America and Europe. It is a zoonosis of rodents and cattle and is occasionally transmitted to humans; infection is more common and more severe in those who are immunocompromised after splenectomy. The causative organisms are the *Plasmodium*-like *Babesia microti* (rodents) and *B. divergens* (cattle).

The incubation period averages 10 days. In patients with normal splenic function, the illness is usually mild. In splenectomized individuals, systemic symptoms are more pronounced, and haemolysis is associated with haemoglobinuria, jaundice, and acute kidney injury. Examination of a peripheral blood smear may reveal the characteristic *Plasmodium*-like organisms.

The standard treatment of patients with severe babesiosis is a combination of quinine 650 mg and clindamycin 600 mg orally three times daily for 7 days. Atovaquone and azithromycin is an alternative combination therapy.

Further reading

Alvarez-Rodriguez A, Jin B-K, Radwanska M, Magez S. Recent progress in diagnosis and treatment of Human African Trypanosomiasis has made the elimination of this disease a realistic target by 2030. *Front Med* (Lausanne) 2022; 9:1037094.
Bollani L, Auriti C, Achille C et al. Congenital toxoplasmosis: the state of the art. *Front Pediatr* 2022; 10:894573.
Porta E, Kalesh K, Steel P. Navigating drug repurposing for Chagas disease: advances, challenges, and opportunities. *Front Pharmacol* 2023; 14:1233253.
Rocha R, Pereira A, Maia C. Non-endemic leishmaniases reported globally in humans between 2000 and 2021-a comprehensive review. *Pathogens* 2022; 11(8):921.
Tripathi H, Bhalerao P, Singh S et al. Malaria therapeutics: are we close enough? *Parasit Vectors* 2023; 16(1):130.
World Health Organization. *Control of the Leishmaniases*. Geneva: WHO; 2010.
http://www.theglobalfund.org. *The Global Fund*.
http://www.who.int/leishmaniasis/resources/en/. *WHO Technical Report 949 Control of the Leishmaniases*.
http://www.who.int/malaria/publications/. *WHO guidelines for the treatment of malaria*.
http://www.wwarn.org. *Up-to-date information on resistance to all antimalarials*.

GASTROINTESTINAL PROTOZOA

The major GI protozoal parasites of humans are shown in **Box 42.64**.

Amoebiasis

Amoebiasis is caused by *E. histolytica*. There are two morphologically identical species of amoeba, which can be distinguished by molecular techniques.

- ***E. histolytica***, which is pathogenic
- ***Entamoeba dispar***, which is non-pathogenic.

Amoebiasis occurs worldwide, although much higher incidence rates are found in LMICs.

The organism exists both as a motile trophozoite and as a cyst that can survive outside the body. Cysts are transmitted by ingestion of contaminated food or water, or they can be spread directly by person-to-person contact, notably between males who have sex with males. Trophozoites emerge from the cysts in the small intestine and then pass on to the colon, where they multiply.

Clinical features

The long-held view that many individuals can carry the pathogen without obvious evidence of clinical disease (asymptomatic cyst passers) is based on misidentification of non-pathogenic *E. dispar*. True *E. histolytica* colonization and carriage is uncommon. In affected people, *E. histolytica* trophozoites invade the colonic epithelium using protozoal cytotoxins and proteolytic enzymes. The parasites continue to multiply and eventually cause frank ulceration of the mucosa. If penetration continues, trophozoites may enter the portal vein, via which they reach the liver and cause intrahepatic abscesses. This invasive form is serious and may even be fatal.

The incubation period of intestinal amoebiasis is highly variable and may be as short as a few days or as long as several months. In some patients, there may be mild chronic diarrhoea, but this may progress to bloody diarrhoea with mucus, sometimes accompanied by systemic symptoms such as headache, nausea, and anorexia. Acute amoebic dysentery, resembling a fulminant ulcerative colitis, may occur and is often misdiagnosed.

Complications are unusual but include toxic dilation of the colon with perforation, severe haemorrhage, amoeboma, and amoebic liver abscess. Amoebic liver abscesses often develop in the absence of a recent episode of colitis. Tender hepatomegaly, a high swinging fever, and profound malaise are characteristic, and early in the course of the disease, localizing symptoms may be minimal. The clinical features are described in more detail in Chapter 23 ('Liver abscess' section).

Box 42.64 Pathogenic human intestinal protozoa

Amoebae
- *Entamoeba histolytica*

Flagellates
- *Giardia intestinalis*

Ciliates
- *Balantidium coli*

Coccidia
- *Cyclospora cayetanensis*
- *Cryptosporidium parvum*
- *Isospora belli*
- *Sarcocystis* spp.

Microspora
- *Enterocytozoon bieneusi*
- *Encephalitozoon* spp.

Diagnosis

Microscopic examination of fresh stool or colonic exudate obtained at sigmoidoscopy (which is rarely performed but will demonstrate 'flask-shaped' ulceration) has been the standard means of diagnosing colonic amoebic infection. To confirm the diagnosis, motile trophozoites containing red blood cells must be identified; the presence of amoebic cysts alone does not imply disease. PCR techniques are also useful. Trophozoites may be seen in colonic biopsies but may be confused with host macrophages if the diagnosis is not suspected.

Results of ELISA-based serological tests are positive in at least 90% of patients with liver abscess and in 60% to 70% with active colitis.

Management

Metronidazole 800 mg three times daily for 5 days is given in amoebic colitis; a lower dosage (400 mg three times daily for 5 days) is usually adequate in liver abscess. Tinidazole is also effective. After treatment of the invasive disease, the bowel should be cleared of parasites with a luminal amoebicide such as paromomycin.

Prevention

Amoebiasis is difficult to eradicate because of the substantial human reservoir of infection. Progress will only be through improved standards of hygiene and sanitation and better access to clean water. Cysts are destroyed by boiling, but chlorine and iodine sterilizing tablets are not always effective.

Giardiasis

Giardia intestinalis is a flagellate (Fig. 42.32) that is found worldwide. It causes small intestinal disease, with diarrhoea and malabsorption. Prevalence is high in many LMICs, and it is the most common parasitic infection in travellers returning to the UK. In certain parts of Europe and some rural areas of North America, large water-borne epidemics have been reported. Person-to-person spread may occur in day nurseries and residential institutions and between males who have sex with males. Those with antibody deficient states such as selective IgA deficiency are at risk. The organism exists as both a trophozoite and a cyst, the latter being the form in which it is transmitted.

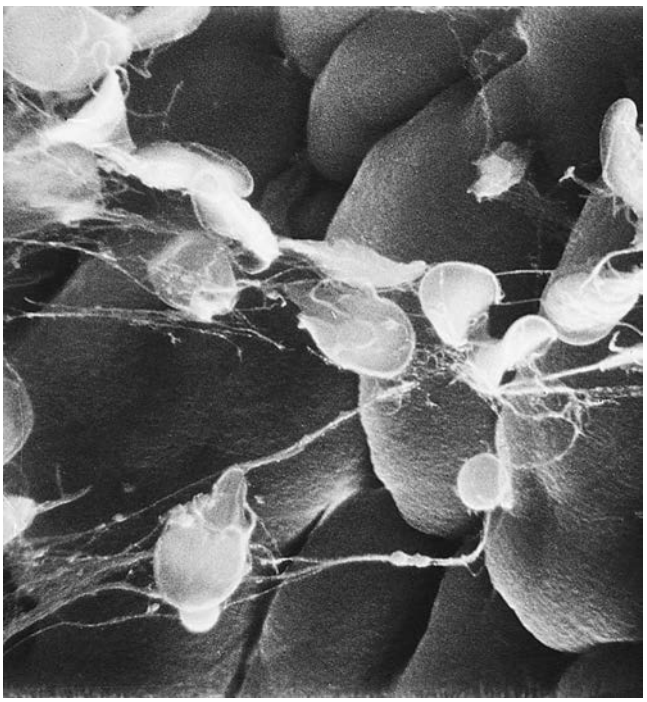


Fig. 42.32 *Giardia intestinalis* on small intestinal mucosa.
(Courtesy of Dr A Phillips.)

The organism sometimes colonizes the small intestine and may remain there without causing detriment to the host. In other cases, severe malabsorption may occur, which is thought to be related to morphological damage to the small intestine. The changes in architecture are usually mild partial villous atrophy; subtotal villous atrophy is rare. The mechanism by which the parasite causes alteration in mucosal architecture and produces diarrhoea and intestinal malabsorption is unknown, though there is evidence that the morphological damage is immune mediated. Bacterial overgrowth has also been found in association with giardiasis and may contribute to fat malabsorption.

Clinical features

Many individuals excreting *Giardia* cysts have no symptoms. Others become ill within 1 to 3 weeks of ingesting cysts; symptoms include diarrhoea, often watery in the early stage of the illness; nausea; anorexia; and abdominal discomfort and bloating. In most patients, these symptoms resolve after a few days, but in others they persist. Stools may then become paler, with the characteristic features of steatorrhoea. If the illness is prolonged, weight loss occurs and can be marked. Chronic giardiasis, frequently seen in low- and middle-income countries, can result in growth retardation in children.

Diagnosis

Both cysts and trophozoites can be found in the stool, but microscopy is inadequate because the parasite may be excreted at irregular intervals. Molecular stool techniques are much more sensitive. The parasite can also be seen in duodenal aspirates (obtained with either an endoscope or a luminal capsule) and in histological sections of jejunal mucosa.

Management

Tinidazole 2 g as a single dose, or metronidazole 400 mg three times daily for 5 to 7 days, will cure the majority of infections, although sometimes a second course is necessary or alternative agents may be required, such as tinidazole–albendazole combinations. Increasing numbers of patients are failing these treatments but can be cured with mepacrine. Immunoglobulin deficiency should be considered in recurrent infections, as well as screening of sexual contacts. Preventive measures are similar to those outlined earlier for *E. histolytica*.

Cryptosporidiosis

This organism is found worldwide, cattle being the major natural reservoir. It has also been demonstrated in supplies of drinking water in the UK. The parasite is able to reproduce both sexually and asexually; it is transmitted by oocysts excreted in the faeces.

In healthy individuals, cryptosporidiosis is a self-limiting illness. Acute watery diarrhoea is associated with fever and general malaise lasting for 7 to 10 days. In immunocompromised patients, especially those with HIV, diarrhoea is severe and intractable (see Chapter 44, 'Cryptosporidiosis' section). Cryptosporidium may also be an important cause of childhood diarrhoea in resource-poor countries, contributing to overall morbidity and mortality.

Diagnosis is usually made by faecal microscopy or PCR, although the parasite can also be detected in intestinal biopsies. Nitazoxanide may be of benefit but is ineffective in those with advanced HIV.

Balantidiasis

Balantidioides coli is the only ciliate that produces clinically significant infection in humans. It is found throughout the tropics, particularly in Central and South America, Iran, Papua New Guinea, and the Philippines. It is usually carried by pigs, and infection is most common in communities that live in close association with swine. Its life cycle is identical to that of *E. histolytica*. *B. coli* causes diarrhoea and sometimes a dysenteric illness, with invasion of the distal ileal and colonic mucosa. Trophozoites rather than cysts are found in the stool. Treatment is with tetracycline or metronidazole.

Blastocystis hominis infection

B. hominis is a strictly anaerobic protozoal organism that inhabits the colon. The pathogenicity for humans remains controversial. Many people appear to excrete the organism with no symptoms at all. There are numerous reports of people with diarrhoea whose symptoms improve after eradication of the parasite, but this could be a spillover effect from, for example, treatment of undiagnosed *Giardia* or alteration of the microbiome. Increasingly evidence suggests that at least some subspecies are a marker of a healthy and diverse microbiota. Treatment is very rarely recommended.

Cyclospora cayetanensis infection

Cyclospora cayetanensis, a coccidian protozoal parasite, infects the epithelium of the small intestine and presents with diarrhoea, usually watery and sometimes chronic. It is endemic in many LMICs, but large-scale outbreaks are also well described in the USA and Europe, associated with contaminated food, both imported and locally grown. Without specific microscopy techniques, the diagnosis may be missed. Infection is usually self-limiting but can be treated with co-trimoxazole.

Microsporidiosis

Protozoa of the phylum Microsporea can cause diarrhoea in patients with HIV and AIDS (see Chapter 44).

Free-living amoebae

Some amoebae, ubiquitous in the environment, can cause severe infections in humans. *Balamuthia mandrillaris*, *Naegleria fowleri*, and various *Acanthamoeba* species can cause a meningoencephalitis. Often the route of infection is across the cribriform plate after swimming or diving in freshwater or from nasal douching in people with pre-existing disruption of the skull base or in normal hosts. Some are associated with skin lesions, and some can involve other organs, especially in immunocompromised hosts. Outcomes of CNS involvement are almost always fatal. *Acanthamoeba* keratitis is a sight-threatening condition seen in contact lens wearers.

Further reading

Haston JC, Cope JR. Amebic encephalitis and menongoencehphalitis: an update on epidemiology, diagnostic methods, and treatment. *Curr Opin Infect Dis* 2023; 36(3):186–191.

Rojas-Velazquez L, Moran P, Serrano-Vazquez A et al. The regulatory function of *Blastocystis* spp. on the immune inflammatory response in the gut microbiome. *Front Cell Infect Microbiol* 2022; 12:967724.

Santos H, Rebello K. An overview of mucosa-associated protozoa: challenges in chemotherapy and future perspectives. *Front Cell Infect Microbiol* 2022; 12:860442.

HELMINTH INFECTIONS

Worm infections are very common in rural areas of low- and middle-income countries, causing much disease in both humans and domestic animals. Worms are frequently imported into industrialized countries. The most common human helminth infections are listed in **Box 42.65**. Eight of the 20 'neglected tropical diseases' identified by the WHO as causing major morbidity in the poorest areas are caused by helminths. Even parasites that, in high-income countries, are regarded as little more than an inconvenience (e.g., roundworm or whipworm) cause huge morbidity in resource-poor settings because of the very high worm burden in children.

Helminths are the largest internal human parasites. They reproduce sexually, generating millions of eggs or larvae. Whereas ***nematodes*** and ***trematodes*** have a mouth and intestinal tract, ***cestodes*** absorb nutrients directly through the outer tegument. All worms are motile, but when the adults are established in their definitive site, they rarely migrate farther. In most cases, adult helminths are long-lived, for example, up to 30 years in the case of schistosomes.

Many helminths have developed complex life cycles, involving more than one host. Both primary and intermediate hosts

Box 42.65 Helminths commonly infecting humans

	Helminth	Common name or disease caused
Nematodes (roundworms)		
Tissue-dwelling worms	*Wuchereria bancrofti*	Filariasis
	Brugia malayi and *Brugia timori*	Filariasis
	Loa loa	Loiasis
	Onchocerca volvulus	River blindness
	Dracunculus medinensis	Dracunculiasis
	Mansonella perstans	Mansonellosis
Intestinal human nematodes	*Enterobius vermicularis*	Threadworm
	Ascaris lumbricoides	Roundworm
	Trichuris trichiura	Whipworm
	Necator americanus	Hookworm
	Ancylostoma duodenale	Hookworm
	Strongyloides stercoralis	Strongyloidosis
Zoonotic nematodes	*Toxocara canis*	Toxocariasis
	Trichinella spiralis	Trichinellosis
Trematodes (flukes)		
Blood flukes	*Schistosoma* spp.	Schistosomiasis
Lung flukes	*Paragonimus* spp.	Paragonimiasis
Intestinal or hepatic flukes	*Fasciolopsis buski*	
	Fasciola hepatica	
	Clonorchis sinensis	
	Opisthorchis felineus	
Cestodes (tapeworms)		
Intestinal adult worms	*Taenia saginata*	Beef tapeworm
	Taenia solium	Pork tapeworm
	Diphyllobothrium latum	Fish tapeworm
	Hymenolepis nana	Dwarf tapeworm
Larval tissue cysts	*Taenia solium*	Cysticercosis
	Echinococcus granulosus	Hydatid disease
	Echinococcus multilocularis	Hydatid disease
	Spirometra mansoni	Sparganosis

are often highly specific to a particular species of worm. In some cases of human infection, humans are the primary host, but in others, humans are a non-specific intermediary or are coincidentally infected. Multiple infections with different helminths are common in endemic areas and can lead to significant morbidity. The WHO recommends regular (usually annual) mass drug treatment for all at-risk people living where soil-transmitted helminths are common. This includes children from the age of 1 year and older, females of reproductive age, and adults in high-risk occupations such as mining and tea picking. There is, however, a lively debate about the strength of evidence supporting this guidance (Box 42.66).

NEMATODES

Tissue-dwelling worms

Filariasis

Several nematodes belonging to the superfamily Filarioidea can infect humans (Box 42.67). The adult worms are long and thread-like, ranging from 2 cm to 50 cm in length; females are generally much larger than males. Larval stages are inoculated into humans by various species of biting flies, each specific to a particular parasite. The adult worms that develop from these larvae mate and produce millions of offspring called microfilariae, which migrate through the blood or skin. These are ingested by feeding flies, in which the remainder of the life cycle takes place. Disease, which may be caused either by the adult worms or by microfilariae, stems from the host immune response to the parasite and is generally characterized by massive eosinophilia. Adult worms are long-lived (10–15 years), and re-infection is common, so disease tends to be chronic and progressive.

Box 42.66 Drugs used in mass treatment of helminth infections

Drug	Infection
Diethylcarbamazine (DEC)	Loiaisis Filariasis
Ivermectin	Loiaisis Filariasis Onchocerciasis Strongyloidiasis
Albendazole	Filariasis (with DEC) Intestinal helminths
Praziquantel	Schistosomiasis

It was Patrick Manson, working on filariasis in 1870s Formosa (now Taiwan), who was the first person to demonstrate a link between a human infectious disease and an arthropod vector. Having demonstrated that filaria were only present in human blood at night, he next conducted a series of experiments on his gardener, Hin Lo, who was infected with filaria. Manson got mosquitoes to feed on Hin Lo while he slept. Then, by dissecting the engorged mosquitoes, he demonstrated the presence of filaria under the microscope in the ones that had taken a blood meal. Sir Patrick Manson later became known as the *father of tropical medicine* but has recently attracted attention for his role in colonial medicine.

Lymphatic filariasis

Lymphatic filariasis may be caused by different species of filarial worm and has a scattered distribution in the tropics and subtropics. More than 120 million people are estimated to be infected, and almost 1 billion are at risk. *Wuchereria bancrofti* is transmitted to humans by *Culex* mosquitoes. Adult female worms are 5 to 10 cm long and live in the lymphatic vessels, disrupting the normal function of the lymphatic system and releasing large numbers of microfilariae into the blood. This generally occurs at night, coinciding with the nocturnal feeding pattern of *Culex* mosquitoes, so 'night bloods' may need to be taken to confirm a diagnosis. Note that non-periodic forms of *W. bancrofti*, transmitted by day-biting species of mosquito, are found in the South Pacific. *Brugia malayi*, and the closely related *Brugia timori*, is very similar to *W. bancrofti* and exhibits the same nocturnal periodicity. The usual vectors are mosquitoes of the genus *Mansonia*.

Many filarial worms coexist with symbiotic *Wolbachia* bacteria, which are, in themselves, a cause of inflammation in the human host and a target for adjunctive treatment.

Clinical features

After the bite of an infected mosquito, the larvae enter the lymphatics and are carried to regional lymph nodes. There, they grow and mature for 6 to 18 months.

Adult worms produce allergic lymphangitis. The clinical picture depends on the individual immune response, which in turn may depend on factors such as age at first exposure. In endemic areas, many people have asymptomatic infection. Sometimes early infection is marked by bouts of fever accompanied by pain, tenderness, and erythema along the course of affected lymphatics. Involvement of the spermatic cord and epididymis is common in Bancroftian filariasis. These acute attacks subside spontaneously in a few days but usually recur. Recurrent episodes cause intermittent lymphatic obstruction, which in time can become fibrotic and irreversible.

Box 42.67 Diseases caused by the filarial worms

Organism	Adult worm	Microfilariae	Major vector	Clinical signs	Distribution
Wuchereria bancrofti	Lymphatics	Blood	*Culex* spp.	Fever Lymphangitis Elephantiasis	Tropics
Brugia timori* and *Brugia malayi	Lymphatics	Blood	*Mansonia* spp.	Fever Lymphangitis Elephantiasis	East and South-east Asia, South India, Sri Lanka
Loa loa	Subcutaneous	Blood	*Chrysops* spp.	'Calabar' swellings Urticaria	West and Central Africa
***Onchocerca* spp.**	Subcutaneous	Skin, eye	*Simulium* spp.	Subcutaneous nodules Eye disease	Africa, South America
Mansonella perstans	Retroperitoneal	Blood	*Culicoides* spp.	Allergic eosinophilia	Sub-Saharan Africa, South America

Obstructed lymphatics may rupture, causing cellulitis and further fibrosis; there may also be chylous (lymph-containing) pleural effusions and ascites. Over time, augmented by repeated episodes of cellulitis and inflammation, there is progressive enlargement, coarsening, and fissuring of the skin, leading to the classical appearances of elephantiasis. The limbs or scrotum may become hugely swollen. Eventually, the adult worms die, but the lymphatic obstruction remains, and tissue damage continues. Elephantiasis takes many years to develop and is only seen in association with recurrent infection in endemic areas.

Occasionally, the predominant features of filarial infection are pulmonary. Microfilariae become trapped in the pulmonary capillaries, generating an intense local allergic response. The resulting pneumonitis causes cough, fever, weight loss, and shifting radiological changes, associated with a high peripheral eosinophil count. This is known as ***tropical pulmonary eosinophilia***.

Diagnosis

The clinical picture in established disease is usually diagnostic, although similar lymphatic damage may occasionally be caused by silicates absorbed through the feet from volcanic soil (podoconiosis). Parasitological diagnosis has traditionally relied on detecting microfilariae in blood films or skin snips. Rapid and sensitive near-patient antigen detection tests are now available for use in endemic areas, but there is a significant false-positive rate when used in travellers with a low pretest probability.

Management

Diethylcarbamazine (DEC) kills both adult worms and microfilariae. Serious allergic responses may occur as the parasites are killed, and particular care is needed when using DEC in areas endemic for loiasis. Mass treatment programmes, using combinations of DEC, ivermectin, and albendazole to target various helminthic infections, are deployed in many parts of the world; the exact regimens depend on local situations. There is currently much interest in using new agents to kill the symbiotic *Wolbachia* bacteria, without which the adult worm will eventually die. However, the best way of incorporating this tactic into the overall management strategy remains unclear. Vector control with insecticide bed nets is helpful in certain areas.

In 2021 the WHO reported that more than US $100 billion of economic loss had been averted through mass treatment campaigns but that 882.5 million people in 44 countries were still living in areas that require preventive chemotherapy to stop the spread of infection. The new neglected tropical diseases roadmap has set a goal of eliminating lymphatic filariasis as a public health problem in 80% of endemic countries by 2030.

Loiasis

Loiasis is found in the humid forests of West and Central Africa. The causative parasite, *Loa loa*, is a small (3–7 cm) filarial worm found in the subcutaneous tissues. The microfilariae circulate in the blood during the day but cause no direct symptoms. The vectors are day-biting flies of the genus *Chrysops*.

Adult worms migrate around the body in subcutaneous tissue planes. Worms may be present for years, frequently without causing symptoms. From time to time, localized, tender, hot, soft tissue swellings (Calabar swellings) occur because of hypersensitivity, often near a joint. These are produced in response to the passage of a worm and usually subside over a few days or weeks. There may also be more generalized urticaria and pruritus. Occasionally, a worm may be seen crossing the eye under the conjunctiva; it may also enter retro-orbital tissue, causing severe pain, but does not lead to blindness. Short-term residents of endemic areas often have more severe manifestations of the disease, presumably because of less opportunity to develop immune tolerance.

Microfilariae may be seen on stained blood films taken around noon, although these are often negative. Serological tests are relatively insensitive and cross-react with other microfilariae. There is usually massive eosinophilia. The Loascope is a useful tool for measuring microfilarial density in a finger-prick blood test taken in the field.

DEC, which was used in early treatment campaigns, can cause severe allergic reactions and encephalopathy associated with parasite killing when the microfilarial burden is high (>20 000/mL) and is being replaced by newer agents. Ivermectin in single doses of 200 to 400 μg/kg is effective but may also cause severe reactions occasionally. Albendazole, which brings about a more gradual reduction in microfilarial load, may be preferable in heavily infected patients. Mass treatment with either DEC or ivermectin can decrease the transmission of infection, but the mainstay of prevention is vector avoidance and control.

Onchocerciasis

Onchocerciasis is also known as river blindness because of the location of the vector and the affected population to rivers. It is the result of infection with *Onchocerca volvulus*, which is transmitted by day-biting flies of the genus *Simulium*. In 2017 the global burden of disease study estimated that there were at least 20.9 million people infected worldwide, of whom 14.6 million had skin disease and 1.15 million had visual loss. Ninety-nine percent of these people live in 31 countries in sub-Saharan Africa, with small foci in Yemen and Central and South America.

Pathogenesis

Infection occurs when larvae are inoculated into humans by the bite of an infected fly. The worms mature in 2 to 4 months and can live for more than 15 years. Adult worms, which can reach lengths of 50 cm (although they are <0.5 mm in diameter), live in the subcutaneous tissues. They may form fibrotic nodules, especially over bony prominences and sites of trauma. Huge numbers of microfilariae are distributed in the skin and may invade the eyes. Live microfilariae cause relatively little harm, but dead parasites may cause severe allergic reactions, with hyaline necrosis and loss of tissue collagen and elastin. In the eye, a similar process causes conjunctivitis, sclerosing keratitis, uveitis, and secondary glaucoma. Choroidoretinitis is also seen occasionally.

Clinical features

Symptoms usually start about 1 year after infection. Initially, there is generalized pruritus, with urticaria and fleeting oedema. Subcutaneous nodules (which can be detected by ultrasonography) start to appear, and in dark-skinned individuals, there is hypo- and hyperpigmentation from excoriation and inflammatory changes ('leopard skin'). Over time, more chronic inflammatory changes appear, with roughened, inelastic skin. Superficial lymph nodes become enlarged, and in the groin, may hang down in loose folds of skin ('hanging groin'). Eye disease, which is associated with chronic heavy infection, usually first manifests as itching and conjunctival irritation. This gradually progresses to more extensive eye disease and eventually to blindness.

Diagnosis

In endemic areas, the diagnosis can often be made clinically, especially if supported by finding eosinophilia on a blood film. To

identify parasites, skin snips taken from the iliac crest or shoulder are placed in saline under a cover slip. After 4 hours, microscopy will show microfilariae wriggling free on the slide. If this is negative, DEC can be applied topically under an occlusive dressing and will provoke an allergic rash in the majority of infected people (modified Mazzotti reaction). Slit-lamp examination of the eyes may reveal the microfilariae. However, skin snip tests are less reliable in lower worm loads and are therefore of limited value in assessing success of treatment. Rapid near-patient serological card tests are under evaluation, and PCR tests are being developed.

Management and prevention

A single dose of ivermectin kills microfilariae and prevents their return for 6 to12 months. There is little effect on adult worms, so annual (or more frequent) retreatment is needed. In patients co-infected with *Loa loa*, ivermectin may occasionally induce severe allergic reactions, including a toxic encephalopathy. When the two infections coexist, field teams may screen the blood of school children and adults using a Loascope to identify individuals carrying a high burden of *Loa loa* microfilariae to be given an alternative drug. Moxidectin has a longer lasting effect than ivermectin and has recently been approved for use in onchocerciasis control programmes.

The WHO Onchocerciasis Control Programme, which started in 1974, had a considerable impact on the disease in West Africa. A combination of vector control measures and, more recently, mass treatment with ivermectin led to a decrease in both infection rates and progression to serious disease. The control programme has been handed over to local governments by the WHO, and local mass treatment programmes are still running in most endemic countries. Eradication may eventually be possible because humans are the only host, but measures are required over a long period because of the longevity of the worm (10–15 years).

Mansonellosis

Mansonella perstans is a filarial worm transmitted by biting midges of the genus *Culicoides*. Small numbers of microfilariae are found in the blood, and although they do not cause serious disease, there may be minor allergic reactions and an eosinophilia.

Dracunculiasis

Infection with the Guinea worm, *Dracunculus medinensis*, occurs when water fleas (copepods) containing the parasite larvae are swallowed in contaminated drinking water. Ingested larvae mature, and the female worm, which can reach more than 1 m in length, migrates through connective and subcutaneous tissue for 9 to 18 months before surfacing on the skin. The uterus of the worm ruptures, releasing larvae that are ingested by the small crustacean water fleas, and the cycle is completed.

The diagnosis is clinical. The traditional management, extracting the worm over several days by winding it around a stick, is probably still the most effective. The worm should not be damaged, and the exit site should be kept clean and dressed. Antibiotics may be needed to control secondary infection.

Prevention and control of dracunculiasis has been one of the success stories in tropical medicine. Large-scale eradication programmes involving removal of water fleas, and thus infective larvae, by chemical treatment or simple filtration have been in place for many years. In 1989 active surveillance in Africa identified nearly 1 million cases of dracunculiasis; similar (though probably more robust) surveillance in 2022 found only 13 cases. Reports of the disease are now confined to four countries: Chad, Ethiopia, Mali, and South Sudan.

Human intestinal nematodes

Adult intestinal nematodes (also referred to sometimes as soil-transmitted helminths or geohelminths) live in the human gut. There are two main types of life cycle, both usually including a soil-based stage. In some species, infection is spread by ingestion of eggs (which often require a period of maturation in the environment); in others, the eggs hatch in the soil, and larvae penetrate directly through the skin of a new host. *Strongyloides* is unusual in that it is the only nematode able to complete its life cycle entirely within the human host without requiring a stage in the environment or an insect vector. Larvae may hatch before leaving the colon and so are able to re-infect the host by penetrating the intestinal wall and entering the venous system.

Although light infestations with intestinal worms cause few symptoms, heavy worm loads cause significant morbidity, especially in children. Cheap, effective, and safe treatments are available, and the WHO promotes mass treatment programmes for those at risk (**Box 42.68**). However, there is a healthy debate about the evidence to support these mass interventions.

Ascariasis (roundworm infection)

Ascaris lumbricoides is a pale-yellow worm, 20–35 cm in length (**Fig. 42.33**). It is found worldwide but is particularly common in poor rural communities, where there is heavy faecal contamination of the immediate environment. Larvae hatch and penetrate the duodenum, migrating through the tissues to the lungs before being expectorated and swallowed. The adult worms live in the small intestine. Ova are deposited in faeces and require a 2- to 4-month maturation in the soil before they are infective.

Infection is usually asymptomatic, although heavy infections are associated with nausea, vomiting, abdominal discomfort, and anorexia. Worms can sometimes obstruct the small intestine, the most common site being the ileocaecal valve. They may also occasionally invade the appendix, causing acute appendicitis, or the bile duct, resulting in biliary obstruction and suppurative cholangitis. Larvae in the lung may produce pulmonary eosinophilia. Heavy infection in children, especially those who are already malnourished, may have significant effects on nutrition and development. Serious morbidity and mortality are relatively rare in ascariasis, but the huge number of people infected means that on a global basis, roundworm infection causes a significant burden of disease, especially in children.

Ascaris eggs can be identified in the stool, and occasionally, adult worms emerge from the mouth or the anus. They may also be seen on barium enema studies. Appropriate drug treatments are shown in **Box 42.68**. Very rarely, surgical or endoscopic intervention may be required for intestinal or biliary obstruction.

Threadworm *(Enterobius vermicularis)*

Enterobius vermicularis is a small (2–12 mm) worm, which is common throughout the world. Larval development takes place mainly in the small intestine and adult worms are normally found in the colon. The gravid female deposits eggs around the anus, causing intense itching, especially at night. Unlike those of *A. lumbricoides*,

Box 42.68 Drugs used for treating human intestinal nematodes[a]

Drug	Dosage	*Ascaris* spp.	Hookworm	*Enterobius* spp.	*Trichuris* spp.	*Strongyloides* spp.
Piperazine	75 mg/kg	++	+	++	–	–
Pyrantel pamoate	10 mg/kg	++	+	++	–	–
Oxantel pamoate	10 mg/kg	++	+	N/A	++	–
Albendazole	400 mg[b]	++	++	++	+	+
Mebendazole	500 mg[c]	++	++	++	+	+
Tiabendazole	25 mg/kg[b]	N/A	N/A	N/A	N/A	++
Levamisole	2.5 mg/kg	++	+	N/A	N/A	–
Ivermectin	200 μg/kg[d]	N/A	N/A	N/A	N/A	++

[a]Single dose unless otherwise stated.
[b]Twice daily for 3 days in strongyloidiasis.
[c]World Health Organization–recommended dose for low- and middle-income countries; in the UK, commonly given as 100 mg single dose for threadworm or 100 mg twice daily for 3 days for whipworm.
[d]Once daily for 2 days.
++, Highly effective; *+*, moderately effective; *–*, ineffective; *N/A*, drug not used for this indication or no data available.

Fig. 42.33 ***Ascaris lumbricoides.*** The worm is approximately 20 cm long.

the eggs do not require a maturation period in soil, and infection is often directly transmitted from anus to mouth via the hands. Eggs may also be deposited on clothing and bed linen and are subsequently either ingested or inhaled. Apart from discomfort and local excoriation, infection is usually harmless.

Ova can be collected either by using a moistened perianal swab or by applying adhesive cellophane tape to the perianal skin. They can then be identified by microscopy.

The most commonly used drugs are mebendazole and piperazine (see **Box 42.68**). However, isolated treatment of an affected person is often ineffective. Other family members (especially small children) may also need to be treated, and the whole family should be given advice about personal hygiene. Two courses of treatment 2 weeks apart may break the cycle of autoinfection.

Whipworm *(Trichuris trichiura)*

Infections with whipworm are common worldwide, especially in poor communities with inadequate sanitation. Adult worms, which are 3 to 5 cm long, inhabit the terminal ileum and caecum, although in heavy infection, they are found throughout the large bowel. The head of the worm is embedded in the intestinal mucosa. Ova are deposited in the faeces and require a maturation period of 3 to 4 weeks in the soil before becoming infective.

Infection is usually asymptomatic, but mucosal damage can occasionally be so severe that there is colonic ulceration, dysentery, or rectal prolapse.

Diagnosis is made by finding ova on stool microscopy or occasionally by seeing adult worms on sigmoidoscopy. Drug treatment is shown in **Box 42.68**.

Hookworm

Hookworm infections, caused by the human hookworms *Ancylostoma duodenale* and *Necator americanus*, are found worldwide. They are relatively rare in high-income countries but very common in areas with poor sanitation and hygiene. Hookworm infection is one of many causes of anaemia in children living in poor rural communities in the tropics. *A. duodenale* is found mainly in East Asia, North Africa, and the Mediterranean, and *N. americanus* is the predominant species in South and Central America, South-east Asia, and sub-Saharan Africa.

Adult worms (which are about 1 cm long) live in the duodenum and upper jejunum, where they are often found in large numbers. They attach firmly to the mucosa using the buccal plate, feeding on blood. Eggs passed in the faeces develop in warm, moist soil, producing infective filariform larvae. These penetrate directly through the skin of a new host and are carried in the bloodstream to the lungs. Having crossed into the alveoli, the parasites are expectorated and then swallowed, thus arriving at their definitive home.

Clinical features

Local irritation as the larvae penetrate the skin ('ground itch') may be followed by transient pulmonary signs and symptoms, often accompanied by eosinophilia. Light infections, especially in a well-nourished person, are often asymptomatic. Heavier worm loads may be associated with epigastric pain and nausea, resembling peptic ulcer disease. Chronic heavy infection, particularly on a background of malnourishment, may cause iron-deficiency anaemia and hypoproteinaemia. Heavy infection in children is associated with delays in physical and mental development.

Diagnosis and management

The diagnosis is made by finding eggs on faecal microscopy. In infections heavy enough to cause anaemia, these are present in large numbers. The aim of treatment in endemic areas is reduction of worm burden rather than complete eradication; albendazole given as a single dose is the best drug (see **Box 42.68**).

Strongyloidiasis

Strongyloides stercoralis is a small (2-mm-long) worm that lives in the small intestine. It is found in many parts of the tropics and subtropics and is especially common in Asia. Eggs hatch in the bowel, and larvae are found in the stool. Usually, these are non-infective rhabditiform larvae, which require a further period of maturation in the soil before they can infect a new host, but sometimes this maturation can occur in the large bowel before leaving the human host. Infective filariform larvae can therefore penetrate directly through the perianal skin, re-infecting the host. In this way, autoinfection may continue for years or even decades. Some war veterans who were imprisoned in the Far East during the Second World War were found to have active strongyloidiasis more than 50 years later. After skin penetration, the life cycle is similar to that of the hookworm, except that the adult worms may burrow into the intestinal mucosa, causing a local inflammatory response.

Clinical features

After skin penetration, *S. stercoralis* causes a similar local dermatitis to hookworm. In autoinfection, this manifests as a migratory linear weal around the buttocks and lower abdomen ***(cutaneous larva currens)***. In heavy infections, damage to the small intestinal mucosa can cause malabsorption, diarrhoea, and even perforation. There is usually a persistent eosinophilia.

Strongyloides hyperinfection syndrome occurs in immunosuppressed patients such as those on long-term corticosteroids, after transplantation, or during cancer treatment. Filariform larvae penetrate directly through the bowel wall in huge numbers and migrate to the lungs. The larvae are accompanied by Gram-negative bacteria from the bowel lumen. Typically, patients with hyperinfection syndrome do not have an eosinophilia, and with the epidemiological exposure long ago, there is often delayed recognition and a fatal outcome.

Diagnosis and management

Motile larvae may be seen on stool microscopy, especially after a period of incubation on charcoal culture. Serological tests are also useful. The best drug for treating strongyloidiasis is ivermectin (200 μg/kg daily for 2 days); albendazole is also effective. In hyperinfection, patients may need a longer course of ivermectin, often given parenterally, which may involve off-licence use of veterinary preparations. Antibiotics against Gram-negative organisms should also be given.

Zoonotic nematodes

A number of nematodes that are principally parasites of animals may also affect humans. The more common ones are described here.

Trichinosis

The normal hosts of *Trichinella spiralis*, the cause of trichinosis, include pigs, bears, and warthogs. Humans are infected by eating undercooked meat from these animals. Ingested larvae mature in the small intestine, where adults release new larvae that penetrate the bowel wall and migrate through the tissues. Eventually, these larvae encyst in striated muscle. Meat inspection policies have reduced the risk of infection from domestic pork, and most infections in Asia, Europe, and North America come from eating wild animal products.

Light infections are usually asymptomatic. Heavier loads of worms produce GI symptoms as the adults establish themselves in the small intestine followed by systemic symptoms as the larvae invade. The latter include fever, oedema, and myalgia. Massive infection may occasionally be fatal, but usually the symptoms subside after the larvae encyst.

The diagnosis can usually be made from the clinical picture, associated eosinophilia, and serological tests. If necessary, it can be confirmed by muscle biopsy a few weeks after infection. Long-standing infection may be revealed by the presence of numerous calcified cysts on X-ray. A 2-week course of albendazole, given early in the course of the illness, will kill the adult worms and decrease the load of larvae reaching the tissues. Analgesia and steroids may be needed for symptomatic relief.

Toxocariasis (visceral larva migrans)

Eggs of the dog roundworm *Toxocara canis* are occasionally ingested by humans, especially children. The eggs hatch, and the larvae penetrate the small intestinal wall and enter the mesenteric circulation, but they are then unable to complete their life cycle in a 'foreign' host. Many are held up in the capillaries of the liver, where they generate a granulomatous response, but some may migrate into other tissues, including the lungs, striated muscle, heart, brain, and eye. In most cases, infection is asymptomatic, and the larvae die without causing serious problems. In heavy infections, there may be generalized symptoms (fever and urticaria) and eosinophilia, as well as focal signs related to the migration of the parasites. Pulmonary involvement may cause bronchospasm and chest X-ray changes. Ocular infection may produce a granulomatous swelling mimicking a retinoblastoma, and cardiac or neurological involvement is occasionally fatal. Rarely, larvae survive in the tissues for many years, causing symptoms long after infection.

Isolation of the larvae is difficult, and the diagnosis is usually made serologically. Albendazole is the most effective treatment.

Cutaneous larva migrans

Cutaneous larva migrans is a common condition caused by the larvae of the non-human hookworms, *Ancylostoma braziliense* and *Ancylostoma caninum*. Like human hookworms, these hatch in warm, moist soil and then penetrate the skin. In humans, they are unable to complete a normal life cycle and instead migrate under the skin for days or weeks until they eventually die. The wandering of the larvae is accompanied by a clearly defined, serpiginous, intensely itchy rash, which progresses at the rate of about 1 cm per day. The rash is normally found on the feet where the patient's skin has had contact with contaminated soil. The diagnosis is purely clinical. Single larvae may be treated with a 15% solution of topical tiabendazole; multiple lesions require systemic therapy with a short course of albendazole or ivermectin.

TREMATODES

Trematodes (flukes) are flat, leaf-shaped worms. They have complex life cycles, often involving freshwater snails and intermediate mammalian hosts. Disease is caused by the inflammatory response to eggs or to the adult worms.

Water-borne flukes

Schistosomiasis

Schistosomiasis is commonly known as Bilharzia, after Theodor Bilharz, who first identified the cause of urinary schistosomiasis in

1851. It affects more than 260 million people worldwide with 9 of 10 of cases occurring in sub-Saharan Africa. Chronic infection causes significant morbidity, and after malaria, it has the greatest socioeconomic impact of any parasitic disease. Schistosomiasis is largely a disease of rural poor individuals but has also been associated with major development projects, such as dams and irrigation schemes.

Parasitology and pathogenesis

There are three species of schistosome that commonly cause disease in humans: *Schistosoma mansoni*, *Schistosoma Haematobium*, and *Schistosoma japonicum*. *Schistosoma mekongi* and *Schistosoma intercalatum* also affect humans but have a very restricted distribution (Fig. 42.34). Eggs are passed in the urine or faeces of an infected person and hatch in freshwater to release the miracidia. These ciliated organisms penetrate the tissue of the intermediate host, a species of water snail specific to each species of schistosome. After multiplying in the snail, large numbers of fork-tailed cercariae are released back into the water, where they can survive for 2 to 3 days. During this time, the cercariae can penetrate the skin or mucous membranes of the definitive host, humans. Transforming into schistosomulae, they pass through the lungs before reaching the portal vein, where they mature into adult worms (the male is about 20 mm long and the female a little larger). Worms pair in the portal vein before migrating to their final destination: mesenteric veins in the case of *S. mansoni* and *S. japonicum* and the vesicular plexus for *S. haematobium*. There they may remain for many years, producing vast numbers of eggs. The majority of them are released in urine or faeces, but a small number become embedded in the bladder or bowel wall, and a few are carried in the circulation to the liver or other distant sites.

The pathology of schistosomiasis varies with species and stage of infection. In the early stages, there may be local and systemic allergic reactions to the migrating parasites. As eggs start to be deposited, there is a local inflammatory response in the bowel or bladder, and ectopic eggs may produce granulomatous lesions anywhere in the body. Chronic heavy infection, in which large numbers of eggs accumulate in the tissues, leads to fibrosis; calcification; and in some cases, dysplasia and malignant change. Morbidity and mortality are related to the duration of infection and worm load, as well as to the species of parasite. Children in endemic areas tend to have the heaviest worm load because of both increased exposure to infection and differences in the immune response between adults and children.

Clinical features

Cercarial penetration of the skin may cause local dermatitis ('swimmer's itch'). After a symptom-free period of 2 to 6 weeks, systemic allergic features may develop, including fever, rash, myalgia, and a patchy pneumonitis (Katayama fever). These phenomena are common in non-immune travellers but are rarely seen in local populations, who are usually exposed to infection from early childhood onwards. If infection is sufficiently heavy, symptoms from egg deposition may start to appear 2 to 3 months after infection.

S. haematobium infection (bilharzia)

The earliest symptom is usually painless terminal haematuria. As bladder inflammation progresses, there are increased urinary frequency and groin pain. Obstructive uropathy may develop, leading to hydronephrosis, recurrent UTIs, and kidney failure. There is a strong association between chronic urinary schistosomiasis and

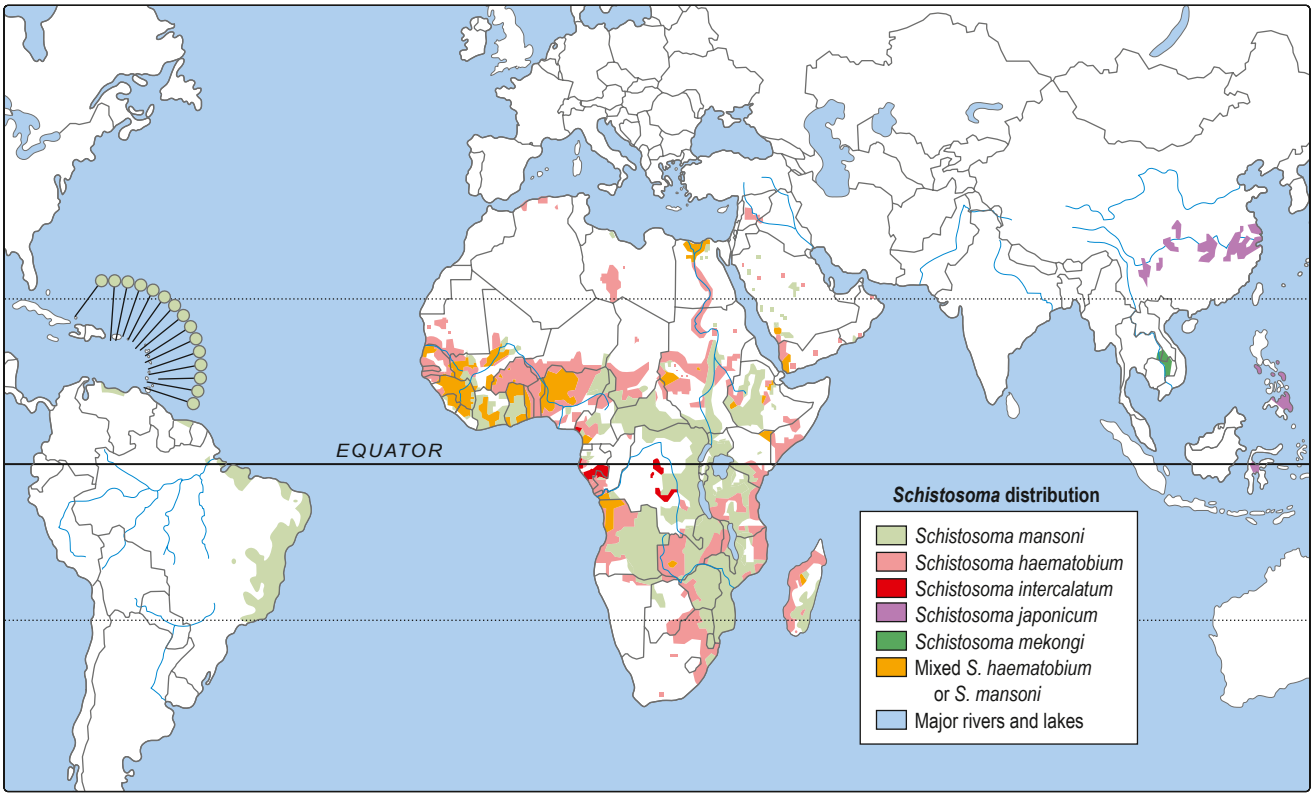


Fig. 42.34 Schistosomiasis: geographical distribution. Map lines delineate study areas and do not necessarily depict accepted national boundaries. *(From Colley DG, Bustinduy AL, Secor WE, King CH. Human schistosomiasis.* Lancet *2014; 383:2253–22264, with permission.)*

squamous cell bladder carcinoma. Ectopic eggs may cause pulmonary or neurological disease. Males may experience haematospermia. In females, *S. haematobium* infection causes genital ulceration, leading to painful sexual intercourse and an increased risk of acquiring HIV infection. If the uterine tubes and ovaries are involved, this may lead to infertility. These complications create a major psychosocial burden for young females already living in poverty.

S. mansoni infection

This usually affects the large bowel. Early disease produces superficial mucosal changes, accompanied by blood-stained diarrhoea. Later, the mucosal damage becomes more marked, with the formation of rectal polyps; deeper ulceration; and eventually, fibrosis and stricture formation. Ectopic eggs are carried to the liver, where they cause an intense granulomatous response. Hepatitis is followed by progressive periportal fibrosis, leading to portal hypertension, oesophageal varices, and splenomegaly. Hepatocellular function is usually well preserved.

S. japonicum infection

Unlike the other species, *S. japonicum* infects numerous other mammals apart from humans. It is similar to *S. mansoni* but infects both the large and small bowel and produces a greater number of eggs. Disease therefore tends to be more severe and rapidly progressive. Hepatic involvement is more common, and neurological involvement is seen in about 5% of patients. *S. japonicum* is now only found in pockets of China, Indonesia, and the Philippines, where it affects children and those working in high-risk occupations such as rice farmers and freshwater fishermen.

Diagnosis

Acute schistosomaisis (Katayama fever) is mainly seen in travellers with recent freshwater exposure in an endemic area. The snails are found in the slow flowing parts of the river or lake, around underwater vegetation near the shore. It is generally a clinical diagnosis. Patients usually have an eosinophilia but may not yet have developed schistosomal antibodies, which are directed against the adult worm and take 2 to 3 months to appear.

When egg deposition has started, the characteristic eggs (with a terminal spine in the case of *S. haematobium* and a lateral spine in the other species) can be detected on microscopy. In *S. haematobium* infection, the best specimen for examination is a filtered mid-day urine sample; parasites may also be found in semen. *S. mansoni* and *S. japonicum* eggs can usually be found on microscopy of faeces or in a rectal snip.

Chronic schistosomiasis may occur in travellers but is more common in citizens of endemic areas who have had significant exposure as children. The presenting syndrome depends on the organs affected and is generally secondary to the inflammatory response produced in response to ectopic egg deposition. Heavy and long-standing infection leads to end-organ damage such as renal failure from obstructive uropathy; pulmonary hypertension, ascites, and variceal bleeding from portal hypertension; and seizures, stroke, and flaccid paralysis from neurological involvement. Corticosteroids are given before praziquantel in spinal and CNS schistosomiasis.

Asymptomatic infection may be picked up when migrants or travellers attend clinic for screening. They generally have an eosinophilia but not always.

Management

The aim of treatment in endemic areas is to decrease the worm load and therefore minimize the chronic effects of egg deposition. It may not always be possible (or even desirable) to eradicate adult worms completely, and re-infection is common. However, a 90% reduction in egg output has been achieved in mass treatment programmes, and in light infections, in which there is no risk of re-exposure, the drugs are usually curative. The most widely used is praziquantel (**Box 42.69**), which is effective against all species of schistosome, and is well tolerated and reasonably cheap. The WHO estimated that in 2021, 251.4 million people required preventive treatment for schistosomiasis, but only 75.3 million people were reported to have been treated. Some of the disparity may have been because of the interruption of control programmes during the COVID pandemic.

Box 42.69 Treatment of trematode infections

Parasite	Drug and dose
Schistosoma mansoni	Praziquantel 40 mg/kg single dose[a]
Schistosoma haematobium	Praziquantel 40 mg/kg single dose[a]
Schistosoma japonicum	Praziquantel 60 mg/kg single dose[a]
***Paragonimus* spp.**	Praziquantel 25 mg/kg 8-hourly for 3 days
Clonorchis sinensis	Praziquantel 25 mg/kg 8-hourly for 1–3 days[b]
***Opisthorchis* spp.**	Praziquantel 25 mg/kg 8-hourly for 1–3 days[b]
Fasciolopsis buski	Praziquantel 25 mg/kg 8-hourly for 1 day
Fasciola hepatica	Triclabendazole 10 mg/kg single dose[c]

[a]May be split to minimize nausea.
[b]Depending on worm load.
[c]Repeated if necessary.

Prevention

Prevention of schistosomiasis is difficult and relies on a combination of approaches. Mass treatment of the population (especially children) decreases the egg load in the community. Health education programmes, the provision of latrines, and access to a safe water supply should decrease contact with infected water. Attempts to eradicate the snail host have generally been unsuccessful, although human-made bodies of water can often be made less 'snail friendly'. Travellers should be advised to avoid potentially infected water and attend for screening 3 months after if exposed.

Food-borne flukes

Many flukes infect humans via ingestion of an intermediate host, often freshwater fish or crustaceans.

Paragonimiasis

More than 20 million people are infected with lung flukes of the genus *Paragonimus*. It is common throughout South and East Asia and is also found in parts of West Africa and Latin America. *Paragonimus* spp. are principally parasites of a variety of mammals that feed on crustacea, and humans can become infected through eating uncooked or poorly prepared shellfish. The adult worms (of which the major species is *Paragonimus westermani*) live in the lungs, producing eggs that are either expectorated or swallowed and passed in the faeces. Miracidia emerging from the eggs penetrate the first intermediate host, a freshwater snail. Larvae released from the snail seek out the second intermediate host, freshwater crustacea, in which they encyst as metacercariae. After consumption by mammalian hosts, cercariae penetrate the small intestinal wall and migrate directly from the peritoneum to the lungs across

the diaphragm. Having established themselves in the lung, the adult worms may survive for 20 years.

The common clinical features are fever, cough, and mild haemoptysis. In heavy infections, the disease may progress, sometimes mimicking pneumonia or pulmonary TB. Ectopic worms may cause signs in the abdomen or the brain.

The diagnosis is made by detection of ova on sputum or stool microscopy. Radiological appearances are variable and non-specific. Treatment is with praziquantel or triclabendazole; the latter, which has a simpler regimen and is supplied free through the WHO, is generally preferred. Prevention involves avoidance of inadequately cooked shellfish.

Liver flukes

Humans can become infected with a variety of liver flukes. Terminology is confusing: three parasite species, *Clonorchis sinensis*, *Opisthorchis felineus*, and *O. viverrini*, are commonly referred to as the human liver flukes, although feline and other mammals are the principal hosts. *C. sinensis* and *O. viverrini* are almost entirely confined to East and South-east Asia, and the distribution of *O. felineus* extends across Central Asia as far as Eastern Europe. Adult worms live in the bile ducts, releasing eggs into the faeces. The parasite requires two intermediate hosts, a freshwater snail and a fish, and humans are infected by consumption of raw fish. The cycle is completed when excysted worms migrate from the small intestine into the bile ducts.

Infection is often asymptomatic, and the flukes can live in the biliary tree for 20 to 25 years. It sometimes causes cholangitis and is associated with biliary carcinoma. The diagnosis is made by identifying eggs on stool microscopy. Treatment is with praziquantel or albendazole, and infection can be avoided by preparing and cooking fish adequately.

Other food-borne flukes

Humans can also be infected with a variety of other animal flukes, notably, the liver fluke *Fasciola hepatica* and the intestinal fluke *Fasciolopsis buski*. Both require a water snail as an intermediate host; cercariae encyst on aquatic vegetation and then are consumed by animals or humans. After ingestion, *F. hepatica* penetrates the intestinal wall before migrating to the liver. During this stage, it causes systemic allergic symptoms. After reaching the bile ducts, it gives rise to similar problems to those caused by the other liver flukes. *F. buski* does not migrate after it excysts and causes mainly bowel symptoms.

Management of trematode infections

For drug treatment, see **Box 42.69**.

CESTODES

Cestodes (tapeworms) are ribbon-shaped worms, which vary from a few millimetres to several metres in length. Adult worms live in the human intestine, where they attach to the epithelium using suckers on the anterior portion (scolex). From the scolex arises a series of progressively developing segments, called proglottids. The mature distal segments contain eggs, which either may be released directly into the faeces or are carried out with an intact detached proglottid. The eggs are consumed by intermediate hosts, after which they hatch into larvae (oncospheres). These penetrate the intestinal wall of the host (pigs, cattle, or other animals) and encyst in the tissues. Humans ingest the cysts in undercooked meat, and the cycle is completed when the parasites excyst in the stomach and develop into adult worms in the small intestine. Infections are usually solitary, but several adult tapeworms may coexist. The exceptions to this life cycle are the dwarf tapeworm, *Hymenolepis nana*, which has no intermediate host and is transmitted from person to person by the faecal–oral route, and *Taenia solium*, which can cause either normal tapeworm infection (when humans are the definitive host) or cysticercosis (when humans are the intermediate host).

Taenia saginata

Taenia saginata, the beef tapeworm, may reach a length of several metres. It is found in all countries where undercooked beef is eaten, although human infection can largely be eradicated by good abattoir hygiene and inspection regimens, and the main burden of disease falls on economically low- and middle-income countries. The adult worm causes few, if any, symptoms and is not a significant cause of chronic health problems. Infection is usually discovered when proglottids are found in faeces or on underclothing, often causing considerable anxiety. Ova may also be seen on stool microscopy. Infection can be cleared with a single dose of praziquantel (10 mg/kg).

Taenia solium and cysticercosis

Taenia solium, the pork tapeworm, is generally smaller than *T. saginata*, although it can still reach 6 metres in length. It is particularly common in South America, South Africa, China, and parts of South-east Asia. As with *T. saginata*, infection is usually asymptomatic. The ova of the two species are identical, but the proglottids can be distinguished on inspection.

Pork tapeworm infection is acquired by eating uncooked pork. Treatment is with praziquantel or niclosamide. Drug treatment should not be accompanied by a purgative, as was previously believed.

Cysticercosis occurs when humans become the intermediate host of the parasite and is caused by cysts rather than the adult worm. It follows the ingestion of eggs from contaminated food and water, not from pork meat. Faecal–oral autoinfection can occur but is rare. Patients with tapeworms do not usually develop cysticercosis, and individuals with cysticercosis do not usually harbour tapeworms. After the ingestion of eggs, the larvae are liberated, penetrate the intestinal wall, and are carried to various parts of the body, where they develop into cysticerci. These are cysts, 0.5 to 1 cm in diameter, containing the scolex of a new adult worm. Common sites for cysticerci include subcutaneous tissue, skeletal muscle, and brain. Superficial cysts may be felt under the skin but usually cause no significant symptoms. Muscle cysts tend to calcify and are often visible on X-rays.

Neurocysticercosis refers to the presence of cysts in the brain. They are generally 5 to 20 mm in diameter and a range from a solitary cyst to many dozen. The cysts are mostly asymptomatic but can give rise to a variety of problems depending on their location, including epilepsy, personality change, hydrocephalus, and focal neurological signs. Contrast enhanced CT or MRI shows the evolution of brain cysts from active colloidal, ring-enhancing lesions with surrounding oedema to nodular, calcified cysts. Serological tests may support the diagnosis but may be negative, and the diagnosis is often clinical based on risk exposure and the characteristic appearances on scans.

Management of cysticercosis

After years of controversy, recent studies suggest that antihelminthic drugs are of benefit in most cases of neurocysticercosis, although their role in other forms of cysticercosis remains unproven. The priority is to control seizures with anticonvulsant medication. Antihelminthic treatment is never urgent and should always be accompanied by corticosteroids to reduce the inflammation caused by dying cysts. Albendazole is the drug of choice, with praziquantel as an alternative for 2 to 3 weeks. Surgery is occasionally required for patients with hydrocephalus.

Diphyllobothrium latum

Infection with the fish tapeworm *Diphyllobothrium latum* is common in northern Europe and Japan, owing to the consumption of raw fish. The adult worm reaches a length of several metres, but like the other tapeworms, usually causes no symptoms. A megaloblastic anaemia (caused by competitive use of vitamin B_{12} by the parasite) may occur. Diagnosis and treatment are the same as for *Taenia* spp.

Hydatid disease

Hydatid disease occurs when humans become an intermediate host of the dog tapeworm, *Echinococcus granulosus*. The adult worm lives in the gut of domestic and wild canines, and the larval stages are usually found in sheep, cattle, and camels. Humans may become infected either from direct contact with dogs or from food or water contaminated with dog faeces. After ingestion, the parasites excyst, penetrate the small intestine wall, and are carried to the liver and other organs in the bloodstream. A slow-growing, thick-walled cyst is formed, inside which further larval stages of the parasite develop. The life cycle cannot be completed unless the cyst is eaten by a dog. Hydatid disease is prevalent in areas where dogs are used in the control of livestock, especially sheep. It is common in Argentina, Peru, Central Asia, China, and parts of East Africa, and about 1 million people are infected worldwide.

Symptoms depend mainly on the site of the cyst. The liver is the most common organ affected (60%) followed by the lungs (20%), kidneys (3%), brain (1%), and bone (1%). The symptoms are those of a slowly growing benign tumour. Pressure on the bile ducts may cause jaundice. Rupture into the abdominal cavity, pleural cavity, or biliary tree may occur. A cyst rupturing into a bronchus may result in its expectoration and spontaneous cure, but if secondary infection supervenes, a chronic pulmonary abscess will form. Focal seizures can occur if cysts are present in the brain. Renal involvement produces lumbar pain and haematuria. Calcification of the cyst occurs in about 40% of cases. The diagnosis and treatment of hydatid liver disease are described in Chapter 23 ('Hydatid disease' section).

A related parasite of foxes, *Echinococcus multilocularis*, causes a similar but more severe infection, alveolar hydatid disease. These cysts are invasive, and metastases may occur; if untreated, the disease is usually fatal. Treatment requires radical surgery as well as antihelminthic treatment.

Further reading

Brunetti E, Tamarozzi F. Watch-and-wait approach for inactive echinococcal cysts: scoping review update since the issue of the WHO-IWGE Expert Consensus and current perspectives. *Curr Opin Infect Dis* 2023; 36(5): 326–332.
Engels D, Hotez PJ, Ducker C et al. Integration of prevention and control measures for female genital schistosomiasis, HIV and cervical cancer. *Bull World Health Organ* 2020; 98(9):615–624.
Gill GV, Bell DR. Strongyloides stercoralis infection in former Far East prisoners of war. *Br Med J* 1979; 2(6190):572–574.
Guzman C, Garcia HH. Cysticercosis Working Group in Peru. Current diagnostic criteria for neurocysticercosis. *Res Rep Trop Med* 2021; 12:197–203.
Hirsch LA, Martin R. *LSHTM and Colonialism: a report on the Colonial History of the London School of Hygiene & Tropical Medicine (1899–c.1960)*. Project Report. London School of Hygiene & Tropical Medicine, London. 2022
Hommes F, Monzó HB, Ferrand RA et al. The words we choose matter: recognising the importance of language in decolonising global health. *Lancet Glob Health* 2021; 9(7):e897–e898.
Hong WD, Benayoud F, Nixon GL et al. AWZ1066S, a highly specific anti-Wolbachia drug candidate for a short-course treatment of filariasis. *Proc Nat Acad Sci* 2019; 116:1414–1419.
Jullien S, Sinclair D, Garner P. The impact of mass deworming programmes on schooling and economic development: an appraisal of long-term studies. *Int J Epidemiol* 2016; 2140–2215.
Kamgno J, Pion SD, Chesnais CB et al. A test-and-not-treat strategy for onchocerciasis in Loa loa-endemic Areas. *N Engl J Med* 2017; 377(21):2044–2052.

ARTHROPOD ECTOPARASITES

Arthropods, which include the arachnid ticks and mites as well as insects, may be responsible for human disease in several ways.

Local hypersensitivity reactions

Local lesions may be caused by hypersensitivity to allergens in arthropod saliva. This common reaction, known as papular urticaria, is non-specific and is seen in the majority of people in response to the bite of a variety of blood-sucking arthropods, including mosquitoes, bugs, ticks, lice, and mites (Fig. 42.35).

Occasionally, tick bites may cause a more severe systemic allergic response. In the USA and Australia but also elsewhere, ticks (e.g., the lone star tick that is also the vector for Rocky Mountain spotted fever) transmit 'alpha-gal' allergy (to alpha-galactose, a carbohydrate found in meat and dairy products). This is also known as ***mammalian meat allergy*** because the allergen is transmitted in tick saliva after feeding on mammals. Affected humans are sensitized to this allergen and react with urticaria and other atopic features after subsequent meat ingestion. The saliva of some tick species contains a neurotoxin, which causes flaccid paralysis in the victim after a bite; this condition is rare but occurs worldwide, most commonly in children.

Resident ectoparasite infections

Most ectoparasites alight on humans only to feed (although some species of lice live in very close proximity to the skin: body lice in clothing and head and pubic lice on human hairs; see Chapter 43, 'Pediculosis pubis' section). However, some types of parasite reside within the skin, causing more specific local lesions.

Scabies

See Chapter 45 ('Scabies' section).

Jiggers

Jiggers is caused by infection with the jigger flea, *Tunga penetrans*, and is common throughout South America and Africa. The pregnant female flea burrows into the sole of the foot, often between the toes. The egg sac grows to about 0.5 cm in size, and then the eggs are discharged on to the ground. The main danger is bacterial infection or tetanus. The flea should be removed with a needle or scalpel and the area kept clean until it heals.

Fig. 42.35 Papular urticaria caused by bedbug bites.

Myiasis

Myiasis is caused by invasion of human tissue by the larva of certain flies, principally the Tumbu fly *Cordylobia anthropophaga* (found in sub-Saharan Africa) and the human botfly *Dermatobia hominis* (Central and South America). The larvae, which hatch from eggs laid on laundry and linen, burrow into the skin to form boil-like lesions; a central breathing orifice may be visible. Secondary infection rarely occurs. Covering the larva with petroleum jelly usually brings it up in search of air, and it can then be easily extracted, although the American species may require surgical extraction.

Systemic envenoming

Many arthropods can cause local or systemic illness through envenoming, that is, injection of venom. A detailed discussion of spider, scorpion, and other bites is beyond the scope of this chapter (see Further reading and online topics).

Vectors of infection

By far the most important role of arthropods in causing human disease is as vectors of parasitic and viral infections. Some of these infections are shown in **Box 42.15** and are discussed in detail elsewhere.

Further reading

Jenkins T, Ahmadi S, Bittenbinder A et al. Terrestrial venomous animals, the envenomings they cause, and treatment perspectives in the Middle East and North Africa. *PLoS Negl Trop Dis* 2021; 15(12):e0009880.

See Elsevier eBooks+ for additional online content on Crimean–Congo haemorrhagic fever, dengue, and Ebola virus disease.

Questions for this chapter can be found on Elsevier eBooks+.

CRIMEAN–CONGO HAEMORRHAGIC FEVER

Somia Iqtadar

Crimean–Congo haemorrhagic fever (CCHF) is a viral haemorrhagic fever caused by the Nairovirus of the Bunyaviridae family and is transmitted to humans by the bite of the *Hyalomma* tick or by direct contact with blood of an infected animal or human. CCHF is a severe disease with a high case fatality rate ranging from 10% to 40%. The disease was first described in the Crimea in 1944 and identified in 1956 in the Congo, hence the name.

Geographical distribution

CCHF is endemic in all of Africa, the Balkans, the Middle East, and Asia. CCHF is now spreading to other areas, including parts of Europe. The World Health Organization estimates that about 3 billion people live in areas at risk of the disease.

Transmission

Ticks are both a reservoir and a vector for the CCHF virus. Numerous wild and domestic animals, such as cattle, goats, sheep, and hares, serve as amplifying hosts for the virus. Animals become infected by the bite of infected ticks, and the virus remains in their bloodstream for about 1 week after infection, allowing the tick–animal–tick cycle to continue when another tick bites. Although a number of tick genera are capable of becoming infected with CCHF virus, ticks of the genus *Hyalomma* are the principal vector (Fig. E42.1).

Population migration with animals contributes to the higher probability of susceptible animals being bitten by infected ticks, thus increasing the risk of transmission to humans who handle the animals.

The CCHF virus is transmitted to people either by tick bites or through contact with infected animal blood or tissues during and immediately after slaughter. The majority of cases have occurred in people involved in the livestock industry, such as agricultural workers, slaughterhouse workers, and veterinarians.

Human-to-human transmission can result from close contact with the blood, secretions, organs, or other bodily fluids of infected persons. (Hospital-acquired infections can also occur because of improper sterilization of medical equipment, reuse of needles, and contamination of medical supplies.)

Fig. E42.1 **A tick.**

Clinical course

CCHF presents in four distinct stages: incubation, pre-haemorrhagic, haemorrhagic, and convalescence.

Incubation period

The length of the incubation period depends on the mode of acquisition. The incubation period after tick bite is usually 1 to 3 days, with a maximum of 9 virus days. The incubation period after contact with infected blood or tissues is usually 5 to 6 days, with a documented maximum of 13 days.

Pre-haemorrhagic stage

This stage lasts about 2 to 4 days on average (range, 1–7 days) and begins with sudden onset of fever (>38°C), chills, myalgia, dizziness, neck pain and stiffness, backache, headache, sore eyes, and photophobia. There may be nausea, vomiting, diarrhoea, abdominal pain, and sore throat early on followed by sharp mood swings and confusion. After 2 to 4 days, the agitation may be replaced by sleepiness, depression, and lassitude, and the abdominal pain may localize to the upper right quadrant, with clinically detectable hepatomegaly and jaundice. Other clinical signs include tachycardia, lymphadenopathy, and splenomegaly.

Haemorrhagic stage

The haemorrhagic stage is usually short (~2–3 days) but can be prolonged up to 2 weeks. It is characterized by petechial rash on internal mucosal surfaces, such as in the mouth and throat, and on the skin. The petechiae may give way to larger rashes called ecchymoses, as well as other haemorrhagic phenomena.

Haematology and blood chemistry commonly show thrombocytopenia, leucopenia and elevated levels of aspartate aminotransferase (AST), alanine aminotransferase (ALT), lactate dehydrogenase, and creatine phosphokinase along with elevated levels of inflammatory cytokines. Coagulation may be affected, with prolonged prothrombin and activated partial thromboplastin times accompanied by a decrease in fibrinogen levels and an increase in the levels of fibrinogen degradation products. The haemorrhagic stage is pronounced in severe cases, with rapid progression to disseminated intravascular coagulation (DIC), sudden liver failure, rapid kidney deterioration, or pulmonary failure and shock. If lethal, death usually occurs in the second week of illness.

Convalescence

In survivors, convalescence generally begins around 9 to 10 days after the onset of illness (range, 9–20 days) and is associated with a return to normal for laboratory parameters. This stage can be prolonged and may be associated with hypotension, tachycardia or bradycardia, polyneuritis, breathing issues, xerostomia, vision and hearing deficiencies, hair loss, and memory loss, among others.

Investigations

CCHF is confirmed by blood sample, performed in specially equipped laboratories with a high biosafety level, if any of the following tests is positive:

- Confirmation of the presence of immunoglobulin (Ig) M or IgG antibodies seroconversion by enzyme-linked immunosorbent assay

- Positive antigen detection test
- Detection of viral nucleic acid in the specimen by reverse transcription polymerase chain reaction (RT-PCR)
- Isolation of the virus.

Patients with fatal disease, as well as in those in the first few days of illness, do not usually develop a measurable antibody response, so diagnosis in these individuals is achieved by virus or RNA detection in blood or tissue samples. Later in the course of the disease, in people surviving, antibodies can be found in the blood, although at this stage, the antigen, viral RNA, and virus are no longer detectable.

Management

When CCHF is suspected, isolation precautions should start, health facility staff alerted, the case immediately reported, blood samples drawn for CCHF diagnostic confirmation, and the treatment protocol listed next started without waiting for confirmation.

Patients should be isolated and cared for using barrier-nursing techniques, including masks, goggles, gloves, gowns, and the proper removal and disposal of contaminated articles.

Treatment

General supportive therapy is the mainstay of patient management in CCHF. Intensive monitoring to guide volume and blood component replacement is recommended.

If there is clinical suspicion of CCHF, a ribavirin treatment protocol needs to be initiated immediately with the consent of the patient or relatives and the attending physician (**Box E42.1**).

Note that ribavirin is not a specific treatment for CCHF viral infection To date, the nucleoside analogue ribavirin is the only direct-acting antiviral that has been widely used clinically in patients with CCHF. However, the use of ribavirin is controversial, with continued debate about whether treatment improves outcome. Pregnancy should be prevented (whether the patient is the female or male partner) for a period 6 months after completing a course of ribavirin.

Favipiravir has also been studied in CCHF and has shown promise in preclinical animal models. However, efficacy data in humans infected with CCHFV are limited.

Prophylaxis

When there has been known direct contact with the blood or secretions of a probable or confirmed case (e.g., needle-stick injury or contact with mucous membranes such as in the eye or mouth), the recommended procedure is to perform baseline blood studies and start the person on the ribavirin protocol.

Household or other contacts of the case who may have had the same exposure to infected ticks or animals or who recall indirect contact with case body fluids should be monitored for 14 days from the date of last contact with the patient or other source of infection by having their temperature taken twice daily. If the patient develops a temperature of 38.5°C or higher, headache, and muscle pains, they should be admitted to hospital and started on ribavirin treatment.

Box E42.1 Treatment protocol for Crimean–Congo haemorrhagic fever disease

High-dose oral ribavirin therapy consists of:

- 2-g loading dose
- 4 g/day in four divided doses (6-hourly) for 4 days
- 2 g/day in four divided doses for 6 days

Prevention and control

It is difficult to prevent or control CCHF infection in animals and ticks because the tick–animal–tick cycle usually goes unnoticed, and the infection is usually not apparent in domestic animals. Furthermore, the tick vectors are numerous and widespread, so tick control with acaricides (chemicals intended to kill ticks) is only a realistic option for well-managed livestock production facilities.

Although an inactivated, mouse brain–derived vaccine against CCHF has been developed and used on a small scale in Eastern Europe, there is currently no safe and effective vaccine widely available for human use. Further research is needed to develop these potential vaccines, as well as to determine the efficacy of the different treatment options, including that of ribavirin and other antiviral drugs.

There is no safe and effective vaccine currently available for human use for CCHF. A new vaccine against deadly CCHF is being trialled in humans for the first time. The ChAdOx2 CCHF vaccine is based on the ChAdOx2 vector, a weakened version of a common cold virus (adenovirus) that has been genetically modified so it is impossible for it to replicate in humans.

In the absence of an approved vaccine, the only way to reduce infection in humans is by raising awareness of the risk factors and educating people about the measures they can take to reduce exposure to the virus.

Public health advice should focus on several aspects, as follows.

Reducing the risk of tick-to-human transmission

- Wear protective clothing (long sleeves, long trousers).
- Wear light-coloured clothing to allow easy detection of ticks on the clothes.
- Use approved acaricides on clothing.
- Use approved repellent on the skin and clothing.
- Examine clothing and skin regularly for ticks; if any are found, remove them safely.
- Seek to eliminate or control tick infestations on animals or in stables and barns.
- Avoid areas where ticks are abundant and seasons when they are most active.

Reducing the risk of animal-to-human transmission

- Wear gloves and other protective clothing while handling animals or their tissues in endemic areas, notably during slaughtering, butchering, and culling procedures in slaughterhouses or at home.
- Quarantine animals before they enter slaughterhouses or routinely treat animals with pesticides 2 weeks before slaughter.

Reducing the risk of human-to-human transmission in the community

- Avoid close physical contact with CCHF-infected people.
- Wear gloves and protective equipment when taking care of ill people.
- Wash hands regularly after caring for or visiting ill people.
- If the patient with CCHF dies, advise the family to follow safe burial practices (**Box E42.2**).

Controlling infection in healthcare settings

Healthcare workers caring for patients with suspected or confirmed CCHF or handling specimens from them should implement standard infection control precautions. These include basic hand hygiene, use of personal protective equipment, safe injection procedures, and safe burial practices.

DENGUE

Somia Iqtadar

Epidemiology

Dengue has emerged as one of the most important mosquito-borne viral diseases of humans. It can carry a significant morbidity and mortality, especially if improperly managed. It is predominantly a disease of urban and semi-urban areas in the tropics and subtropics. With climate change and global warming, the mosquito vector has extended its range from tropical regions deeper into subtropical zones. The spread of dengue fever therefore illustrates how global trade (e.g., allowing the transport of the mosquito vectors in used tyres), increasing travel within and between countries (permitting the movement of viraemic individuals), urban overcrowding (being conducive to multiple infections from a single infected mosquito), and ineffective vector-control strategies have supported a pandemic in the modern era.

Virology

Infection is caused by dengue virus, flavivirus of family Flaviviridae. It is transmitted by *Aedes aegypti* and *Aedes albopictus*. There are four distinct serotypes: DEN-1, -2, -3, and -4. Each episode of infection induces life-long protective immunity to the homologous serotype but confers only partial and transient protection against subsequent infection by the other three serotypes. Secondary infection (by another serotype) is a major risk factor for dengue haemorrhagic fever (DHF), mainly due to antibody-induced enhancement.

Epidemiological studies have identified young age, female sex, high body mass index, virus strain, and genetic variants of the human major histocompatibility complex class I–related sequence B and phospholipase C epsilon 1 genes as risk factors for severe dengue.

Spectrum of disease

The incubation period (time between viral invasion and onset of clinical disease) for dengue infection is generally between 4 and 7 days (range, 3–14 days). The spectrum of clinical manifestations varies from totally asymptomatic to undifferentiated febrile illness (viral syndrome), dengue fever (DF) and DHF, including dengue shock syndrome (DSS) (Fig. E42.2).

Box E42.2 Safe burial practices for patients with viral haemorrhagic fever

- Long, thick rubber gloves or two pairs of surgical gloves should be used for washing the body for burial.
- The dead body should be sprayed with 1:10 liquid bleach solution and then wrapped in the winding sheet.
- The winding sheet should be sprayed with bleach solution.
- The body should then be placed in a plastic bag, which should be sealed with adhesive tape.
- The transport vehicle should be disinfected, and all of the deceased person's clothing should be burned.

Revised dengue classification

Clinical course of dengue infection

After the incubation period, the illness begins abruptly. The subsequent clinical course of dengue disease can be highly variable but may be broadly divided into three phases: febrile, critical, and convalescent (Fig. E42.3).

Febrile phase

Typically, this begins with the sudden onset of high-grade fever. This acute febrile phase generally lasts for 2 to 7 days and is often accompanied by facial flushing, skin erythema, generalized body ache, myalgia, arthralgia, and headache. There is often associated anorexia, nausea, and vomiting. Some patients may have a sore throat with congested pharynx and conjunctivae. These initial

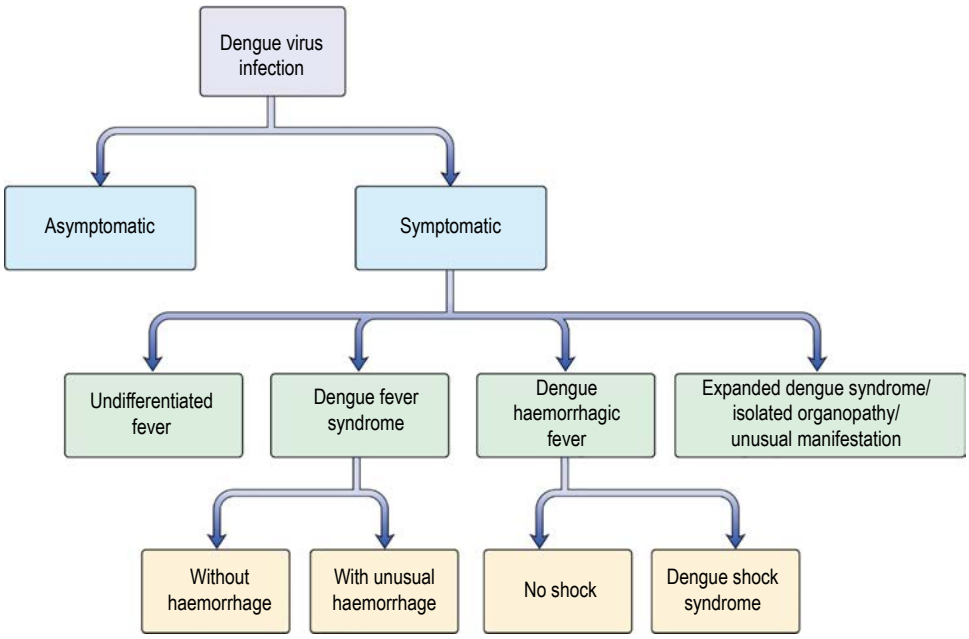


Fig. E42.2 Classification of dengue infection. *(From World Health Organization.* Comprehensive Guidelines for Prevention and Control of Dengue and Dengue Haemorrhagic Fever, *revised and expanded edition. World Health Organization, Regional Office for South-East Asia; 2011.)*

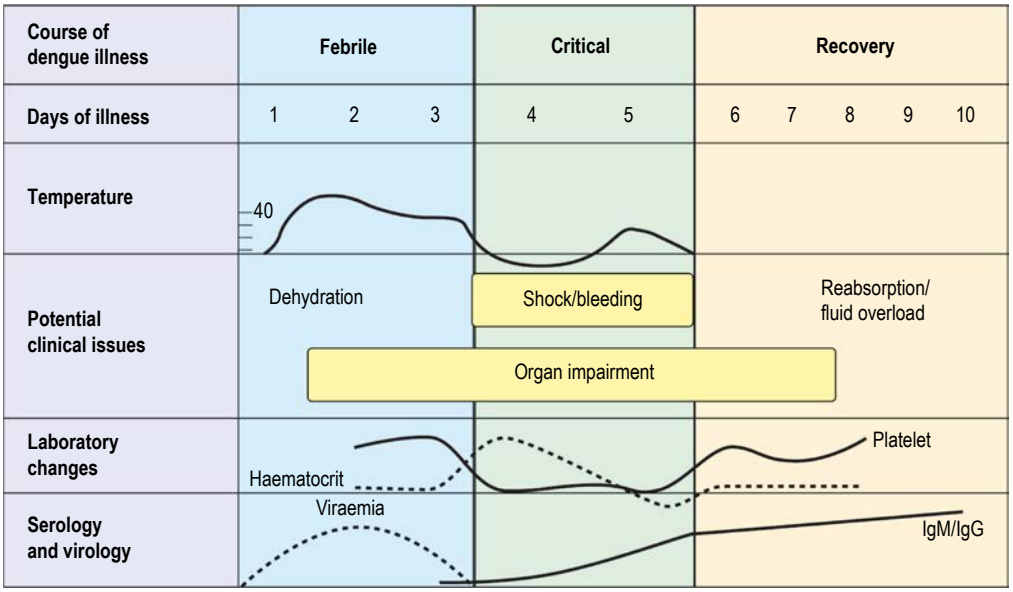


Fig. E42.3 Clinical course of dengue virus infection. *Ig,* Immunoglobulin. *(Simmons CP, et al. Current concepts: Dengue*. N Engl J Med. *2012 Apr;366:1423-32. Tricou V, Minh NN, Farrar J, et al. Kinetics of viremia and NS1 antigenemia are shaped by immune status and virus serotype in adults with dengue*. PLoS Negl Trop Dis. *2011 Sep;5(9):e1309*. https://doi.org/10.1371/journal.pntd.0001309. *Duyen HT, Ngoc TV, Ha do T, et al. Kinetics of plasma viremia and soluble nonstructural protein 1 concentrations in dengue: differential effects according to serotype and immune status*. J Infect Dis. *2011 May 1;203(9):1292-300*. https://doi.org/10.1093/infdis/jir014.)

clinical features are common to both patients who have simple DF and those who will go on to develop DHF subsequently.

On physical examination, mild haemorrhagic manifestations such as mucosal bleeding, petechial hemorrhages and a positive tourniquet test result are equally common in both DF and DHF. Vaginal bleeding is also seen frequently in young adult females. Uncommonly, however, massive vaginal or gastrointestinal bleeding may occur during this phase. Hepatomegaly is common, but tender hepatomegaly is highly suggestive of DHF.

Investigations

The earliest haematological abnormality is a progressive decrease in the total white cell count (WCC). This non-specific viral response should alert the physician to the possibility of dengue, particularly during a dengue epidemic. Dengue virus protein (NS1 antigen) and PCR test results are generally positive during this early stage.

Critical phase

Toward the end of the febrile phase—around the time of defervescence (usually between the third and fifth days of illness but may be up to the seventh day)—a few patients enter a stage of increased capillary permeability. Unlike in other routine viral infections, in which the patient's condition tends to improve with defervescence, in DHF at this point, depending on the capillary leak, the patient can proceed in two different clinical directions. Those without significant plasma leak gradually convalesce, but those who develop major plasma leak may deteriorate in the face of critical loss of volume.

The critical phase typically lasts for about 24 to 48 hours. During this stage, varying degrees of circulatory disturbances can develop. In less severe cases, these changes are minimal and transient. Many of these patients recover with routine oral fluid and electrolytes or even with non-specific management at home. In more severe forms of plasma leakage, significant volume depletion occurs. A compensatory response in the form of increased sympathetic drive kicks in. Now the patient becomes restless, with cold, clammy skin; a rapid, thready pulse; and prolonged capillary refill time. As the diastolic blood pressure rises (increased sympathetic tone) in the face of unchanged systolic pressure (caused by vasoconstriction), the pulse pressure narrows. Abdominal pain; persistent vomiting; restlessness; altered conscious level; clinical fluid accumulation; mucosal bleed; and a tender, enlarged liver are the clinical warning signs of severe dengue, with an increased possibility of rapid progression to shock. In this stage, if fluid resuscitation is not instituted promptly and appropriately, the patient can progress rapidly to profound and irreversible shock. It is important to note that haemoconcentration (a rising trend of haematocrit [Hct] from the baseline value) and thrombocytopenia are usually detectable even before the subsidence of fever and the onset of shock. It gives credence to the notion that a low-grade plasma leak has started before haemodynamic compromise has become clinically apparent. The Hct level correlates well with the amount of plasma volume loss and disease severity. The Hct may not, however, truly represent the volume loss in cases of frank haemorrhage or early and excessive fluid replacement or when Hct determination is wrongly timed. The biochemical abnormalities usually observed include leucopenia with a relative lymphocytosis, prolonged prothrombin time or activated partial thromboplastin time, elevated transaminases (typically, aspartate AST more than three times alanine ALT), hypoproteinaemia, and hypoalbuminaemia.

Convalescent phase

Plasma leak stops within 24 to 48 hours of time of onset and is followed by reabsorption of extravascular fluid. The patient's general wellbeing improves, appetite returns, gastrointestinal symptoms abate, haemodynamic status stabilizes (WCCs improve followed

Fig. E42.4 **The convalescent rash of dengue fever.**

by improvement in platelet count), and a diuresis ensues. Some patients may have the classical rash of 'isles of white in a sea of red' (Fig. E42.4). Some may experience generalized pruritus. Bradycardia and electrocardiographic changes are not uncommon during this stage. It is important to note that during this phase, the Hct level may drop further because of resorption-associated haemodilution. The recovery of platelet count is typically preceded by recovery of WCC.

Pathophysiology of plasma leakage

The initial febrile phase of dengue disease mimics the picture of common acute viral illnesses. Toward the end of the febrile illness, however, features of increased capillary permeability and plasma leak are witnessed in a small subset of patients. Although it might well be speculated that increased plasma permeability is present in all cases of dengue disease, the plasma leak remains small and is clinically compensated in most patients. Only in a small percentage does this condition cross the critical threshold of compensatory mechanisms to become clinically significant.

This acute increase in vascular permeability beyond the critical limit of compensatory responses is the primary pathophysiological event that differentiates DHF and DSS from uncomplicated DF. This leak is responsible for loss of plasma into the extravascular compartment, giving rise to the haemoconcentration; hypovolaemic state; and, in extreme conditions, shock. In common with the classical volume-depleted state, it leads to reflex tachycardia and generalized vasoconstriction as a result of increased sympathetic activity. This compensated stage of shock is characterized by the responses initiated by the patient to overcome the stress of volume depletion. Physiological mechanisms (neural, hormonal, and biochemical) are triggered to compensate for the stress of hypovolaemia. The baroreceptors in the arteries detect a drop in blood pressure and cause the release of adrenaline (epinephrine) and noradrenaline (norepinephrine). These catecholamines cause vasoconstriction and tachycardia, the body's attempt to increase the blood pressure. Vasoconstriction results in tissue hypoxaemia and anaerobic glycolysis, which in turn give rise to acidosis. Acidosis-mediated respiratory centre stimulation causes hyperventilation (tachypnoea). By hyperventilating and washing out CO_2, the body is indirectly trying to correct the acidosis. Vasoconstriction can affect the kidneys, stimulating the renin–angiotensin–aldosterone axis to cause further vasoconstriction and conservation of fluid via increased renal reabsorption. Vasoconstriction that started in the splanchnic circulation can spread to kidneys, gastrointestinal tract, and other organs to divert blood preferentially to the heart, lungs, and brain. This reduction in renal circulation (glomerular filtration rate [GFR]) manifests itself in the form of reduced urine production.

Pathophysiological events that occur during DSS are exactly like those of 'classical hypovolaemic shock' caused by blood loss (or volume loss during gastroenteritis). The only difference is that in DHF, the volume lost to the serosal cavities is rich in proteins and is available for reabsorption (through the lymphatics) subsequently.

The clinical manifestations of vasoconstriction in various systems are as follows:

- ***Skin.*** This is cold and clammy, with pallor and a delayed capillary refill time (>2 seconds).
- ***Cardiovascular system.*** There are a raised diastolic blood pressure and a narrowing of pulse pressure, an effect of increased sympathetic outflow (≤30 mmHg).
- ***Renal system.*** As a result of reduced GFR, urinary output drops (<0.5 mL/kg per hour).
- ***Gastrointestinal system.*** There are nausea, vomiting, and abdominal pain.
- ***Central nervous system.*** As a function of reduced cerebral perfusion, there may be lethargy, restlessness, apprehension, impaired consciousness, and combative behaviour.
- ***Respiratory system.*** Tachypnoea (respiratory rate >20 breaths/min in adults) is a feature of shock.

The molecular mechanism responsible for the increased vascular permeability seen during DHF and DSS is still not well understood. It appears that immune hyperdrive results in massive overproduction of cytokines ('cytokine storm') caused by aberrant activation of T lymphocytes and disturbances of the homeostatic system involving regulatory T cells. High concentrations of mediators of inflammation, including C3a; C5a; tumour necrosis factor; interleukin (IL)-2, IL-6, and IL-10; interferon; and histamine have also been noted.

A second infection with a heterotypic dengue virus may impart an increased risk of developing DHF. Antibody-dependent enhancement of viral replication is believed to be responsible for this phenomenon. Dengue virus is released from the host cells in two forms. The immature form, unlike mature particles, is incapable of entering new host cells. Subneutralizing concentration of the cross-reacting antibody from the previous infection may opsonize these immature virus particles and allow the entry and replication of immature virion in the macrophage or mononuclear cells. This is mediated through Fc epitope-recognizing domains on the macrophages. T-cell activation is also enhanced. Profound T-cell activation with cell death during acute dengue infection may suppress or delay viral elimination, leading to the higher viral loads and enhanced immunopathology found in patients with DHF.

Expanded dengue syndrome

In recent years, with geographical spread and greater involvement of adults, the term 'expanded dengue syndrome' has been used to describe unusual manifestations of dengue infection. These include neurological, hepatic, renal, hematological, and other isolated organ involvement, which could be explained as complications of severe profound shock or associations with underlying host conditions, diseases, or co-infections. Some important manifestations are shown in Box E42.3.

Box E42.3 Expanded dengue syndrome (unusual or atypical manifestations of dengue)

System	Unusual manifestations
Neurological	Febrile seizures in young children Encephalopathy Encephalitis, aseptic meningitis Intracranial haemorrhages, thrombosis Mononeuropathies, polyneuropathies, Guillain–Barré syndrome Transverse myelitis
Gastrointestinal or hepatic	Hepatitis. fulminant hepatic failure Acalculous cholecystitis Acute pancreatitis
Renal	Acute renal failure Haemolytic uraemic syndrome Acute tubular necrosis
Cardiac	Conduction abnormalities Myocarditis Pericarditis
Respiratory	Acute respiratory distress syndrome Pulmonary haemorrhage Pleural effusion
Musculoskeletal	Myositis with raised creatine phosphokinase Rhabdomyolysis
Lymphoreticular/ bone marrow	Infection-associated haemophagocytic syndrome Haemophagocytic lymphohistiocytosis Idiopathic thrombocytopenic purpura Disseminated intravascular coagulation
Eye	Macular haemorrhage Conjunctival bleed Impaired visual acuity Optic neuritis
Others	Post-infectious fatigue syndrome, depression, hallucinations, psychosis Alopecia

Reproduced from Comprehensive Guidelines for Prevention and Control of Dengue and Dengue Haemorrhagic Fever, WHO (2011).

Box E42.4 Differential diagnosis of dengue haemorrhagic fever

Organism	Disease
Arboviruses	Chikungunya fever Zika
Other viral diseases	Crimean–Congo haemorrhagic fever (nairovirus) Measles Rubella Other viral exanthems Epstein–Barr virus Enteroviruses Influenza Hepatitis A Hantavirus
Bacterial diseases	Meningococcaemia Leptospirosis Typhoid fever Melioidosis Scarlet fever
Parasitic diseases	Malaria

Differential diagnosis of dengue illness

It is easy to misdiagnose dengue in non-endemic settings because the clinical features are non-specific and mimic many other diseases. Differential diagnoses are shown in **Box E42.4**.

Investigations

Laboratory investigations of DF include disease-monitoring laboratory tests and diagnostic tests.

Disease-monitoring laboratory tests

- Full blood count
  - WCC. At the beginning of the febrile phase, the WCC is usually normal but decreases rapidly as the disease progresses. It may show a relative lymphocytosis; this non-specific trend of leucopenia should raise the suspicion of possible dengue infection in appropriate settings.
  - Hct. Haemoconcentration, as depicted by a rising Hct, is a marker of plasma leakage in dengue infection and helps to differentiate between DF and DHF. Significant blood loss and early fluid replacement may mask this trend. It is of paramount importance to obtain a baseline Hct in the early febrile phase of disease. It will become useful later for early recognition of plasma leak (detection of a rising Hct level).
  - Platelet count. This is perhaps the most common (and most maligned) laboratory investigation in dengue infection. In the early febrile phase, the platelet count is usually within the normal range but decreases rapidly as the disease progresses to the late febrile phase or at defervescence and may remain low for the first few days of recovery. There is a significant correlation between disease severity and thrombocytopenia; however, thrombocytopenia does not directly predict bleeding.
- Liver function tests (LFTs). Abnormal LFTs in the form of elevated transaminases is common and is characterized by greater elevation of the AST compared with the ALT. The frequency and degree of elevation of the liver enzymes are higher with DHF than with DF.

  Key indicators are as follows:
- Leucopenia followed by progressive thrombocytopenia is suggestive of dengue infection.
- A rising Hct accompanying progressive thrombocytopenia is suggestive of DHF.
- In the absence of a baseline Hct level, a Hct value greater than 40% in adult females and greater than 46% in adult males should raise the suspicion of plasma leakage.

Diagnostic tests

Confirmatory laboratory tests include antibody detection (serology), virus isolation, detection of viral genetic materials (PCR), and detection of dengue virus protein (NS1 antigen). Dengue viraemia in a patient is short-lived, typically occurs 2 to 3 days before the onset of fever and lasts for 4 to 7 days of illness. During this period the dengue virus, its nucleic acid (PCR RNA) and circulating viral antigen (NS1) can be detected. In the early stages of the disease (within up to 6 days of onset), virus isolation or viral nucleic acid or antigen detection can be used to diagnose infection. At the end of the acute phase of infection, immunological tests (IgM and IgG antibodies) are the methods of choice for diagnosis.

Management of dengue infection

There is no currently available medication to combat the dengue virus after host invasion. The mainstay of dengue infection management remains symptomatic and supportive.

Box E42.5 Warning signs of impending severe dengue fever

- No clinical improvement or worsening of the patient's condition just before or during transition to the afebrile phase (defervescence) or as the disease progresses
- Persistent vomiting
- Severe abdominal pain
- Lethargy or restlessness
- Bleeding: epistaxis, melena, haematemesis, menorrhagia, dark-coloured urine (haemoglobinuria), or haematuria
- Dizziness
- Pale, cold, and clammy hands and feet
- Low or no urine output for 4–6 hours

Patients with DF should maintain an adequate oral intake of fluid to avoid dehydration. Paracetamol can be used for fever and myalgia. Aspirin and other non-steroidal anti-inflammatory agents should be avoided because of the risk of bleeding complications.

It is important to identify those who are at risk of developing severe disease and in need of hospitalization because prompt and appropriate management of severe cases in hospital can effectively reduce the mortality rate associated with dengue infections. The parameters that need to be monitored include general condition, vital signs, serial Hct, full blood count, urine output, and presence of any warning signs (**Box E42.5**).

In DHF, aggressive intravascular volume repletion with intravenous crystalloids and colloids is indicated to prevent or reverse hypovolaemic shock. In mild cases, oral rehydration may be sufficient. Intravenous fluids are indicated if the patient is vomiting or if the Hct continues to rise by 10% to 20% despite oral rehydration. Isotonic crystalloid solutions should be used throughout the critical period, except in very young infants in whom 0.45% saline may be used. In obese patients, ideal body weight should be used as a guide to calculating fluid requirements. Blood transfusions may be needed only for patients with severe bleeding manifestations. Most patients with DSS respond to 10 mL/kg fluid in children or 300 to 500 mL in adults given over 1 hour or as a bolus. The rate of fluid administration is then reduced gradually to 7, 5, 3, then 1.5 mL/kg/h with regular monitoring of clinical condition, vital signs, urine output, and Hct. In patients with profound shock, if shock is not reversible after the initial fluid bolus, this can be repeated, and colloids and blood transfusion should be considered. Fluid that is lost into tissue spaces during the period of plasma leakage is subsequently reabsorbed rapidly. Thus, care should be taken to avoid hypervolaemia and pulmonary oedema. Intravenous fluids should generally be discontinued after 24 to 48 hours. There is no evidence that corticosteroids are of benefit in patients with DSS.

Patients can be discharged from hospital when they have been afebrile for more than 24 hours and have normal oral intake and urine output, rising platelet counts, and stable Hct and vital signs.

Prevention

Prevention is dependent on control of the vector, *A. aegypti*. The most effective preventive strategy is mosquito control to avoid artificial stagnation of water to reduce breeding sites. Mosquito barriers, such as insecticide-impregnated mosquito nets and insect repellents, are needed for patients until fever subsides to prevent the vector mosquito from biting and acquiring the virus.

A novel technique in which selected strains of *Wolbachia* have been introduced into *A. aegypti* to prevent transmission of dengue virus by the vector is now used in many countries with good results. *Wolbachia* prevents dengue transmission by either directly blocking the virus or by decreasing the lifespan of the vector.

Two vaccines to protect against dengue are available for use. Dengvaxia a three-dose vaccine is approved for use in people of 9 to 16 years with laboratory confirmation of previous dengue virus infection in those living in dengue endemic areas.

Qdenga is approved in the EU, UK, Brazil, Argentina, Indonesia, and Thailand.

EBOLA VIRUS DISEASE

Somia Iqtadar

Ebola virus disease (EVD) and its late complication, Ebola haemorrhagic fever (EHF), are loosely grouped as belonging to the viral haemorrhagic fevers but are distinguished by their severity and fatality. EVD can infect humans and a range of non-human primates (monkeys, gorillas, and chimpanzees), dogs, and pigs.

The causative agent of EVD is Ebola virus, which belongs to the Filoviridae family, genus Ebola virus; it is a pleomorphic, negative-sense RNA virus. The first Ebola virus species was discovered in 1976 in what is now the Democratic Republic of the Congo near the Ebola River, hence its name. Since then, outbreaks have appeared sporadically around this geographical area. The 2014 outbreak in West Africa was the largest and most complex Ebola outbreak to date; it spread rapidly, causing many casualties.

Five subspecies of Ebola virus have been identified to date. Four of these have been implicated in human infection: Zaire Ebola virus, Sudan Ebola virus, Taï Forest Ebola virus (formerly Côte d'Ivoire Ebola virus), and Bundibugyo Ebola virus. The fifth, Reston Ebola virus, has been identified only in non-human primates and has not caused verified EVD in humans.

The natural reservoir host of Ebola viruses remains unknown. However, on the basis of available evidence and the nature of similar viruses, researchers believe that the virus is zoonotic (animal borne), with bats being the most likely reservoir.

Transmission

The manner in which the virus is first acquired by the index case at the start of an outbreak remains speculative. It appears to be through contact with an infected animal, but when the infection becomes established in a human host, it may spread from person to person very efficiently via faecal–oral, parenteral, or air-borne splashes. Ebola virus enters the host through:

- Direct contact with the blood or secretions (saliva, stools, vomitus) of an infected person
- Exposure to objects (e.g., needles) that have been contaminated with infected secretions
- Eating the meat of infected animals.

The viruses that cause EVD are often spread between families and friends because these individuals come into close contact with infectious secretions. During outbreaks of EVD, the disease can spread rapidly within healthcare settings and to healthcare professionals if barrier nursing and appropriate protective equipment, such as masks, gowns, and gloves, are not used.

Proper cleaning and disposal of instruments, such as thermometers, needles, and syringes, are also important, as is the disposal of highly infectious bodily waste and any bedding and equipment soiled with it. Without adequate sanitation, quarantine, and sterilization of items and instruments that cannot be disposed of, virus transmission can continue and amplify an outbreak.

Clinical features

When infection occurs, symptoms usually begin abruptly. Symptoms of EVD typically include fever, headache, joint and muscle aches, diarrhoea, vomiting, abdominal pain, rash, red eyes, cough, sore throat, and internal and external bleeding.

Clinical course

After the incubation period, which ranges from 2 to 21 days, the patient may show the following clinical features:

- ***Days 1 and 2.*** The patient often presents with non-specific flu-like symptoms with fever up to 39°C, profuse sweating, malaise, frontal and temporal headache, myalgia, retrobulbar pain, and conjunctival injection. Gastrointestinal symptoms dominate the clinical picture, with nausea, profuse vomiting, watery diarrhoea, and diffuse abdominal pain being the most common. Within 2 days, blood may appear both in vomitus and in stools. Respiratory symptoms such as sore throat and cough often follow the initial presentation. Relative bradycardia is the only sign at this stage.
- ***Days 3 and 4.*** A sore, tender throat with dysphagia becomes the predominant symptom in addition to the gastrointestinal features already described. The soft palate becomes erythematous. Dehydration usually starts to be clinically evident at this stage. Lymphadenopathy involving occipital, cervical, and axillary lymph nodes is seen in some patients.
- ***Days 5 to 7.*** By now, about 50% of patients begin to display a haemorrhagic diathesis with spontaneous epistaxis, gingival haemorrhage, gastrointestinal and genital bleeding (in females), haematuria, and bleeding at injection sites. Subconjunctival haemorrhages are also seen frequently. Signs include an erythematous, non-itchy rash, spreading from the face and buttocks to the trunk and arms, which may go on to become papular or maculopapular within 24 hours. Haemorrhagic lesions often become confluent within hours.
- ***Days 8 to 16.*** At this stage, dehydration and hypotension become severe in the absence of supportive care. Oedema may be present, and shock syndrome may prevail. Most deaths occur around day 12 with clinical evidence of multiorgan failure, involving the kidney and liver in particular. Central nervous system features, including fits and coma, may also be seen. Persistent hiccups, although rare, carry a poor prognosis. Among those who recover, the rash usually disappears around day 12; palmar and plantar desquamation occurs on days 14 to 16. Complications may include orchitis, residual hepatic dysfunction, transverse myelitis, or uveitis.

Differential diagnosis

Other viral haemorrhagic fevers, such as dengue and CCHF, and other conditions, such as acute surgical abdomen, malaria, leukaemia, and enteric fever, are the most important differential diagnoses.

Pathophysiology

Symptoms may appear anywhere from 2 to 21 days after exposure to Ebola virus, though an incubation period of 8 to 10 days is most common. Typically, Ebola virus infection runs its course within 14 to 21 days.

Cytokines are released when the virus invades reticuloendothelial cells, which initiates inflammatory responses that are not very protective. These cytokines produce initial non-specific flu-like symptoms such as fever, myalgia, and malaise. This early phase of infection is characterized by thrombocytopenia, leucopenia, and a pronounced lymphopenia. Neutrophilia develops after several days, as do elevations in AST and ALT. As the infection progresses, the cytopathic effects of infection and cytokine release result in DIC, bleeding, and capillary leak caused by microvascular endothelial cell damage and compromised vascular integrity. Bilirubin may be normal or slightly elevated. With the onset of anuria, blood urea nitrogen and serum creatinine increase. Terminally ill patients may develop a metabolic acidosis that may contribute to the often-observed tachypnoea.

Cutaneous manifestations are frequently marked. Bleeding from mucous membranes and puncture sites is reported in 40% to 50% of cases and maculopapular rashes in approximately 50%. Other cutaneous manifestations include petechiae, purpura, ecchymosis, and haematomas (especially around needle injection sites). All infected patients show some extent of coagulopathy and features of impaired perfusion. In general, the development of haemorrhagic symptoms is indicative of a poor prognosis.

However, contrary to popular belief, haemorrhage contributes little to the hypovolaemia and shock associated with EVD; instead, death supervenes because of multiple organ dysfunction syndrome caused by capillary leak, fluid redistribution, hypotension, DIC, and focal tissue necrosis.

Diagnosis

Diagnosing EVD in an individual who has been infected for only a few days is difficult because the early symptoms, such as fever, abdominal pain, red eyes, and a skin rash, are non-specific and are often seen in individuals with many different viral infections.

However, if a person demonstrates these symptoms along with a history of travel to an endemic area or contact with someone who has EVD, the individual must be considered as a suspected case of EVD. The patient should be isolated, appropriate investigations should be performed, and public health professionals should be notified. Samples from the patient can then be collected and tested to confirm infection.

Investigations

These are shown in **Box E42.6**.

Management

Therapeutics

There are currently two treatments approved by the US Food and Drug Administration (FDA) to treat EVD caused by the Ebola virus, species Zaire ebolavirus, in adults and children.

The first drug approved in October 2020, Inmazeb, is a combination of three monoclonal antibodies (mAbs) atoltivimab, maftivimab, and odesivimab.

The second drug, Ebanga, is a single mAb and was approved in December 2020.

Box E42.6 Laboratory tests used in the diagnosis of Ebola virus disease

Timeline of infection	Diagnostic tests available
Within a few days of symptom onset	NP-antigen capture ELISA testing IgM ELISA PCR (RT-PCR) Virus isolation
Later in the disease course or after recovery	IgM and IgG antibodies
Retrospectively in deceased patients	Immunohistochemistry testing RT-PCR Virus isolation

ELISA, Enzyme-linked immunosorbent assay; *Ig,* immunoglobulin; *NP,* nucleoprotein; *RT-PCR,* reverse transcription polymerase chain reaction.

mAbs are proteins produced in a laboratory or other manufacturing facility that act like natural antibodies to stop a germ such as a virus from replicating after it has infected a person. These particular mAbs bind to a portion of the Ebola virus's surface called the glycoprotein, which prevents the virus from entering a person's cells.

Both of these treatments were evaluated in a randomized controlled trial during the 2018 to 2020 Ebola outbreak in the Democratic Republic of the Congo. The overall survival rate was much higher for patients receiving either of the two treatments. Neither Inmazeb nor Ebanga has been evaluated for efficacy against species other than Zaire ebolavirus.

Supportive care

Whether or not other treatments are available, basic interventions can significantly improve chances of survival when provided early. These are referred to as supportive care, and include:

- Maintenance of fluid and electrolyte balance
- Maintenance of blood pressure and oxygen saturation to prevent tissue hypoxia
- Early detection and treatment of superimposed infections.

Patients can produce infectious virions for prolonged periods. Therefore, strict barrier isolation in a room away from traffic patterns must be maintained throughout the illness. Early notification of public health professionals and contact screening are essential. The patient's urine, stool, sputum, and blood, along with any objects that have come in contact with the individual or their bodily fluids (e.g., laboratory equipment), are all highly contagious and must be incinerated or disinfected with a 0.5% sodium hypochlorite solution. Those who have died of EVD should be buried promptly, with standard protocols observed for contagion limitation.

Prevention

Because the disease is highly infectious, prevention of EVD presents many challenges. Strict application of isolation measures and barrier nursing is the cornerstone of management of all contagious viral haemorrhagic fevers, including EVD.

A safe and effective vaccine is an important tool to protect and prevent the introduction and spread of Ebola.

There are two licensed Ebola vaccines. Both vaccines are safe and protective against the species Zaire ebolavirus only.

Ebola Zaire vaccine (rVSV-ZEBOV) Ervebo is genetically engineered to express a glycoprotein from Zaire ebolavirus to provoke a neutralizing immune response. It was licensed in November 2019 by the European Medicines Agency and prequalified by WHO. The US FDA licensed the vaccine in December 2019. It is recommended by the Strategic Advisory Group of Experts on Immunization as part as a broader set of Ebola outbreak response tools.

A second new vaccine is delivered in two doses called Zabdeno (Ad26.ZEBOV) and Mvabea (MVA-BN-Filo) for individuals 1 year and older.

The vaccine is delivered in two doses: Zabdeno is administered first, and Mvabea is given approximately 8 weeks later as a second dose. This prophylactic two-dose regimen is therefore not suitable for an outbreak response in which immediate protection is necessary.

For individuals at imminent risk of exposure to Ebola (e.g., healthcare professionals and those living in or visiting areas with an ongoing EBOV outbreak) who completed the Zabdeno and Mvabea two-dose vaccination regimen, a Zabdeno booster vaccination should be considered if more than 4 months have passed since the second dose was administered.

43 Sexually transmitted infections

Janet D. Wilson

Clinical skills 1519
History 1519
Examination in the sexual health clinic 1519
Investigations 1519
Sexually transmitted infections 1519
Asymptomatic screening 1520
Investigations for symptomatic patients 1523
Management, prevention, and control of STIs 1523
Biomedical prevention 1524
Specific infections 1525

CORE SKILLS AND KNOWLEDGE

Sexually transmitted infections (STIs) are among the most common causes of illness in the world and remain endemic in all societies. They rank among the top five disease categories for which adults seek healthcare worldwide.

The medical specialty of genitourinary medicine encompasses the diagnoses and treatment of STI and human immunodeficiency virus (HIV). It also includes the public health control of these infections through screening, partner notification and contract tracing, and health promotion, as well as provision of vaccines and other biomedical interventions to protect against STI and HIV acquisition in high-risk populations. The specialty is multidisciplinary and includes physicians, specialist nurses, health advisers, psychologists, microbiologists, virologists, pharmacists, and public health clinicians.

Key STI skills include:

- Taking a sexual history and obtaining information for partner notification
- Performing a sexual health examination and investigations and correctly interpreting results
- Discussing prevention interventions to ensure that people can avoid acquiring or transmitting STIs and HIV.

The best place to experience the care of people with STIs is in the sexual health clinic or via specialists services for those living with HIV (see Chapter 44).

CLINICAL SKILLS

History

Patients presenting with possible STIs may be anxious, embarrassed, and concerned about confidentiality. Clinicians must be alert to these issues and respond sensitively. The clinical setting must ensure privacy and confidentiality. The clinical history is tailored to the presentation encountered, but many of the points in **Box 43.1** are likely to be relevant.

Examination in the sexual health clinic

The purpose and procedure of the examination should be explained and informed consent obtained. Examination should be in a private room with a lockable door and a chaperone present. The patient should be provided with a gown or paper drapes to cover areas not being examined. Ensuring the patient is relaxed and in a comfortable position will result in a better experience and enable a thorough examination with properly taken swabs. The process of anogenital examination is shown in **Figs. 43.1 to 43.3**. If the person is asymptomatic, clinical examination may not be indicated.

Investigations

Although investigation is guided by the history and examination, multiple infections may coexist, and some may be asymptomatic. The minimum investigations are tests for chlamydia, gonorrhoea (including samples from extra-genital sites if indicated by the sexual history), syphilis, and HIV. HIV antibody testing should be performed on an 'opt-out' basis. Screening for hepatitis viruses should be performed in those at increased risk. The recommended investigations for different clinical presentations are shown in **Boxes 43.2 to 43.8**. The recommended tests for specific infections are detailed in this chapter.

SEXUALLY TRANSMITTED INFECTIONS

The World Health Organization (WHO) estimated in 2020 that more than 1 million people acquire an STI every day. Each year, there are an estimated 374 million new infections in people aged 15 to 49 years with one of the four major curable STIs (chlamydia, gonorrhoea, syphilis, and trichomoniasis) worldwide. These include 7.1 million cases of syphilis, 82 million cases of gonorrhoea, 129 million cases of chlamydia, and 156 million cases of trichomoniasis.

Box 43.1 Taking a clinical history in patients with a sexually transmitted infection

Symptom history relevant to sexually transmitted infections

- STIs may be asymptomatic
- Genital symptoms
  - Urethral discharge in males
  - Vaginal discharge in females
  - Lower abdominal pain in females
  - Genital lumps
  - Genital ulceration
  - Genital itching
  - Rectal symptoms
- Systemic symptoms: fever, skin rash, joint pains, eye symptoms

Sexual history

- Number and types of recent sexual contacts, with dates
- Sex of partners
- Regular or casual sexual contact. If regular, how long have they been having sex?
- Use of condoms
- Any known symptoms or STI diagnosis in the partner
- Sexual practices (e.g., insertive or receptive; vaginal, anal, pharyngeal, oro-anal)
- Whether the patient has ever paid or been paid for sex
- Country of origin of sexual partners

Previous history of STIs and ongoing risk exposure

- Previous STIs, including dates and treatment received
- HIV testing and results
- Hepatitis A and B, and HPV vaccination status
- Risk factors for ongoing acquisition of HIV and STIs
  - Frequent partner change
  - Unprotected sex
  - Use of drugs or alcohol to a level that reduces safer sex
  - Current or history of injecting drugs by patient and by any sexual partners
  - GBMSM
  - Sexual partners from countries with high HIV or other STI prevalence

Offer risk reduction advice and make patients aware of post-exposure HIV prophylaxis after sexual exposure and pre-exposure prophylaxis.

Wider history

- General medical history
- Medication history, including recent antibacterial or antiviral treatment
- Menstrual, obstetric, and cervical cytology screening histories
- Details of current contraception
- Any past or current drug or alcohol misuse

GBMSM, Gay, bisexual, and other males who have sex with males; *HIV,* human immunodeficiency virus; *HPV,* human papillomavirus; *STI,* sexually transmitted infection.

In 2016 more than 490 million people were estimated to be living with genital herpes, and more than 300 million females were estimated to have human papillomavirus (HPV) infections, the primary cause of cervical cancer. The estimates suggest that most of these infections are in low-income countries.

In England there has been an overall increasing trend in STIs over the past decade before a drop occurred in 2020 and 2021 during the COVID-19 pandemic. However, by 2022 cases of gonorrhoea and infectious syphilis (primary, secondary, and early latent) had exceeded numbers in 2019 to 79 268 (16.1% increase) and 8693 (8.1% increase) cases, respectively. There were further increases between 2022 and 2023 and the number of gonorrhoea diagnoses in 2023 was the largest annual number reported since records began in 1918, and the number of syphilis diagnoses was the largest since 1948. Gonorrhoea is increasing in people of all ages, but the increase is highest among people aged 15 to 24 years. Infectious syphilis is increasing in both heterosexual persons, as well as gay, bisexual, and other males who have sex with males (GBMSM) but in 2023 there was a larger proportional increase in syphilis diagnoses among heterosexual persons than GBMSM. Between 2021 and 2022, chlamydia diagnoses increased by 24.3% from 160 279 to 199 233, and diagnoses of first episode genital herpes increased 13.8% from 21 892 to 24 910. STIs are associated with frequent partner change and inconsistent use of condoms. Increased travel and recreational drug and alcohol use are also implicated. The only decreasing trend is in diagnoses of first episode genital warts which decreased by 8.5% from 28 497 to 26 079 between 2021 and 2022 due to the established HPV vaccination program. Amongst the largely vaccinated age group of 15 to 17 year olds diagnoses remained low at 104 in 2022 and 107 in 2023.

Despite high global incidence and substantial morbidity, STIs remain neglected in clinical practice, public health, and research. There have been rapid increases in antimicrobial resistance in *Neisseria gonorrhoeae* and *Mycoplasma genitalium*, with the possibility of untreatable cases. STIs often coexist, and the inflammation caused by one STI can increase the risk of acquisition of others; most infections increase the acquisition and transmission of HIV.

ASYMPTOMATIC SCREENING

All STIs can be asymptomatic, and multiple infections frequently coexist; hence, many people without symptoms attend sexual health clinics to seek sexual health checks and advice. Sexual health checks offer individual and public health benefits because treatment of asymptomatic infections reduces the period of infectivity, prevents long-term complications, and reduces onward transmission.

Increasingly, sexual health clinics are making asymptomatic sexual health checks available online because routine examination of those with no symptoms yields few additional diagnoses. A short electronic medical history is self-completed to confirm the absence of symptoms, and a sexual history and risk assessment for STIs, HIV, and other blood-borne viruses determines which investigations are needed. The self-sampling kit is posted to the person with instructions of how to take samples and package them safely. These are posted back to the service for testing, and results are returned electronically.

The minimum investigations, even if the person is asymptomatic, are tests for chlamydia, gonorrhoea (including samples from extragenital sites if indicated by the sexual history), syphilis, and HIV. HIV antibody testing should be performed on an 'opt-out' basis. If it is declined, the reasons why should be documented because it may need to be discussed again at a follow-up visit if the patient is at high risk for HIV infection.

Hepatitis A screening and vaccination

Screening for hepatitis viruses should be performed in those at increased risk.

A test for antibodies to hepatitis A, with vaccination if negative, should be performed in:

- GBMSM
- People who inject drugs (PWID)

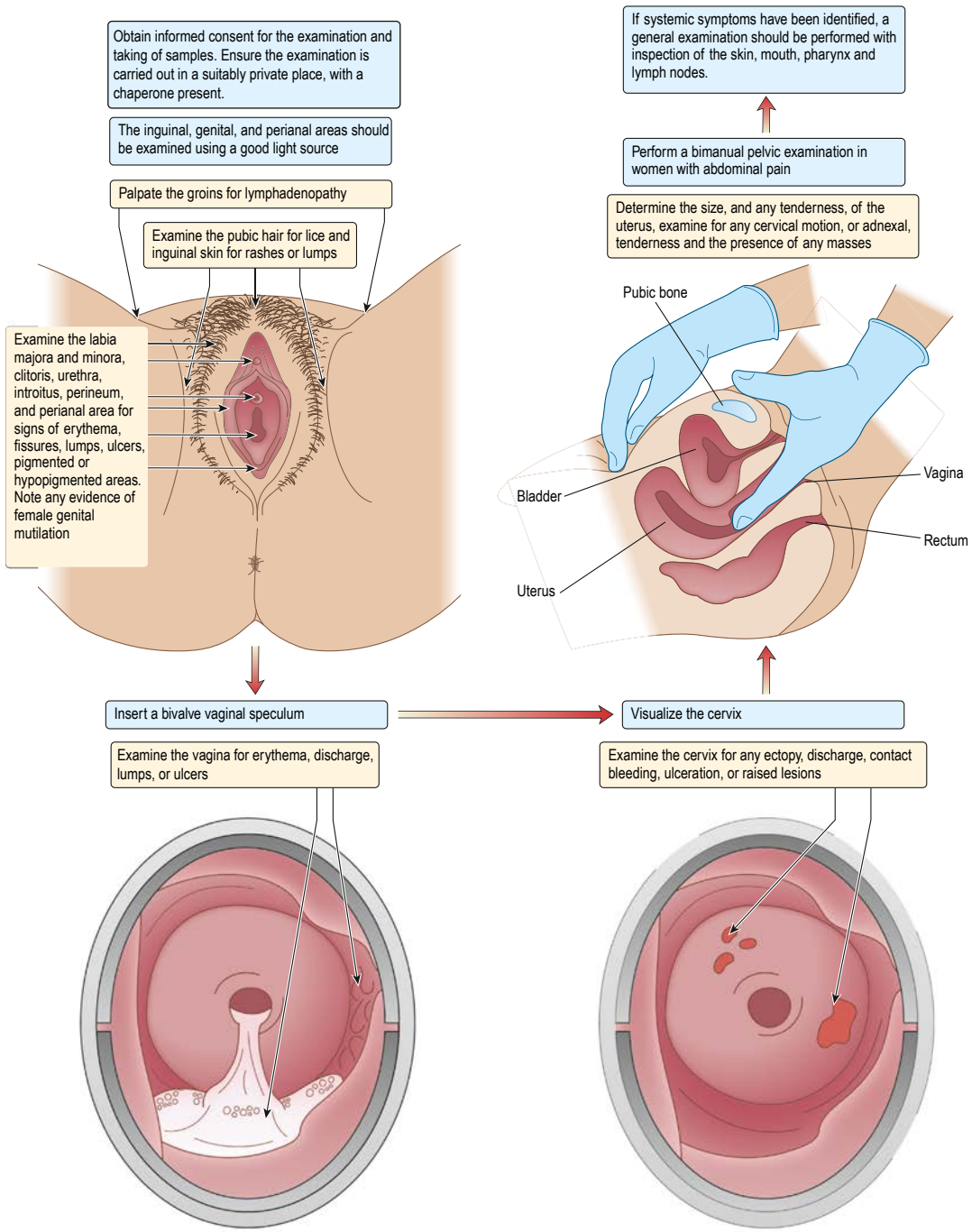


Fig. 43.1 **Examination of the vulva, vagina, and cervix.**

- People with chronic hepatitis B or C
- People living with HIV (PLWH)

Hepatitis B screening and vaccination

The recommended screening test in those who have not been immunized is hepatitis B core antibody (anti-HBc), which is a marker of past infection, or hepatitis B surface antigen (HBsAg), which identifies currently infected individuals. Screening is recommended for:

- Sexual partners of those who are HBsAg positive
- GBMSM and their sexual partners
- People who have been sexually assaulted
- Sex workers
- PWID and their sexual partners
- Individuals, or those with a partner, from countries with intermediate or high prevalence of chronic hepatitis B (defined as 2% or more and including all countries outside Western Europe, North America, and Australasia)
- Needle-stick injury recipients
- People with chronic hepatitis C
- PLWH
- Those born to mothers infected with hepatitis B.

Patients who are negative and have ongoing increased risk should be offered vaccination.

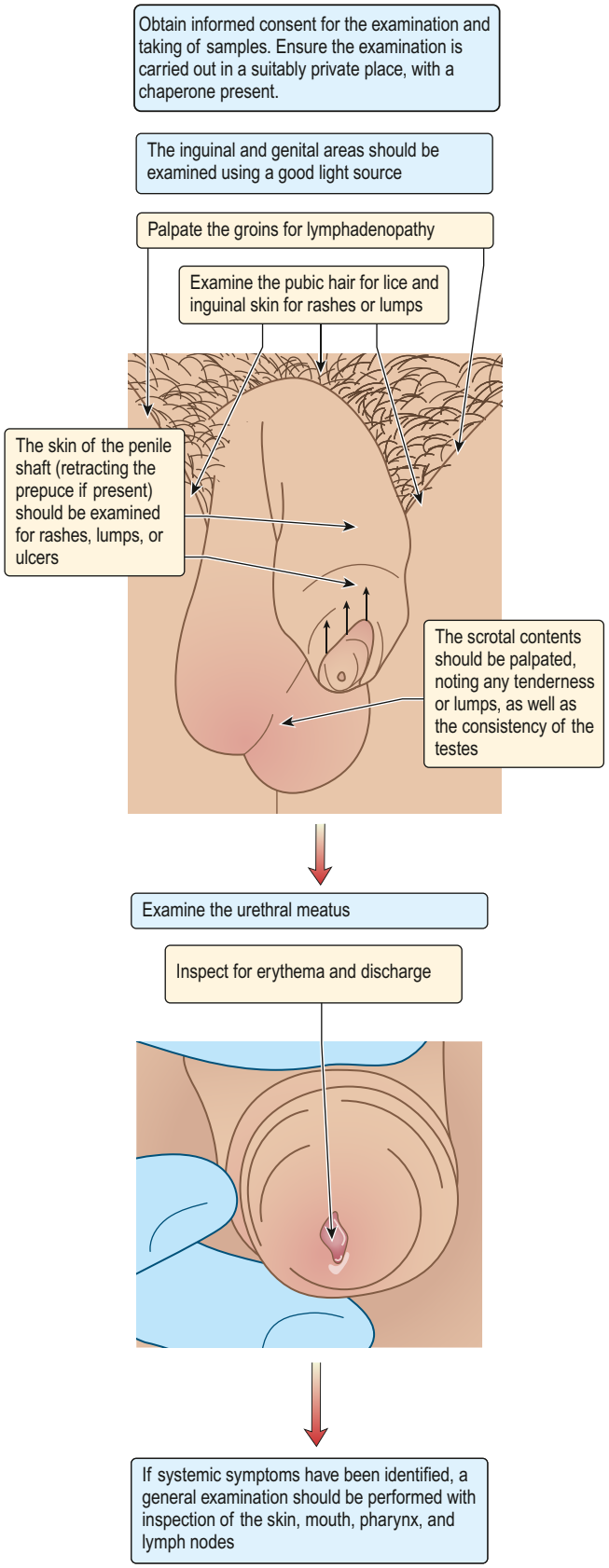


Fig. 43.2 **Examination of the penis, penile meatus, and scrotal contents.**

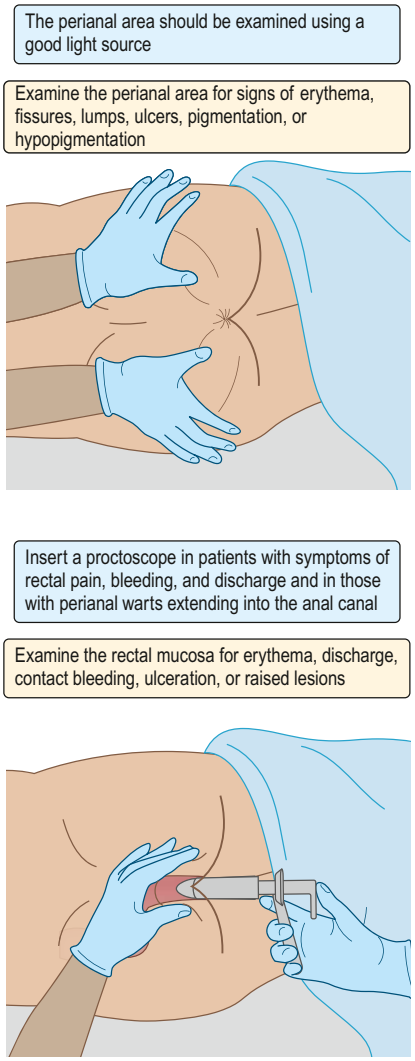


Fig. 43.3 **Examination of the perianal area and anal canal.**

Hepatitis C screening

Screening for antibodies to hepatitis C is recommended in:

- Sexual partners of persons who are hepatitis C virus (HCV) positive
- PWID
- GBMSM who are taking or are eligible for pre-exposure prophylaxis (PrEP) for HIV (see Chapter 44)
- GBMSM with additional risk factors, e.g., sex associated with trauma or injury, recreational drug use, rectal lymphogranuloma venereum (LGV)
- Sex workers
- Individuals from countries with intermediate or high prevalence of chronic hepatitis C (as defined earlier for hepatitis B)
- Needle-stick injury recipients
- People with chronic hepatitis B
- PLWH
- Those born to mothers infected with HCV.

Box 43.2 Urethritis

Presentation
- Males: urethral discharge and dysuria
- Females: usually asymptomatic; may cause mild dysuria

Infective causes

Usually divided into gonococcal (caused by *Neisseria gonorrhoeae*) or non-gonococcal
- *Chlamydia trachomatis*[a]
- *Mycoplasma genitalium*[b]
- *Ureaplasma urealyticum*
- *Neisseria gonorrhoeae*
- Trichomonas vaginalis[c]
- Herpes simplex virus[c]
- UTI[d]
- Adenoviruses (often associated with conjunctivitis)[c]

Non-infective causes
- Non-specific urethritis (where no cause can be identified)
- Physical or chemical trauma
- Urethral stricture

Features that help discriminate between causes
- Urethral stricture
- Colour of discharge: green-yellow with gonorrhoea
- Urinary symptoms: dysuria, urinary frequency, nocturia, or haematuria may suggest UTI
- Testicular pain or swelling: gonorrhoea, chlamydia, and UTI can cause epididymo-orchitis
- Other symptoms: genital sores suggest HSV, rash, or sore eyes suggest adenovirus

Investigations
- Microscopy of urethral discharge
- NAAT and culture for *N. gonorrhoeae*
- NAAT for *C. trachomatis*
- NAAT for *M. genitalium*
- Serology for syphilis and HIV
- Tests for *T. vaginalis* and HSV are not usually performed routinely
- No commercial tests are available for *U. urealyticum*
- A MSU if symptoms are suggestive of UTI

[a]The most common cause of non-gonococcal urethritis (NGU).
[b]The second most common cause of NGU.
[c]Rarer cause of NGU.
[d]Most frequent non-sexually transmitted cause of NGU.
HSV, Herpes simplex virus; *MSU,* midstream specimen of urine; *NAAT,* nucleic acid amplification test; *UTI,* urinary tract infection.

Box 43.3 Vaginal discharge

Infective causes

Vaginal infections
- BV[a]
- *Candida albicansa*[a]
- *Trichomonas vaginalis*[b]

Cervical infections
- *Chlamydia trachomatis*
- *Neisseria gonorrhoeae*
- Herpes simplex virus

Non-infective causes
- Physiological discharge or cervical ectopy[c]
- Cervical polyps
- Neoplasms
- Retained products (e.g., of conception, tampons)
- Chemical irritation

Features that help discriminate between causes
- Odour: BV and *T. vaginalis* can have an offensive odour.
- Vulval itching or soreness: *Candida* spp. and *T. vaginalis* can both cause vulval irritation.
- Other symptoms: dysuria suggests UTI, abdominal pain may suggest spread of gonorrhoea or chlamydia into the upper genital tract, and intermenstrual or postcoital bleeding that persists after treatment may warrant gynaecology assessment.

Investigations
- Microscopy of vaginal discharge for BV, *Candida* spp., and *T. vaginalis*
- Culture for *Candida* spp.
- NAAT for *T. vaginalis*
- NAAT for *N. gonorrhoeae* and *C. trachomatis*
- Serology for syphilis and HIV

[a]The most common causes of altered vaginal discharge, but *C. trachomatis* and *N. gonorrhoeae* infections must also be considered.
[b]The most common cause worldwide but rarer in high-income countries.
[c]Common diagnosis but requires exclusion of infective causes.
BV, Bacterial vaginosis; *HIV,* human immunodeficiency virus.

INVESTIGATIONS FOR SYMPTOMATIC PATIENTS

These depend on the clinical presentation but should always include tests for chlamydia, gonorrhoea, syphilis, and HIV. The recommended investigations for these presentations are shown in **Boxes 43.2 to 43.8**.

MANAGEMENT, PREVENTION, AND CONTROL OF STIs

The treatment of specific conditions is covered in the appropriate sections. Most sexual health clinics give directly observed therapy when possible or dispense medication directly to the patient to improve adherence and prevent onward transmission.

Tracing the sexual partners of patients (partner notification) is crucial in controlling spread of STIs. The aims are to identify and treat those with infections (particularly those with asymptomatic infections) to break the chain of transmission. Appropriate antibiotic therapy may be offered to recent sexual partners of those known to have active infections (epidemiological treatment). Interviewing people about their sexual partners requires tact and sensitivity, and specialist health advisers are available for this in sexual health clinics.

Many STI diagnoses are in young people (16–24 years). Prevention starts with education and information. People often begin sexual activity at young ages, so education programmes need to include school pupils, as well as young adults. Education of health professionals is also required. Primary public health programmes aimed at informing groups most at risk of STIs, including GBMSM, sex workers, young people, and Black and minority ethnic groups to help modify sexual behaviour, should be implemented at local and national levels. Vaccination against hepatitis A and B should be given to those at increased risk, and HPV vaccine should be given to those eligible through the national HPV vaccination programme for GBMSM delivered through sexual health clinics. Secondary interventions include prompt diagnosis and management of STIs to reduce complications and onward transmission. The National Chlamydia Screening Programme in England aims to provide earlier detection and treatment for chlamydia for sexually active females younger than 25 years old by providing easy access to chlamydia testing online and in community settings.

Reducing the numbers of sexual partners, avoiding overlapping relationships, using condoms correctly and consistently, and avoiding sex if symptoms are present can reduce the risks of acquiring

Box 43.4 Lower abdominal pain

Infective causes

- PID[a,b]
  - *Chlamydia trachomatis*[c]
  - *Neisseria gonorrhoeae*
  - *Mycoplasma genitalium*
  - BV
- UTI

Non-infective causes

- Ectopic pregnancy
- Acute appendicitis
- Endometriosis
- Irritable bowel syndrome
- Neoplasms
- Torsion or haemorrhage of ovarian cyst

Features that help discriminate between causes

- Site, character, and duration of pain
- Vaginal discharge, postcoital or intermenstrual bleeding, deep dyspareunia
- Date of last menstrual period and contraception. Is pregnancy possible?
- Dysuria, urinary frequency, nocturia or haematuria
- Nausea, vomiting, diarrhoea, or constipation

Investigations

- Microscopy of vaginal discharge for BV and *T. vaginalis*
- NAAT for *T. vaginalis*
- NAAT and culture for *N. gonorrhoeae*
- NAAT for *C. trachomatis*
- NAAT for *M. genitalium*
- Serology for syphilis and HIV
- Pregnancy test to exclude ectopic pregnancy
- MSU for culture if urinary symptoms

[a] Most common in sexually active individuals younger than 25 years of age.
[b] Symptoms of abnormal vaginal discharge or abnormal bleeding and bacterial vaginosis (BV) are usually present with pelvic inflammatory disease (PID).
[c] The most common cause of PID.
HIV, Human immunodeficiency virus; *MSU,* midstream specimen of urine; *NAAT,* nucleic acid amplification test; *UTI,* urinary tract infection.

Box 43.5 Genital lumps

Infective causes

- HPV, anogenital warts[a,b]
- Molluscum contagiosum[a,b]
- *Sarcoptes scabiei* (scabies): excoriated lesions
- *Treponema pallidum:* secondary condylomata lata[c]

Non-infective causes

- Normal anatomical variants: papillae, sebaceous glands, skin tags
- Sebaceous cysts
- Neoplasms or premalignant disease (e.g., perianal, vulval, and penile intraepithelial neoplasia)

Features that help discriminate between causes

- Lumps: anogenital warts can occur on mucous membranes and keratinized skin; molluscum contagiosum and scabetic papules only occur on keratinized skin.
- Single or multiple lesions: painful single lesions suggest malignant or premalignant conditions, anogenital warts, molluscum contagiosum and syphilitic condylomata lata are usually multiple
- Itch: scabetic itchy papules typically present on the scrotum, penile shaft, and labia majora, as well as classic body areas.
- Pain: neoplasms and premalignant conditions are often sore or painful.
- Length of time: if present before first sexual activity, likely to be a normal anatomical variant.
- Lumps or rashes elsewhere: scabetic itchy papules are usually visible on classic areas of body; widespread skin rash of secondary syphilis usually present with condylomata lata.

Investigations

- NAAT for *Neisseria gonorrhoeae*
- NAAT for *Chlamydia trachomatis*
- Serology for syphilis and HIV
- Clinical diagnoses of anogenital warts (HPV testing not appropriate) and molluscum contagiosum

[a] The most common infective cause followed by molluscum contagiosum.
[b] Mucous membrane lumps suggest anogenital warts; molluscum contagiosum does not occur on mucous membranes.
[c] Moist papules are rare in secondary syphilis but easily misdiagnosed as genital warts.
HPV, Human papillomavirus; *NAAT,* nucleic acid amplification test.

and transmitting an STI. For those who change their sexual partners frequently, regular sexual health checks (approximately 3-monthly) are advisable. If people develop symptoms, they should be encouraged to seek medical advice as soon as possible to reduce complications and spread to others.

Biomedical prevention

Prevention of HIV acquisition is part of a sexual health service. The universal offering of HIV testing contributes to reducing the pool of PLWH who are undiagnosed because many new infections are transmitted from people who are unaware of their HIV infection. It is possible to reduce acquisition of HIV infection after high-risk sexual exposure if triple anti-retroviral therapy (ART) is started as soon as possible. This is termed 'HIV post-exposure prophylaxis (PEP)'. Initiation within 72 hours reduces dissemination and replication of virus in tissues. The recommended first-line PEP regimen is tenofovir disoproxil 245 mg/emtricitabine 200 mg with raltegravir 1200 mg once daily for 28 days.

Tenofovir disoproxil 245 mg/emtricitabine 200 mg can also be used as HIV PrEP for people who are not infected with HIV. This has been shown to be highly effective at preventing HIV acquisition in a large real-world study conducted from sexual health clinics in England. With more than 24 000 participants, the PrEP Impact trial demonstrated that HIV PrEP can reduce HIV acquisition by 86% over several years of use. HIV PrEP has been routinely provided to people at higher risk of acquiring HIV at UK sexual health clinics since October 2020 (see also Chapter 44).

Further reading

Migchelsen SJ, Enayat Q, Harb AK et al. *Sexually Transmitted Infections and Screening for Chlamydia in England, 2022*. UK Health Security Agency: London; 2023 Available at: https://www.gov.uk/government/statistics/sexually-transmitted-infections-stis-annual-data-tables/sexually-transmitted-infections-and-screening-for-chlamydia-in-england-2022-report#:~:text=in%202022%20there%20were%20392%2C453,(COVID%2D19)%20pandemic.
Mitjà O, Suñer C, Giacani L et al. Treatment of bacterial sexually transmitted infections in Europe: gonorrhoea, Mycoplasma genitalium, and syphilis. *Lancet Reg Health Eur*. 2023; 34:100737.
Sullivan AK, Saunders J, Desai M et al. HIV pre-exposure prophylaxis and its implementation in the PrEP Impact Trial in England: a pragmatic health technology assessment. *Lancet HIV*. 2023; 10:e790–806.
Waltmann, A, Chen, JS, Duncan, JA. Promising developments in gonococcal vaccines. *Curr Opin Infect Dis.* 2024; 37:63–69.
World Health Organization. *Sexually Transmitted Infections (STIs)* Newsletter 10 July 2023 Available at https://www.who.int/news-room/fact-sheets/detail/sexually-transmitted-infections-(stis)#:~:text=STIs%20have%20a%20profound%20impact,and%20trichomoniasis%20(156%20million.

Box 43.6 Genital ulceration

Infective causes

HSV types 1 and 2[a]
- Primary
- Recurrent

Treponema pallidum[b]
- Primary chancre
- Secondary mucous patches
- Tertiary gumma

Herpes zoster[c]
LGV[d]
Mpox[d]
Chancroid[e]
Donovanosis (granuloma inguinale)[e]

Non-infective causes
- Trauma
- Aphthous ulceration (e.g., Behçet's syndrome)[f]
- Lichen sclerosus
- Erosive lichen planus
- Fixed drug eruptions
- Stevens–Johnson syndrome
- Crohn's disease
- Neoplasms

Features that help distinguish between causes
- Pain: HSV, herpes zoster, and non-infective causes are usually painful.
- Single or multiple: HSV and herpes zoster are usually multiple.
- Length of time: it is more likely to be non-infective if present for several weeks.
- External dysuria: this is a classic symptom of HSV in females.
- Ulcers or rash elsewhere: patients with syphilis, mpox, and non-infective dermatological causes usually have skin signs elsewhere.
- Systematic symptoms: these can occur with primary HSV, early syphilis, herpes zoster, mpox, non-infective dermatological causes, and Crohn's disease.

Investigations
- Ulcer swab for PCR for HSV-1, HSV-2, and *T. pallidum*
- NAAT for *Neisseria gonorrhoeae*
- NAAT for *Chlamydia trachomatis*
- Serology for syphilis and HIV (with repeat syphilis serology after a 3-month window period if appropriate)
- In GBMSM, consider:
  Ulcer swab for *C. trachomatis* NAAT with genotyping for LGV if positive mpox PCR

[a] The most common cause of multiple painful ulcers.
[b] Syphilis ulcers are usually single and non-painful.
[c] Rarer than herpes simplex virus (HSV); characterized by a unilateral presentation.
[d] Lymphogranuloma venereum (LGV) and mpox are rare except in gay, bisexual, and other males who have sex with males (GBMSM).
[e] Chancroid and donovanosis extremely rare.
[f] Most common non-infective cause, usually similar lesions in mouth.
HIV, Human immunodeficiency virus; *NAAT,* nucleic acid amplification test; *PCR,* polymerase chain reaction.

Box 43.7 Genital itching

Infective causes
- *Candida albicans*[a]
- *Trichomonas vaginalis*
- Phthirus pubis (pubic lice)[b]
- Sarcoptes scabiei (scabies)[c]

Non-infective causes[d]
- Irritant dermatitis
- Genital eczema
- Lichen sclerosus
- Lichen planus

Features that help distinguish between causes
- Site and duration of itch: itching confined to genital area with *Candida* spp. and *T. vaginalis.*
- Vaginal discharge, offensive odour, or vulval swelling: suggest *Candida* spp., *T. vaginalis*, or possibly irritant dermatitis.
- Itching or rash elsewhere or any known skin problems: pubic lice, scabies, and non-infective causes.

Investigations
- Microscopy of vaginal discharge for candida and *T. vaginalis*
- Culture for *Candida* spp.
- NAAT for *T. vaginalis*
- NAAT for *Neisseria gonorrhoeae* and *Chlamydia trachomatis*
- Serology for syphilis and HIV

[a] The most frequent symptom in females is vulval itching rather than vaginal discharge.
[b] Relatively rare but easily visible in the hair area.
[c] May present with excoriated papules caused by excessive scratching.
[d] Genital dermatoses are common causes of genital itching, frequently misdiagnosed as *Candida* spp.
HIV, Human immunodeficiency virus; *NAAT,* nucleic acid amplification test.

SPECIFIC INFECTIONS

Human immunodeficiency virus and acquired immunodeficiency virus

This is discussed in Chapter 44.

Hepatitis B

This is discussed in Chapter 23. Sexual contacts should be screened and immunized if they are not immune.

Chlamydia trachomatis

Chlamydia trachomatis is the most common STI in the UK; up to 10% of sexually active people younger than the age of 25 years are infected. It infects the urethra, endocervix, rectum, pharynx, and conjunctiva. Up to 50% of males and 80% of females may be asymptomatic, so it is frequently unrecognized and therefore untreated. Pelvic inflammatory disease (PID) is the main complication of *C. trachomatis* in females, leading to tubal infertility, ectopic pregnancy, and chronic pelvic pain, causing significant morbidity and cost to health services.

Clinical features

The incubation period is up to 14 days.

In males, *C. trachomatis* can cause an anterior urethritis with a mucopurulent urethral discharge (worse on waking, when there may be crusting at the meatus) and dysuria. The infection can ascend along the vas deferens, leading to epididymo-orchitis.

In females, the most common site of infection is the endocervix, but the urethra can also be infected. Symptoms may include increased vaginal discharge, dysuria, postcoital or intermenstrual bleeding, and lower abdominal pain. Examination of the cervix may reveal mucopurulent cervicitis or contact bleeding. Ascending infection into the uterus and fallopian tubes causes endometritis and acute salpingitis (PID). In pregnancy, *C. trachomatis* is associated with preterm birth, postpartum infection, and neonatal mucopurulent conjunctivitis or pneumonia caused by vertical transmission during vaginal delivery.

Rectal infection, through receptive anal sex, or via spread from the cervix or via the vagina in females may be asymptomatic but

Box 43.8 Rectal symptoms

Presentation
- Symptoms of proctitis are rectal pain, mucopurulent anal discharge, rectal bleeding, constipation, and tenesmus

Infective causes

Rectal infections
- *Neisseria gonorrhoeae*[a]
- *Chlamydia trachomatis*[a]
- LGV[b]
- HSV[c]
- *Treponema pallidum*[c]
- Mpox[b]

Enteric infections
- *Shigella* spp.[b]
- *Campylobacter* spp.
- *Entamoeba histolytica*
- *Salmonella* spp.
- *Escherichia coli*

Non-infective causes
- Ulcerative colitis
- Crohn's disease
- Neoplasms

Features that help distinguish between causes
- Irritation or, pain around the anus or rectum: suggest infective causes of proctitis.
- Anal discharge or bleeding: suggest infective causes of proctitis or neoplasm.
- Abdominal pain, diarrhoea, or blood in stool: suggest infective causes of enteritis or inflammatory bowel disease.
- Feeling of incomplete bowel evacuation or constipation: suggest infective causes of proctitis or neoplasm.
- Systemic symptoms: occur with primary HSV, early syphilis, mpox, enteritis, and inflammatory bowel disease.

Investigations
- Proctoscopy: mucosal erythema, oedema, contact bleeding, mucopurulent discharge, ulceration in severe cases
- Microscopy of a rectal smear, rectal culture for *N. gonorrhoeae*
- NAAT for *N. gonorrhoeae*
- NAAT for *C. trachomatis* (if positive for *C. trachomatis*, genotyping for LGV should be requested)
- Swab for HSV-1, HSV-2
- *T. pallidum* PCR
- Serology for syphilis and HIV
- LGV should be considered in GBMSM with suspected inflammatory bowel disease, as the clinical presentation can be very similar
- Cultures for enteric pathogens should be performed if symptoms of enteritis are present

[a] The most common sexually transmitted causes of proctitis.
[b] Outbreaks occurring in Europe in gay, bisexual, and other males who have sex with males (GBMSM).
[c] Rarer causes than *N. gonorrhoeae* and *C. trachomatis*; mainly seen in GBMSM.
HSV, Herpes simplex virus; *LGV,* lymphogranuloma venereum; *NAAT,* nucleic acid amplification test; *PCR,* polymerase chain reaction.

can cause proctitis. Reactive arthritis (see Chapter 32) can occur with *C. trachomatis* infection, particularly in human leucocyte antigen B27–positive individuals.

Diagnosis

Nucleic acid amplification tests (NAATs) are the diagnostic tests of choice with sensitivities of 95% to 99%. In males, first-voided urine (FVU) samples are used. In females, vulvovaginal swabs (VVSs) are used. FVU samples in males and self-taken VVSs are non-invasive, meaning they can be obtained by the patient without the need for an examination, so they are ideal for asymptomatic chlamydia screening. GBMSM who practice receptive anal sex and receptive oral sex should have rectal and pharyngeal swabs for *C. trachomatis* NAAT performed.

Management

Doxycycline 100 mg orally twice daily for 7 days is recommended for uncomplicated infection; oral azithromycin 1 g stat followed by 500 mg/day for 2 days is used if doxycycline is contraindicated. Azithromycin is recommended in pregnancy and lactation in WHO, US, UK and Australian guidelines. Doxycycline is contraindicated in pregnancy. Longer courses of antibiotics are required for complicated infections (see PID and epididymo-orchitis later). Epidemiological treatment before test result availability is usually offered to those who have had recent sex with someone who has confirmed *C. trachomatis* infection because the infection rate can be up to 75%. Abstinence from sex for at least 7 days or until treatment is completed should be advised. Sexual contacts must be traced, notified, and treated because many infections are asymptomatic. A follow-up consultation (usually by telephone) should assess adherence to therapy and partner notification. A routine test of cure is not necessary except in pregnancy, in complicated infections, and if symptoms persist or re-infection is suspected. NAAT results may remain positive for some time after treatment because they detect non-viable organisms, so a test of cure should be deferred until 4 weeks after completion of treatment.

Gonorrhoea

Neisseria gonorrhoeae is a Gram-negative intracellular diplococcus (see Fig. 43.4) that infects the urethra, endocervix, rectum, pharynx, and conjunctiva. Up to 10% of males and 50% of females are asymptomatic. Co-infection with *C. trachomatis* is common, occurring in 20% to 40% of gonococcal infections.

Clinical features

The incubation period is 2 to 14 days, and most symptoms in males occur between 2 and 5 days.

In males, *N. gonorrhoeae* can cause anterior urethritis with mucopurulent or purulent urethral discharge and dysuria (Fig. 43.5). The discharge can be profuse, causing staining of underwear. Complications include ascending infection, leading to epididymo-orchitis.

In females, symptoms may include increased vaginal discharge, dysuria, postcoital or intermenstrual bleeding, and lower abdominal pain. Examination of the cervix may reveal mucopurulent or purulent cervicitis or contact bleeding. Complications include Bartholin's abscesses, and ascending infection into the uterus and fallopian tubes causes endometritis and acute salpingitis (PID). In pregnancy, *N. gonorrhoeae* is associated with preterm birth, postpartum infection, and neonatal purulent conjunctivitis caused by vertical transmission during vaginal delivery.

Rectal infection, through receptive anal sex, or via spread from the cervix or via the vagina in females, may be asymptomatic but can cause proctitis. *N. gonorrhoeae* septicaemia (disseminated gonococcal infection [DGI]) is a rare complication presenting as fever, tenosynovitis, arthritis, and characteristic erythematous skin lesions with necrotic centres.

Diagnosis

NAATs are the diagnostic test of choice for *N. gonorrhoeae* because they have better sensitivity than culture. However, as *N. gonorrhoeae* antimicrobial resistance is increasing and ceftriaxone-resistant isolates have been identified in the UK and globally, culture on selective media should be performed before any treatment for

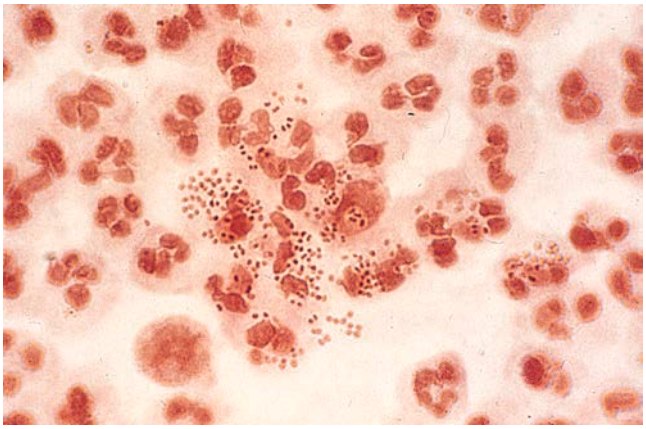


Fig. 43.4 ***Neisseria gonorrhoeae.*** Gram-negative intracellular diplococci. *(Courtesy Dr B. Goh.)*

Fig. 43.5 ***Neisseria gonorrhoeae.*** Mucopurulent urethral discharge.

N. gonorrhoeae being given. The specimens of choice for *N. gonorrhoeae* NAATs are FVU samples in males and VVSs in females. Self-taken VVSs are as sensitive at detecting *N. gonorrhoeae* as clinician-taken VVSs. FVU samples are less sensitive than VVSs in females. FVU samples in males and self-taken VVSs in females are non-invasive and therefore ideal for asymptomatic screening. GBMSM who practice receptive anal and receptive oral sex should have rectal and pharyngeal swabs for *N. gonorrhoeae* NAATs. Urethral swabs in males and endocervical swabs in females and rectal and pharyngeal swabs are used for culture.

Microscopy of Gram-stained urethral and endocervical secretions may demonstrate intracellular, Gram-negative diplococci, allowing rapid diagnosis (see Fig. 43.4). Sensitivity ranges from 90% to 95% in urethral specimens from symptomatic males to 37% to 50% in endocervical specimens. Microscopy should not be used for pharyngeal specimens because non-pathogenic *Neisseria* spp. colonize the mouth and look identical to *N. gonorrhoeae* microscopically. The sensitivity of blood and synovial fluid cultures is poor, so NAATs from the genital tract, rectum, and pharynx remain the tests of choice for investigations of DGI.

Management

Treatment should be given to those who have positive microscopy, a positive NAAT, or a positive culture for *N. gonorrhoeae*. Antibiotic-resistant strains of *N. gonorrhoeae* are increasing, so gonococcal culture for antibiotic susceptibility testing should be taken before treatment. Resistance to all the commonly used antibiotics has been reported in the UK, but ceftriaxone resistance is very rare. When antimicrobial susceptibility is not known before treatment, single-dose ceftriaxone 1 g intramuscularly (i.m.) is recommended in the UK. If confirmed susceptibility at all suspected sites of infection is known in advance of treatment, oral ciprofloxacin 500 mg single dose can be used. If there is a history of penicillin anaphylaxis or established cephalosporin allergy, gentamicin 240 mg i.m. should be used in combination with oral azithromycin 2 g. Ceftriaxone, gentamicin, and azithromycin can be used in pregnancy. Longer courses of antibiotics are required for complicated infections (see PID and epididymo-orchitis later) and DGI. Epidemiological treatment (treatment based on known exposure) pending test results can be offered to those who have had sex within the past 14 days with those who have confirmed *N. gonorrhoeae* infection. Abstinence from sex for at least 7 days or until treatment has been completed should be advised. All sexual contacts should be notified, tested, and treated if positive. A follow-up assessment and test of cure using *N. gonorrhoeae* NAAT should take place 14 days after treatment. This can be done using postal testing.

Non-gonococcal urethritis and *Mycoplasma genitalium*

Non-gonococcal urethritis (NGU) in males is characterized by a urethral discharge and dysuria. There are several causes, many of which are sexually transmitted; the most common of these is *C. trachomatis*, discussed earlier. Up to 25% of NGU cases may be caused by *Mycoplasma genitalium*. Other causes are *Ureaplasma urealyticum*, *Trichomonas vaginalis*, adenoviruses, and herpes simplex virus (HSV) 1 and 2, in that order of frequency. Studies investigating the cause of NGU have consistently identified no known cause in more than 60% of cases. Non–sexually transmitted causes of NGU include urinary tract infections (UTIs) and strictures.

Clinical features

The incubation period is usually 2 to 3 weeks. The main symptom is a mucopurulent urethral discharge (worse on waking when there may be crusting at the meatus). Dysuria is common but not universal. Discomfort or itching within the urethra may be present.

Diagnosis

Microscopy of Gram-stained urethral secretion showing five or more polymorphonuclear leucocytes per high-power (×1000 oil-immersion lens) field is diagnostic. Symptomatic males with no microscopic evidence of urethritis should be re-examined and tested after holding their urine overnight. Males with symptoms of urethritis should have NAATs for *N. gonorrhoeae*, *C. trachomatis*, and *M. genitalium* on an FVU sample. *M. genitalium*–positive specimens should be tested for macrolide resistance mutations because a high percentage of infections are macrolide resistant. Testing for *T. vaginalis* and HSV is only usually performed in persistent or recurrent NGU. There is no commercial test available for *U. urealyticum*. A midstream specimen of urine (MSU) should be taken if symptoms are suggestive of a UTI (urinary frequency, nocturia, urgency, haematuria).

Management

First-line therapy for NGU is oral doxycycline 100 mg twice daily for 7 days. Abstinence from sex for at least 7 days should be advised.

All sexual contacts should be notified, tested, and treated. If the patient is *M. genitalium* positive, and macrolide sensitive, azithromycin 1 g orally as a single dose followed by 500 mg/day for 2 days should be started within 2 weeks of finishing doxycycline. If the patient is *M. genitalium* positive and macrolide resistant, oral moxifloxacin 400 mg/day for 7 days should be given. This additional treatment is not needed if the *M. genitalium* NAAT result is negative. Follow-up is indicated if *C. trachomatis* or *M. genitalium* is confirmed or if symptoms persist.

Recurrent or persistent non-gonococcal urethritis

This is defined as persistent or recurrent symptomatic urethritis occurring 30 to 90 days after treatment of acute NGU and occurs in 10% to 20% of cases. The aetiology is probably multifactorial, but *M. genitalium* and *T. vaginalis* are likely causes. It is necessary to ensure that there is still microscopic evidence of urethritis, that there was good adherence to NGU treatment with sexual abstinence, and that sexual partners were also treated. If there is no microscopic evidence of urethritis, patients should be reassured and further antibiotic therapy avoided. If microscopy confirms persistent or recurrent NGU, subsequent treatment needs to cover *M. genitalium* and *T. vaginalis*. The recommended treatment is azithromycin 1 g orally as a single dose followed by 500 mg/day for 2 days with oral metronidazole 400 mg twice daily for 5 to 7 days. If the patient is *M. genitalium* positive and macrolide resistant or if treatment with azithromycin has already failed, oral moxifloxacin 400 mg/day for 7 days should be given. Current sexual partners of those infected with *M. genitalium* should also be tested and treated if positive with the same antibiotic regimen as their sexual partner. Abstinence from sex until both have finished their treatment should be advised. All who were *M. genitalium* positive should have a test of cure 5 weeks after the start of treatment.

Pelvic inflammatory disease

PID results when infections ascend from the cervix or vagina into the upper genital tract. It includes endometritis, salpingitis, tubo-ovarian abscess, and pelvic peritonitis. It is most frequent in sexually active females aged younger than 25 years. *C. trachomatis* and *N. gonorrhoeae* account for about 25% of UK cases and *M. genitalium* for 10% to 13% of cases. The anaerobic bacteria associated with bacterial vaginosis (BV) have also been implicated in incident PID. However, even in laparoscopically proven PID, often no bacterial cause is found. PID has serious long-term sequelae because of tubal damage and pelvic adhesions, resulting in infertility, increased risk of ectopic pregnancy, and chronic pelvic pain. The risk of sequelae increases with more severe and multiple episodes of PID. Tubal infertility occurs in 10% to 12% after one episode of PID and increases to 50% to 60% after three or more episodes. The risk of ectopic pregnancy is increased 6- to 10-fold, and abdominal or pelvic pain for longer than 6 months occurs in 18%.

Clinical features

The onset of symptoms often occurs in the first part of the menstrual cycle. Lower abdominal pain, which is usually bilateral, is the most common symptom. Other symptoms include increased vaginal discharge, irregular bleeding, deep dyspareunia, and dysuria. There may be a mucopurulent cervical discharge with contact bleeding. Lower abdominal tenderness and adnexal and cervical motion tenderness on bimanual examination are the most common signs.

Diagnosis

The diagnosis is usually made on the clinical findings of lower abdominal pain, with increased vaginal discharge, abnormal bleeding, and cervical motion or adnexal tenderness on bimanual examination supporting the diagnosis. However, clinical diagnosis has a specificity of only 65% to 70%. The differential diagnosis includes ectopic pregnancy, acute appendicitis, endometriosis, UTI, and irritable bowel disease.

Investigations should include:

- Microscopy of Gram-stained vaginal discharge for BV
- Microscopy of an endocervical specimen for evidence of inflammatory cells and intracellular, Gram-negative diplococci suggestive of gonorrhoea
- NAAT and culture for *N. gonorrhoeae*
- NAAT for *C. trachomatis* and *M. genitalium*
- Pregnancy test to exclude ectopic pregnancy
- An MSU sample if UTI symptoms (urinary frequency, nocturia, urgency, haematuria).

Management

Early diagnosis and treatment reduce the risk of long-term sequelae, so empirical treatment should be started before microbiology results are known. A broad-spectrum antibiotic regimen is needed to cover the main bacterial causes. The recommended first-line treatment is single-dose ceftriaxone 1 g by intramuscular injection, with oral doxycycline 100 mg twice daily and oral metronidazole 400 mg twice daily for 14 days. If *M. genitalium* positive, oral moxifloxacin 400 mg/day for 14 days is the first-line treatment. Abstinence from sex for at least 14 days should be advised. All sexual contacts should be notified, tested, and treated.

Those with moderate or severe clinical findings should be reviewed within 2 to 3 days to ensure that they are improving on treatment. Lack of response requires further investigation and possible admission for intravenous therapy, surgical intervention, or both. Symptom resolution, adherence to therapy, and partner notification should be reviewed after 2 to 4 weeks.

Epididymo-orchitis

Acute epididymo-orchitis is a clinical syndrome consisting of pain, swelling, and inflammation of the epididymis that can extend into the testis. It is caused mainly by extension of infection from the urethra or the bladder. In males younger than 35 years, *C. trachomatis* and *N. gonorrhoeae* are the main causes, and *M. genitalium* may also be implicated. In males older than 35 years, it is more commonly a complication of a UTI. Mumps is a cause of epididymo-orchitis in non-immune males. The most common differential diagnosis is torsion of the spermatic cord, which is a urological emergency.

Clinical features

The typical presentation is subacute onset of unilateral scrotal pain and swelling. Although there may be symptoms of urethral discharge and dysuria, these are often absent. On examination, there is tenderness and usually palpable swelling of the epididymis. There may also be tenderness and swelling of the testicle, with oedema and erythema of the overlying scrotal skin. A urethral discharge may be present.

Diagnosis

The diagnosis is usually based on clinical features. The main differential diagnosis is testicular torsion.

Investigations

These include a NAAT and culture for *N. gonorrhoeae* and a NAAT for *C. trachomatis* and *M. genitalium*. If microscopy of Gram-stained urethral secretions shows 5 or more polymorphonuclear leucocytes per high-power (×1000 oil-immersion lens) field, this indicates the diagnosis is NGU. If intracellular Gram-negative diplococci are present, this is suggestive of *N. gonorrhoeae*. A urine dipstick test and an MSU should be done to confirm or exclude UTI.

Management

Because empirical treatment should be started before microbiology results are known, a broad-spectrum antibiotic regimen is needed to cover the main bacterial causes. In younger males, in whom an STI is the likely diagnosis, the recommended regimen is single-dose ceftriaxone 1 g i.m. with oral doxycycline 100 mg twice daily for 14 days. If the patient is *M. genitalium* positive, oral moxifloxacin 400 mg/day for 14 days should be given. When a UTI is the more likely diagnosis, oral ofloxacin 200 mg twice daily for 14 days should be prescribed. Abstinence from sex for at least 14 days should be advised. All sexual contacts should be notified, tested, and treated. Patients should be reassessed after 3 days if there is no improvement in their symptoms because this requires further investigations. All should be seen after 2 weeks to assess symptom resolution, adherence to therapy, and partner notification.

Bacterial vaginosis

BV is the most frequent cause of vaginal discharge during reproductive years. The historical prevalences are 9% of those attending cervical cytology screening and 15% of pregnancies. BV develops when *Gardnerella vaginalis* and other BV-associated bacteria (BVAB) displace the normal lactobacilli-dominant vaginal flora. There is increasing evidence that BV is associated with sex with new and multiple partners and that reducing unprotected sex may reduce incident and recurrent infections. BV in pregnancy is associated with an increased risk of miscarriage and preterm birth. BV also increases the risk of acquisition and transmission of HIV and other STIs and is associated with incident PID. Up to 50% of those with BV have no symptoms. It is not necessary to look for or treat asymptomatic BV.

Clinical features

The symptoms are an increased vaginal discharge and offensive fishy odour. On examination, there is a creamy-white homogeneous discharge, which may be slightly frothy (caused by the volatile amine production by the bacteria, which is responsible for the characteristic odour) (Fig. 43.6). Visible inflammatory changes are not seen with BV.

Diagnosis

The most accurate method of diagnosis is microscopy of Gram-stained vaginal discharge because the characteristic pattern of the BV bacteria is easily distinguished from the normal lactobacilli-dominant vaginal flora.

Diagnosis using clinical criteria is possible but is less sensitive and specific. All three of the following should be present for the diagnosis to be made:

- Characteristic creamy-white homogeneous vaginal discharge
- Raised vaginal pH of greater than 4.5 (measured using narrow-range pH paper)
- Characteristic fishy odour (which can be released by mixing vaginal discharge with 10% potassium hydroxide)

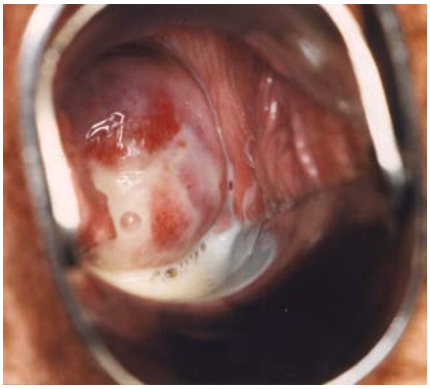


Fig. 43.6 Bacterial vaginosis. Homogeneous creamy-white vaginal discharge, which may be slightly frothy and adherent to the vaginal walls.

Commercial NAATS are available that detect combinations of BVAB, and these have similar sensitivity and specificity to microscopy. Many of the clinical symptoms and signs of BV are similar to *T. vaginalis*, so a *T. vaginalis* NAAT should be performed on those with recurrent BV to exclude undiagnosed *T. vaginalis* infection.

Management

The recommended treatment is oral metronidazole 400 mg twice daily for 7 days. A single dose of metronidazole 2 g can also be used but is less effective. Alternative topical treatments are intravaginal metronidazole 0.75% gel for 5 nights or intravaginal clindamycin 2% cream for 7 nights. High-dose metronidazole should be avoided in pregnancy, but the 7-day oral course can be safely prescribed, as can either of the intravaginal regimens.

BV recurrences are frequent, with about 50% experiencing a recurrence within 12 months of completing metronidazole therapy. Studies have identified male carriage of BVAB on the penis and urethra, suggesting that male partners may be re-infecting females with BVAB. A recent trial of simultaneous treatment of the male partner with oral metronidazole for 7 days did not reduce the rate of recurrence over placebo. However, small pilot trials of simultaneous treatment of the male partner with oral metronidazole for 7 days *plus* clindamycin cream applied to the penile skin have been associated with reduced recurrence in females; randomized trial data are awaited.

Candidiasis

Candida spp. are ubiquitous organisms and are not classified as STIs. Vulvovaginal *Candida* infection is extremely common with about 75% females having least one episode of symptomatic candidiasis in their lifetime. Predisposing factors for symptomatic infection include pregnancy, diabetes, broad-spectrum antibiotics, corticosteroids, and immunosuppression. *Candida albicans* causes 90% of vaginal yeast infections, with *Nakaseomyces glabrata* and other *Candida* species causing the remainder. Transient penile colonization (with penile rash) may occur following sex because of direct inoculation from the vagina.

Clinical features

In females, the main symptom is vulval itching. An increased thick, white vaginal discharge; vulval burning; external dysuria; and superficial dyspareunia may also be present. On examination, there may be the typical white, curdy, adherent plaques on the vaginal walls, but the amount and consistency of the discharge may look normal (Fig. 43.7).

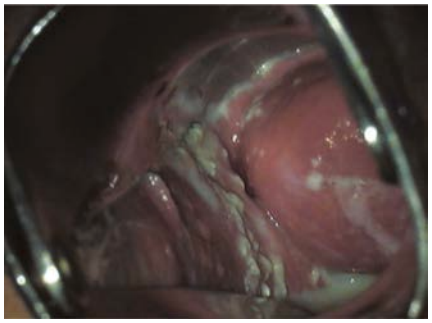


Fig. 43.7 Candidiasis. Adherent plaques of discharge on the vaginal walls.

In males, there may be a transient penile irritation and rash immediately after sex, but some experience more persistent balanoposthitis. On examination, there may be erythema of the foreskin and glans penis or a spotty, red, itchy rash on the glans with an accumulation of white discharge under the foreskin. In severe cases, there may be fissuring and phimosis of the foreskin.

Diagnosis

Microscopy of a Gram-stained vaginal or subpreputial smear identifies fungal pseudohyphae and spores in about 50% of cases of candidiasis. Culture of vaginal or subpreputial swabs has almost 100% sensitivity. Diabetes should be excluded in males with severe balanoposthitis.

Management

There are several short-course oral and intravaginal antifungal agents available, all with efficacies of 80% to 85%. Recommended treatments are the oral triazole drugs (e.g., fluconazole 150 mg as a single dose or itraconazole 200 mg twice daily for 1 day) and intravaginal imidazole pessaries or creams (e.g., clotrimazole pessary 500 mg as a single dose, miconazole vaginal ovule 1.2 g as a single dose, econazole pessary 150 mg nightly for 1–3 nights). These treatments can be supplemented with antifungal cream applied to the vulva. Males can be treated with either oral therapy or topical antifungal cream. Nystatin pessaries 200 000 units nightly for 14 nights are the treatment of choice for *N. glabrata* and other non-*albicans* yeasts. Intravaginal treatments are safe in pregnancy, but oral antifungal agents should be avoided.

Recurrent candidiasis (four or more symptomatic episodes in 1 year) occurs in up to 5% of females. It frequently requires weekly oral fluconazole 150 mg or clotrimazole pessary 500 mg for up to 6 months to prevent recurring symptoms. There is no evidence that treatment of sexual partners reduces recurrence, so partners do not need treatment unless they also have symptoms.

Trichomoniasis

T. vaginalis infection is the most common STI worldwide, though it is rarer in Western Europe and Australasia. The organism is a flagellated protozoon that is sexually transmitted. In females, it infects the vagina and urethra, and in males the urethra and subpreputial sac. *T. vaginalis* in pregnancy is associated with an increased risk of preterm birth and low birth weight, and it increases the risk of acquisition of HIV.

Clinical features

In females, 10% to 50% are symptomatic. The most common symptoms are an increased purulent vaginal discharge and malodour. There may also be vulval pruritus, external dysuria, and dyspareunia. On examination, there may be vulval erythema with inflamed vaginal mucosa. The discharge is yellow or grey and frothy and can be profuse. The cervix may have multiple small haemorrhagic areas, which have given rise to the description 'strawberry cervix'.

In males, the majority have no symptoms, although they may experience urethral discharge, irritation, and dysuria.

Diagnosis

Phase-contrast microscopy of vaginal discharge identifies the motile protozoa in up to 50% of infected females. The sensitivity of microscopy of urethral discharge in males is much lower. Culture detects 70% to 80% of infections. NAATs are the diagnostic tests of choice with VVSs in females and FVU in males detecting more than 90% of infections.

Management

The treatment of choice is oral metronidazole 400 mg twice daily for 7 days. Single-dose oral metronidazole 2 g can be used but is less effective. Because *T. vaginalis* infects areas beyond the vagina (e.g., the urethra), intravaginal metronidazole gel has poor cure rates and should not be used. Abstinence from sex for at least 7 days should be advised. All sexual contacts should be notified, tested, and treated. Tests of cure are recommended only if the patient remains symptomatic after treatment or if symptoms recur.

Occasionally, *T. vaginalis* can become resistant to metronidazole and other nitroimidazoles. This is usually relative rather than absolute resistance and may be overcome by high-dose metronidazole or tinidazole therapy.

Human papillomavirus (anogenital warts)

Anogenital warts are painless, benign, epithelial tumours, and a common STI. The causative agents are human papillomavirus (HPV) types 6 and 11. Genital HPV infection is acquired by direct skin-to-skin contact during sex with a person who has either clinical or subclinical infection. Subclinical infection is very common and 15% to 64% of those infected will develop warts.

Warts caused by HPV 6 and 11 do not undergo malignant transformation. The main oncogenic HPV types are 16 and 18. These lead to subclinical infection, not genital warts, and cause anogenital pre-malignant conditions and cancers. Neonates may acquire HPV from an infected birth canal, which may result either in anogenital warts or laryngeal papillomatosis.

Clinical features

Anogenital warts have a long incubation period; the average is 3 months, but it can extend to years. The warts first appear at sites of trauma during sex. In males, this is around the prepuce and glans; from there, they can spread to the urethra and down the penile shaft. In females, they usually start at the fourchette and then spread to the vulva and perineum (Fig. 43.8). Perianal lesions are common and not necessarily associated with receptive anal sex because subclinical HPV often affects the whole anogenital area. However, perianal lesions do occur most frequently in GBMSM. Warts on mucous membranes tend to be soft and non-keratinized, whereas those on the hair-bearing skin tend to be firm and keratinized. Warts tend to increase in size and number during pregnancy and in patients who are immunosuppressed.

Diagnosis

The diagnosis is made on clinical appearance. HPV testing is not indicated for diagnosing anogenital warts. The main differential

Fig. 43.8 **Genital warts on the fourchette and perineum.**

diagnoses are molluscum contagiosum and the condylomata lata of secondary syphilis. Atypical lesions should be biopsied, particularly in older patients, because pre-malignant and malignant lesions can look similar to warts. Investigations should include NAAT for *N. gonorrhoeae* and *C. trachomatis* and serology for syphilis and HIV because co-infection with other STIs is common.

Management

Several treatments are available for patients with anogenital warts but all have significant failure and relapse rates. Topical podophyllotoxin (0.5% solution or 0.15% cream used twice daily for 3 consecutive days per week) acts as a cytotoxic agent and is recommended first-line. If warts are still present after 8 weeks of treatment, imiquimod, which enhances the local immune response when applied to skin infected with HPV (5% cream used daily, three times a week), is recommended. If warts persist after 16 weeks of imiquimod, ablative treatments such as cryotherapy (1–2-weekly for a maximum of 4 treatments) or electrocautery should be used. Podophyllotoxin and imiquimod have the advantage of being self-applied home therapies.

Podophyllotoxin and imiquimod are not used in pregnancy. Poorer responses to treatment can occur in pregnancy, those living with HIV, and those who are immunosuppressed.

Smoking is associated with reduced clearance of HPV, so smoking cessation advice should be given. The use of condoms should be advised to protect against transmission of HPV infection and genital warts. Follow-up is recommended if warts are still present at the end of treatment.

Prevention and vaccination

There are very effective vaccines against HPV:

- A quadrivalent vaccine protects against types 6, 11, 16, and 18, which covers the most common high-risk types and those that cause genital warts.
- A vaccine protecting against nine virus types, which contains the same HPV antigens as the quadrivalent vaccine plus five additional high-risk types (31, 33, 45, 52, and 58), is currently used in the national vaccination programme in the UK.

In the UK, vaccination of females aged 12 to 13 years commenced in 2008 to prevent cervical cancer. Because the oncogenic genital HPVs are transmitted through sex, it also provided some indirect protection to males. However, increasing evidence of high levels of HPV-associated cancers and genital warts in GBMSM (who gain no indirect benefit from female vaccination) prompted a successful pilot study to assess the feasibility and cost-effectiveness of a national HPV vaccination programme for GBMSM delivered through sexual health clinics. In 2018 the programme for GBMSM aged 45 years or younger attending sexual health or HIV clinics was introduced, and in 2019 vaccination of males aged 12 to 13 years also commenced. HPV vaccines are highly effective, and only one dose is needed to confer life-long protection if given before the age of 25 years. If given between 25 to 45 years, two doses are needed 6 months to 2 years apart. People who are immunosuppressed (e.g., PLWH) require three doses ideally given within a 12-month period. Between 2015 and 2019, diagnoses of genital warts in 15- to 17-year-olds declined by 91% in females and 81% in males in England.

Molluscum contagiosum

Molluscum contagiosum is a large DNA virus. It causes small (typically 2–5 mm in diameter), benign, smooth papules with central umbilication. It is spread via direct skin-to-skin contact. When transmitted sexually, the lesions are usually multiple and present on the labia majora, penile shaft, pubic region, lower abdomen, and upper inner thighs.

Diagnosis

The diagnosis is made on the characteristic clinical appearance. Because this is a sexually acquired condition, investigations for other STIs should include NAAT for *N. gonorrhoeae* and *C. trachomatis* and serology for syphilis and HIV.

Management

Molluscum infection is often self-limiting, resolving naturally within 6 to 18 months on average in immunocompetent individuals. Treatment options, if required, are cryotherapy or podophyllotoxin cream, which has the advantage of being a self-applied home therapy. Patients should be advised about the risks of auto-inoculation of the virus and discouraged from shaving or waxing pubic hair to prevent further spread. No routine follow-up or partner notification is required unless other STIs are identified.

Herpes simplex

Genital herpes is the most common cause of genital ulceration in all countries worldwide. The peak incidence for primary infection is in 16- to 24-year-old individuals. Females acquire the infection more frequently, probably because of the larger surface area of susceptible mucous membrane on the vulva. Transmission occurs from the mucous membrane of a person who is shedding HSV, many of whom will be asymptomatic. Only about 20% of people with serological evidence of genital herpes give a clinical history of herpes, suggesting that many individuals have subclinical infection.

Genital herpes can be due to HSV type 1 or type 2. It is possible to be co-infected with both. HSV-1 infection may be spread from an infected genital tract or from orolabial lesions via orogenital sex. HSV-2 is almost always transmitted via genital-to-genital contact. In the UK, more than 50% of primary HSV cases are caused by HSV-1.

During the primary infection, the virus ascends the peripheral sensory nerves supplying the area of inoculation and establishes latency in the dorsal root ganglia, thus allowing future reactivation and recurrences.

Clinical features

The ***initial episode*** is the first occurrence of either HSV-1 or HSV-2. This is subdivided according to prior exposure to the other HSV type.

Primary genital infection is the first ever exposure to either HSV type 1 or 2. It typically presents with multiple painful, shallow ulcers (Fig. 43.9). There is usually tender inguinal lymphadenopathy and systemic symptoms of viraemia, including fever, myalgia, and headache. In females, vulval pain and external dysuria occur. Cervical ulcers may be present and can have the appearance of a malignancy. Rectal infection may lead to severe proctitis with pain and bleeding; this is mainly seen in GBMSM. The lesions start to heal over a period of 10 to 21days, even without treatment. Neurological complications can include aseptic meningitis and autonomic neuropathy, leading to urinary retention. Primary infection can also be asymptomatic.

Non-primary genital infection occurs in people with previous HSV-1 or HSV-2 who then acquire the other type of genital HSV. There is some cross-protection from the prior HSV infection, resulting in a milder illness than in primary infection. Non-primary genital infections are more likely to be asymptomatic than primary infections.

Recurrent genital herpes is due to reactivation of previous HSV-1 or HSV-2 infection. HSV-2 recurs more frequently than HSV-1. The median recurrence rate in the subsequent year after a primary or non-primary infection is about one recurrence for HSV-1 and about four recurrences for HSV-2. The recurrences may be preceded by a prodrome of tingling, itching, or pain in the area. On examination, there are usually a few ulcers confined to a small area, and systemic symptoms are rare. Recurrences are not always symptomatic but are potentially infectious because of subclinical viral shedding. Long-term studies show that symptomatic recurrences and subclinical viral shedding gradually decrease with time.

The clinical presentation of primary infection in patients who are immunosuppressed and in pregnancy is usually more severe, with increased frequency of symptomatic and subclinical recurrences. Rarely, the infection can disseminate, causing a systemic life-threatening condition.

Genital herpes increases the acquisition and transmission of HIV and is a significant attributable risk in the spread of HIV. Many people with recurrent HSV develop psychological and psychosexual problems and fear rejection on disclosure of their infection to sexual partners.

Diagnosis

HSV DNA detection using polymerase chain reaction (PCR) on a swab taken from the ulcer is the diagnostic method of choice. This can distinguish between HSV-1 and HSV-2. Tests for other STIs should be performed, including NAAT for *N. gonorrhoeae* and *C. trachomatis* and serology for syphilis and HIV.

Blood tests for HSV type-specific antibodies can be used to diagnose prior HSV-1 and HSV-2 infections when the clinical history is suggestive of genital herpes but confirmation by HSV DNA detection has not been possible. The presence of HSV-2 antibodies is indicative of genital herpes, but the presence of HSV-1 antibodies cannot differentiate between genital and orolabial infections.

Fig. 43.9 Herpes simplex rash on the penis. *(Courtesy Dr B. Goh.)*

Management

Initial episode

Recommended oral antiviral therapies are aciclovir 400 mg three times daily, valaciclovir 500 mg twice daily, or famciclovir 250 mg three times daily, all for 5 days. Aciclovir is the drug of choice in pregnancy and breast-feeding. Supportive measures include cleansing with saltwater; topical anaesthetic agents; and in females with multiple ulcers and severe dysuria, immersing the genitals in a bath or a bowl of warm water to urinate. If lesions are already healing, antiviral therapy will have little added effect. Secondary bacterial infection occasionally occurs and should be treated with antibiotics.

The natural history of HSV infection should be explained, including recurrences, subclinical viral shedding, and the potential for sexual transmission with both of these. Patients should be advised to avoid sex during the prodrome and recurrences. Subclinical viral shedding is most common during the first 12 months after initial HSV-2 infection and in those with frequent symptomatic recurrences. Condoms and suppressive treatment reduce the risk of transmission from subclinical viral shedding, but neither completely prevents it. Consequently, disclosure should be advised in all relationships.

Recurrence

The appropriate management depends on the number and severity of recurrences. Because recurrences tend to be less severe and self-limiting, they can sometimes be managed with saltwater bathing and topical anaesthetic agents.

When recurrences are infrequent but severe, episodic antiviral therapy, started early by the patient, reduces the duration and severity but does not reduce the number of recurrences. Recommended episodic regimens of oral agents are shorter: aciclovir 800 mg three times daily for 2 days, famciclovir 1 g twice daily for 1 day or valaciclovir 500 mg twice daily for 3 days. Alternatively, 5-day courses can be used: aciclovir 400 mg three times daily, valaciclovir 500 mg twice daily, or famciclovir 125 mg twice daily.

Suppressive treatment

All people with recurrent HSV experience a reduction in recurrences on suppressive therapy. The decision whether to start suppressive treatment depends on the number of recurrences and the physical and psychological impact of these versus the inconvenience of taking daily treatment. Recommended oral regimens are aciclovir 400 mg twice daily, valaciclovir 500 mg/day, or famciclovir 250 mg twice daily for 12 months. Therapy should then be discontinued to assess the frequency of recurrences. If they are still frequent, suppressive treatment can be restarted. It is safe to continue suppressive treatment over many years. Frequent recurrences are associated with psychological and psychosexual morbidity; support and counselling are often needed.

Herpes simplex virus in pregnancy

The main risk of HSV in pregnancy is vertical transmission. Despite antiviral treatment, neonatal HSV has high morbidity and mortality.

An initial episode of genital HSV in late pregnancy poses the highest risk of transmission, and elective caesarean delivery is recommended, particularly for infection within 6 weeks of the expected delivery date. Daily suppressive therapy (oral aciclovir 400 mg three times daily) is recommended from 32 weeks' gestation until birth for those with initial episodes earlier in pregnancy to reduce recurrence, subclinical viral shedding, and the need for a caesarean delivery. In those at high risk of preterm delivery, aciclovir 400 mg twice daily should be started from 22 weeks' gestation followed by aciclovir 400 mg three times daily from 32 weeks'. The risk of neonatal herpes with recurrent HSV is small, even if lesions are present at the time of delivery, therefore caesarean delivery is not indicated for recurrent RSV, but daily suppressive aciclovir throughout the third trimester is recommended.

Syphilis

Syphilis is a systemic disease caused by *Treponema pallidum*, a motile spirochaete. It is mainly transmitted by direct contact with an infectious lesion and enters the new host through breaches in squamous or columnar epithelium, usually during sex. Primary infection of non-genital sites may occur but is rare. It can also be transmitted by infected blood products and vertically during pregnancy. Syphilis is classified as acquired or congenital.

Acquired syphilis is further subdivided into primary, secondary, and early latent (all also referred to as early or infectious syphilis, indicating that infection has been acquired during the previous 2 years); late latent (infection for more than 2 years); and tertiary syphilis (the most destructive stage, which includes cardiovascular and neurological syphilis and gummatous lesions of any organ).

Congenital syphilis is subdivided into early (diagnosed within the first 2 years of life) and late (diagnosed after the age of 2 years).

The incidence varies significantly with geographical location. It is more common in low- and middle-income countries. In high-income countries, it is mainly confined to GBMSM. The number of cases of infectious syphilis in England increased by 260% between 2013 and 2022, with most cases in GBMSM. Syphilis still affects large numbers of pregnancies worldwide. In 2016 an estimated 988 000 pregnancies were infected with syphilis worldwide, resulting in more than 350 000 adverse birth outcomes, including 200 000 stillbirths and newborn deaths. The global rate of congenital syphilis was 425 cases per 100 000 livebirths in 2020, which is way above the WHO's target of 50 cases per 100 000 livebirths.

Clinical features

Primary syphilis

Between 9 and 90 days (mean, 21 days) after exposure, a papule develops at the site of inoculation. This ulcerates to become a painless, firm ulcer (chancre). There is usually also painless regional lymphadenopathy. The primary lesion may go unnoticed, especially if it is on the cervix or within the rectum. Healing occurs spontaneously within 3 to 8 weeks.

Secondary syphilis

This occurs in 25% with untreated primary syphilis. Between 3 and 10 weeks after the appearance of the primary lesion, septicaemia leads to constitutional symptoms, including fever, sore throat, malaise, and arthralgia, though any organ may be affected. Meningitis, cranial nerve palsies (particularly 8th nerve palsy with hearing loss), uveitis, interstitial keratitis, and retinal involvement may all occur and represent early neurological disease. Hepatitis, nephritis, and arthritis have all been described.

Common signs include:

- Widespread skin rash (present in 75%), which can involve the whole body, including the palms and soles, typically, a non-itchy, maculopapular rash that may have a coppery colour (**Fig. 43.10**)
- Generalized lymphadenopathy (present in 50%)
- Condylomata lata, which are moist, wart-like plaques found in the perianal area and other moist body sites
- Mucosal lesions in the mouth and on the genitalia presenting as distinct mucous patches or becoming confluent to form 'snail-track ulcers'.

Without treatment, the symptoms and signs of secondary syphilis resolve but may recur, especially in the first year of infection.

Latent syphilis

Latent syphilis is diagnosed based on reactive syphilis serology in someone who has no symptoms and has not been treated. It is divided into:

- Early latent: within 2 years of infection acquisition (1 year in the USA); sexual transmission is possible
- Late latent: present for 2 or more years; sexual transmission does not occur

Latent syphilis may persist for years or may even be life-long.

Tertiary syphilis

About one-third of people with untreated latent syphilis develop tertiary syphilis within 5 to 30 or more years of contracting the infection. Gummatous syphilis (with inflammatory, granulomatous, destructive lesions) is the most benign and commonly involves the skin and bones, but lesions can occur in any organ. Cardiovascular syphilis causes aortitis, aortic regurgitation, aneurysm of the ascending aorta, and stenosis of the coronary artery ostia. Neurosyphilis causes chronic meningovascular damage and endarteritis of the small vessels of the brain and spinal cord, presenting as 'general paralysis of the insane' and tabes dorsalis.

Syphilis in pregnancy and congenital syphilis

Syphilis can be transmitted across the placenta at any stage of pregnancy. The risk of transmission is dependent on the stage of maternal infection and can be up to 100% in early syphilis and even up to 10% with late infection. The WHO estimates that untreated early syphilis in pregnancy results in rates of second-trimester miscarriage or stillbirth of 25%, preterm birth before 32 weeks' gestation of 13%, neonatal death of 11%, and congenital syphilis among the infants born of 20%. Detection and treatment of syphilis early

Fig. 43.10 Rash of secondary syphilis on the palms. *(Courtesy Dr B. Goh.)*

in pregnancy can prevent congenital syphilis and neonatal death at term and reduce adverse pregnancy outcomes.

Signs of early congenital syphilis occur in the neonatal period and include a rash, condylomata lata, mucous patches, nasal discharge, hepatosplenomegaly, and periostitis. Late congenital syphilis (after 2 years of age) can present with neurological or gummatous lesions but also includes the 'stigmata of congenital syphilis' which occur because early damage to developing structures, particularly teeth and bones. These include Hutchinson's teeth, sabre tibia, bossing of the frontal and parietal bones, and saddle nose.

Diagnosis

T. pallidum infection cannot be cultured, but it can be identified by dark-ground microscopy of secretions from a primary chancre or condylomata lata. Sensitivity is dependent on highly skilled microscopists. Most laboratories can test swabs from these lesions for *T. pallidum* using PCR. However, serological testing remains the main laboratory diagnosis.

Most laboratories use a treponemal enzyme immunoassay to detect immunoglobulin (Ig) G and IgM as a screening test. If the result of this is positive, a further treponemal test and a non-treponemal test are performed.

Treponemal tests

T. pallidum haemagglutination and *T. pallidum* particle agglutination assays are highly specific for treponemal disease but usually remain positive for life, even after treatment, so they are unable to differentiate between prior treated infection and re-infection.

Non-treponemal tests

The Venereal Disease Research Laboratory (VDRL) and rapid plasma reagin (RPR) tests become positive after about 6 weeks of infection. They are quantifiable tests that can be used to monitor treatment response and evidence of re-infection. They are not specific to syphilis, and false-positive results may occur in other conditions, particularly in other infections and autoimmune diseases.

Examination of the cerebrospinal fluid for evidence of late neurosyphilis is indicated in patients with positive syphilis serology who demonstrate neurological signs and symptoms.

Management

Treponemicidal levels of an antibiotic are required for at least 7 days in patients with early syphilis to cover the slow division time of the organism (30 hours). In those with late syphilis, treponemes may divide even more slowly, so a longer duration therapy is required. Ideally, a long-acting penicillin should be given i.m.

Early syphilis (primary, secondary, and early latent)

- Benzathine penicillin G 2.4 MU i.m. single dose
- Penicillin allergy: oral doxycycline 100 mg twice daily for 14 days

Jarisch–Herxheimer reaction (mild fever, malaise, and headache) is caused by release of inflammatory cytokines and occurs within 8 hours of treatment in 50% of patients with primary syphilis and up to 90% with secondary syphilis.

Late latent, cardiovascular[a], and gummatous syphilis

- Benzathine penicillin G 2.4 MU i.m. three doses at weekly intervals
- Penicillin allergy: oral doxycycline 100 mg twice daily for 28 days

[a]Forty to 60 mg oral prednisolone daily for 3 days should be started 24 hours before antibiotic treatment for cardiovascular and neurosyphilis.

Neurosyphilis (both early and late stages of infection)[a]

- Procaine penicillin 2.4 MU i.m. daily plus probenecid 500 mg orally four times daily for 14 days
- Penicillin allergy: doxycycline 200 mg twice daily for 28 days

Pregnancy

- Penicillin can be safely used in pregnancy; doxycycline should be avoided.

Syphilis and human immunodeficiency virus

The diagnosis and management of syphilis in people living with HIV remains unaltered; however, if it is untreated, it may advance more rapidly than in HIV-negative patients, and there is a higher incidence of neurosyphilis.

Prognosis and follow-up

Those being treated for early syphilis should abstain from sex for at least 14 days, and sexual contacts must be traced and investigated. There should be regular follow-up within the first year using repeat VDRL/RPR titres to establish the 'fourfold fall', which demonstrates adequately treated syphilis.

The prognosis of syphilis depends on the stage at which the infection is treated. Early syphilis has an excellent outlook, but when permanent damage has occurred in those with tertiary syphilis, it is irreversible, though further progression will be halted.

Lymphogranuloma venereum

LGV is an STI caused by the invasive serovars L1, L2, and L3 of *C. trachomatis*. It is endemic in several tropical areas, including southern Africa, India, South-east Asia, and the Caribbean, where it accounts for 2% to 10% of cases of genital ulcer disease. It used to be rare in Western Europe but since 2003 has become endemic among GBMSM, particularly in those living with HIV. There is no evidence of transmission of LGV within heterosexuals in Europe. The main presentation in GBMSM is with rectal symptoms; genital infections are rarely seen. LGV should be considered in GBMSM with suspected inflammatory bowel disease because the clinical presentation can be very similar, and the histological findings of LGV proctitis are similar to those of other causes of granulomatous proctitis, such as Crohn's disease. It is frequently associated with other STIs and acute hepatitis C infection.

Clinical features

There are three clinical stages. The primary lesion may be transient and is frequently unnoticed. In the genital area, it takes the form of a painless papule or shallow ulcer appearing at the area of inoculation 7 to 30 days after exposure. The main presentation in GBMSM is proctitis with symptoms of rectal pain, mucopurulent discharge, rectal bleeding, constipation, and tenesmus. Systemic symptoms such as fever and malaise may occur.

Secondary lesions are enlarged, tender regional lymph nodes. With genital LGV, they are usually unilateral and affect the inguinal and femoral nodes. When both are involved, the 'groove sign' develops because of the inguinal ligament separating the two enlarged lymph node systems. The nodes may become matted with bubo formation, which may rupture. With rectal infection, the affected nodes are in the pelvis, so they are not usually clinically apparent.

The tertiary stage is a chronic inflammatory response with tissue destruction. In the rectum, it can cause fistulae, strictures, and granulomatous fibrosis, mimicking Crohn's disease. There may also

be scarring of the genital area, and destruction of local lymph nodes can lead to genital lymphoedema.

Diagnosis

A swab should be taken from the genital ulcer base or from the rectal mucosa for *C. trachomatis* NAAT. If this is positive, genotyping for LGV should be performed. Testing for the other causes of genital ulcers (see **Box 43.6**) or rectal symptoms (see **Box 43.8**) should include a swab for HSV and *T. pallidum* PCR and serology for syphilis, which should be repeated after the 3-month window period. An mpox PCR test should also be considered if there are other features suggestive of this such as rash elsewhere and marked inguinal lymphadenopathy. A NAAT for *N. gonorrhoeae* and *C. trachomatis* on FVU (or a VVS) and serology for HIV should be performed. Serology for hepatitis C should be considered in GBMSM in view of the frequent co-infection of acute hepatitis C with LGV.

Management

The first-choice treatment is oral doxycycline 100 mg twice daily for 21 days or azithromycin 1 g weekly for 3 weeks. Patients should be advised to abstain from sex until completion of treatment. Sexual contacts should be notified, examined, tested, and treated. Follow-up should continue until all symptoms and signs have resolved, which is usually by 3 to 6 weeks.

Mpox

Mpox is a zoonotic infection, caused by the monkeypox virus, which is endemic in West and Central Africa. Before 2022, cases in the UK and Europe had been very rare, limited to people travelling from countries where mpox is endemic or contacts of imported cases. In May 2022, a case of mpox was identified in London, UK, in someone with recent travel history from Nigeria, where mpox is endemic. Shortly afterward, several other cases were confirmed in London with no travel links to mpox endemic countries, and mpox was made a notifiable disease in England in June 2022. Further cases were confirmed throughout the UK, Europe, and North America, in GBMSM who had recent sexual contact with new or multiple partners but no travel to endemic countries. Soon many countries and regions had reported infections, and most were from Clade IIb, B 1 lineage. In July 2022 the WHO declared the outbreak a public health emergency of international concern (PHEIC) because of the rapid spread around the world through new modes of transmission, which were not well understood. This triggered a global response to the outbreak to reduce the spread of the infection by rapid diagnosis and partner notification and by increasing immunity through administering pre-exposure smallpox vaccination, where available, to GBMSM. Smallpox vaccines provide about 85% effectiveness against mpox. Some immunological response to the first vaccine occurs within 2 weeks, which is still within the mpox incubation period, so post-exposure vaccination was also offered within 4 days of exposure to try to modify the severity of disease.

In May 2023 the WHO declared an end to the PHEIC because there had been steady progress in controlling the spread of the disease. By December 2023 there had been 92 976 confirmed cases with 172 deaths in 116 countries. Although there has been a global decline in cases, the threat has not disappeared, and there is still a high risk of mpox being imported from one country to another.

Clinical features

Mpox is transmitted by skin-to-skin contact or from body fluids including respiratory secretions. It can also be spread through fomites (e.g., clothing, bedding, towels) used by someone with mpox.

After an incubation period of 5 to 21 days, a rash appears often starting on the face and then spreading elsewhere, including the mouth, genitals, and anal area. The rash evolves from papules to blisters or pustules to scabbing blisters. This is accompanied by fever and systemic symptoms such as headache, myalgia, and malaise. Inguinal lymphadenopathy is often present with the nodes being larger than is usually seen with genital herpes and syphilis (which are the main differential diagnoses).

Diagnosis

Suspected mpox lesions should be swabbed for mpox PCR test. Testing for the other causes of genital ulcers (see **Box 43.6**) should include a swab for HSV and *T. pallidum* PCR tests and serology for syphilis, which should be repeated after the 3-month window period. A NAAT for *N. gonorrhoeae* and *C. trachomatis* on FVU (or a VVS) and serology for HIV should be performed.

Management

The symptoms from mpox are usually self-limiting and clear after 2 to 4 weeks. They can be managed with supportive care and pain relief. The patient should be advised to isolate at home in a separate room from family and pets until the rash has healed. Mpox in people living with HIV and other immunocompromised people can be severe and hospital admission, and antivirals may be needed.

Close household and non-household contacts of confirmed cases do not need to isolate but should avoid sexual or intimate contact with others for 21 days. Post-exposure smallpox vaccination can be offered if within 4 days of the contact to reduce transmission.

Chancroid

Chancroid is caused by *Haemophilus ducreyi*. It used to be one of the most common causes of genital ulcers worldwide, but its incidence has now decreased markedly. It is extremely rare in most countries except for North India and Malawi.

Clinical features

Chancroid has a short incubation period of 3 to 7 days. Tender papules develop at the site of inoculation, which rupture into painful, ragged-edged ulcers with necrotic bases that bleed easily. The usual sites of infection are the prepuce, glans penis, labia minora, and vulval fourchette. There is often painful inguinal lymphadenopathy, which can develop into large buboes that suppurate.

Diagnosis and management

Detection of *H. ducreyi* DNA using PCR is the most sensitive diagnostic test, but there are no commercial assays available for this. A 'probable diagnosis' may be made if the patient has the appropriate clinical presentation, without evidence of syphilis or HSV.

Testing for the other causes of genital ulcers should be undertaken (see **Box 43.6**) and should include an ulcer swab for HSV and *T. pallidum* PCR, an ulcer swab for *C. trachomatis* NAAT with genotyping for LGV if positive and serology for syphilis, which should be repeated after the 3-month window period. A NAAT for *N. gonorrhoeae* and *C. trachomatis* on FVU (or a VVS) and serology for HIV should also be performed.

Single-dose regimens include ceftriaxone 250 mg i.m. or azithromycin 1 g orally. Multiple-dose regimens are oral ciprofloxacin 500 mg twice daily for 3 days or oral erythromycin 500 mg four times daily for 7 days. Multiple-dose regimens should be used in PLWH because treatment failures have been reported with single-dose therapy. Patients should be advised to abstain from sex for at least

7 days and be followed up at 3 to 7 days, when the ulcers should be healing. PLWH should be monitored closely because healing may be slower. Sexual partners should be notified, examined, tested, and treated epidemiologically because asymptomatic carriage has been reported.

Donovanosis

Donovanosis (also known as granuloma inguinale) is exceedingly rare and is confined to a few countries in South-east Asia, southern Africa, parts of India, and South America. It is caused by *Klebsiella granulomatis*.

Clinical features

Nodules at the site of inoculation develop into friable, non-painful ulcers or hypertrophic lesions that increase in size.

Diagnosis and management

The diagnosis is made on the presence of Donovan bodies using Giemsa or silver stains in scrapings or biopsies of the lesions. Donovan bodies are the encapsulated intracellular Gram-negative rods of *K. granulomatis* visible within mononuclear cells. Screening for all other STIs should be undertaken.

Antibiotic treatment should be given for a minimum of 3 weeks and until the lesions have completely healed. Oral regimens include azithromycin 1 g weekly or 500 mg/day or doxycycline 100 mg twice daily. Patients should be advised to abstain from sex for at least 3 weeks and be followed up until the lesions have fully resolved. Sexual partners should be notified, examined, and treated if necessary.

Pediculosis pubis

The pubic louse *(Phthirus pubis)* can attach tightly to the pubic and coarse body hair. It can also attach to the eyelashes and eyebrows. It is host specific and is transferred by close bodily contact. Although infestation may be asymptomatic, the most common complaint is of itch caused by hypersensitivity to the louse bites.

Diagnosis and management

Lice may be seen on the pubic and body hairs. They resemble small scabs or freckles but can be seen moving. The eggs (nits) are laid at the hair base and are strongly adherent to the hairs. Screening for other STIs should include NAAT for *N. gonorrhoeae* and *C. trachomatis* and serology for syphilis and HIV.

Treatment should be applied to all areas of the body, including facial hair if present. Permethrin 1% should be left on for 10 minutes, and malathion 0.5% should be left on for 12 hours. A second application is usually advised after 7 days. Permethrin is safe in pregnancy. Sexual partners should be examined and treated if infected.

Further reading

BASHH Summary Guidance on Testing for Sexually Transmitted Infections, 2023 www.bashhguidelines.org/media/1334/bashh-summary-guidance-on-testing-for-stis-2023.pdf. This includes guidance for all trans people, including those with both binary and non-binary identities.

Bennett C, Rebafka A, Carrier J, Cook S, Edwards D. Impact of primary and recurrent genital herpes on the quality of life of young people and adults: a mixed methods systematic review. *JBI Evid Synth*. 2022; 20:1406–1473.

BMJ Best Practice Mpox available at https://bestpractice.bmj.com/topics/en-gb/1611/epidemiology.

British Association for Sexual Health and HIV (BASHH). *BASHH Clinical Effectiveness Guidelines*. BASHH; http://www.bashh.org/.

Chow EPF, Carter A, Vickers T, et al. Effect on genital warts in Australian female and heterosexual male individuals after introduction of the national human papillomavirus gender-neutral vaccination programme: An analysis of national sentinel surveillance data from 2004–2018. *Lancet Infect Dis*. 2021; 12:1747–1756.

Moseley P, Bamford A, Eisen S, et al. Resurgence of congenital syphilis: new strategies against an old foe. Lancet Infect Dis. 2024; 24:e24–e35.

Questions for this chapter can be found on Elsevier eBooks+.

44 Human immunodeficiency virus

Jane Anderson and Chloe M. Orkin

Clinical skills 1537
History 1537
Examination 1537
Investigation 1538
Background 1538
History of HIV 1538
Epidemiology 1538
Acquisition 1539
Viral structure 1540
Pathogenesis 1540
Symptomatic HIV infection 1543
HIV-related end-organ disease 1543
Conditions associated with immunosuppression 1544
HIV testing 1545
Clinical approach to the person living with HIV 1545
Monitoring 1546
Management of HIV 1547
Anti-retroviral therapy 1547
Monitoring anti-retroviral therapy 1551
Complications and long-term safety of anti-retroviral therapy 1552
Pregnancy 1553
Post-exposure prophylaxis 1554
HIV-associated disease 1554
Fungal infections 1554
Protozoal infections 1557
Viral infections 1557
Bacterial infections 1559
Infections caused by other organisms 1560
HIV-associated cancers 1561
Prevention and control of HIV 1561
Pre-exposure prophylaxis 1563

CORE SKILLS AND KNOWLEDGE

Major advances in the treatment of patients living with HIV infection have led to an increase in global prevalence, even as rates of new infection fall. Where people have consistent access to effective anti-retroviral therapy (ART), HIV infection has been transformed from a progressively debilitating life-limiting illness characterized by opportunistic infections into a chronic condition requiring life-long multidisciplinary care.

Key skills in this specialty include:

- Assessing people living with HIV and its consequences, including level of immunosuppression, as well as recognition of opportunistic infection and AIDS-defining conditions
- Understanding the principles of ART and the holistic approach to management of people living with HIV infection
- Understanding the role of coexisting clinical conditions on morbidity and mortality and health-related quality of life
- Discussing prevention interventions to ensure that people can avoid acquiring or transmitting HIV.

The best place to experience the care of people living with HIV is within specialist HIV services. In the UK, these are in secondary and tertiary care settings, commonly co-located with sexual health, genitourinary medicine, and infectious diseases. People hospitalized with HIV related conditions (e.g., people with late-stage HIV infection and opportunistic infections) offer opportunities to learn about advanced disease.

CLINICAL SKILLS

History

History taking in human immunodeficiency virus (HIV) medicine encompasses a patient's risk factors for HIV, including sexual history, current health status, and medications, as well as an understanding of barriers to healthcare access and treatment adherence. Presenting symptoms may include those of acute disease and advanced infection (**Box 44.1**).

Examination

Clinical examination focuses on overall health status and the detection of HIV-related complications when immunosuppression is present. Examination is likely to be dictated by the presenting symptoms but may also include examination of the following:

- Skin for dermatological manifestations of HIV (e.g., fungal infection, Kaposi's sarcoma [KS])
- Lymph nodes for lymphadenopathy caused by infection or malignancy
- Mouth for oral lesions associated with HIV (e.g., candidiasis, hairy leukoplakia, cytomegalovirus ulcers)
- Cardiovascular system for signs of cardiovascular disease that can be exacerbated by HIV and its treatments
- Respiratory system for infection
- Abdomen for organomegaly
- Genitourinary tract for genital lesions (e.g., ulcers, warts, or signs of other sexually transmitted disease)
- Neurological examination for cognition and for clinical signs of opportunistic disease in the central nervous system

Box 44.1 Symptoms relevant to HIV

Acute HIV infection

- Fever
- Malaise
- Myalgia
- Lethargy
- Sore throat
- Maculopapular rash

Advanced HIV with immunosuppression

- Systemic symptoms: fever, weight loss, anorexia
- Symptoms related to HIV end-organ damage: cognitive impairment, skin disease, weight loss, gastrointestinal disturbance, heart failure, kidney failure
- Symptoms of immunosuppression and associated infection or malignancy: fever, cough, breathlessness, diarrhoea, skin changes

- Mental health which can affect risk and adherence (e.g., depression, anxiety, substance abuse)
- Overall health: malnutrition and wasting can be seen in advanced disease.

Investigation

- Testing for HIV in the UK is:
  - Targeted (to those with risk factors for HIV and clinical presentations consistent with HIV)
  - Universal (e.g., all presentations to the emergency department, all pregnancies)
- Investigations for people living with HIV include:
  - Assessment of immunosuppression (viral load, CD4 count)
  - Tests of drug resistance
  - Investigations for complications of immunosuppression (e.g., infection, malignancy)
  - Investigations for side-effects of the medications used in HIV (e.g., liver function, kidney function, diabetes, and cardiovascular disease screening)

BACKGROUND

Human immunodeficiency virus (HIV) is a retrovirus that preferentially invades and destroys cells of the immune system by replication, in particular T lymphocytes. If untreated, the result is a state of profound immunodeficiency, leaving the individual susceptible to potentially fatal infections and neoplasms, acquired immunodeficiency syndrome (AIDS), and death. After it is acquired, HIV cannot be removed from the body, and no cure currently exists. Effective anti-retroviral treatment suppresses viral replication, preserves immune function, and prevents onward transmission. Although commonly used, the imprecise and potentially confusing term 'HIV/AIDS' should be avoided. AIDS is a clinical syndrome and HIV the causative agent. Most people diagnosed with HIV in the UK take effective treatment and do not have AIDS. A person-first approach to language is important, especially in HIV care. 'People living with HIV' is the preferred terminology.

HISTORY OF HIV

The history of HIV is shown in **Box 44.2**.

Box 44.2 The history of HIV

1981: Five reports made of *Pneumocystis carinii* pneumonia in California.
1982: First documentation made of AIDS.
1983: Causative agent identified and named lymphadenopathy-associated virus (LAV) in France. (Francoise Barre Sinnoussi and Luc Montagnier were later awarded the Nobel Prize for Medicine.)
1984: HTLV-III identified and shown to be identical to LAV.
1985: First commercial blood tests for the virus licensed in the UK, enabling better understanding of the major routes of transmission. This led to public health messaging and interventions, including condoms, heat treatment of blood products, needle exchange, and opiate replacement programmes.
1986: The International Committee on the Taxonomy of Viruses term the causative virus human immunodeficiency virus (HIV), replacing the interchangeable use of LAV and HTLV-III.
1987: The first HIV-specific drug, AZT (zidovudine), an inhibitor of reverse transcriptase, is licensed and introduced.
Early 1990s: Trials show that using two antiviral drugs in combination improves outcomes.
1994: Use of AZT in pregnancy is shown to significantly reduce vertical transmission of HIV.
1996: Three-drug combination therapy (HAART) using compounds from different drug classes (reverse transcriptase inhibitors and PIs) is shown to effectively suppress viral replication, reducing AIDS and mortality.
1999: Universal offer of an HIV test to all pregnant females is implemented in the UK.
2000: Universal access to effective therapy to combat HIV worldwide is adopted by the UN General Assembly.
2000–2010: Emphasis on HIV testing, shifting toward an opt-out model rather than an opt-in one. Equitable and universal access to diagnosis and treatment becomes a focus of global attention.
2014: UNAIDS set the 90-90-90 targets: 90% of those living with HIV know their HIV status, 90% of those diagnosed are taking effective therapy, and 90% of those on effective therapy having fully suppressed virus, aiming to end the global epidemic by 2030. These targets remain unachieved.
2019: UK national HIV Action Plan establishes the goals of achieving zero new HIV infections, AIDS, and HIV-related deaths in England by 2030. Key components are prevention of acquisition, prompt diagnosis, and improving the health and quality of life of people living with HIV.

AIDS, Acquired Immunodeficiency syndrome; *HAART,* highly active anti-retroviral therapy; *HTLVT-III,* human T-lymphotropic virus type III; *PI,* protease inhibitor; *UNAIDS,* United Nations Programme on HIV/AIDS.

EPIDEMIOLOGY

Since the first description of AIDS in 1981, more than 85.6 million people are estimated to have acquired HIV globally, and at least 40 million people have died. At the start of 2023, 39 million people worldwide were living with HIV infection of whom 29.8 million were accessing ART, up from 7.7 million in 2010.

Highly active ART has dramatically reduced mortality and morbidity for those who are able to access consistent care, transforming HIV from a universally fatal infection into a long-term, manageable condition. Additionally, those taking effective ART in whom viral suppression is maintained below the limits of detection ('undetectable') cannot pass the virus on to sexual partners, characterized as undetectable = untransmissible (U = U).

In 2014, the United Nations Programme on HIV/AIDS (UNAIDS) set global targets for a scale up in HIV diagnosis and treatment with the ambition of ending the AIDS epidemic by 2030. These '90-90-90' ambitions are that:

- 90% of all people living with HIV know their HIV status.

- 90% of people with diagnosed HIV infection receive sustained ART.
- 90% of all people receiving ART achieve viral suppression.

Very few countries reached the 90-90-90 by 2020, when the targets were revised upward to '95-95-95'. Data from 2022 show that of people living with HIV across the world, 86% (73%–98%) knew their status, 76% (65%–89%) were receiving treatment, and 71% (60%–83%) had suppressed viral loads. Despite significant scale up of diagnosis and treatment, the global response remains suboptimal with HIV continuing to have an impact, including some of the world's most vulnerable populations.

Globally, UNAIDS data show that new HIV infections have been reduced by 59% since the peak in 1995, and deaths by 38% since 2010. In 2022, around 1.3 million people newly acquired HIV compared with approximately 3.2 million people in 1995.

Although HIV is a global pandemic, there is considerable geographical variation, particularly in populations most affected, mode of transmission, and viral subtypes. Globally, 46% of all new HIV infections were in females in 2022. There remains considerable geographical diversity: across Eastern Europe, Central Asia, North Africa, and the Middle East, new infection rates continue to rise. HIV is a major contributor to the global burden of disease and remains in the top 10 causes of disability-adjusted life-years and death.

HIV in sub-Saharan Africa

Sub-Saharan Africa remains the region most significantly affected by HIV. Although southern and eastern Africa are home to 6.2% of the global population, this figure includes more than half of the world's population living with HIV, and 43% of new HIV infections each year occur in this region. The predominant route of HIV transmission in the region is through heterosexual sexual intercourse, with females being disproportionately affected. In 2022 females aged 15 to 24 years were more than three times more likely to acquire HIV than their male peers.

HIV in high-income countries

Although the rate of new diagnoses is falling in many high-income countries, the overall prevalence of HIV is rising. In the UK, new diagnoses declined from a peak in 2005 and plateaued in 2021 during the COVID-19 (severe acute respiratory syndrome coronavirus 2 [SARS-CoV-2]) pandemic, with an uptick of new HIV diagnoses in 2022. Falling death rates coupled with continuing new infections mean that the total number of people living with HIV continues to rise. In 2021, 107 000 people in the UK were estimated to be living with HIV of whom around 5000 remain undiagnosed. Of those diagnosed, 99% were on treatment and 98% virally suppressed. Of those diagnosed with HIV in the UK, gay, bisexual, and other males who have sex with males (GBMSM) and heterosexual populations from sub-Saharan Africa are the two largest groups of people living with the infection; 30% are female.

Declining infection rates are largely driven by reductions in the GBMSM community because of a combination of factors, including condom use, increased and repeat testing, rapid initiation of effective therapy for those with positive HIV test results, and greater use of pre-exposure prophylaxis (PrEP) (see later). In heterosexual males and females, the rates of decline are not as impressive.

As mortality rates fall, the population living with HIV is becoming older. In the UK in 2022, 50% of adults accessing HIV care were aged 50 years or older compared with 28% in 2013. With increasing age, HIV care is complicated by multimorbidity and polypharmacy.

Late diagnosis is the most common cause of HIV-related morbidity and mortality in the UK. Although the number of people diagnosed late (CD4 count <350 cells/mm^3 within 3 months of diagnosis) declined from 52% in 2004 to 44% in 2022, this figure remains high. Reducing late diagnosis and undiagnosed HIV is critical to both individuals and public health. Increasing testing access in both sexual health and other care settings such as emergency medicine and primary care is a critical intervention, as is HIV testing in patients presenting with clinical conditions that are associated with HIV (indicator conditions). The UK Government HIV Action Plan has mobilized resources to expand HIV testing, although at the time of writing, this remains confined to areas of high HIV prevalence.

The changing face of the epidemic

With effective ART, increasing numbers of people living with HIV are living with long-term, sustained viral suppression. In countries where ART is consistently accessible, AIDS-defining illnesses have given way to non-communicable conditions as the major health problems for people living with HIV. Although chronic health conditions accumulate during the life course, higher levels of multimorbidity and polypharmacy occur in people living with HIV at a younger age compared with the general population, including frailty and its associated disabilities. Leading causes of hospital admission of people living with HIV are respiratory, psychiatric, cardiovascular, renal, and neurological disorders. Data from the UK show that 75% of those living with HIV have at least one other long-term condition, including mental health conditions, hypertension, lipid disorders, and diabetes. The prevention, identification, and treatment of co-morbidities through an integrated, outcomes-focused, person-centred approach are central to HIV care.

ACQUISITION

Routes of acquisition

Even though HIV can be isolated from a wide range of body fluids and tissues, most infections are transmitted either sexually via semen and cervical secretions or via blood. Transmission risk correlates to the HIV viral load, which is highest in acute infection and can be decreased to undetectable by effective ART. HIV-associated stigma and discrimination; gender-based violence; and in some countries, the legal position with respect to LGBTQ+ and other human rights for those at high risk can restrict access to appropriate services and increase the risks of transmission and disengagement from care.

There is no evidence that HIV is spread by casual social contact (kissing, holding hands), household contact (plates, cutlery, towels), or insects such as mosquitoes and bed bugs.

Sexual transmission (vaginal and anal)

Heterosexual sexual transmission accounts for most global infections. HIV is readily transmitted from male to female genital tracts and through anal sex. Coexistent sexually transmitted infections (STIs), especially genital ulceration, enhance transmission. However, when the viral load is undetectable there is zero risk of sexual transmission. Male circumcision has been shown to reduce both acquisition and transmission. In the UK, 29% of new diagnoses in 2022 were reported among GBMSM and 28% among females who have sex with males. In central and sub-Saharan Africa, the epidemic has always been heterosexual with most adults living with HIV in these regions being female. Central Asia, Eastern Europe,

the Middle East, and North Africa continue to see expanding epidemics, largely through heterosexual transmission and opioid use, exacerbated by high rates of STIs and cultural and legal constraints affecting those at most risk.

Vertical transmission (transplacental, perinatal, breast-feeding)

Globally, vertical transmission remains the most common route of HIV acquisition in children. European studies suggest that with no intervention, 15% of babies born to mothers living with HIV are likely to acquire HIV, although rates of up to 40% have been reported in Africa and the USA. Increased likelihood of transmission is associated with advanced disease in the mother, higher maternal viral load, prolonged rupture of membranes, and chorioamnionitis. Although transmission can occur *in utero*, most transmissions take place perinatally. Breast-feeding has been shown to double the risk of vertical transmission in the pre-treatment era; however, the risk is very low for females using effective ART. In high-income countries, combined interventions to reduce transmission, including opt out HIV screening during pregnancy, early use of ART, and replacement infant feeding have led to a dramatic fall in the numbers of children acquiring HIV. The vertical transmission rate has been below 0.4% in England since 2012, reducing from 2.86% in 2000 to 2001. The lack of universal access to these interventions in resource-poor countries where 90% of vertical infections occur is a major health inequality. Vertical transmission rates are now included in global monitoring frameworks to assess progress toward eliminating the HIV epidemic.

Contaminated blood, blood products, and organ donation

Screening of blood and blood products was introduced in 1985 in Europe and North America. Before this, HIV infection was associated with the therapeutic use of blood transfusion and clotting factors (in haemophilia). In some parts of the world where blood products may not be screened and in areas where the rate of new HIV infections is very high, transfusion-associated transmission continues to occur.

Contaminated needles

The practice of sharing needles and syringes for intravenous (i.v.) drug use continues to be a major route of HIV transmission in parts of South-east Asia, Latin America, and Eastern Europe. Successful education and needle exchange schemes have significantly reduced the rate of transmission by this route in the UK.

Iatrogenic transmission from reuse of needles and syringes in low- and middle-income countries is reported. Healthcare workers have a risk of approximately 0.3% after a single needle-stick injury from a person living with HIV, although data were obtained before widespread use of effective ART and the associated reduction in viral load.

VIRAL STRUCTURE

HIV belongs to the lentivirus group of retroviruses. There are two types, HIV-1 and HIV-2. HIV-1 is the most frequently occurring strain globally. HIV-2 is almost entirely confined to West Africa and European countries with colonial links to West Africa, particularly France and Portugal. HIV-2 has only 40% structural homology with HIV-1. Although associated with immunosuppression and AIDS, HIV-2 is more indolent than HIV-1. It is also less responsive to ART. The structure of HIV is shown in Fig. 44.1.

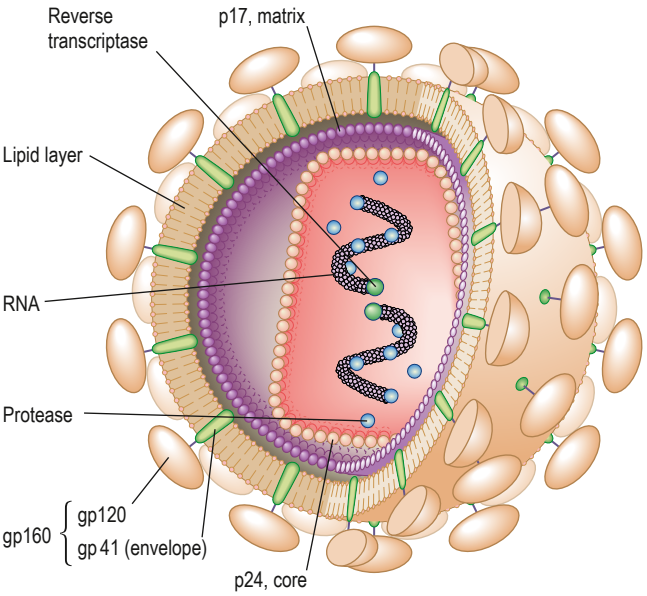


Fig. 44.1 Structure of HIV. Two molecules of single-stranded RNA are shown within the nucleus. The reverse transcriptase polymerase converts viral RNA into DNA (a characteristic of retroviruses). The protease includes integrase (p32 and p10). p24 (core protein) concentrations can be used to monitor HIV disease. p17 is the matrix protein. Glycoprotein 120 (gp120) is the outer envelope glycoprotein, which binds to cell surface CD4 molecules. gp41, a transmembrane protein, influences infectivity and cell fusion capacity.

Retroviruses are characterized by possession of the enzyme reverse transcriptase, which allows viral mRNA to be transcribed into DNA and incorporated into the host genome. Reverse transcription is an error-prone process that when combined with high viral turnover leads to considerable genetic variation and a diversity of viral subtypes or clades across the world. Based on DNA sequencing, HIV-1 is divided into four distinct strains, which represent four independent cross-species transfers: three (M, N, and O) are based on the chimpanzee-related strains of simian immunodeficiency virus (SIV), and one (P) may represent chimpanzee-to-gorilla-to-human transmission.

Group M (major) is the lineage that accounts for at least 98% of infections worldwide. Within the M group, there is a high degree of diversity, with subtypes (or clades), denoted A to K. Whereas there is a predominance of subtype B in Europe, North America, and Australia, areas of Central and sub-Saharan Africa have multiple M subtypes, clade C being most common. Recombination of viral material generates an array of circulating recombinant forms, which increase the genetic diversity. Recombinant forms are more common than non-recombinant forms in certain areas.

PATHOGENESIS

The relationship between the virus and the host defines pathogenesis. At the time of initial exposure, the virus is transported by dendritic cells from mucosal surfaces to regional lymph nodes where permanent infection is established, usually by one ‘founder virus’. The HIV surface glycoprotein 120 (gp120) is recognized by the CD4 T-cell, which is the main cell susceptible to infection. The interaction

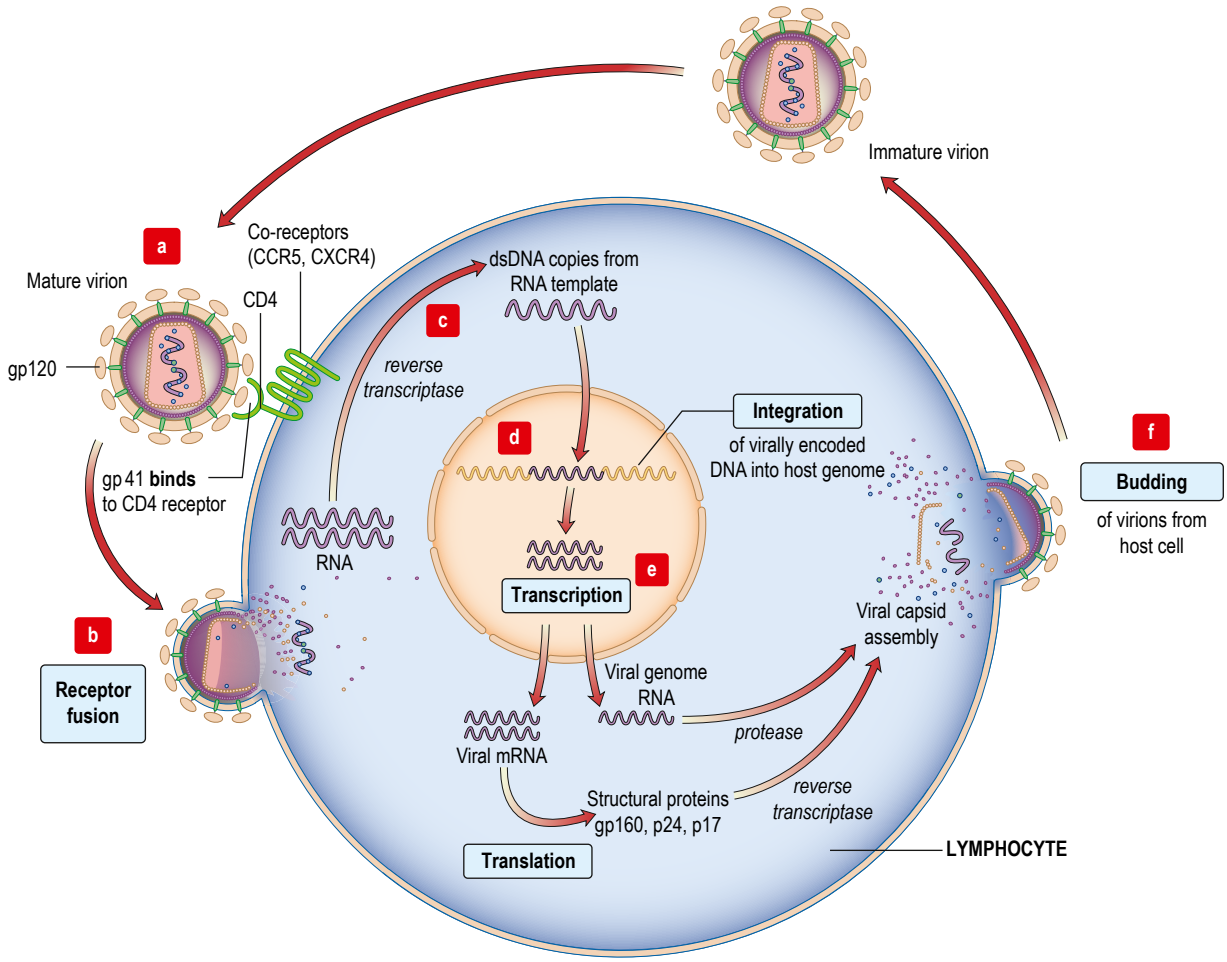


Fig. 44.2 Human immunodeficiency virus entry and replication in CD4 T lymphocytes. (a) ***Binding:*** the virus binds to host CD4 receptor molecules via the envelope glycoprotein gp120 and co-receptors CCR5 and CXCR4. **(b)** ***Fusion:*** a subsequent conformational change results in fusion between glycoprotein 41 (gp41) and the cell membrane. **(c)** ***Reverse transcription:*** entry of the viral capsid is followed by uncoating of the RNA. DNA copies are made from both RNA templates. DNA polymerase from the host cell leads to the formation of double-stranded DNA (dsDNA). **(d)** ***Integration:*** in the nucleus, virally encoded DNA is inserted into the host genome. **(e)** ***Transcription:*** regulatory proteins control transcription (an RNA molecule is now synthesized from the DNA template). **(f)** ***Budding:*** the virus is re-assembled in the cytoplasm and budded out from the host cell.

between CD4 and HIV gp120 surface glycoprotein, together with host chemokine CCR5 or CXCR4 co-receptors, facilitates entry into host cells (Fig. 44.2). Early infection is characterized by CCR5 CD4 memory T lymphocytes infection and depletion in the gastrointestinal (GI) tract. This compromises mucosal immune function, allowing microbial lipopolysaccharides to enter the circulation. HIV infection independent of CD4 receptors can occur in glial cells, monocytes, macrophages, and renal epithelial cells, leading to end-organ manifestations.

Studies of viral turnover have demonstrated a virus half-life of approximately 6 hours. To maintain plasma viraemia, 10^8 to 10^9 virus particles are released and cleared daily. Virus production by infected cells lasts for about 2 days and is probably limited by the death of the cell.

Progressive loss of activated CD4 T-cells due to killing by CD8 T-cells is a key factor in the immunopathogenesis of HIV. The production of broadly neutralizing antibodies occurs between 12 and 20 weeks after infection. These are usually clade specific but, in some people can be effective against several viral subtypes. Cell-mediated immunodeficiency from CD4 T-cell depletion leaves the person open to infections with intracellular pathogens, especially encapsulated bacteria.

HIV is also associated with immune activation, which leads to a long-term inflammatory state and is a key driver of co-morbidity. T-cell activation is observed from the earliest stages of infection, which leads to an increase in the numbers of CD4-bearing target cells susceptible to destruction. HIV itself is considered a pro-inflammatory state, with co-pathogens (e.g., cytomegalovirus [CMV]) and microbial translocation also implicated. Inflammation is demonstrated by raised concentrations of inflammatory cytokines (e.g., interleukin-6, D-dimer, and coagulation cascade activation). These pro-inflammatory phenomena contribute to HIV-associated end-organ damage, and co-morbidities, including cardiovascular disease, neuro-inflammation, and certain malignancies.

The spectrum of untreated HIV disease is broad; manifestations result from HIV-associated immune dysfunction, inflammation, co-morbidity, and co-infection. Most people living with HIV in high-income settings begin treatment early in the course of infection. With early initiation of ART and without immunosuppression, progression to an AIDS-defining event is rare. Without treatment,

Box 44.3 AIDS-defining conditions

- Candidiasis of the bronchi, trachea, or lungs
- Candidiasis, oesophageal
- Cervical carcinoma, invasive
- Coccidioidomycosis, disseminated or extrapulmonary
- Cryptococcosis, extrapulmonary
- Cryptosporidiosis, chronic intestinal (1-month duration)
- CMV disease (other than of the liver, spleen, or nodes)
- CMV retinitis (with loss of vision)
- Encephalopathy, HIV-related
- Herpes simplex, chronic ulcers (1-month duration), or bronchitis, pneumonitis, or oesophagitis
- Histoplasmosis, disseminated or extrapulmonary
- Isosporiasis, chronic intestinal (1-month duration)
- Kaposi's sarcoma
- Lymphoma, Burkitt
- Lymphoma, immunoblastic (or equivalent term)
- Lymphoma (primary) of brain
- *Mycobacterium avium-intracellulare* complex or *Mycobacterium kansasii*, disseminated or extrapulmonary
- *Mycobacterium tuberculosis*, any site
- *Mycobacterium*, other species or unidentified species, disseminated or extrapulmonary
- *Pneumocystis jirovecii* pneumonia
- Pneumonia, recurrent
- Progressive multifocal leucoencephalopathy
- *Salmonella* septicaemia, recurrent
- Toxoplasmosis of brain
- Wasting syndrome caused by HIV

CMV, Cytomegalovirus.

Box 44.4 Summary of the Centers for Disease Control and Prevention classification of HIV

Absolute CD4 count (/mm³)	A	B	C
	Asymptomatic or persistent generalized lymphadenopathy or acute seroconversion illness	HIV-related conditions[a], not A or C	Clinical conditions listed in AIDS surveillance case definition (see Box 44.3)
>500	A1	B1	C1
200–499	A2	B2	C2
<200	A3	B3	C3

[a]Examples of category B conditions include bacillary angiomatosis, candidiasis (oropharyngeal), constitutional symptoms, oral hairy leukoplakia, herpes zoster involving more than one dermatome, idiopathic thrombocytopenic purpura, listeriosis, pelvic inflammatory disease (especially if complicated by tubo-ovarian abscess), and peripheral neuropathy.
AIDS, Acquired immunodeficiency syndrome.

immunosuppression is progressive, opportunistic infections and tumours that meet the criteria for the diagnosis of AIDS (Box 44.3) are invariable, and the disease is fatal.

Several international HIV classifications exist, the two most frequently used being the WHO and Centers for Disease Control and Prevention classifications (Box 44.4). These classifications are used in public health monitoring and surveillance. Both classifications include a combination of immunological markers (e.g., CD4 count) correlated to the presence or absence of clinical AIDS-defining conditions.

Early disease: incubation and acute seroconversion illness

Early infection refers to the 6-month period after HIV acquisition and is marked by significant viral replication, resulting in very high concentrations of viraemia in the plasma and genital tract and consequently by high infectivity. This time is increasingly recognized as a key period for intervention, both to reduce long-term sequelae and to cut onward transmission.

The 2 to 4 weeks immediately after infection may be silent, both clinically and serologically. Alternatively, some people present with a self-limiting acute viral illness, which may be indistinguishable from glandular fever, which occurs 3 to 12 weeks after exposure. Symptoms include headache; fever; arthralgia; myalgia; lethargy; lymphadenopathy; sore throat; mucosal ulcers; and occasionally, a transient, faint pink, maculopapular rash. Neurological symptoms include photophobia; myelopathy; neuropathy; Guillain-Barré syndrome; and in rare cases, aseptic meningitis. The illness lasts up to 3 weeks, and recovery is usually complete. This is a crucial point for diagnosis of HIV; however, HIV is frequently not considered in the differential diagnosis in patients presenting with such non-specific viral symptoms, which is why in the UK routine HIV testing is recommended for presentations of acute mononucleosis.

Laboratory abnormalities in the early phase of infection include:

- Lymphopenia
- Atypical reactive lymphocytes on blood film
- Thrombocytopenia
- Raised liver transaminases
- Low CD4 lymphocytes
- Reversed CD4-to-CD8 ratio

Viral testing

Although antibodies to HIV may be absent during the earliest stages of infection, concentrations of circulating viral RNA are high, and p24 core protein may be detectable. HIV RNA polymerase chain reaction (PCR) may be positive within 10 days of exposure, with p24 antigen emerging a few days later. With increasingly sensitive assays, most people develop detectable antibodies within the first 2 to 4 weeks of early infection. Immunoglobulin (Ig) M antibodies followed by IgG antibodies usually appear at 2 to 6 weeks after infection. If early or acute HIV infection is suspected but standard diagnostic antibody test results are negative, repeat testing in 7 days and referral for expert advice are recommended.

In infants born to females living with HIV, maternal HIV antibody is passively transferred to the infants across the placenta and may persist in the absence of infection for up to 18 months. Nucleic acid testing is the preferred approach in infants younger than 18 months of age.

HIV antibody responses in early infection may be altered in people using antiviral PrEP, making RNA detection a more reliable indicator of early infection.

Clinical latency

The rate of clinical progression of untreated HIV is variable. Most people living with HIV are asymptomatic for years. While undiagnosed and untreated, the virus continues to replicate, leading to immunosuppression and risk of onward transmission. In most people, a gradual decline in CD4 count occurs over a period of approximately 10 years before progression to symptomatic disease or AIDS (Fig. 44.3). Others progress much more rapidly, with high concentrations of viral RNA and a rapid decline in CD4 count over 2 to 5 years. 'Long-term non-progressors' may continue with a normal CD4 count for decades. Within this group, a small subpopulation of 'elite suppressors' maintain a persistently undetectable viral load (usually defined as below 50 cpml) and a high CD4 count without taking ART. Older age is associated with more rapid progression. Sex and pregnancy do not appear to influence progression.

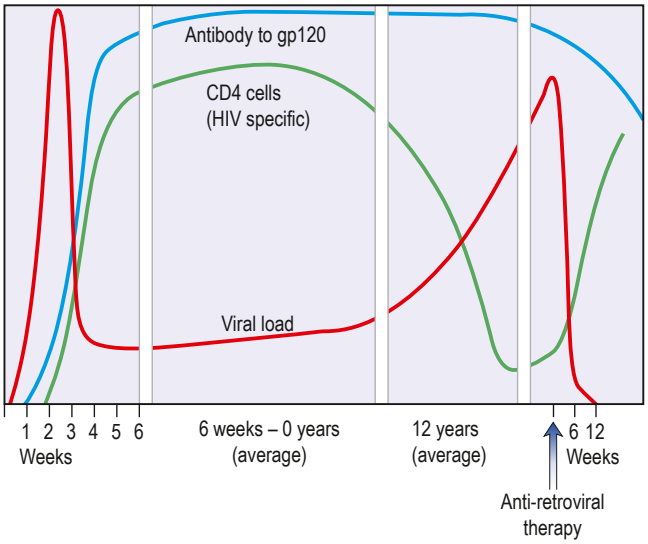


Fig. 44.3 The immune response to HIV (seroconversion).
gp120, Glycoprotein 120.

Persistent generalized lymphadenopathy

A subgroup of patients with asymptomatic infection has persistent generalized lymphadenopathy (PGL). This is defined as lymphadenopathy at two or more extra-inguinal sites for more than 3 months in the absence of other causes. The nodes are usually symmetrical, firm, mobile, and non-tender and may be accompanied by splenomegaly. Biopsy is rarely indicated but if undertaken reveals an architecture of follicular hyperplasia and endothelial proliferation. The presence of PGL does not seem to affect disease progression, although as the disease progresses, nodes may disappear.

SYMPTOMATIC HIV INFECTION

As immunosuppression worsens and the viral load rises, an array of symptoms and signs occurs. The clinical picture is caused by

- HIV-related end-organ disease
- HIV-associated immunosuppression

HIV-related end-organ disease

Neurological disease

Infection of the nervous tissue occurs at an early stage, and the likelihood of clinical neurological involvement increases with advanced disease. Manifestations include the following:

- AIDS dementia complex (ADC)
- HIV-associated neurocognitive disorder
- Progressive multifocal leucoencephalopathy (PML)
- Neuropathy
- Aseptic meningitis

These conditions have become much less frequent since the introduction of ART. The pathogenesis is thought to be caused by the release of neurotoxic products by HIV and cytokine abnormalities secondary to immune dysregulation. Viral suppression in the plasma is essential to control central nervous system (CNS) manifestations. Very unusually resistance may develop in the CNS compartment, necessitating switching or intensifying ART.

Box 44.5 Mucocutaneous manifestations of HIV infection

Cutaneous

- Dry skin and scalp
- Onychomycosis
- Seborrhoeic dermatitis
- Tinea: cruris, pedis
- Pityriasis: versicolor, rosea
- Folliculitis
- Acne
- Molluscum contagiosum
- Warts
- Herpes zoster: multidermatomal, disseminated
- Papular pruritic eruption
- Scabies
- Ichthyosis
- Kaposi's sarcoma

Mucous membranes

- Candidiasis: oral, vulvovaginal
- Oral hairy leukoplakia
- Aphthous ulcers
- Herpes simplex: genital, oral, labial
- Periodontal disease
- Warts: oral, genital

ADC has varying degrees of severity, ranging from mild memory impairment and poor concentration to severe cognitive deficit, personality change, and psychomotor slowing. Changes in affect are common, and depressive or psychotic features may be present. In severe cases, cerebral atrophy is shown on computed tomography (CT). T2-weighted magnetic resonance imaging (MRI) shows white matter lesions of increased density. Electroencephalography shows non-specific changes consistent with encephalopathy. The cerebrospinal fluid (CSF) is usually normal, although the protein concentration may be raised. Histologically, the spinal cord may show vacuolar myelopathy. Patients with mild neurological dysfunction may be more sensitive to a superimposed insult: fever, metabolic disturbance, or psychotropic medication may lead to a marked deterioration in cognitive functioning.

Sensory polyneuropathy is seen in advanced HIV infection, mainly in the legs and feet, although the hands may also be affected. Severe forms cause intense pain, which disrupts sleep, impairs mobility, increases frailty, and reduces quality of life. ***Autonomic neuropathy*** leads to postural hypotension and diarrhoea.

Eye disease

Ocular manifestations are common with advanced disease:

- Cytomegalovirus (CMV) retinitis is sight threatening and has a characteristic 'pizza retina' appearance on fundoscopy (see Fig. 44.4).
- Benign retinal cotton wool spots are rarely troublesome and should not be confused with CMV retinitis.
- Anterior uveitis can present as an acute red eye. It can occur *de novo* but also secondary to drugs (e.g., rifabutin therapy for mycobacterial infection) and infection (e.g., syphilis).
- Retinal disease can occur with *Pneumocystis* infection, toxoplasmosis, syphilis, and lymphoma, and the eye may be the site of first presentation.

Referral to ophthalmology for ocular manifestations is an urgent priority.

Mucocutaneous manifestations

The skin is commonly affected in people living with HIV (Box 44.5) because both dendritic and Langerhans cells are target cells for HIV:

- Pruritus is common at all stages of HIV. Generalized dry, itchy, flaky skin is typical, and hair may become thin and dry. An intensely pruritic papular eruption favouring the extremities may be found, particularly in patients of African ethnicities.
- Eosinophilic folliculitis presents with urticarial lesions, particularly on the face, arms, and legs.
- Otitis externa may be recurrent.

- Cutaneous drug reactions are common (e.g., rashes secondary to sulphur-containing drugs; Fig. 45.33).
- Delayed hypersensitivity is a good indicator of cell-mediated immunity. It is frequently reduced or absent, even before clinical immunosuppression.
- Aphthous ulceration can be severe, recurrent, and refractory to treatment. It can impair the patient's ability to eat. Biopsy may be needed to exclude other causes of ulceration. Topical steroids are used, and resistant cases may respond to thalidomide.

Haematological complications

These are common in advanced HIV infection:

- Lymphopenia progresses as the CD4 count falls.
- Anaemia of chronic HIV infection is usually mild, normochromic, and normocytic.
- Neutropenia is common and usually mild.
- Isolated thrombocytopenia may occur early in infection and be the only manifestation of HIV. Platelet counts are often moderately reduced but can fall dramatically to 10 to 20 × 10^9/L leading to easy bleeding and bruising. Megakaryocytes are increased in the bone marrow, but their function is impaired. Circulating antiplatelet antibodies lead to peripheral destruction. Effective ART alone usually resolves thrombocytopaenia. However, in some cases, it is refractory and may require additional therapies such as i.v. immunoglobulin therapy, corticosteroids, and platelet transfusion.
- Pancytopenia may occur because of underlying opportunistic infections (e.g., *Mycobacterium avium-intracellulare* [MAI], disseminated CMV) or malignancies (e.g., lymphoma).
- Myelotoxic drugs used in HIV include ganciclovir and valganciclovir (neutropenia), systemic chemotherapy (pancytopenia), co-trimoxazole (agranulocytosis), and rifabutin (neutropenia, thrombocytopenia).

Gastrointestinal effects

Weight loss and diarrhoea are common in people with advanced HIV. Wasting is attributed to a direct effect of viraemia on metabolism but may also be a consequence of anorexia. People living with HIV who are not on ART have lower body mass indexes (BMIs) than people on HIV treatment.

HIV enteropathy with varying degrees of villous atrophy causes chronic diarrhoea in the absence of other pathogens. Achlorhydria is reported in people with advanced HIV disease, affecting absorption, nutrition, and drug pharmacodynamics; the mechanism is unclear.

Kidney complications

HIV-associated nephropathy (HIVAN) is a cause of significant kidney disease. It is most frequently seen in people of Black ethnicities. It causes a collapsing variant of focal segmental glomerulosclerosis, along with a tubulopathy due to direct viral infection of tubular cells. The kidneys may be hyperechoic on ultrasound examination. HIVAN requires urgent ART and specialist renal care. The course is usually progressive, and dialysis may be required. Other kidney pathology in HIV includes ***HIV-immune complex kidney disease***, which can present with a range of glomerular abnormalities, IgA nephropathy, and thrombotic microangiopathy.

Many nephrotoxic drugs are used in the management of opportunistic infections, for example, foscarnet, amphotericin B, pentamidine, and sulfadiazine. Tenofovir (an anti-retroviral drug) is associated with renal impairment and very rarely with Fanconi's syndrome (dysfunction of the proximal renal tubule leading to glycosuria, phosphaturia, and aminoaciduria).

Respiratory complications

Pneumocystis jiroveci pneumonia (PJP) is the commonest pulmonary complication in advanced HIV (see later). Pneumonia is more common in HIV compared than in the general population (see Box 44.13). Pneumococcal, influenza, and SARS-CoV-2 vaccinations are recommended (see Box 44.14).

Lymphoid interstitial pneumonitis is very uncommon in adults but occurs in children. There is an infiltration of lymphocytes, plasma cells, and lymphoblasts in alveolar tissue. Epstein–Barr virus (EBV) may also be present. Lymphocytic, lymphoblastic, and plasma cell infiltration leads to a dry cough, which may be confused with *Pneumocystis* infection based on comparable reticulonodular radiograph findings. ART and corticosteroids are the mainstays of treatment.

Sinusitis is seen in advanced infection. Antibacterial therapy and topical steroids are usually effective, but some patients require surgical intervention. A similar process is seen in the middle ear, which can lead to chronic otitis media.

Endocrine complications

Endocrine abnormalities have been reported in people living with HIV and include reduced concentrations of testosterone, early menopause, and adrenal insufficiency. The latter mainly occurs in those with advanced disease and is also associated with CMV co-infection. Gluco- and mineralocorticoid replacement may be needed.

Cardiac complications

People living with HIV are twice as likely to develop cardiovascular disease as the general population. Ischaemic heart disease is associated with viraemia and occurs when ART is taken intermittently. The increased risk in postmenopausal females may be overlooked. Management of cardiovascular risk factors, including hyperlipidaemia, hypertension, diabetes, obesity, and smoking, is required. The use of statins has been shown to reduce the incidence of major cardiac events and death in people living with HIV even when atherosclerotic cardiovascular disease risk is low. Statins (preferably pitavastatin or atorvastatin) should be offered to all people living with HIV aged older than 40 years, irrespective of lipid profile or estimated CVD risk. Statin provision is a priority for those aged older than 40 with a 10-year cardiovascular disease risk greater than 5%.

Cardiomyopathy is associated with HIV and may progress to congestive cardiac failure. Lymphocytic and necrotic myocarditis are described. In untreated disease, direct infection of the myocardium, opportunistic infections, and nutritional disorders all contribute. For people on ART, ischaemic cardiomyopathy is more common. Specialist cardiology assessment is useful in determining the aetiology.

Conditions associated with immunosuppression

Immunodeficiency leads to opportunistic disease (Box 44.6). Many of the organisms causing clinical disease in HIV are latent or ubiquitous in the environment. Opportunistic illnesses are frequently caused by organisms that

- Are not considered pathogenic in the absence of HIV
- Cause unusually severe disease in HIV

Box 44.6 Major HIV-associated pathogens

Protozoa
- *Toxoplasma gondii*
- *Cryptosporidium parvum*
- *Microsporidia* spp.
- *Leishmania donovani*
- *Isospora belli*

Viruses
- Cytomegalovirus
- Herpes simplex
- Varicella zoster
- Human papillomavirus
- Human herpes virus-8
- JC polyomavirus

Fungi and yeasts
- Pneumocystis jirovecii
- *Cryptococcus neoformans*
- *Candida* spp.
- Dermatophytes (*Trichophyton* spp.)
- *Aspergillus fumigatus*
- *Histoplasma capsulatum*
- *Coccidioides immitis*

Bacteria
- *Salmonella* spp.
- *Mycobacterium tuberculosis*
- *Mycobacterium avium-intracellulare*
- *Streptococcus pneumoniae*
- *Staphylococcus aureus*
- *Haemophilus influenzae*
- *Moraxella catarrhalis*
- *Rhodococcus equi*
- *Bartonella quintana*
- *Nocardia* spp.

- Are oncogenic, e.g., tumours with a viral aetiology.

Susceptibility increases with advancing immunosuppression. Mechanisms of increasing susceptibility to infection include:
- Defective T-cell function against protozoa, fungi, and viruses
- Impaired macrophage function against intracellular bacteria such as *Mycobacteria* and *Salmonella* spp.
- Defective B-cell immunity against capsulated bacteria such as *Streptococcus pneumoniae* and *Haemophilus* spp.

CD4 T-lymphocyte numbers are used as markers to predict risk of infection. For example, PJP and toxoplasmosis occur with CD4 counts less than 200 cells/mm^3, and *M. avium* complex (MAC) and CMV occur with CD4 counts less than 100 cells/mm^3. People with CD4 counts of more than 200 cells/mm^3 are at low risk of ***AIDS-defining conditions*** (see **Box 44.3**). This hierarchy of infection allows for appropriate intervention with prophylactic drugs.

Presentations may be atypical in advanced HIV because the inflammatory response is impaired. Examples are the absence of meningism (neck stiffness) in cryptococcal meningitis or a paucity of clinical findings in early PJP. Multiple pathogens may coexist. Radiological findings may be helpful because indirect serological tests are frequently unreliable. Diagnosis may require specimens from the appropriate site for culture or histology.

The clinical consequences of HIV-related immune dysfunction depend on at least three factors:
- ***Previous microbial exposure.*** Many clinical infections are caused by reactivation of latent infection.
- ***Pathogenicity of organisms encountered.*** High-grade pathogens, such as *Mycobacterium tuberculosis*, *Candida* spp., and the herpesviruses, are clinically relevant even when immunosuppression is mild. Less virulent organisms cause disease in those with more advanced immunodeficiency. Organisms usually controlled by intact cell-mediated immunity are a particular clinical problem.
- ***Degree of host immunosuppression.*** When patients are severely immunocompromised (CD4 count <100 cells/mm^3), disseminated infections with organisms of low virulence (e.g., MAI, *Cryptosporidium* spp.) can establish themselves. These infections are very resistant to treatment, mainly because there is no functioning immune response to clear organisms. The mpox mortality rate in people with advanced HIV infection (CD4 count <100 cells/mm^3) is 27%.

HIV TESTING

HIV testing is a key intervention in treatment and prevention of HIV. A range of settings and modalities helps maximize testing opportunities. Infection is diagnosed either by detection of virus-specific antibodies (anti-HIV) or by direct identification of viral material. Modern 'fourth-generation' tests, which are now the standard of care in many countries, test for anti-HIV antibodies and the p24 antigen simultaneously.

HIV antibodies

This is the most used marker of infection. Routine tests are based on enzyme-linked immunosorbent assay techniques. An initial screening test should be with a high-sensitivity fourth-generation assay and then confirmed using a second high specificity diagnostic platform of equivalent sensitivity. Rapid diagnostic tests are frequently used as point-of-care testing (POCT). All positive HIV POCT results must be followed by laboratory testing for confirmation. Up to 3 months (mean 6 weeks) may elapse between infection and antibody detection (serological latency, 'window period'). Antibodies to HIV have no protective function and persist for life.

As with all IgG antibodies, anti-HIV cross the placenta. Babies born to females living with HIV have detectable maternal antibodies at birth; these antibodies are lost gradually over the first 18 months of life in babies without HIV. Anti-HIV antibody is therefore not a reliable marker vertical transmission before 18 months of age.

The Serological Testing Algorithm for Recent HIV Seroconversion (STARHS) (also termed Recent Infection Treatment Algorithm [RITA]) identifies recently acquired infections by pairing a more sensitive and a less sensitive (detuned) test. A positive result on the sensitive test and a negative 'detuned' test are indicative of recent infection, and positive results on both tests point to an infection that is more than 130 days old. The major application of this is in epidemiological surveillance and monitoring.

Viral p24 antigen

The p24 antigen (p24ag) is a HIV viral capsid protein that can be detected in blood shortly after infection. It usually disappears 8 to 10 weeks after infection (see **Fig. 44.3**). It is helpful in identifying individuals with recent infection who have not yet mounted an antibody response. Anti-p24 antibodies (anti-p24) can be detected in the early weeks of infection and through the asymptomatic phase. They are frequently lost as the disease progresses.

In people using PrEP, the detection of viral antigens or antibody responses may be delayed, masking early infection.

Genome detection assays

Nucleic acid–based assays are used to aid diagnosis in situations in which serological tests are not reliable. These include use in the babies of those living with HIV in pregnancy and in early infection when antibody may not be present. They are also used for subtyping HIV variants, including for medicolegal reasons.

CLINICAL APPROACH TO THE PERSON LIVING WITH HIV

Initial assessment

People are newly diagnosed with HIV in many clinical settings. Referral should be made to the clinical team responsible for HIV care. In the UK, HIV care is usually undertaken by infectious disease

Box 44.7 Baseline assessment of asymptomatic, newly-diagnosed HIV

History
- Coexisting conditions: medical, sexual, and reproductive, psychiatric
- Knowledge and beliefs about HIV, including treatment and transmission
- Full drug history, including recreational drug use
- Vaccination history
- Social circumstances: relationships, disclosure, support

Examination
- General physical examination including weight, height, BMI, blood pressure, and waist circumference

Investigations

Haematology
- Full blood count, differential count, and film

Biochemistry
- Kidney function, including eGFR
- Liver blood tests
- Fasting serum lipid profile: total cholesterol, HDL
- Fasting blood glucose and haemoglobin A_{1c}
- Serum bone profile
- Urinalysis
  - Dipstick for blood, protein, and glucose
  - Urine protein:creatinine ratio

Immunology
- Lymphocyte subsets
- HLA-B*5701 status

Virology
- HIV antibody (confirmatory)
- HIV viral load
- HIV genotype and subtype determination
- Hepatitis A IgG
- Hepatitis B surface antigen and full profile
- Hepatitis C antibody (followed by hepatitis C RNA testing if antibody-positive, and confirmation of antibody-positive status if RNA-negative)

Microbiology
- Toxoplasmosis serology
- Syphilis serology
- Tuberculosis status
- Screen for other STIs

Other
- Cervical cytology
- Chest X-ray if indicated
- 10-year cardiovascular risk assessment
- Fracture risk assessment

BMI, Body mass index; *eGFR*, estimated glomerular filtration rate; *HDL*, high-density lipoprotein; *HLA*, human leucocyte antigen; *Ig*, immunoglobulin; *STI*, sexually transmitted infection.

or sexual health services. An HIV specialist clinician should review all patients with new diagnoses within 2 weeks or on the same day if the patient is unwell.

A full medical history, physical examination, and laboratory evaluation should be undertaken at diagnosis to determine the stage of infection, evaluate co-morbidities and co-infections, and assess mental health. The initial assessment should include the patient's socioeconomic situation, relationships, family and social support networks, and substance misuse. Contact tracing and partner notification should be initiated and managed by the HIV team in all cases, including children who require testing. Baseline investigations depend on the clinical setting, but those appropriate for an asymptomatic person in the UK are shown in **Box 44.7**.

HIV resistance testing (genotyping)

Sequencing of the viral genotype at diagnosis establishes the viral subtype and identifies resistance to ART. The results are used to guide the selection of ART agents. Genotyping cannot be performed in people with an undetectable viral load because replicating virus is needed.

Monitoring

Regular monitoring is required with the schedule determined by clinical, virological, and treatment stages. All reasonable efforts should be made to encourage people to take ART, but in rare cases when people choose not to initiate therapy, monitoring should take place according to the degree of immunosuppression and no less frequently than two to four times per year.

Immunological monitoring

CD4 lymphocyte counts and the **CD4:CD8 ratio** are monitored. The absolute CD4 count and its percentage of total lymphocytes fall as HIV progresses, conferring risk of HIV-related pathology; patients with CD4 counts less than 200 cells/mm^3 or CD4 percentage less than 10% are at greatest risk. Rapidly falling CD4 counts are an indication for urgent referral and initiation of ART. Factors other than HIV (e.g., smoking, exercise, intercurrent infections, diurnal variation) also affect CD4 numbers. When stable on ART with a CD4 above 350 cells/mm^3, CD4 counts are performed annually.

Virological monitoring

The **HIV viral load** (HIV RNA) is the number of copies of HIV RNA in the blood. It is monitored at least twice per year. The aim of therapy is to suppress viral replication to achieve an undetectable viral load defined as less than 50 copies/mL. Viral load tests include quantitative branched DNA (bDNA) and reverse transcriptase PCR. Very sensitive tests can report 'target not detected (0 copies/mL)'. For the purposes of HIV transmission risk, the World Health Organization's 'Say Zero' campaign uses a threshold viral load less than 200 copies/mL; there is negligible risk of transmission at less than 1000 copies/mL.

Transient increases in viral load (viral blips) may occur after immunizations (e.g., against influenza, SARS-CoV-2 and pneumococcus) or during episodes of acute intercurrent infection (e.g., tuberculosis [TB]). Repeat viral load measurements should not be carried out within 1 month of these events. People should be reassured about the risks of transmission to partners according to the values given earlier.

HIV RNA is the standard marker of anti-retroviral treatment efficacy (see later). An uncontrolled viral load in a person on ART indicates either problems with adherence or viral resistance.

HIV viral load should be measured when ART is initiated and again at 1, 3, and 6 months after starting. Ideally, the viral load falls 10-fold after 1 month, with maximum effect at 10 to 12 weeks. When switching ART, viral load should be monitored within 6 weeks. Viral load should be monitored three monthly in the first year and 6 monthly for people stable on ART.

By about 6 months after seroconversion to HIV, the viral set-point for an individual is established. There is a correlation between HIV RNA concentration and long-term prognosis, independent of the CD4 count. People with a viral load consistently above 100 000 copies/mL have a 10 times higher risk of progression to clinical AIDS over the next 5 years than those with counts consistently below 10 000 copies/mL.

Monitoring other aspects of health

Regular assessment is required, including intercurrent medical problems (including mental health), medications, vaccinations, recreational drug use, sexual history, reproductive decision-making, cervical cytology, and social situation (including support networks, employment, benefits, food security, and accommodation). Regular review includes advice on reducing HIV transmission, sexual and reproductive health, testing opportunities for STIs and blood-borne viruses, and cervical screening. Information is required to allow people to make informed choices about childbearing and breastfeeding. Diagnostic testing should be offered to sexual partners and existing family members as necessary.

Regular assessments of BMI, waist circumference, metabolic and bone markers, blood pressure, and cardiovascular risk are required. Dietary advice should be available when weight problems exist. General health promotion advice including exercise, smoking cessation, alcohol, diet, and drug misuse should be given, considering the increased risk of cardiovascular and metabolic co-morbidities associated with HIV and its treatment. Hormone replacement therapy should be considered for menopausal females.

Life expectancy for people who are diagnosed early during HIV infection and have life-long access to effective ART life approaches that of the general population. Yet **health-related quality of life** reported by people living with HIV is consistently lower than that of the general population and below that of people with other long-term conditions. Lower health-related quality of life scores are associated with poor clinical and virological outcomes. Underlying causes are multiple and complex with contributors including pain; anxiety and depression; impaired cognitive function; sleep disturbance; co-morbidity; and social, welfare, and spiritual concerns. Both internalized and external HIV associated stigma have a detrimental impact on quality of life. Effective healthcare approaches must recognize and address the full range of symptoms and wellbeing challenges faced by people living with HIV. HIV-validated patient-reported outcome measures are useful tools to achieve person-centred care that addresses the diversity and complexity of the concerns of people living with HIV.

HIV-associated stigma is a major driver of health inequalities globally, undermining every aspect of HIV-related prevention, treatment, and care. **Stigma** and **discrimination** are experienced at many levels ranging from discriminatory policies and laws; prejudice; social exclusion; and internalized stigma that leads to feelings of shame, low self-esteem, and fear of discrimination. HIV-associated stigma frequently intersects with other forms of stigma, including that associated with sexual orientation, gender, race, migration status, and age, compounding impact on mental and physical wellbeing. In females, gender-based violence should be explicitly discussed. Stigma within health and care settings is regularly reported by people living with HIV, preventing people from seeking care and getting timely and appropriate treatment. Examples include refusal of treatment, disrespect from healthcare providers, longer waiting times, and excessive infection control measures. Interventions to combat stigma are essential and include educational, empowerment, and skills-building approaches. At a health systems level, it is crucial to recognize, report, and investigate examples of discrimination, including lack of privacy and breaches of confidentiality. Designing and delivering health and care services in partnership with service users and ensuring that services are accessible to those who may already be underserved and marginalized will further improve outcomes for people living with HIV.

Ageing brings distinctive challenges to the experience of living with HIV with **frailty** developing earlier than for those without HIV. Higher rates of age-related co-morbidities occurring at an earlier age in people living with HIV add complexity to clinical care, increase rates of polypharmacy, and have an impact on physical and mental wellbeing. Higher rates of social isolation, loneliness, and economic hardship are experienced by older people living with HIV, who are also often dealing with intersectional stigma. Providers of care for older adults have limited knowledge and experience of the needs of older people living with HIV. At the same time, HIV clinical services need to evolve to meet the needs of their patients as they age, including the inclusion of frailty assessments in HIV care pathways.

MANAGEMENT OF HIV

Effective ART has transformed outcomes for people living with HIV, extending life expectancy toward that of the general population for those who are diagnosed before the onset of immunosuppression, reducing morbidity, and eliminating onward transmission for people who are virally suppressed. The aims of current management are to maintain viral suppression while maximizing long-term wellbeing, with focused prevention and treatment of HIV-associated co-morbidities (**Box 44.8**).

People living with HIV have a chronic, potentially stigmatizing, and sometimes unpredictable condition. Specialist clinical management must address co-morbidities, medication side-effects, complex drug interactions, and adherence (including the emergence of resistant viral strains when adherence has not been optimal). Even with complete viral suppression, ART does not fully restore health; treated infection is associated with a variety of co-morbidities, including cardiovascular disease, cancer, and impaired mental health.

There is no cure for HIV. Cure related research is still in its infancy with a focus on eliminating the viral reservoirs that current ART therapy is unable to achieve.

ANTI-RETROVIRAL THERAPY

The key practical principles of prescribing ART are given in **Box 44.9**.

ART continues to evolve and improve. Increased potency, reduced toxicity, single-tablet formulations, compounds with different mechanisms of action, and an improved understanding of drug resistance have all incrementally improved outcomes.

In people without drug resistance, the combined use of three drugs has been the gold-standard ART regimen. However, emerging evidence supports the use of certain two-drug therapies in people without drug resistance as a way of reducing cumulative toxicity while maintaining viral suppression. The first long-acting two-monthly injectable two-drug therapy is now available in some countries, including the UK, for some people who are virally suppressed. This regimen is less robust with respect to drug resistance but safe with careful selection based on clinical criteria and is preferred by some people who find it less stigmatizing and more convenient. Better clinical outcomes have been closely linked to physician expertise and the numbers of patients under direct care.

With increasing numbers of medication available as generics ('off-patent'), drug costs are falling. This necessitates cost-effectiveness assessment of novel agents to justify use. Regularly updated treatment guidelines are produced by the British HIV

Box 44.8 An approach to the sick patient living with HIV

Potential problems
- Adverse drug reactions
- Drug–drug interactions
- Presentation or complications of malignancy
- Immune reconstitution phenomenon
- Infection in an immunocompromised host
- Acute opportunistic infections
- Organic or functional brain disorders
- Non–HIV-related pathology

History
- Anti-retroviral drugs and prophylaxis
- Previous HIV-related pathology
- Potential sources of infectious agents: food hygiene, pets, contacts with acute infections, contact with tuberculosis, STIs
- Recreational drugs, travel
- Secure confidentiality: ask who is aware of HIV diagnosis

Examination
- Signs of adverse drug reactions (e.g., skin rashes, oral ulceration)
- Signs of disseminated sepsis
- Clinical evidence of immunosuppression (e.g., oral *Candida* infection, oral hairy leukoplakia)
- Neurological signs or meningism
- Evidence of altered mental state: organic or functional
- Genitalia (e.g., herpes simplex, syphilis, gonorrhoea)
- Fundi (e.g., cytomegalovirus retinitis)
- Mouth (e.g., *Candida* infection)
- Lymphadenopathy

Investigation
- Full blood count and differential count (lymphocyte subsets and viral load may be affected by intercurrent illness)
- Kidney function tests
- Liver blood tests
- Plasma glucose
- Blood gases, including acid–base balance
- Blood cultures, including specimens for mycobacterial culture
- Microscopy and culture of available or appropriate specimens: stool, sputum, urine, cerebrospinal fluid
- Malaria screen in recent travellers from endemic areas
- Serological tests for cryptococcal antigen, toxoplasmosis; save serum for viral studies
- Chest X-ray
- CT or MRI scan of the brain if there are focal neurological signs and *always* before lumbar puncture

CT, Computed tomography; *MRI,* magnetic resonance imaging; *STI,* sexually transmitted infection.

Association (BHIVA), European AIDS Society, and the Department of Health and Human Services in the USA. Up-to-date treatment guidance should always be used.

Starting anti-retroviral therapy

ART should be offered to everyone diagnosed and living with HIV irrespective of CD4 count. It should be started as soon as the patient is ready, which may be on the day of diagnosis. There is clear evidence of the benefits of ART on morbidity and mortality at any CD4 count, together with the impact of treatment on prevention, reducing onward transmission to zero when there is viral suppression.

ART should be offered and ideally started immediately in all people presenting with symptomatic acute HIV, especially if they have neurological symptoms. This stage of disease may represent a unique opportunity for therapy because there is less viral diversity, and the host immune capacity is still intact. There is evidence to show that the viral load can be reduced substantially by effective therapy at this point.

In people with an AIDS-defining condition or an opportunistic illness, ART should be initiated as soon as possible and within 2 weeks of starting treatment for intercurrent conditions except when there is the potential for or diagnosed TB meningitis or cryptococcal meningitis.

The (U=U campaign is based on evidence that people who are virally suppressed cannot transmit HIV. This should be clearly communicated with every person living with HIV because it has been shown to improve mental health and adherence to treatment. Language is important, and clinician should say ‘zero risk’, not ‘negligible risk’.

Involving people living HIV fully in therapeutic decision-making is essential for success. Treatment-support resources, including local, independent, and peer-support resources, are available and should be provided. Although HIV specialists manage and prescribe ART, it is advisable for people living with HIV to share their status with their general practitioners and other healthcare professionals. When an individual prefers not to do this, the benefits and potential harm (especially drug–drug interactions) should be clearly explained and the decision revisited regularly.

Although the benefits of ART in HIV infection are indisputable, treatment is life-long and requires a commitment to take daily oral treatment or to attend on-time for injectable treatment. Contemporary ART regimens are tolerable and safe. Multimorbidity and polypharmacy occur earlier than in the general population.

Not all people living with HIV choose to take ART. In this case, capacity and mental health should be assessed and support offered when needed. Sexual partners of people choosing not to take ART should be advised of prevention interventions, including PrEP (see later).

Which anti-retroviral therapy to start

ART uses a combination of drugs. Recommended treatments contain either three or two different drugs. In many instances, these drugs are co-formulated.

Three drug regimens include two nucleoside/nucleotide reverse transcriptase inhibitors (NRTIs) in combination, with a drug from another class, usually an integrase inhibitor. In specific circumstances, a boosted protease inhibitor (bPI) or a non-nucleoside reverse transcriptase inhibitor (NNRTI) is used. Recommendations for first-line ART are summarized in **Box 44.10**.

Nucleoside/nucleotide reverse transcriptase inhibitors

NRTIs are nucleoside analogues that inhibit the synthesis of DNA by reverse transcription and act as DNA chain terminators. They require three phosphorylation steps for activity. They were the earliest drug class to be used against HIV (e.g., zidovudine [AZT]). Nucleotide reverse transcriptase inhibitors (NtRTIs) have a similar mechanism of action but require only two intracellular phosphorylation steps for activity.

Historically, two N(t)RTIs have been combined to provide the ‘backbone’ of an ART regimen together with a third drug. Fixed-dose NRTI-containing drug combinations are available, reducing pill burden. More recently, oral, and injectable NRTI-sparing two-drug therapies have now been developed for people who are virally suppressed. The choice of which NRTIs should form the backbone of therapy is influenced by efficacy, toxicity, and cost.

Common NRTIs include:

Box 44.9 Prescribing anti-retroviral drugs: practice points

Characteristics of anti-retroviral drugs	Practice points
Long-term adherence is key to success	Make treatment decisions in partnership with the patient. Check for factors that may compromise accurate adherence. Ensure that the proposed drug regimen fits with the patient's lifestyle. Clarify that the patient understands requirements and reasons for strict adherence. Check access to appropriate storage conditions for some agents.
Should not be stopped suddenly	Make sure that mechanisms are in place to ensure adequate drug supplies (e.g., regular clinic appointments, repeat prescriptions, home delivery of medications). Be aware of unexpected time away from home (e.g., holidays, intercurrent hospital admissions, immigration detention, police detention). If there is an urgent medical indication to stop, obtain advice from a specialist physician or pharmacist.
Can be compromised by the introduction of other medications, including other ARVs and vice versa	Always check potential interactions before adding new agents (see www.hiv-druginteractions.org). Use care with DAA therapies in hepatitis C (expert advice required). Use care with enzyme inducers (e.g., rifampicin, rifabutin, warfarin, and nevirapine), which reduce effective concentrations. Remember that methadone levels may be reduced by efavirenz. Some ART may block metabolism of other agents causing toxicity (e.g., steroids, statins. Therapeutic drug monitoring may be necessary
Can interact adversely with some herbal, complementary, and recreational agents	Check potential interactions before adding new agents; see www.hiv-druginteractions.org. Herbal remedies that induce cytochrome P450 (e.g., St. John's wort and Chinese herbal remedies may reduce ART concentrations. Therapeutic drug monitoring may be necessary.
May produce additive toxicities when given with other medications	For example: • Corticosteroid (inhaled and systemic) concentrations may be elevated • Muscle toxicity with statins • Hepatotoxicity with anti-tuberculosis medication • Increased myelosuppression with chemotherapy or high-dose co-trimoxazole
Are associated with a range of adverse drug reactions, which may be confused with other pathology	For example, rash, fever, nausea, and diarrhoea may all be caused by intercurrent pathology or anti-retroviral drugs.
May exacerbate co-morbidities	Examples include hepatic dysfunction caused by hepatitis B and C, increased cardiovascular risk, and osteoporosis.

ART, Anti-retroviral therapy; *ARV,* anti-retroviral; *DAA,* direct-acting antiviral.

- *Tenofovir disoproxil fumarate (TDF)* has been the mainstay of most ART regimens for many years. It is metabolized in the kidneys. Side-effects include renal impairment, osteopenia, and osteoporosis, especially in older adults. Fanconi's syndrome is very rare. Monitoring includes kidney function (serum creatinine, estimated glomerular filtration, urine protein:creatinine ratio), and serum phosphate (for hypophosphatemia). It has a low liability for drug interactions, with the exception of drugs that reduce creatinine clearance.
- *Tenofovir alafenamide (TAF)* is a prodrug of tenofovir. Its use results in lower plasma concentrations of tenofovir than TDF, which reduces the potential for kidney and bone toxicity. Many different fixed-dose formulations incorporate TAF with other anti-retroviral classes. TAF comes in two doses (10 and 25 mg) because it can interact with drugs metabolized by the CYP3A4 pathway. When given with an enzyme inhibitor (e.g., ritonavir), the dose must be lowered. It should not be co-administered with enzyme-inducing drugs such as rifampicin.
- *Emtricitabine (FTC)* and *lamivudine (3TC)* can be considered interchangeably. They are very well-tolerated but have uncommonly been known to cause nail and skin darkening in those with dark skin.
- *Abacavir* is associated with hypersensitivity in at-risk individuals (human leucocyte antigen [HLA]-B*5701)

Co-formulated NRTI-containing combinations include:

- Truvada: TDF/FTC or TDF/lamivudine
- Descovy: TAF/FTC
- Kivexa: abacavir plus lamivudine

Older NRTI and NtRTIs have been associated with mitochondrial toxicity because of their effect on human mitochondrial DNA polymerase. This manifests as myopathy, neuropathy, and hepatic steatosis with lactic acidosis.

Integrase strand transfer inhibitors

These drugs act as selective inhibitors of HIV integrase, blocking viral replication by preventing insertion of HIV DNA into the human DNA genome. They are gold-standard first-line therapy and are effective in viral resistance. They have superior tolerability and safety profiles to those of other classes (e.g., NNRTIs and bPIs; see later). Integrase strand transfer inhibitors (INSTIs) such as bictegravir and dolutegravir are therefore agents of choice. Older integrase inhibitors include raltegravir and elvitegravir.

- *Bictegravir* is a second-generation INSTI with a good side-effect profile and has a higher genetic barrier to resistance than raltegravir. A fixed-dose tablet of bicgravir co-formulated with TAF and FTC is available (Biktarvy).
- *Dolutegravir* is a second-generation INSTI with a good side-effect profile and a higher genetic barrier to resistance than raltegravir. Dolutegravir is associated with a small increase in serum creatinine because it blocks tubular secretion. It is an important

Box 44.10 Initial anti-retroviral therapy regimens: choice of initial therapy and preferred regimens

Recommended as initial treatment for most people living with HIV (grade 1A)	
Regimen	**Specific details**
Bictegravir/emtricitabine/tenofovir AF	
Dolutegravir plus emtricitabine/tenofovir AF or emtricitabine/tenofovir DX	Bone and renal caveats for tenofovir DX
Dolutegravir/lamivudine	No baseline lamivudine resistance Baseline viral load <500 000 copies/mL and CD4 count >200 cells/mm^3
	No active hepatitis B infection and if at risk of hepatitis B, hepatitis B virus immune
Dolutegravir/lamivudine/abacavir	HLA-B*5701 negative and estimated 10-year risk of CVD <10%
Recommended as initial treatment in certain clinical situations (grade 1A)	
Regimen	**Specific details**
Darunavir plus cobicistat or ritonavir plus emtricitabine plus tenofovir AF or tenofovir DX	Bone and renal caveats for tenofovir DX
Doravirine plus emtricitabine or lamivudine plus tenofovir AF or tenofovir DX	Bone and renal caveats for tenofovir DX
Efavirenz plus emtricitabine or lamivudine plus abacavir or tenofovir AF or tenofovir DX	May be a first-line choice in pregnancy and for people on TB treatment but not recommended outside these scenarios
Raltegravir plus emtricitabine plus tenofovir AF or tenofovir DX	Baseline viral load <100 000 copies/mL Bone and renal caveats for tenofovir DX

CVD, Cardiovascular disease; *HLA,* human leucocyte antigen; *TB,* tuberculosis; *Tenofovir AF,* tenofovir alafenamide; *tenofovir DX,* tenofovir disoproxil.
Adapted from British HIV Association. BHIVA guidelines on antiretroviral treatment for adults living with HIV-1 2022 (2023 interim update). https://www.bhiva.org/HIV-1-treatment-guidelines.

component of regimens in drug resistance when it is dosed twice daily. Fixed-dose tablets co-formulated with lamivudine (Dovato) and with abacavir and lamivudine (Triumeq) are available.

- *Cabotegravir* is a second-generation INSTI with a good side-effect profile and a higher genetic barrier to resistance than raltegravir. It forms part of a complete injectable 2-monthly regimen together with rilpivirine (NNRTI; see later). There is also an oral formulation available, which can be used during the month of oral-lead in before injection together with oral rilpivirine.
- *Elvitegravir* is metabolized via the cytochrome P450 (CYP450) pathway, requiring co-administration of a CYP450 blocker to secure adequate plasma concentrations, thus increasing drug interaction potential. Single-tablet co-formulations exist with tenofovir, FTC, and cobicistat (Stribild) and with TAF, FTC, and cobicistat (Genvoya).
- *Raltegravir* was the first licensed compound in this drug class. It can be used as first-line therapy and as part of a regimen to treat people with resistance to other classes. It has a favourable side-effect profile and few drug interactions with the exception of cation-containing compounds and antacids. Newer generation drugs have higher genetic barriers to resistance.

Co-formulated NRTI-containing combinations also containing an INSTI include the following:

- Biktarvy (TAF/FTC plus bictegravir)
- Dovato: lamivudine plus dolutegravir
- Triumeq (abacavir/lamivudine and dolutegravir)

Non-nucleoside analogues

NNRTIs bind to reverse transcriptase to inhibit its function. Older NNRTIs are metabolized through the CYP450 pathway, including CYP3A4 and CYP2D6, and consequently have a high liability for drug interactions. Older NNRTIs are associated with rashes (including Stevens–Johnson syndrome) and elevation of liver enzymes.

- *Doravirine* is a second-generation NNRTI with a better tolerability profile, fewer drug interactions, and retention of some activity against viruses resistant to other NNRTIs. As a result, it is a recommended option for first-line treatment in the European AIDS Clinical Society guidelines.
- *Rilpivirine* is a second-generation NNRTI with a low generic barrier. It needs to be taken with a meal and cannot be taken with antacids or proton pump inhibitors. Its predominant clinical use is as part of the 2-monthly regimen together with cabotegravir, though fixed-dose single-tablet oral combinations containing rilpivirine together with FTC and TAF (Odefsey) or TDF (Eviplera) are available.
- *Etravirine* is a second-generation NNRTI with some activity against drug-resistant NNRTI strains. It is sometimes useful in the treatment of ART-experienced patients. Rash is the most common adverse effect.

Other older NNRTIs such as efavirenz and nevirapine are not recommended in the UK, but people who attend from abroad may be taking them.

- *Efavirenz* has CNS side-effects, including dysphoria and insomnia. It is associated with an increased risk of suicide.
- *Nevirapine* carries a small risk of CD4-related serious hepatic or cutaneous toxicity. It is not used in females with a CD4 count above 250 cells/mm^3 or males with a CD4 count of more than 400 cells/mm^3.

Protease inhibitors

Protease inhibitors (PIs) inhibit the production of functional viral proteins and enzymes and impede viral maturation, leading to production of dysfunctional viral particles. PIs have high potency and high genetic barriers and can be co-prescribed with reverse transcriptase inhibitors. However, these drugs must be prescribed with 'boosters' such as ritonavir or cobicistat to increase drug exposure because of significant metabolism via CYP450 and CYP3A4. The main disadvantages of PIs are the increased risks of drug interactions and lipid abnormalities.

- Boosted ***atazanavir*** and ***darunavir*** are generally dosed once a day. They are most often used in people with resistance; darunavir may be used twice a day to overcome phenotypic resistance. Both can cause diarrhoea and lipid abnormalities. Atazanavir

increases unconjugated bilirubin levels and may produce icterus.
- ***Symtuza*** is a once-daily, fixed-dose single tablet containing darunavir, cobicistat, FTC, and TAF and is the only PI-based single-tablet regimen.

Entry inhibitors, co-receptor blockers, and capsid inhibitors

Drugs from these mechanistic groups are used when there is a need to construct an unusual regimen because of virological resistance or intolerance.
- *Maraviroc* is a chemokine receptor antagonist that blocks cellular CCR5 receptor entry by CCR5 tropic strains of HIV. Tropism assays must be done to establish that the patient is carrying a CCR5 tropic virus. It is dosed twice daily, and the dose must be adjusted based on other co-prescribed ART because it is a substrate of the CYP450 CYP3A4 pathway.
- *Enfuvirtide* is a fusion inhibitor. It is an injectable peptide derived from HIV gp41 that inhibits gp41-mediated fusion of HIV with the target cell. It has a low genetic barrier. It must be injected twice daily, which limits utility. Side-effects relate to the subcutaneous route of administration in the form of injection site reactions.
- *Fostemsavir* is an entry inhibitor that binds to the viral envelope protein gp120 adjacent to the gp120-CD4+ binding site. It is dosed orally twice daily and is a very useful adjunct in patients who are highly treatment experienced with triple class resistance. Its use has significantly improved likelihood of success in this situation and has a low adverse event rate.
- *Lenacapavir* is a first-in-class capsid inhibitor with a triphasic mechanism of action. It can be dosed orally daily or weekly or can be injected 6 monthly. It is licensed for the treatment of people with three-class resistance in whom a regimen cannot otherwise be constructed. It has few side-effects and has produced excellent outcomes for people with no other treatment options. The genetic barrier appears to be low.
- *Ibalizumab* is a post-attachment inhibitor that prevent the HIV gp120 protein from changing its shape to engage with co-receptors. It is a monoclonal antibody that requires i.v. administration every 14 days. It has been shown to be effective in people with multidrug-resistant HIV, for which it is licensed.

Monitoring anti-retroviral therapy

Success rates for initial therapy using modern ART are very high. By 4 weeks of therapy, the viral load should have dropped by at least 1 log 10 copies/mL and by 12 to 24 weeks should be fully suppressed (<50 copies/mL). Slower immunological recovery is seen when treatment is initiated in advanced infection (low CD4 count and late diagnosis) and in older adults.

People are reviewed 2 to 4 weeks after starting therapy to check adherence and for adverse effects. The viral load is measured at 1, 3, and 6 months after starting ART, and a suboptimal response at any time point demands a full assessment and possible change in therapy.

When clinically stable and virologically suppressed on therapy, the viral load should be routinely measured every 6 months. When the viral load is below 50 copies/mL and the CD4 count has been above 350 cells/mm^3 for at least 12 months, monitoring frequency falls to annually or even longer (**Box 44.11**).

Regular clinical assessment (6–12-monthly) should include adherence, tolerability, BMI, waist diameter, blood pressure, and urinalysis. Patients should be monitored for drug toxicity and co-morbidities: liver blood tests, kidney function, bone profile, and a metabolic assessment that includes fasting lipids and haemoglobin A_{1C}. Males who have sex with males (MSM) should be tested for syphilis and offered sexual health screening three-monthly or if symptomatic. Cervical screening should be annual.

Box 44.11 Monitoring patients on anti-retroviral therapy

History
- Adherence to medication
- Evidence of ART toxicity or intolerance
- Documentation of any new clinical conditions since last assessment
- Drug history, including all co-medications: prescribed, recreational, over the counter, complementary, and herbal

Physical examination
- Blood pressure
- Weight
- HIV viral load
- Lymphocyte subsets
- Full blood count
- Urinalysis if tenofovir is being used
- Liver blood tests
- Kidney function
- Fasting lipid profile
- Fasting blood glucose or haemoglobin A_{1c}

Additional investigations (depending on the clinical picture)
- HIV genotype if there is evidence of virological failure
- Therapeutic drug monitoring in certain situations
- DEXA scan and calculation of fracture risk (e.g., FRAX score)
- Cardiovascular risk assessment

ART, Anti-retroviral therapy; *DEXA,* dual-energy X-ray absorptiometry; *FRAX,* Fracture Risk Assessment Tool.

Drug resistance

Resistance to ART (**Box 44.12**) results from mutations in the viral genome. HIV has a rapid turnover, with 10^8 replications occurring per day. The error rate is high, resulting in genetic diversity within the population of virus in an individual, which includes drug-resistant mutants when drug pressure is insufficient. When drugs only partially inhibit virus replication, there is selection pressure for the emergence of drug-resistant strains. The rate at which resistance develops depends on the genetic barrier of the drug (how many mutations are needed before resistance occurs). Drugs such as NNRTIs and some NRTIs have low genetic barriers, and a single mutation will confer high-level resistance. Conversely, PIs and second-generation INSTIs have high genetic barriers. It is possible to transmit drug-resistant virus and for someone to be diagnosed with a highly resistant virus, though this is very uncommon.

HIV anti-retroviral drug resistance testing is routine at diagnosis and before starting therapy and when therapy is failing (viral load > 200 copies/mL on more than one consecutive occasion). The tests involve PCR sequencing of areas of the viral genome that indicate drug susceptibility. Such assays are limited by the starting concentration of virus and their ability to detect minority strains. For results to be useful when therapy is failing, samples must be analyzed when the patient is on therapy because when the selection pressure of therapy is withdrawn, wild-type virus becomes the predominant strain, and resistance mutations may no longer be detectable.

Box 44.12 Mechanisms and implications of HIV drug resistance

1. HIV replicates rapidly and inaccurately. Replication in the presence of ARV drugs leads to a selection pressure for mutations that can survive (i.e., select for drug resistance).
2. Specific point mutations in the viral reverse transcriptase, protease, and integrase genes correlate with reduced drug sensitivity and can be identified by virus genotyping.
3. Inadequate drug concentrations fail to suppress viral replication and viral load and precipitate drug resistance.
4. Inadequate drug concentrations can result from poor adherence, altered GI tract absorption, increased drug breakdown, and drug–drug interactions.
5. Some drugs, especially NNRTIs and integrase inhibitors, have a low genetic barrier (i.e., a small number of mutations can result rapidly in high levels of resistance).
6. Stopping drugs with long half-lives can leave a subtherapeutic drug tail for long enough for resistant strains to develop.
7. In patients on stable therapy, the introduction of new drugs that impact cytochrome P450 pathways can lead to dangerous alterations in ARV drug concentrations.
8. Therapeutic drug monitoring may be a useful investigation for some drugs in some circumstances.
9. Without drug selection pressure, wild-type virus reasserts itself, and resistant variants no longer make up the major circulating viral strains and may not be found on investigation. Genotyping should, if possible, be carried out on specimens obtained while the patient is on therapy.
10. Resistant variants survive and are archived. They reappear if the drug selection pressure is re-introduced. This means that all previous genotypes need to be considered when assessing virological failure and planning new therapy.

ARV, Anti-retroviral; *GI,* gastrointestinal; *NNRTI,* non-nucleoside reverse transcriptase inhibitor.

Databases containing nearly all published HIV and protease sequences, with their associated resistance patterns, are maintained in real time by Stanford University (see Further reading). Phenotypic assays provide a more direct measure of susceptibility, but the complexity of these assays is a barrier to clinical use.

Drug interactions with anti-retroviral therapy

Many ART drugs are victims, perpetrators, or substrates of drug–drug interactions.

Although chelators and chemotherapy may affect first-line INSTI concentrations, there is a lower risk of drug interactions because they largely are metabolized via glucuronidation.

Boosted PIs are potent inhibitors of CYP450 and CYP3A4 pathways, and some NNRTIs are potent inducers. When the drugs are victims, this can lead to subtherapeutic ART concentrations and resistance to treatment. Conversely, when ART leads to enzyme inhibition, supra-therapeutic or subtherapeutic concentrations of concomitant medication may lead to toxicity or under-dosing (e.g., hormone replacement therapy, contraception). This effect applies to naturopathic (e.g., St John's wort) and allopathic medications (e.g., statins, chemotherapy, anticoagulants, rifamycin). An exhaustive list of potential interactions can be found using the online drug interaction tool maintained by Liverpool University (see Further reading).

Therapeutic drug monitoring of peak and trough plasma concentrations may be useful in certain settings.

Adherence

People's beliefs about their own immune systems, trust in 'Western medicine', stigma, faith, and fear of disclosure may affect their perceptions about the personal need for medicines. Adherence to treatment is pivotal to success. With the older drugs, adherence below 95% was associated virological and failure. However, newer ARTs are more forgiving. Irregular adherence facilitates the emergence of drug-resistant variants and virological treatment failure.

Factors implicated in poor adherence may be associated with the medication, the individual, or the provider:

- ***Medication-related factors:*** side-effects, pill burden, inconvenience
- ***Patient factors:*** motivation, psychological wellbeing, family and social support, fear of disclosure, trust in allopathic medicine, health beliefs
- ***Provider factors:*** health literacy around HIV and ART, education of clinicians in adherence support techniques

Supporting adherence is a key part of clinical care, and peer mentorship is very useful.

Acceptability, tolerability, and an assessment of adherence over the previous 2 weeks should be documented at each visit. Adherence over previous 2 weeks is a good proxy for general adherence. Questions about adherence should be phrased in a neutral non-judgmental way to facilitate discussions about difficulties. Provision of multidisciplinary support for adherence is an expected standard of care. Digital aids such as apps, alarms, and dosette boxes can be helpful.

Anti-retroviral therapy failure

Persistent viral replication, CD4 decline, and disease progression indicate ART failure. It is caused by a variety of factors, including poor adherence, limited drug potency, and drug or food interactions compromising drug absorption or effectiveness. Penetration into compartments such as the CNS can occasionally be an issue and necessitate a change in ART regimen.

Switching anti-retroviral therapy

A rise in viral load, toxicity, tolerability concerns, co-morbidities, coinfections, and emergence of better therapies are all reasons to review current treatment. If the patient has a viral load below the limit of detection and a change needs to be made because of intolerance to a particular drug, then within-class switches should be considered before cross-class switches. Simplification to recommended two-drug therapies can be considered.

Virological failure is defined as two consecutive viral loads of more than 200 copies/mL in a previously fully suppressed patient on ART. It requires testing for drug resistance while the person continues to take ART. If resistance occurs, a prompt change in treatment is warranted because maintaining a regimen that is failing leads to resistance. At least two drugs to which there is no resistance should be prescribed, if possible from new classes. The use of bPIs used twice daily in a regimen for multidrug resistance is common.

Stopping anti-retroviral therapy

Stopping ART is not recommended unless there is severe toxicity. Switching therapy is preferable to stopping ART. It is important to note that some drugs may have long half-lives. Injectable cabotegravir and rilpivirine can persist for many months and up to 1 year. Treatment interruptions are not helpful to cardiovascular or virological outcomes.

Complications and long-term safety of anti-retroviral therapy

As efficacy and tolerability of anti-retroviral drugs (ARVs) have improved and longevity for people living with HIV has increased to approximate that of the general population, long-term safety has

become increasingly important. Although the side-effect profiles of modern ART have significantly improved, unwanted effects still sometimes occur. Whereas some are acute and associated with initiation of medication, others emerge after longer-term exposure.

Allergic reactions

Allergic reactions occur with greater frequency in people living with HIV infection, especially in the context of advanced disease in which there is immune dysregulation. Abacavir is associated with a potentially fatal hypersensitivity reaction, strongly linked to the presence of HLA-B*5701, usually within the first 6 weeks of treatment. Rechallenge with abacavir can be fatal and is contraindicated. In the UK, routine screening for the HLA-B*5701 allele has reduced the incidence of abacavir hypersensitivity. Allergies to first-generation NNRTIs such as nevirapine and efavirenz (often in the second or third week of treatment) usually present with a widespread maculopapular pruritic rash, often with a fever and disordered liver blood tests. Reactions can resolve, even with continuing therapy, but drugs should be stopped immediately in any patient with mucous membrane involvement or severe hepatic dysfunction.

Lipodystrophy and metabolic syndrome

A syndrome of lipodystrophy occurred in people who received older ART drugs, including NRTIs such as D4T (stavudine), DDI (didanosine), and AZT (zidovudine), together with first-generation PIs in the 1980s and early 1990s. It was characterized by:

- A loss of subcutaneous fat in the arms, legs, and face (lipoatrophy)
- Deposition of visceral, breast, and local fat
- Raised total cholesterol, high-density lipoprotein cholesterol, and triglycerides
- Insulin resistance with hyperglycaemia
- An increased risk of cardiovascular events.

Management includes dietary advice, an increase in exercise, and statins and fibrates reduce circulating lipids. Simvastatin is contraindicated because of drug interactions with PIs. Injections with facial fillers can help with facial fat loss and fat loss that is painful (e.g., feet, buttocks).

This syndrome does not occur with contemporary ART.

Mitochondrial toxicity and lactic acidosis

Mitochondrial toxicity from older drugs in the nucleoside analogue class (mainly stavudine and didanosine) leads to increased lactate and lactic acidosis, with fatalities reported. NRTIs inhibit gamma-DNA polymerase and other enzymes that are necessary for normal mitochondrial function. Symptoms are often vague and insidious and include anorexia, nausea, abdominal pain, and general malaise. Venous lactate concentration is raised, and the anion gap is typically widened. This is a serious condition, requiring immediate cessation of ART and provision of appropriate supportive measures until normal biochemistry is restored. All patients should be alerted to the possible symptoms and encouraged to attend hospital promptly.

Bone metabolism

Although untreated HIV is believed to have a direct impact on bone metabolism, a fall in bone mineral density is associated with starting ART and is most noticeable in tenofovir-containing regimens. Prevalence varies between study populations, geographical locations, and with dietary intake. Hypothesized mechanisms include direct effects of ART on bone or an indirect renal-mediated effect. Elevated urinary phosphate or low serum phosphate is indicative. Fragility fractures can occur. TAF is preferred when a decrease in bone mineral density has occurred because this appears to have less effect on bone health. Fracture risk (e.g., Fracture Risk Assessment Tool [FRAX] score) should be assessed at baseline in patients older than 50 years of age and postmenopausal females or when other risk factors for bone loss exist. Dual-energy X-ray absorptiometry scans can be used where available.

Immune reconstitution inflammatory syndrome

Immune reconstitution inflammatory syndrome (IRIS) is a paradoxical inflammatory reaction that occurs on initiating ART, usually in people who are profoundly immunosuppressed when beginning therapy. As their immune systems recover, they mount an inflammatory response to a range of pathogens with exacerbation of symptoms and new or worsening clinical signs. Examples include unusual mass lesions or lymphadenopathy associated with mycobacteria, deteriorating radiological appearances associated with TB infection, inflammatory retinal lesions in association with CMV, deterioration in liver function in chronic hepatitis B carriers, and vigorous vesicular eruptions with herpes zoster. Most concerning are IRIS lesions that can obstruct the airway and CNS disease, especially TB and cryptococcal meningitis. To avoid this situation, certain pathogens, in particular *M. tuberculosis* and *Cryptococcus* spp., should be treated for several weeks to reduce the microbiological burden before ART is initiated. Corticosteroids may be needed. In the recent multi-country mpox outbreak, cases of IRIS in advanced HIV disease were reported.

Pregnancy

In the UK, the overall vertical HIV transmission rates are 1% for females diagnosed before delivery and 0.1% for females on ART with a viral load of less than 50 copies/mL. Risk of vertical transmission increases with higher maternal viral load. Management of HIV in pregnancy requires close collaboration between obstetrical, medical, and paediatric teams. The aim is to deliver a healthy baby without HIV to a healthy parent, without prejudicing future treatment options. Effective ART in pregnancy is important to both maternal and fetal outcomes. Drug concentrations may change because of changes in volume of distribution so ART guidelines in pregnancy may differ.

Females who are taking an effective ART regimen at conception should continue this treatment. Pregnant females not already on ART, including elite controllers, should start an effective three-drug ART during pregnancy and be advised to continue life-long treatment with appropriate monitoring and support. For the small numbers of females with HIV who do not wish to take ART treatment for their own health, ART should be offered to reduce transmission to the fetus from the beginning of the second trimester if the viral load is greater than 30 000 copies/mL. Consideration should be given to starting ART earlier in pregnancy if the viral load is greater than 100 000 copies/mL. Stopping ART after delivery is not recommended. Mode of delivery is reviewed with viral load results at 36 weeks' gestation. Pre-ART and delivery by caesarean section to reduce vertical transmission risk have been recommended. This is no longer the case, and for females on effective ART with a viral load below 50 copies/mL at 36 weeks' gestation, vaginal delivery can be supported in the absence of obstetric contraindications. If the viral load is 400 copies/mL or greater at 36 weeks' gestation, caesarean delivery is recommended. The baby should receive post-exposure prophylaxis with zidovudine for 2 weeks postpartum.

Guidelines still advise that replacement infant feeding in people living with HIV should be offered to reduce transmission to the fetus; however, the risk of transmission through breastmilk for females with an undetectable viral load is very low. BHIVA advises supporting people who wish to breast-feed or chest-feed with information and frequent clinical and viral load monitoring for both the mother and baby.

Post-exposure prophylaxis

It is possible to prevent HIV acquisition after exposure and before infection is established. This includes the occupational exposure of healthcare workers and sexual exposure. Animal models provide biological plausibility for use of ART as post-exposure prophylaxis (PEP), but there have been no prospective trials to inform the best approach. Each exposure is evaluated on a case-by-case basis, weighing risk of infection with treatment benefit. The risk of acquisition of HIV after exposure is dependent on the viraemia in the person with HIV. When the source person has an undetectable viral load (<200 copies/mL), then PEP is not indicated because HIV cannot be transmitted.

PEP is most beneficial within the first 24 hours but may be useful up to 72 hours after possible exposure. In the UK, the recommended first-line PEP regimen is tenofovir disoproxil 245 mg/FTC 200 mg with raltegravir 1200 mg once daily for a minimum of 28 days with monitoring for side-effects. A blood test for HIV test should be done at the time that PEP is dispensed. Rapid POCT can also be useful in this setting. PEP after sexual exposure should be a trigger for discussion of PrEP (see later).

Toward a cure

There is increasing interest in the possibilities of a cure for HIV. This could be either:

- Functional: long-term control of the virus without the use of drugs
- Sterilizing: elimination of the HIV reservoir

Reports of therapeutic possibilities exist, including a small number of people living with HIV who underwent stem cell transplants for treatment of coexisting malignancies. Cure studies involving kick-and-kill strategies and the use of monoclonal antibodies are being explored. These approaches are in the early clinical phases.

HIV-ASSOCIATED DISEASE

Opportunistic infections occur with immunosuppression evidenced by a CD4 count below 200 cells/mm^3. This occurs in people who are diagnosed late, people who have disengaged with HIV care or are not adherent with ART, people who have chosen not to take ART, and in regions without access to ART.

In the UK, improved HIV outcomes because of early diagnosis and widespread uptake of ART mean that the types of infection seen in the context of HIV have altered with a lower prevalence of 'classic' opportunistic infections, such as *Pneumocystis* pneumonia and CMV (**Box 44.13**), but an increase in community-acquired infections such as *S. pneumoniae* and *Haemophilus influenzae*.

Immune reconstitution with ART may produce unusual responses to opportunistic pathogens and confuse the clinical picture. Thus, prevention and treatment of opportunistic infections remain integral to the management of HIV infection.

Avoidance of infection

Exposure to certain organisms can be avoided. Food hygiene to reduce exposure to *Salmonella* spp., toxoplasmosis, and *Cryptosporidium* spp. Condom use reduce exposure to HSV, hepatitis B and C viruses, and human papillomaviruses (HPVs). Use of CMV-negative blood products reduces CMV infections. Travel-related infection can be minimized with appropriate prophylaxis and vaccinations.

Immunization strategies

Guidance on the appropriate use of vaccines in HIV is shown in **Box 44.14**. Immunization may be less effective in immunosuppressed persons.

Chemoprophylaxis

In the absence of a normal immune response, many opportunistic infections are difficult to eradicate using antimicrobials, and the recurrence rate is high. Primary and secondary chemoprophylaxis can be used to reduce the incidence of some opportunistic infections. Advantages must be balanced against potential toxicity, drug interactions, treatment burden, and cost.

Primary prophylaxis is effective in reducing the risk of *Pneumocystis jirovecii*, TB, cryptococcus, toxoplasmosis, and MAI.

Primary prophylaxis is *not* usually recommended against CMV, herpesviruses, and fungi other than cryptococcus.

With the introduction of ART and immune reconstitution, ongoing chemoprophylaxis can be discontinued when the CD4 counts is above 200 cells/mm^3 and the viral load is undetectable (<200 copies/mL). Long-term secondary prophylaxis may be needed after cryptococcal and herpes zoster infections.

Fungal infections

Pneumocystis jirovecii infection

P. jirovecii is the most common opportunistic infection causing pneumonia (see **Box 44.13**). It can also rarely cause extrapulmonary disease. Disseminated PJP infection occurs when the CD4 count falls below 200 cells/mm^3. The introduction of effective ART and primary prophylaxis in people with CD4 counts of less than 200 cells/mm^3 has significantly reduced the incidence in the UK.

The organism damages the alveolar epithelium, which impedes gas exchange and reduces lung compliance. The onset is often insidious over a period of weeks, with increasing shortness of breath (usually on exertion), non-productive cough, fever, and malaise. Clinical signs include tachypnoea and tachycardia. The patient is usually hypoxic and desaturates on exercise. Fine crackles are heard on auscultation, although in mild cases, there may be no auscultatory abnormality. Although the chest radiograph may be normal in early infection, typical appearances are of bilateral perihilar interstitial infiltrates progressing to confluent alveolar shadows throughout the lungs. High-resolution CT scans of the chest demonstrate a characteristic ground-glass appearance even when there is little to see on the chest radiograph. Definitive diagnosis is obtained through bronchoalveolar lavage and by PCR amplification of the fungal DNA or microscopic identification with silver staining or immunofluorescent techniques.

Treatment should be instituted as early as possible. First-line therapy for severe infection (P_aO_2 <8 kPa or type 2 respiratory failure) is with i.v. co-trimoxazole (120 mg/kg daily in divided doses) for 21 days. Up to 40% of people will have an adverse

Box 44.13 Some causes of pneumonia in immunocompromised patients

Pathogen	Affected patient population	Clinical features
Pneumocystis jirovecii	Impaired cell-mediated immunity: CD4 count <200 cells/mm^3, long-term corticosteroid use, immunosuppressant drugs	*Pneumocystis* pneumonia, hypoxia on exertion Chest radiograph: perihilar ground-glass shadowing, cysts
Non-tuberculous mycobacterial species	Impaired cell-mediated immunity: CD4 count usually <200 cells/mm^3 Structural lung disease: bronchiectasis, cystic fibrosis, severe COPD	Varied clinical presentation (see Chapter 20): Non-specific fevers, cough, malaise Lymphadenopathy or hepatosplenomegaly Chest CT: nodules, cavitation, thickened airways, 'tree-in-bud' small airways
***Nocardia* spp.**	Impaired cell-mediated immunity: CD4 count usually <150 cells/mm^3, chronic corticosteroid use, after solid organ transplantation, after stem cell transplantation Structural lung disease: bronchiectasis, cystic fibrosis, severe COPD	Acute, subacute, or chronic pneumonia Multiple radiographic presentations, including lobar infiltrates, abscesses, cavities, pleural effusion, pulmonary nodules
***Aspergillus* spp.** ***Zygomycetes* spp.** ***Penicillium* spp.**	Prolonged neutropenia: after chemotherapy for haematological malignancy, after stem cell transplant (myeloablative transplants are at particular risk) Impaired cell-mediated immunity: graft-versus-host disease, immunosuppressant therapy Chronic granulomatous disease	Invasive fungal pneumonia: cough ± haemoptysis, pleuritic pain, fever CT findings: cavitating consolidation, 'tree-in-bud', nodules with ground-glass halo, air-crescent sign (caused by lung necrosis in late stages)
***Cryptococcus* spp.**	Impaired cell-mediated immunity: CD4 count usually <200 cells/mm^3, chronic corticosteroid use, after solid organ transplantation, after stem cell transplantation	Non-specific respiratory symptoms: fever, cough, breathlessness Usually associated with neurological involvement, but isolated pulmonary disease does occur Progresses to disseminated disease in immunocompromised patients Chest CT: cavities, nodules, infiltrates
Histoplasma capsulatum	Impaired cell-mediated immunity: CD4 count usually <50 cells/mm^3	Fever, fatigue, weight loss Cough and dyspnoea are the most common respiratory symptoms Chest radiograph can be normal in disseminated disease Chest CT: diffuse reticular or reticulonodular infiltrates, miliary infiltrates, occasionally mediastinal lymphadenopathy
***Coccidioides* spp.** ***Paracoccidioides* spp.** ***Blastomyces* spp.**	Impaired cell-mediated immunity: CD4 count usually <50 cells/mm^3	Often presents as disseminated disease (chest radiograph frequently normal) Focal or diffuse pneumonia CT findings: diffuse reticulonodular infiltrates, consolidation, nodules (multiple or single), cavities, mediastinal lymphadenopathy
Cytomegalovirus (uncommon; usually presents with retinitis or colitis)	Impaired cell-mediated immunity: CD4 count usually <50 cells/mm^3, after transplantation on immunosuppressant therapy	Cough, dyspnoea, and fever Chest CT: reticular or ground-glass opacities, alveolar infiltrates, nodules or nodular opacities, pleural effusions
Respiratory syncytial virus, human metapneumovirus, influenza, parainfluenza	Impaired cell-mediated immunity: CD4 count usually <200 cells/mm^3, chronic corticosteroid use, after solid organ transplantation, after stem cell transplantation	Cough, dyspnoea, fever Chest CT: ground-glass opacification, air-space shadowing, 'tree-in-bud', airway dilation and wall thickening
Toxoplasma gondii	Impaired cell-mediated immunity: CD4 count usually <100 cells/mm^3	Isolated pulmonary or disseminated disease Dry cough, fever, dyspnoea Chest CT: appearance similar to that of *Pneumocystis jirovecii* or fungal pneumonia, pleural effusion

COPD, Chronic obstructive pulmonary disease; *CT,* computed tomography.

drug reaction, including a typical allergic rash. If the patient is sensitive to co-trimoxazole, i.v. pentamidine (4 mg/kg per day), clindamycin and primaquine, atovaquone, or dapsone and trimethoprim can be given for the same duration. In severe disease, systemic corticosteroids reduce mortality rates. Deterioration can be rapid. Continuous positive airways pressure (CPAP) or mechanical ventilation may be required. Pneumothorax is a common complication of invasive ventilation, and CPAP is preferable where feasible. When ventilation occurs, the course is often protracted. ART should be initiated as early as possible because IRIS is not a feature of PJP.

Secondary prophylaxis is required when the CD4 count is below 200 cells/mm^3. The usual regimen is co-trimoxazole 960 mg three times a week or 480 mg/day. Patients sensitive to sulphonamide are given dapsone, pyrimethamine, or nebulized pentamidine. The last protects only the lungs and does not penetrate the upper lobes particularly efficiently; hence, if relapses occur on this regimen, they may be either atypical or extrapulmonary.

Box 44.14 Use of vaccines in adults living with HIV

Vaccine	Type	Regimen[a]
SARS-CoV-2	All are suitable	Initial doses: two (increase to three doses if CD4 count <200 cells/mm^3, HIV-related illness in the preceding 12 months, if not on ART, or on ART with detectable viral load); boosters as available
Hepatitis A	Inactivated	Non-immune, at-risk patients
Hepatitis B	Subunit	All non-immune patients
Human papillomavirus	Virus-like particle	Three doses for GBMSM (or equivalent risk, i.e., females and heterosexual males living with HIV) to age 45 years
Influenza	Inactivated	Annual vaccination for all
Meningococcus	Conjugated or recombinant protein	At-risk groups, age criteria
Pneumococcus	Conjugated and polysaccharide	Conjugated vaccine for all, polysaccharide additionally for at-risk groups
Pertussis	Acellular multicomponent	Pregnancy
Measles, mumps, and rubella[b]	Live attenuated	All non-immune patients
Varicella (chickenpox)[b]	Live attenuated	All non-immune patients
Herpes zoster (shingles)[b]	Live attenuated	At-risk groups (age criteria)

[a]Vaccination regimens are different in people living with human immunodeficiency virus (HIV) because of suboptimal immune responses requiring higher or more frequent dosage.
[b]Live attenuated viruses should not be given to patients with CD4 counts <200 cells/mm^3 and should be used with caution in those with CD4 counts <350 cells/mm^3.
GBMSM, Gay, bisexual, and other males who have sex with males; *SARS-CoV-2,* severe acute respiratory syndrome coronavirus 2.
Adapted from British HIV Association. Guidelines on the use of vaccines in adults living with HIV.

Cryptococcosis

The most common presentation of cryptococcal infection in the context of HIV is meningitis, although pulmonary and disseminated infections can also occur. The organism *Cryptococcus neoformans* is widely distributed, often in bird droppings, and is usually acquired by inhalation. The onset may be insidious with non-specific fever, nausea, and headache. As the infection progresses, raised intracranial pressure occurs with cranial nerve palsies, visual symptoms, confusion, and loss of consciousness. Seizures or other focal neurological presentations may occur. Neck stiffness and photophobia may be absent because these signs depend on the inflammatory response of the host, which, in this setting, is abnormal.

Diagnosis is made on examination of the CSF (after exclusion space-occupying pathology on CT or MRI imaging scan). Manometry should be performed at the time of lumbar puncture is done because raised intracranial pressure can be life-threatening. Indian ink staining shows the organisms directly. CSF cryptococcal antigen is positive at variable titre and remains present after treatment, although concentrations decrease. Culture should be sent.

Factors associated with a poor prognosis include a high organism count in the CSF, a low white blood cell count in the CSF, and impaired consciousness at presentation.

Usual treatment includes:

- Induction: i.v. liposomal amphotericin B (4.0 mg/kg per day) plus oral flucytosine (25 mg/kg per day) for 14 days. i.v. fluconazole (400 mg/day) is useful if kidney function is impaired or if amphotericin side-effects are troublesome. Steroids should be avoided.
- Consolidation: fluconazole (400 mg once daily) for 8 weeks.

ART initiation should be deferred for at least 4 weeks after the start of treatment because of the risk of IRIS. Long-term maintenance therapy with fluconazole should be continued for 12 months. This can be reviewed if the viral load is consistently undetectable and the CD4 count is consistently above 100 cells/mm^3 over a period of at least 3 months.

Candidiasis

Mucosal infection with *Candida* spp. occurs in the mouth and oesophagus and may be severe enough to cause odynophagia and dysphagia. *Candida albicans* is the usual organism, although infections with *Candida krusei* and *Candida glabrata* also occur. Pseudomembranous candidiasis, consisting of creamy plaques in the mouth and pharynx, is easily recognized, but erythematous *Candida* appears as reddened areas on the hard palate or as atypical areas on the tongue. Angular cheilitis can occur in association with either form. Vulvovaginal *Candida* infection may be difficult to treat.

Fluconazole or itraconazole is the agent of choice for treatment. Liquid formulations may be useful if swallowing is painful or impaired. Nystatin pastilles may also be helpful in managing pain. Disseminated *Candida* infection is uncommon in the context of HIV infection, but if present, fluconazole is the preferred drug; amphotericin, voriconazole, posaconazole, or caspofungin can also be used. Effective ART is the most successful strategy for managing candidiasis in persons living with HIV.

Aspergillosis

Infection with *Aspergillus fumigatus* is rare in people living with HIV unless there are coexisting factors such as lung pathology, neutropenia, transplantation, or glucocorticoid use. Spores are air borne and ubiquitous. After inhalation, lung infection proceeds to haematogenous spread to other organs. Sinus infection can occur. Voriconazole is the preferred treatment, with liposomal amphotericin B (3 mg/kg i.v. daily) an alternative. Caspofungin is also effective.

Histoplasmosis, blastomycosis, coccidioidomycosis, and *Talaromycosis* spp.

These fungal infections are geographically restricted but should be considered in patients who have travelled to areas of high risk. The most common manifestation is with pneumonia, which may

be confused with *P. jirovecii* in its presentation (see earlier). Systemic infections are also reported, particularly with *Talaromyces* (previously *Penicillium marneffei*), which can produce papular skin lesions. Treatment is with amphotericin B.

Protozoal infections

Toxoplasmosis

Toxoplasma gondii most commonly causes encephalitis and cerebral abscess in the context of HIV, usually as a result of reactivation of latent asymptomatic infection. The incidence depends on seropositivity to toxoplasmosis in the population. High antibody levels are found in France (90% of the adult population). About 25% of the adult UK population is seropositive to *Toxoplasma* spp.

Clinical features include focal neurology, seizures, fever, headache, and confusion. Focal neurological signs occur in more than 50% of cases. Eye involvement with chorioretinitis may also be present. In most cases of IgM, serology is positive as well as IgG. Typically, contrast-enhanced MRI of the brain shows ring-enhancing lesions and is more sensitive than CT.

Diagnosis is by characteristic radiological findings on MRI. Single-photon emission CT may also help differentiate toxoplasmosis from primary CNS lymphoma. Empirical anti-toxoplasmosis therapy is indicated when toxoplasmosis is suspected, and radiological improvement within 3 weeks is considered diagnostic. The differential diagnosis includes cerebral lymphoma, PML tuberculoma, and cryptococcal infection.

Treatment is with pyrimethamine for at least 6 weeks (loading dose, 200 mg; then 50 mg/day), combined with sulfadiazine (2–4 g/day for 2 days followed by 500–1000 mg four times per day) and folinic acid (leucovorin) (5–25 mg with each dose of pyrimethamine). Clindamycin is used if sulfa drugs cause hypersensitivity. Alternatives include trimethoprim plus sulfamethoxazole, atovaquone, and pyrimethamine plus azithromycin.

ART should be initiated as soon as the patient is clinically stable, approximately 2 to 4 weeks after acute treatment has begun, to minimize the risk of IRIS.

Cryptosporidiosis

Cryptosporidium parvum cysts attach to the epithelium of the small bowel wall, causing secretion of fluid into the gut lumen and failure of fluid absorption. Self-limiting acute diarrhoea may be seen in an immunocompetent individual. In advanced HIV infection, it can cause severe and progressive watery diarrhoea, which may be associated with anorexia, abdominal pain, and nausea and vomiting. It is also associated with sclerosing cholangitis. The cysts are seen on stool specimen microscopy using Kinyoun acid-fast stain and are readily identified in small bowel biopsy specimens. Nitazoxanide and paromomycin may have some effect; however, immune restoration with effective ART is the treatment of choice and usually leads to remission. In the ART era, infection is rare.

Microsporidiosis

Enterocytozoon bieneusi and *Septata intestinalis* are causes of diarrhoea. Spores can be detected in stools using a trichrome or fluorescent stain that attaches to the chitin of the spore surface. Immune restoration with ART is the treatment of choice and can have a dramatic effect.

Leishmaniasis

Leishmaniasis occurs in immunosuppressed individuals who have been in endemic areas, including South America, tropical Africa, and much of the Mediterranean. Symptoms are frequently nonspecific and consist of fever, malaise, diarrhoea, and weight loss. Splenomegaly, anaemia, and thrombocytopenia are significant findings. Leishmania amastigotes live and multiple in tissues and may be seen on bone marrow biopsy or splenic aspirates. Serological tests for *Leishmania* are not reliable, and biopsy may be required. Treatment is with liposomal amphotericin plus ART. Relapse is common without ART, and long-term secondary prophylaxis may be required in this setting.

Viral infections

Hepatitis B

Hepatitis B infection does not appear to influence the natural history of HIV or responses to anti-HIV treatment. However, in HIV co-infected patients:

- Rates of hepatitis B e antigen (HBeAg) clearance and anti-E (HBeAb) seroconversion are significantly reduced.
- HBV viral load is higher.
- The risk of developing chronic infection is increased.
- Disease may reactivate if treatment for hepatitis B is withdrawn in resolved or controlled hepatitis B infection.

Liver disease occurs most commonly in those with high HBV DNA levels indicative of continuing replication. Detection and quantification of HBV DNA are markers of viral activity and predictive of responses to treatment. Baseline HBV resistance testing may guide interventions in those previously exposed to antiviral drugs. Hepatitis B is treatable, and people living with HIV and hepatitis B should always receive ART that contains either TDT/TAF plus FTC/3TC (see earlier). Initiation of ART may result in a flare of HBV because of improved immune function.

Hepatitis C

Hepatitis C co-infection is associated with:

- More rapid progression of HIV infection
- Blunted CD4 responses to ART
- More likely and more rapid hepatitis C progression
- Higher hepatitis C viral load
- Worse drug-related hepatotoxicity.

Assessment of those with both HIV and hepatitis C virus (HCV) requires full clinical and laboratory evaluation, as well as ultrasound and FibroScan to stage the liver disease. For HCV, both viral load and genotype will influence therapeutic decision-making.

Potent direct-acting antivirals (DAAs) have revolutionized HCV care, and the infection can be cured with short courses of oral therapy. All people living with HIV and HCV should commence ART. ART should be started first in most instances, ideally 4 to 6 weeks before starting treatment for HCV. NRTIs and INSTIs are compatible with DAAs for HCV, but drug interactions should be monitored with other ART classes. The Liverpool website provides comprehensive advice on drug interactions with ART (see Further reading). HCV cure outcomes are similar in individuals with co-infection and are dependent on stage of disease, HCV genotype, and co-administered ART, with response rates of more than 90%.

Acute hepatitis C has increased in people living with HIV, particularly MSM. Epidemics have been reported in several countries, including the UK. UK guidelines recommend that patients are offered DAAs for HCV in acute infection if:

- Decrease in HCV RNA is less than $2^{\log 10}$ at week 4 after diagnosis
- HCV RNA is positive at week 12 after diagnosis.

Immune responses to HCV do not protect against re-infection, and high rates of re-infection have been reported after both therapeutic and spontaneous clearance. Patients who initially clear HCV but have evidence of re-emergent virus should be assessed for relapse or re-infection. Relapse is treated as for chronic HCV, and re-infection is managed as for acute hepatitis C.

Hepatitis B and C virus co-infection

Because of the comparable routes of transmission of hepatitis viruses and HIV, co-infection is common, particularly in MSM, people who inject drugs, and those who acquired HIV through blood products. A higher prevalence of hepatitis viruses is found in those living with HIV infection compared with the general population: of 38 million people living with HIV worldwide, about 12% (5 million) also have chronic viral hepatitis. All those with newly diagnosed HIV should be screened for hepatitis A, B, and C. Hepatitis A and B vaccination should be offered to those without immunity.

As treatment advances have improved HIV prognosis, hepatitis virus co-infection has become an increasingly significant cause of morbidity and mortality, making the management of HIV and hepatitis co-infection a key aspect of clinical care. People living with HIV and hepatitis co-infection are more likely to have rapid liver disease progression than people with mono-infection. Liver disease should be staged. Appropriate monitoring for complications of portal hypertension and hepatocellular carcinoma screening should be performed in those with cirrhosis. People living with HIV and liver disease should be jointly managed by HIV clinicians together with hepatologists. For people with end-stage liver disease, care should be based in specialist centres, ideally with links to a transplant unit.

In co-infected patients, the hepatotoxicity associated with specific ART may be potentiated.

Cytomegalovirus (CMV)

CMV has historically been the cause of considerable morbidity, especially in later stages of disease when the CD4 count is consistently below 100 cells/mm^3. The availability of ART has dramatically altered the epidemiology of CMV in those living with HIV: most people start ARVs early and are therefore not at risk. Disease presentations include retinitis, colitis, oesophageal ulceration, encephalitis, and pneumonitis. CMV also causes polyradiculopathy and adrenalitis. Arteritis may be the major pathogenic mechanism.

CMV retinitis occurs when the CD4 count is below 50 cells/mm^3. Routine fundoscopy should be carried out to look for evidence of early infection. Although usually unilateral to begin with, the infection may progress to involve both eyes. Presenting features depend on the area of retina involved. Loss of vision is most common when there is macular involvement. Other symptoms include floaters, field loss, scotomata, orbital pain, and headache. Examination of the fundus (Fig. 44.4) reveals haemorrhages and exudates, which follow the vasculature of the retina: a so-called 'pizza pie' appearance. Retinal detachment and papillitis may occasionally occur. Features are highly characteristic, and the diagnosis is made clinically. If untreated, retinitis spreads within the eye, destroying the retina within its path. All persons with symptoms of visual disturbance should have a thorough examination with the pupils dilated; if no evident pathology is seen, a specialist ophthalmological opinion should be sought.

Treatment for CMV retinitis should be started as soon as possible, with either oral valganciclovir (900 mg twice daily), i.v. ganciclovir (5 mg/kg twice daily), or i.v. foscarnet (90 mg/kg twice daily) given for at least 3 weeks or until retinitis is quiescent. Ganciclovir can be given directly into the vitreous cavity, but regular injections are required. A sustained-release implant of ganciclovir is available that can be surgically inserted into the affected eye. Cidofovir is available for use if other drugs are contraindicated. The major side-effect of ganciclovir is myelosuppression; foscarnet and cidofovir are nephrotoxic. If immunosuppression is not reversed, reactivation is common, leading to blindness. Maintenance therapy may be required until ART is instituted and immune competence restored.

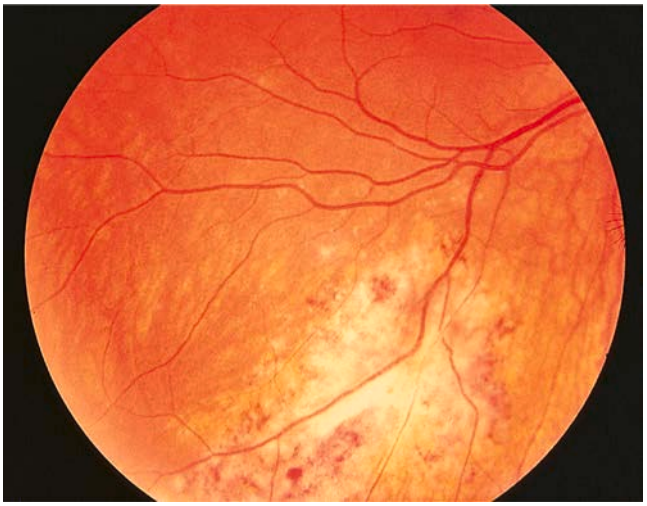


Fig. 44.4 **Untreated cytomegalovirus retinitis.**

CMV colitis presents with abdominal pain (generalized or in the left iliac fossa), diarrhoea that may be bloody, generalized abdominal tenderness, and a low-grade fever. Dilated large bowel may be seen on and abdominal radiograph. Sigmoidoscopy shows a friable or ulcerated mucosa. Histology shows characteristic 'owl's eye' cytoplasmic inclusion bodies (Fig. 20.14). Treatment is with i.v. ganciclovir (5 mg/kg twice daily) for 14 to 28 days. When the patient is stable, optimization of ART improves symptoms, and the histological changes are reversed. Restoration of immune competence with ART removes the need for maintenance therapy.

CMV oesophagitis occurs because non-colonic sites along the GI tract are prone to CMV infection, leading to ulceration of the oesophagus, usually in the lower third, and causing odynophagia.

CMV encephalopathy has clinical similarities to that caused by HIV itself, though with a more aggressive course. ***CMV polyradiculopathy*** can affect the lumbosacral roots, leading to muscle weakness and sphincter disturbance. CSF has an increase in white blood cells, which, surprisingly, are almost all neutrophils. Although progression may be arrested by anti-CMV medication, functional recovery may not occur. The diagnosis is based on MRI and CSF viral PCR. Treatment is with ganciclovir. ART should be started within 2 weeks of CMV treatment.

Herpes infections

Herpes simplex primary infection occurs with greater frequency and severity, presenting in an ulcerative rather than vesicle form in immunosuppressed individuals. Genital, oral, and occasionally disseminated oral or genital infection can occur. Viral shedding may be prolonged in comparison with immunocompetent patients.

Varicella zoster infection can occur at any stage of HIV but tends to be more aggressive and longer lasting in patients who are immunosuppressed. Multidermatomal zoster may occur.

Treatment with valaciclovir is usually effective. Frequent recurrences need suppressive therapy. Aciclovir-resistant strains (usually caused by thymidine kinase-deficient mutants) may respond to foscarnet.

Human herpesvirus 8 (HHV-8) is the causative agent of KS (see later).

EBV is more likely to be latent in people living with HIV. There are increased EBV titres in oropharyngeal secretions and high numbers of EBV-infected B cells. The normal T-cell response to EBV is depressed in HIV. EBV is associated with *primary cerebral lymphoma* and *non-Hodgkin's lymphoma* (see later). *Oral hairy leukoplakia* caused by EBV is a sign of immunosuppression first noted in HIV but now recognized in other conditions. It appears intermittently on the lateral borders of the tongue or the buccal mucosa as a pale, ridged lesion. Although it is usually asymptomatic, patients may find it unsightly and occasionally painful. The virus can be identified histologically and on electron microscopy. ART is helpful. There is a variable response to treatment with aciclovir.

Human papillomavirus infection

HPV produces genital, plantar, and occasionally oral warts, which may be slow to respond to therapy and recur repeatedly. HPV is also associated with cervical, vaginal, and anal intraepithelial neoplasia, which in time may progress to squamous cell carcinomas of the cervix, vulvo-vagina, and rectum. HPV vaccination should be offered to those living with HIV.

Polyomavirus infection

JC virus, a member of the papovavirus family, infects oligodendrocytes, leading to demyelination particularly within the white matter of the brain, causing PML. Clinical features are progressive neurological or intellectual impairment, often including hemiparesis or aphasia. The condition is usually inexorably progressive, but a stuttering course may be seen. Radiologically, the lesions are usually multiple and confined to the white matter. They do not enhance with contrast and do not produce a mass effect. MRI (Fig. 44.5) reveals enhanced signal on T2-weighted images of the lesions. MRI appearance and JC virus detection by PCR in a CSF sample are usually sufficient for diagnosis and avoid the need for brain biopsy for tissue diagnosis. There is no specific treatment. Effective ART leads to clinical and radiological remission.

SARS-CoV-2 (COVID-19) infection

The incidence of SARS-CoV-2 appears to be equivalent that of the general population. Adverse outcomes are more commonly reported in association with CD4 counts below 350 cell/mm^3, people with untreated HIV, and those with a detectable viral load. Patients living with HIV who are hospitalized because of COVID-19 infection are younger and have more co-morbidities than the general population. Early anti-SARS-CoV-2 treatment may be indicated for people living with HIV, especially if untreated or with a poor immune response to ART. Pre-existing ART should be continued during hospital admissions, although the route of administration may need to change if oral treatment is not possible in critical illness. Drug–drug interactions must be considered. Viral blips may be seen, and the CD4 count may fall, necessitating consideration of appropriate prophylaxis for other opportunistic infections.

Vaccination against SARS-CoV-2 is recommended for all those living with HIV. Vaccine choice is dependent on local availability and guidelines. In situations in which vaccine supplies may be limited, those with CD4 counts below 350 cells/mm^3 should be prioritized. Securing continuity of HIV treatment and care in the event of isolation or lockdown measures is crucial. Long-term symptoms after COVID-19 should be addressed as for the general population with appropriate referrals to specialist services.

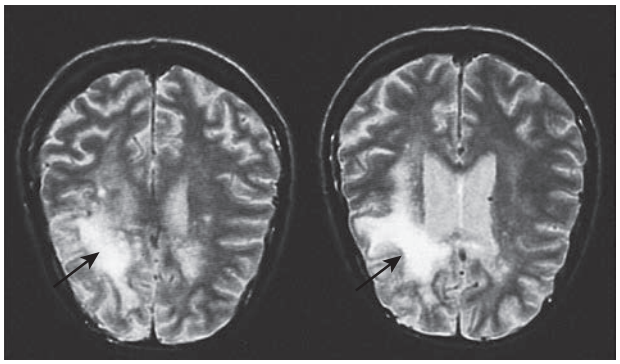


Fig. 44.5 Magnetic resonance imaging scans showing progressive multifocal leucoencephalopathy *(arrows)*.

Mpox infection

Mpox is caused by an *Orthopoxvirus* related to variola. It is transmitted by person-to-person contact and occasionally by fomites. An outbreak in 2022 occurred outside the usual endemic areas of central and west Africa, disproportionately affecting MSM and linked with sexual contact. All those with mpox should have a full STI screen including HIV testing.

Cutaneous, mucosal, and anogenital vesicular or ulcerating lesions are accompanied by fever, lymphadenopathy, and a generalized rash. In people with uncontrolled HIV or a CD4 count below 200 cells/mm^3, the clinical picture may be more severe with multisystem manifestations. PCR detection of mpox DNA from lesions is the diagnosis test of choice.

Treatment is symptomatic for pain and fever alongside skin lesion care, isolation, and contact tracing. In severe cases, hospitalization and antiviral therapy may be indicated. Oral tecovirimat (600 mg twice daily for those weighing 40–120 kg; three time daily for those weighing >120 kg). Brincidofovir and cidofovir may also be effective. Vaccinia immune globulin i.v. has been used for people who are unable to mount an adequate immune response. Mpox IRIS has been seen with the introduction of ART, and care must be taken with the introduction of ART in those who are not yet taking or restarting ART. Vaccination should be offered to all people living with HIV. PEP with mpox vaccination should be given as soon as possible and ideally within 4 days of exposure.

Influenza infection

Influenza virus A is no more frequent in persons living with HIV but has been associated with an increased severity and complication rate when the CD4 counts are less than 200 cells/mm^3. People living with HIV should be vaccinated annually. Prophylaxis with oseltamivir is used in unvaccinated individuals with CD4 counts below 200 cells/mm^3.

Bacterial infections

Bacterial infection in HIV is common. Cell-mediated immune responses normally control infection against intracellular bacteria such as mycobacteria. The abnormalities of B-cell function associated with HIV lead to infections with encapsulated bacteria because reduced production of IgG2 cannot protect against the

polysaccharide coat of such organisms. These functional abnormalities may be present well before there is a significant decline in CD4 numbers so bacterial sepsis may be seen in the early stages of HIV infection. *S. pneumoniae*, *H. influenzae*, and *Moraxella catarrhalis* infections are examples. Bacterial infection is often disseminated and although usually amenable to standard antibiotic therapy, may recur. Long-term prophylaxis may be required for frequently recurring infection.

Mycobacterium tuberculosis

Many parts of the world that have a high prevalence of TB also have a high prevalence of HIV infection, both of which are increasing. TB can cause disease with minimal immunosuppression, so it can appear in both early and advanced HIV. HIV-related TB frequently represents reactivation of latent TB, but there is also clear evidence of newly acquired infection and hospital-acquired spread in populations living with HIV.

The pattern of disease differs with the degree of immunosuppression:

- People with relatively well-preserved CD4 counts have a clinical picture comparable to that in the general population with predominantly pulmonary infection.
- In advanced HIV, pulmonary disease may be atypical, and extrapulmonary disease is more common, including CNS, lymph nodes, bone marrow, and liver infection. TB bacteraemia may be present. CNS disease is the most severe and concerning presentation and can be intracranial or affect the spinal cord.

The diagnosis depends on demonstrating the organisms in appropriate tissue specimens. The response to tuberculin testing is blunted in individuals living with HIV and is unreliable. Sputum microscopy results may be negative even in those with pulmonary infections; culture techniques are the best diagnostic tool.

M. tuberculosis infection usually responds well to standard treatment regimens. Multi- and extensive drug-resistant TB are emerging problems; adherence to anti-tuberculous therapy is crucial. Secondary prophylaxis may be required after treatment.

Testing for latent TB infection with interferon-gamma release assays (IGRAs) should be offered to all people living with HIV in areas of high TB prevalence and to all those with additional risk factors for TB regardless of local prevalence. Testing for TB should take place irrespective of CD4 count. For those who are IGRA positive and in whom active TB has been excluded, latent TB infection should be treated with 6 months of isoniazid plus pyridoxine or 3 months of isoniazid plus rifampicin plus pyridoxine.

Drug–drug interactions between anti-retroviral and anti-tuberculous medications are complex and caused by enzyme induction or inhibition. Rifampicin is a potent inducer of CYP450. PIs are contraindicated with rifampicin. Drugs that can be used with rifampicin include efavirenz and dolutegravir (twice daily). FTC/3TC and TDF can also be used, but TAF-containing drugs should be avoided. Rifabutin has a weaker effect on CYP450 and may be substituted for rifampicin if PIs are required for HIV therapy; rifabutin can be used with reduced doses of PIs given three times per week. Co-prescribing for HIV and TB is complex and should always been done with an HIV specialist.

IRIS can occur because of the improvement in immune function in patients with TB starting ART. Paradoxical inflammatory reactions include exacerbation of symptoms, new or worsening clinical signs, and deteriorating radiological appearances. This usually occurs in the first few weeks after initiation of ART. In people living with HIV and TB:

- Allowing at least 2 weeks of anti-tuberculous therapy before commencing ART reduces the risk of IRIS.
- ART should then be started as soon as is practicable and within 4 weeks of initiation of TB treatment.
- In people with a CD4 counts less than 50 cells/mm^3, ART should be offered within 2 weeks.
- In CNS disease, ART should be delayed until 8 weeks after initiation of TB treatment because worsening of CNS disease could lead to fatality.
- Steroids may be needed to manage IRIS.

Mycobacterium avium-intracellulare infection

Atypical mycobacteria, particularly MAI, generally appear only in the very late stages of HIV infection with profound immunosuppression. MAI is a saprophytic organism of low pathogenicity that is ubiquitous in soil and water. Entry may be via the GI tract or lungs with dissemination via infected macrophages.

Clinical features include fever, malaise, weight loss, anorexia, sweats, and anemia. GI symptoms may be prominent with diarrhoea and malabsorption. Abnormal liver blood test results and a low haemoglobin may be seen. A high index of suspicion is needed. Direct examination and culture of blood, lymph node, bone marrow, or liver indicate the diagnosis most reliably. MAI is typically resistant to standard anti-tuberculous treatment. Rifabutin in combination with clarithromycin or azithromycin can reduce the burden of organisms and ameliorate symptoms in some. Ethambutol, rifabutin, and clarithromycin is a common combination. Addition of amikacin may be needed. Primary prophylaxis with rifabutin or azithromycin may delay the appearance of MAI in people with CD4 counts below 50 cells/mm^3, but no corresponding increase in survival has been shown.

Salmonella spp.

Salmonella infections are much less frequent with the advent of ART. Organisms are usually acquired orally and cause GI disease or disseminated infection. *Salmonella* osteomyelitis and cystitis have been reported. The diagnosis is from blood and stool cultures. Most isolates are sensitive to oral ciprofloxacin 500 mg twice daily for 5 days. Education on food hygiene should be provided.

Other bacterial infections

Folliculitis, abscesses, and cellulitis are common and are usually caused by ***S. aureus***.

Periodontal disease is a chronic disease mediated by immunoinflammatory factors. In the early stages, oral bacteria interact with periodontal tissue, triggering an inflammatory response that leads to destruction of the periodontium, with pain and damage to the gums. It is more common in smokers. Therapy is with local debridement and systemic antibiotics.

Infections caused by other organisms

Strongyloides infection is caused by a nematode found in tropical areas and may cause a hyper-infection syndrome in people living with HIV. Larvae are produced, which invade through the bowel wall and migrate to the lung and occasionally to the brain. Albendazole or ivermectin is used for treatment.

Scabies may be much more severe in HIV infection and cause disseminated atypical, crusted papular lesions known as 'Norwegian scabies', from which mites are readily demonstrated. Superadded staphylococcal infection may occur. Treatment with conventional agents, such as lindane, may fail. Ivermectin has been used to good effect.

HIV-associated cancers

HIV is pro-inflammatory state that increases the risk of cancer. In the UK, the incidence and prevalence of malignancies are higher in people living with HIV, with cancers occurring at younger ages and with higher rates of mortality compared with the general population. The most tightly associated malignancies are KS, non-Hodgkin's lymphoma, and cervical cancer, which are AIDS-defining illnesses. Hepatoma and Hodgkin's lymphoma, although not AIDS defining, are significantly more common. Anal cancer is also more common than in the general population.

Although AIDS-defining malignancies have significantly decreased with the advent of ART, the increased longevity of people living with HIV increases the risks for development of cancers that are associated with older age. Non–AIDS-defining cancers (NADCs) are now among the leading causes of death among people living with HIV. ART may provide some prevention benefits for NADCs.

Kaposi's sarcoma

KS in association with HIV (epidemic KS) behaves more aggressively than in the general population (endemic KS). The incidence has fallen significantly since the introduction of ART. HHV-8 is implicated in the pathogenesis. KS skin lesions are characteristically purple and well circumscribed, occurring in multiple sites with a predilection for the tip of the nose, the genitals, and the feet. KS is a multicentric tumour consisting of spindle cells and vascular endothelial cells, which together form slit-like spaces in which red blood cells become trapped. This process is responsible for the characteristic purple hue of the tumour. In addition to the skin lesions, KS affects the lymph nodes, the lung, and the GI tract, giving rise to a wide range of symptoms and signs. Most people with visceral involvement also have skin or mucous membrane lesions. Visceral KS carries a worse prognosis than disease confined to the skin. KS can affect the eyelids and conjunctiva, causing periorbital oedema (Fig. 44.6).

ART is the treatment for KS. Local radiotherapy gives good results in discrete skin lesions and is helpful in lymph node disease. In aggressive disease, systemic chemotherapy with liposomal anthracyclines is indicated.

Lymphoma

Lymphomas are relatively common in people living with HIV, especially those with low CD4 counts and in whom the viral load has been uncontrolled for a substantial period. Non-Hodgkin, large B-cell type is most common. Lymphomas are frequently extranodal and often affect the brain, lung, and GI tract. Many of these tumours are strongly associated with EBV, with evidence of expression of latent gene antigens such as Epstein–Barr nuclear antigens (EBNA 1–6), some of which are involved in the immortalization of B cells and drive a neoplastic pathway. HIV-associated lymphomas are commonly aggressive. Patients often present with systemic 'B' syndromes and progress rapidly despite chemotherapy. Primary cerebral lymphoma is variably responsive to radiotherapy but overall carries a poor prognosis.

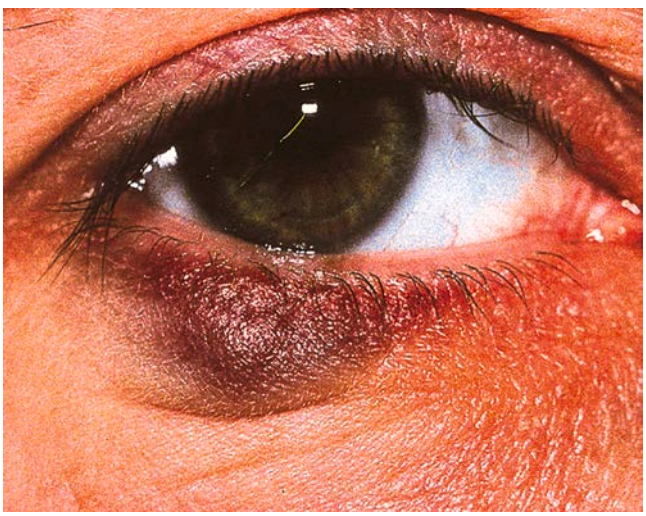


Fig. 44.6 **Kaposi's sarcoma of the eyelid.**

Cervical cancer

For females with HIV, there is an increased risk of cervical cancer caused by oncogenic subtypes of HPV. The risk is significantly reduced by effective ART. Enhanced surveillance and early intervention are indicated with annual cervical cytology to monitor for pre-malignant changes.

Anal cancer

Anal cancer risk is significantly elevated in people living with HIV, particularly among MSM. Females with precancerous cervical changes and cervical cancer are also at increased risk. Anal cancer is strongly linked to infection with HPV, with HPV 16 and 18 being the most common subtypes detected in almost 90% of cases. The incidence of precancerous change is reduced by HPV vaccination. Early detection and treatment of high grade anal intraepithelial neoplasia through high-resolution anoscopy has been shown to modify progression to invasive cancer.

PREVENTION AND CONTROL OF HIV

HIV is a treatable chronic condition with a life expectancy that can match that of the general population in those who start treatment early. Initiating treatment when infection is more advanced compromises clinical outcomes and is the most important predictor of morbidity and mortality. In 2020, people diagnosed late in England were found to be 17 times more likely to die within 1 year of their diagnosis compared with prompt diagnosis. Of the 100 000 people living with HIV in 2022, about 5% were undiagnosed, and 44% of people newly diagnosed were diagnosed late (CD4 count <350 cells/mm^3).

Long-term engagement with effective treatment delivers the best clinical outcomes for people living with HIV and when the viral load is undetectable prevents onward transmission. In 2022, the UK Health Security Agency estimated around 15 000 people in England had transmissible levels of virus, of whom 24% were undiagnosed, 35% were not retained in care, and 24% were on treatment with no reported viral load.

Expanding HIV testing

Expansion of testing within and beyond sexual health services is a key intervention to achieve good clinical outcomes and a reduction in new infections. In 2021, the UK government committed to end new HIV transmissions in the UK by 2030 with expansion of HIV testing as a key intervention

UK guidelines list clinical settings where HIV testing should be universally offered. Guidance also includes a list of clinical situations and diagnoses **(indicator conditions)** that are highly predictive of HIV infection and in which HIV testing is recommended (Box 44.15). Testing is recommended for all new registrants in primary

Box 44.15 HIV testing: where and who to test

Universal HIV opt-out testing is recommended for all patients attending the following settings:

- Sexual health services
- Antenatal services
- Termination of pregnancy services
- Addiction and substance misuse services
- Healthcare services for tuberculosis, hepatitis B, hepatitis C, and lymphoma
- Primary and secondary healthcare, including emergency departments, in areas of high and extremely high HIV seroprevalence
- Individuals commencing chemotherapy or immunosuppressive or immunomodulatory therapy in line with speciality guidelines

HIV testing should be routinely recommended to the following individuals:

- All patients diagnosed with an STI
- Sexual partners of those living with HIV
- Males who have disclosed sexual contact with other males
- Female sexual contacts of males who have sex with males
- People of Black African heritage
- People reporting current or prior history of injecting drug use
- Sex workers
- Transwomen
- People in prison
- Persons from a country of high HIV prevalence (>1%)
- Persons who report sexual contact abroad or in the UK with individuals from countries of high HIV prevalence
- Individuals known to have or have had a mother living with HIV and who do not have documented HIV-negative status
- Patients presenting for healthcare when HIV is a differential diagnosis

HIV-associated indicator conditions

Respiratory

- Tuberculosis, *Pneumocystis jirovecii* pneumonia, bacterial pneumonia, aspergillosis

Neurology

- Cerebral toxoplasmosis, primary cerebral lymphoma, cryptococcal meningitis, progressive multifocal leucoencephalopathy, aseptic meningitis or encephalitis, cerebral abscess, space-occupying lesion of unknown cause, Guillain–Barré syndrome, transverse myelitis, peripheral neuropathy, dementia

Dermatology

- Kaposi's sarcoma, severe or recalcitrant seborrhoeic dermatitis, severe or recalcitrant psoriasis, multidermatomal or recurrent herpes zoster

Gastroenterology

- Persistent cryptosporidiosis, oral candidiasis, oral hairy leukoplakia, chronic diarrhoea of unknown cause, weight loss of unknown cause, *Salmonella* spp. infection, *Shigella* spp. infection, *Campylobacter* spp. infection, hepatitis B infection, hepatitis C infection

Oncology

- Non-Hodgkin's lymphoma, anal cancer, anal intraepithelial dysplasia, lung cancer, seminoma, head and neck cancer, Hodgkin's lymphoma, Castleman's disease

Gynaecology

- Cervical cancer, vaginal intraepithelial neoplasia, cervical intraepithelial neoplasia (grade 2 or above)

Haematology

- Any unexplained blood dyscrasia, including thrombocytopenia, neutropenia, and lymphopenia

Ophthalmology

- Cytomegalovirus retinitis; infective retinal diseases, including herpesviruses and *Toxoplasma* spp.; any unexplained retinopathy

Ear, nose, and throat

- Lymphadenopathy of unknown cause, chronic parotitis, lymphoepithelial parotid cysts

Other

- Mononucleosis-like syndrome (primary HIV infection), pyrexia of unknown origin, any lymphadenopathy of unknown cause, any STI

STI, Sexually transmitted infection.

Adapted from: British HIV Association/British Association for Sexual Health and HIV/British Infection Association adult HIV testing guidelines; 2020 and Palfreeman A, Sullivan A, Rayment M et al. British HIV Association/British Association for Sexual Health and HIV/British Infection Association adult HIV testing guidelines 2020. HIV Med *2020; 21(Suppl 6):1–26.*

care and for patients admitted to acute medical care in areas where HIV seroprevalence is more than 2 per 1000 population.

Expanded opt-out testing for blood-borne viruses (HIV, hepatitis B and C) in emergency departments in areas of high HIV prevalence was implemented in 2022 to reach communities underserved by existing testing opportunities. Early results show a significant number of people newly diagnosed with all three viruses, as well as people already diagnosed but not accessing effective care. Opt-out testing for blood-borne viruses is implemented in all adult prisons in England for new arrivals.

Point-of-care HIV antibody tests that are sensitive, specific, and easy to use; are readily available; can test either blood or oral fluids; and give results within minutes. Home sampling and home testing approaches have increased testing rates in some populations.

All reactive POCTs should be followed up with confirmatory serological assays, and arrangements must be in place to ensure that people who are found to be HIV positive have rapid routes into specialist clinical care and to peer support.

Strategies for reducing HIV transmission

HIV is a preventable infection, yet new infections continue to occur. Combinations of behavioral, biomedical, and structural approaches are required, with specific interventions appropriate for the population at risk.

Models of the epidemic in the UK suggest that most new infections are transmitted from people who are unaware of their HIV status or have recently acquired HIV. Reduction of the high rates of undiagnosed infection through scale-up of testing and early initiation of effective treatment is therefore central to an effective prevention response.

For those at high risk of acquisition, additional emphasis on the need for repeat testing is required because those with early infection and a high viral load are particularly infectious.

Vaccine development has been hampered by the genetic variability of the virus and the complex immune response that is required from the host, with disappointing results from trials of candidate agents.

Consistent, correct condom use and education for behavior change are key strategies.

Provision of clean injecting equipment for people who inject drugs has been successful in the countries where it has been implemented.

Medically performed circumcision in heterosexual males without HIV has been demonstrated in studies in Africa to reduce female-to-male transmission of HIV by at least 50% and reduce male-to female transmission.

Use of effective ART to maintain a viral load below the limits of detection by people living with HIV has been demonstrated to reduce onward transmission risk to zero. These data have been used to support the 'U=U' message campaign, reinforcing the benefits of testing and early treatment.

Partner notification schemes are important in ensuring appropriate testing and interventions are available for those who may be at risk of HIV acquisition. Availability and accessibility of confidential HIV testing provide an opportunity for health education to reduce risk.

Pre-exposure prophylaxis

The use of ART by people who are not living with HIV but who are at risk of acquisition is called PrEP. This consists of oral TDF/FTC taken either daily or on demand before potential HIV exposure. TAF/FTC can also be used. Injectable PREP using cabotegravir every 2 months has been developed, but this not available on the National Health Service (NHS) at the time of writing. It is superior to oral PrEP and preferred by those who entered clinical trials. When used consistently, PrEP has been shown to be highly successful and safe, with a reduction in the risk of HIV acquisition by up to 90%. Based on these results, PrEP, with appropriate safety monitoring, is recommended as an effective HIV prevention strategy. Work is still ongoing to scale up access for all those who could benefit from PrEP within the NHS. PrEP uptake rates are highest among gay and bisexual males and lowest for females and heterosexuals of Black African and Asian ethnicities.

Global strategies

Understanding and changing behaviour is crucial but notoriously difficult; sex, HIV, and AIDS carry many taboos. Stigma, poverty, punitive homophobic and transphobic legislation, social unrest, and war all contribute to the spread of HIV. Ongoing investment in and prioritization of HIV requires political will (not always readily available) if progress is to be sustained.

Further reading

AIDS info. *Global Data on HIV Epidemiology and Response*. UNAIDS https://aidsinfo.unaids.org/.
Bekker LG, Beyrer C, Mgodi N et al. HIV infection. *Nat Rev Dis Primers*. 2023; 9(1):42.
BHIVA/BASHH/BIA Adult HIV Testing Guidelines 2020. https://www.bhiva.org/HIV-testing-guidelines.
British HIV Association guidelines on antiretroviral treatment for adults living with HIV-1 2022 (2023 interim update) https://www.bhiva.org/HIV-1-treatment-guidelines.
Collini P, Mawson RL. A new era of HIV care for age-associated multimorbidity. *Curr Opin Infect Dis*. 2023; 36(1):9–14.
Department of Health and Social Care. *Towards Zero: the HIV Action Plan for England—2022 to 2025*. https://www.gov.uk/government/publications/towards-zero-the-hiv-action-plan-for-england-2022-to-2025/towards-zero-an-action-plan-towards-ending-hiv-transmission-aids-and-hiv-related-deaths-in-england-2022-to-2025.
European AIDS Clinical Society Guidelines version 12.0. 2023. https://www.eacsociety.org/guidelines/eacs-guidelines/.
Gandhi RT, Bedimo R, Hoy JF et al. Antiretroviral drugs for treatment and prevention of HIV infection in adults: 2022 recommendations of the International Antiviral Society–USA panel. *JAMA*. 2023; 329(1):63–84.
Gibas KM, Kelly SG, Arribas JR et al. Two-drug regimens for HIV treatment. *Lancet HIV*. 2022; 9(12):e868–83.
Guaraldi G, Milic J, Cascio M et al. Ageism: the-ism affecting the lives of older people living with HIV. *Lancet HIV*. 2024; 11(1):e52–9.
Henning RJ, Greene JN. The epidemiology, mechanisms, diagnosis and treatment of cardiovascular disease in adult patients with HIV. *Am J Cardiovasc Dis*. 2023; 13(2):101–121.
Jones HT, Levett T, Barber TJ. Frailty in people living with HIV: an update. *Curr Opin Infect Dis*. 2022; 35(1):21–30.
Kirby T. UK emergency department testing for HIV expanded. *Lancet HIV*. 2024.
Lee MJ, Fidler S, Frater J. Immunotherapeutic approaches to HIV cure and remission. *Curr Opin Infect Dis*. 2022; 35(1):31–41.
Mody A, Sohn AH, Iwuji C, Tan RK, Venter F, Geng EH. HIV epidemiology, prevention, treatment, and implementation strategies for public health. *Lancet*. 2024; 403(10425):471–92.
Stanford University HIV Drug Resistance Database. Available at: http://hivdb.stanford.edu.
UK Health Security Agency: *HIV testing, PrEP, new HIV diagnoses and care outcomes for people accessing HIV services: 2023 report*. https://www.gov.uk/government/statistics/hiv-annual-data-tables/hiv-testing-prep-new-hiv-diagnoses-and-care-outcomes-for-people-accessing-hiv-services-2023-report.
UK Health Security Agency. *UK Standards for Microbiology Investigations.* Laboratory diagnosis of HIV infection. Issued 19th December 2023.
UNAIDS: https://www.unaids.org/en.
University of Liverpool. HIV drug interactions. Available at: https://www.hiv-druginteractions.org/checker.
Waters L, Winston A, Reeves I et al. BHIVA guidelines on antiretroviral treatment for adults living with HIV-1 2022. *HIV Med*. 2022; 23 Suppl 5:3–115.
WHO 2020 Global Health Estimates 2020: *Deaths by Cause, Age, Sex, by Country and by Region, 2000-2019*. Geneva, World Health Organization; 2020.

Questions for this chapter can be found on Elsevier eBooks+.

45 Dermatology

Anthony Paul Bewley, Malvina Cunningham, Anshoo Sahota, and Catherine Anne Harwood

Clinical skills 1565

Background 1566

Skin's structure and function 1570

Investigation of skin disease 1572

Principles of dermatological therapy 1572

Topical therapy 1572
Phototherapy 1573
Systemic therapy 1574

Dermatological emergencies 1574

Common rashes 1574

Acne and related disorders 1574
Eczema and dermatitis 1577
Psoriasis 1584
Other common papulosquamous rashes 1586
Urticaria and angio-oedema 1587

Skin infections 1589

Bacterial infections 1589
Mycobacterial infections 1591
Viral infections 1592
Fungal infections 1595
Parasites and infestations 1598
Tropical dermatoses 1600
Human immunodeficiency virus and the skin 1600

Skin signs of systemic disease 1602

Multisystem inflammatory disorders 1602
Pruritus 1604
Endocrine disease 1604
Metabolic disease 1605
Haematological disease 1605
Amyloidosis 1605
Chronic liver disease 1606
Chronic kidney disease 1606
Autoimmune rheumatic disease 1606
Disorders of blood vessels and lymphatics 1607
Systemic malignant disease 1610
Genetic disease 1610

Photosensitivity disorders 1611

Blistering and bullous skin disease 1611

Immunobullous diseases 1612
Mechanobullous diseases (epidermolysis bullosa) 1614

Skin tumours 1614

Benign skin tumours 1614
Dysplastic and premalignant skin lesions 1616
Malignant skin tumours 1618

Disorders of pigmentation 1625

Hypopigmentation 1625
Hyperpigmentation 1626

Nail disorders 1626

Hair disorders 1627

Hair loss (alopecia) 1627
Increased hair growth 1628

Birth marks and neonatal rashes 1628

Port-wine stain (naevus flammeus) 1628

Drug eruptions 1629

Psychodermatology 1630

CORE SKILLS AND KNOWLEDGE

Skin diseases are common, and their burden on a population level is significant. Across Europe, they have an annual incidence of more than 40% with the commonest being fungal infection, acne, and eczema. In the UK, skin cancers are more common than the sum total of every other type of cancer. Skin diseases cause considerable distress, and some can be life-threatening.

Dermatology is primarily an outpatient specialty, although inpatient treatment is sometimes needed for severe inflammatory or infectious skin disease when associated with systemic symptoms. The demand for specialist care exceeds supply, and teledermatology has become an important means of triaging referrals and offering remote access to specialist advice. Dermatologists work in multidisciplinary teams with oncologists, pathologists, radiotherapists, and plastic or maxillofacial surgeons to provide diagnosis and treatment.

Key skills include:

- Examining the skin and describing common skin lesions and rashes accurately
- Understanding the management of common skin conditions, including acne, eczema, psoriasis, skin infections, and malignancies
- Recognizing life-threatening skin conditions, including skin cancers, severe drug reactions, and aggressive skin and soft tissue infections
- Understanding the role of topical therapies, phototherapy, and systemic drugs in the treatment of patients with skin diseases.

Dermatology is best learned by interviewing and examining patients in outpatient clinics and on surgical lists or in phototherapy and daycare units.

CLINICAL SKILLS

Skin conditions generally consist of ***lesions*** (lumps and bumps) and ***rashes***. Diagnosis is by taking a detailed history (Box 45.1) and examining the whole skin surface, as well as the hair, nails, and mucosal sites. Diagnostic accuracy can be increased by palpation (to assess depth of lesions) and by use of a ***dermatoscope*** (a handheld magnifying device with strong and sometimes polarized light). A dermatoscope (Fig. 45.1) can be used to assess the pigment and blood vessel components of skin conditions (Box 45.2). Irregular pigmentation and irregular or tortuous blood vessels can be seen in skin cancers. Benign conditions usually have regular pigment and blood vessels. Rashes are less often assessed by dermoscopy, but some specific features of rashes are described.

Box 45.1 Taking a history in dermatology

History of the presenting complaint

- Time course: duration, changes in distribution or appearance over time
- Local symptoms: itching, pain and tenderness, bleeding
- Involvement of mouth, nails, hair, or genitalia
- Provoking factors: temperature, sunlight, medication, stress
- Features of systemic inflammatory disease: malaise, fever
- Symptoms in other systems: eyes, joints, internal organs

Risk factors

- History of autoinflammatory or autoimmune disease
- Current and recent medications including supplements and recreational substances, previous history of drug allergy
- Family history of skin disease (e.g., psoriasis, skin cancer) and atopy
- Skin type, cumulative sunlight exposure, use of sunbeds
- Time spent in tropical countries or history of travel

Psychological factors[a]

- Depression and anxiety. Stressful events.
- Disturbed body image; effect on relationships and self-confidence

[a]Screening tools for psychological distress, such as the Patient Health Questionnaire-9 and Generalized Anxiety Disorder 7-Item, may be useful in assessment.

Box 45.2 Dermatoscopy terms

Dermoscopic terms	Seen in these conditions
Lines	Melanocytic naevi and melanoma
Reticular	Dermatofibroma, melanoma, and BCC
White	BCC
Leaf-like	Seborrhoeic keratosis
Cerebriform	Lentigo maligna
Angulated	
Clods	Growing naevi
Circumferential	Seborrhoeic keratosis
Brown, yellow, or orange	BCC
Blue	Melanoma
Pink	Haemangioma
Red/purple	
Dots	Melanoma, lentigo maligna
Grey	Seborrhoeic keratosis, actinic keratosis
White	
Circles	**Circles**
White	SCC, lentigo maligna
Pseudopods	Reed naevus
Structureless	**Structureless**
Brown-black	Dysplastic naevus, melanoma
Blue	Melanoma
Pink	Melanoma
White	Melanoma, dermatofibroma
Red	Actinic keratosis

BCC, Basal cell carcinoma; *SCC,* squamous cell carcinoma.

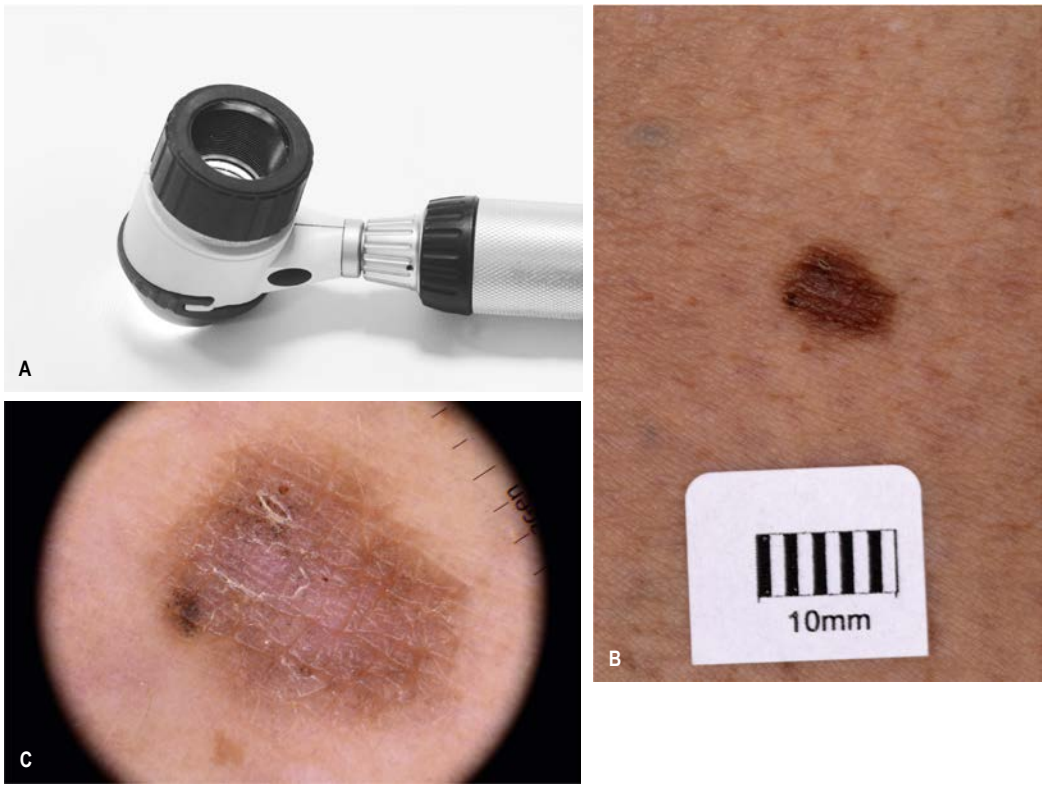


Fig. 45.1 **(A)** A dermatoscope. **(B)** Lesion without use of dermatoscope. **(C)** Same lesion seen through a dermatoscope. (*With kind permission from NHS Lanarkshire Medical Illustration Service.*)

The ***distribution*** (Box 45.3 and Fig. 45.2) and ***pattern*** (Fig. 45.3) of disease should be described. An algorithmic approach can be used to describe lesions and rashes (Fig. 45.4). Terminology used to describe the ***appearance*** of various types of rashes and lesions is discussed in Fig. 45.5.

Further reading

Kittler H, Marghoob AA, Argenziano G et al. Standardization of terminology in dermoscopy/dermatoscopy: Results of the third consensus conference of the International Society of Dermoscopy. *J Am Acad Dermatol.* 2016; 74(6):1093–1106.

Box 45.3 Describing the distribution of skin disease

- **Generalized** (consider systemic cause) or **localized** (consider external cause)
- **Unilateral** or **bilateral**
- **Symmetrical** or **asymmetrical**
- **Acral** (primarily affecting the extremities, i.e., fingers, toes, nose, ears, and penis) or **truncal** (primarily affecting the torso)

BACKGROUND

Skin diseases affect people of all ages and are among the most common human ailments. Although rarely fatal, many are chronic and constitute a leading cause of disability worldwide. The World Health Organization (WHO) International Classification of Disease (ICD) 10 lists more than 1000 skin or skin-related diseases, but most burden comes from a few diseases (Fig. 45.6). These include eczema or dermatitis; acne; psoriasis; urticaria; and superficial fungal, bacterial (impetigo, cellulitis, abscesses), and viral infections (warts, molluscum contagiosum).

People with skin conditions might seek advice from various sources, including online resources, pharmacists, primary care practitioners, and specialists in intermediate or secondary care. The estimated number of people in European countries with skin conditions and the commonest of those skin conditions are summarized in Box 45.4.

The psychosocial impact of chronic skin disease should not be underestimated, including its effect on the family, work, education, recreation, and relationships. The dermatology life quality index (DLQI) is a simple validated tool for assessing health-related quality

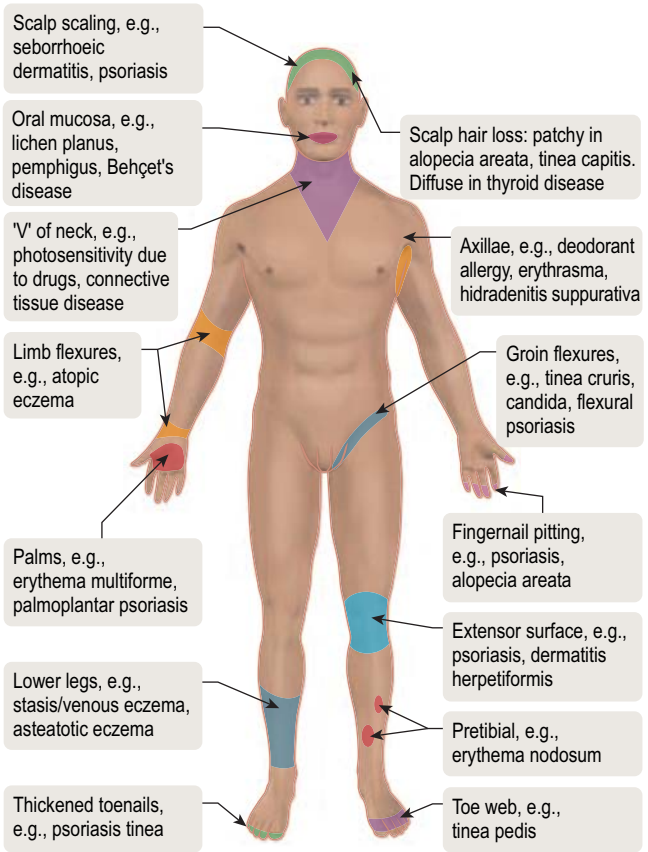


Fig. 45.2 Pattern of disease in different dermatological conditions.

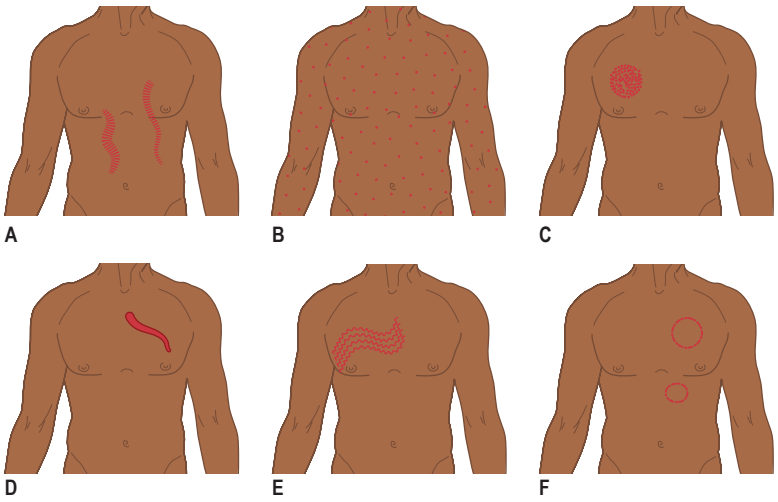


Fig. 45.3 Describing patterns in skin disease. **(A)** Serpiginous. **(B)** Scattered. **(C)** Grouped. **(D)** Linear. **(E)** Whorled. **(F)** Annular.

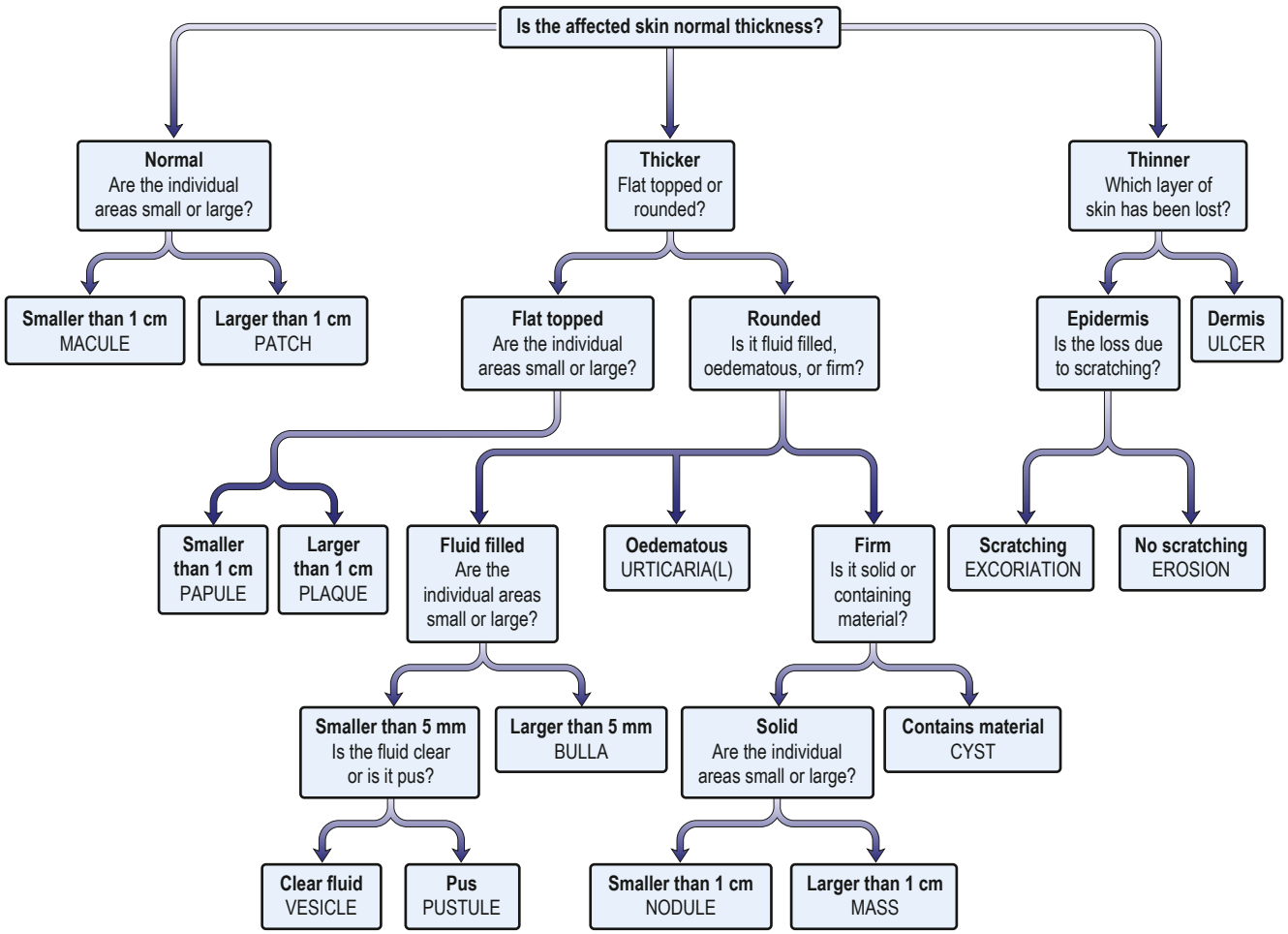


Fig. 45.4 Diagnostic approach to skin disease.

Nodule: a circumscribed, elevated lump in the skin greater than 1 cm. It can arise from the epidermis or dermis. This example is of nodular prurigo.

Papule: a circumscribed, slightly elevated lesion smaller than 1 cm diameter. This example is dermatosis papulosa nigra.

Macule: a completely flat lesion, altered in colour and smaller than 1 cm. This example is of solar lentigines.

Weal: a weal is due to localised dermal oedema. It is typically seen in urticaria.

Purpura: the extravasation of blood into the dermis. It is often due to injury to the blood vessels from inflammation.

Ecchymosis: bruising in the skin related to skin fragility or blood clotting disorders.

Crusting: a combination of an exudate (serous fluid) in combination with scaling. This example is impetigo.

Scaliness: looks like dry skin and indicates that the epidermis is affected by the condition. This example is seborrhoeic dermatitis.

Lichenification is the exaggeration of skin markings as a result of rubbing or scratching the skin. It is typically seen in chronic eczema.

Plaque: a circumscribed, palpable, raised lesion larger than 1 cm. It can arise from either epidermis or dermis. This example is of psoriasis.

Hyperpigmentation is darkening of the skin and is most often due to excessive melanin. It can also be due to deposition of iron or other pigments in the dermis.

Hypopigmentation is partial loss of skin colour. Depigmentation is total loss of colour as in this example of vitiligo.

Fig. 45.5, cont'd

A pustule is a circumscribed lesion containing pus. The pus can be infectious or it can be sterile and due to inflammation as in this example of pustular psoriasis.

A vesicle is a fluid-filled blister smaller than 5 mm. Vesicles can be solitary but are more likely to be multiple. This example is pompholyx.

A bulla is a fluid-filled blister bigger than 5 mm. This example is due to extreme insect bite reactions.

Ulcers typically involve loss of epidermis and dermis. This example is of pyoderma gangrenosum.

Erosions occur when the epidermis is damaged and lost but the dermis is intact.

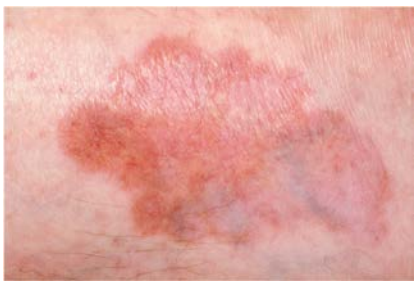
Excoriation is skin erosion due to scratching. In skin of colour there is loss of epidermal colour in the excoriated areas.

Petechiae are tiny red marks due to leakage from small blood vessels in the dermis. They do not blanche on pressure. This example is due to infective endocarditis.

Fissures are splits or cracks in epidermis and can be most commonly seen in the thick stratum corneum of the palms and soles.

Erythema is redness due to increased dermal blood flow. It blanches on pressure.

Telangiectasia are blood vessels which have enlarged enough to be visible. They may be part of a solitary lesion or be multiple as part of a systemic condition.



Atrophy is thinning of the dermis. In this example it is due to necrobiosis lipoidica with blood vessels visible through the atrophic dermis.

Fig. 45.5 Terminology describing types of lesions. (*Nodule image from Paller, Mancini.* Hurwitz Clinical Pediatric Dermatology: A Textbook of Skin Disorders of Childhood and Adolescence, *6th edn. Papule image from Neville, B., Damm, D. D., Allen, C. M., et al. [2009].* Oral and maxillofacial pathology *[3rd ed.]. St Louis: Saunders. Macule image from* Andrews' Diseases of the Skin Clinical Atlas, *2nd edn. Weal image reproduced with permission from ©DermNet dermnetnz.org 2024. Purpura image courtesy Kalman Watsky, MD. From Piette WW. Purpura: mechanisms and differential diagnosis. In: Bolognia J, Schaffer J, Cerroni L, eds.* Dermatology. *4th ed. St. Louis: Elsevier; 2018:377. Ecchymosis image from Lynch VA, Duval JB.* Forensic Nursing Science. *2nd edn. St. Louis: Mosby; 2011. Crusting image from Hurwitz S.* Clinical Pediatric Dermatology: A Textbook of Skin Disorders of Childhood and Adolescence, *5th edn. St. Louis: Elsevier; 2016. Scaliness image courtesy Department of Dermatology, School of Medicine, University of Utah. Lichenification image courtesy of Julie V. Schaffer, MD. Copyright Year 2025. Plaque image courtesy of Julie V. Schaffer, MD. Copyright Year 2025. Bulla image courtesy of Julie V. Schaffer, MD. Copyright Year 2025. Ulcers image courtesy of Christopher Sayed, MD. Erosions image courtesy of Kalman Watsky, MD. Copyright Year 2025. Excoriation image courtesy of Steven Binnick, MD. Copyright Year 2020. Petechiae image from Fowler VG Jr, Bayer AS. Infective endocarditis. In Goldman L, Ausiello D, eds.* Cecil Medicine, *23rd edn. Philadelphia: Saunders; 2008. Copyright Year 2024. Fissures image courtesy of the Institute of Dermatology, London, UK. Copyright Year 2020. Erythema image courtesy of Frank C. Powell, MD. Copyright Year 2025. Telangiectasia image courtesy of Jeffrey P. Callen, MD. Copyright Year 2025. Atrophy image from the collection of the late NP Smith, MD, the Institute of Dermatology, London, UK. Copyright Year 2020.*)

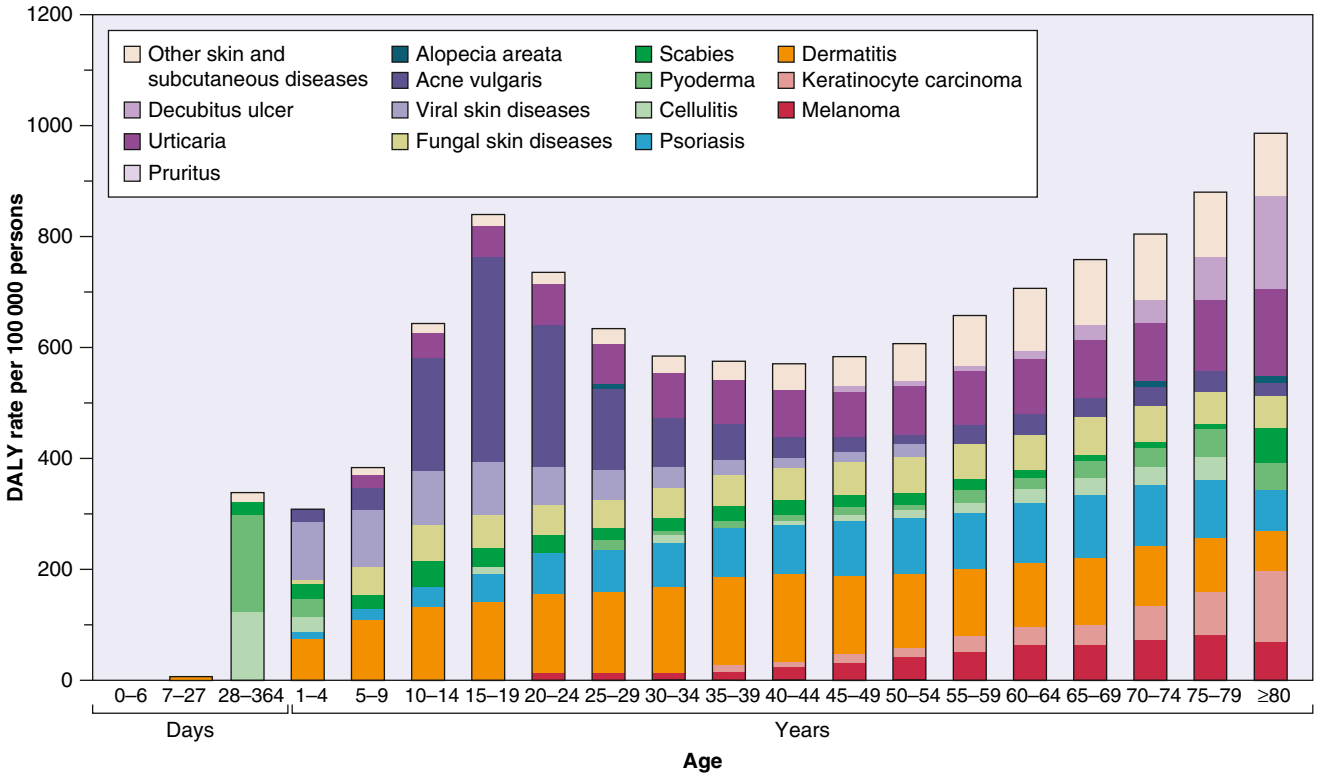


Fig. 45.6 Age distribution of skin and subcutaneous disease burden. Disability-adjusted life year (DALY) rate per 100 000 persons from 15 skin disease categories throughout the human lifespan. (*From Karimkhani C, Dellavalle RP, Coffeng LE et al. Global skin disease morbidity and mortality: an update from the Global Burden of Disease Study 2013.* JAMA Dermatol *2017; 153:406–412.*)

of life for patients with skin disease. Some skin diseases can have a greater impact on quality of life than even cardiovascular disease and cancers.

SKIN'S STRUCTURE AND FUNCTION

Understanding the structure and function of the skin helps make sense of the signs and symptoms of skin disease and can help to guide treatment.

The skin is composed of three distinct layers: the epidermis, dermis, and subcutis (**Fig. 45.7**). An overview of the functions of the skin is provided in **Box 45.5**.

Epidermis

The epidermis is the outermost layer. It functions as a renewable dynamic barrier against the environment and pathogens and plays an essential role in homeostasis. Its thickness ranges from 0.05 to 1.5 mm, depending on body site, and it is tethered to the underlying dermis by the dense protein network of the basement

membrane zone (BMZ; see Fig. 45.45). Inherited or autoimmune-induced deficiencies of BMZ proteins lead to fragility and blistering of the skin and mucous membranes (see 'Blistering and bullous skin disease' section). The epidermis is the source of the melanin that gives rise to skin colour.

Barrier functions of the epidermis

The epidermis is a ***stratified squamous epithelium*** composed mostly of ***keratinocytes*** (keratin-synthesizing cells) that originate from a proliferating ***basal layer***. They differentiate and migrate out to the surface to be shed as dead, flattened ***squames***. Epidermal differentiation is a highly complex process involving 'switching on and off' genes that regulate protein and lipid synthesis. Basal keratinocytes synthesize the ***keratin*** filaments of the cytoskeleton and desmosomal proteins (desmoplakin, desmoglein) for cell-to-cell cohesion. ***Lipids*** are synthesized in the granular layer and secreted to form a watertight intercellular layer. ***Filaggrin*** is expressed in the outer stratum corneum, where it acts as a 'natural moisturizing factor', keeping the skin hydrated and pliable.

The stratum corneum provides most of the skin's barrier function, and deficiency in filaggrin leads to 'leaky', dry skin that cracks, allowing entry of allergens and pathogens. Filaggrin gene mutations are common and have been identified as a major risk factor for atopic eczema and other atopic diseases. Changes in lipid and protease activity in the outermost epidermis lead to shedding (desquamation). The keratinocyte's journey from basal cell to skin surface normally takes about 30 days but is considerably faster in people with psoriasis and other conditions, leading to thickened, scaly skin.

Melanin-synthesizing ***melanocytes*** reside in the basal layer and transfer ultraviolet (UV)-protecting pigment to adjacent keratinocytes. The epidermis also contains nerve endings and sensory cells ***(Merkel cells)***.

Box 45.4 Prevalence of skin disease in European countries, including the UK

	12-month prevalence (%)
At least one skin complaint	50.9
Fungal infection	8.9
Eczema	5.5
Acne	5.4
Psoriasis	3.9
Rosacea	2.0
Urticaria	1.0
Non-melanoma skin cancer	1.1
Vitiligo	0.8
Malignant melanoma	0.6

M.A. Richard, C. Paul, T. Nijsten, P. Gisondi, C. Salavastru, C. Taieb, M. Trakatelli, L. Puig, A. Stratigos, EADV burden of skin diseases project team. Prevalence of most common skin diseases in Europe: a population-based study. Journal of the European Academy of Dermatology and Venereology, *2022;Volume 36, Issue 7 p. 1088–1096.*

Box 45.5 Functions of the skin

- Physical barrier against friction and shearing forces
- Protection against allergens, chemicals, and ultraviolet radiation
- Protection against infection (innate and acquired immunity)
- Prevention of excessive water loss or absorption
- Temperature regulation
- Ultraviolet-induced synthesis of vitamin D
- Wound healing
- Social and interpersonal function

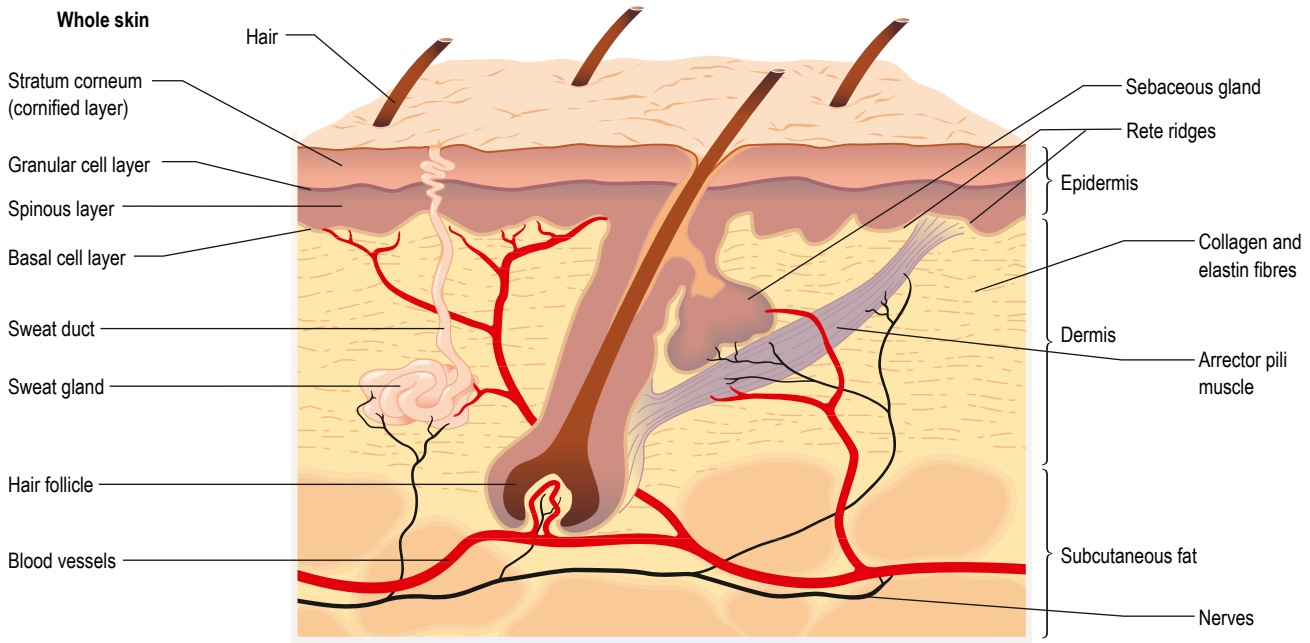


Fig. 45.7 The structure of the skin.

Immune functions of the epidermis

Keratinocytes also secrete cytokines, such as interleukins (ILs), interferon-gamma, and tumour necrosis factor-alpha (TNF-α), in response to injury; these cytokines stimulate wound healing and protect against microbial invasion. They also produce antimicrobial peptides that form part of the innate immune system. Deficiency of these peptides may increase susceptibility to staphylococcal and viral infection in people with atopic eczema. The epidermis contains a network of ***Langerhans cells*** (dendritic cells) that capture and process antigens for presentation to T lymphocytes.

Dermis

The dermis is a tough matrix of ***collagen*** and ***elastin fibres***, which give it strength, surrounded by a gel-like ground substance containing hyaluronic acid, proteoglycans, and glycoproteins, which forms a water-based environment that allows diffusion of molecules and provides compressive strength. It also contains cells (fibroblasts, mast cells, lymphocytes, and dendritic cells), blood and lymphatic vessels, nerves, muscle, and ***appendages*** (sweat glands, sebaceous glands, and hair follicles). The upper ***papillary dermis*** has finger-like projections that contain terminal capillary networks. The lower ***reticular dermis*** is thicker and denser.

Eccrine sweat glands are found throughout the skin, except for the mucosal surfaces, and carry out thermoregulatory sweating. ***Apocrine sweat glands*** are found in the axillae and anogenital area and do not function until puberty. ***Sebaceous glands*** are also inactive until puberty, when they excrete the oily substance sebum, under the influence of androgens. They are most numerous on the 'greasy' (seborrhoeic) sites of the central face, scalp, and upper torso. Sebum flows on to the skin surface via the pilosebaceous duct of the hair follicle. It contributes to skin barrier function and has antimicrobial actions.

The skin is richly innervated with ***sensory fibres*** (transmitting touch, pain, itch, vibration, pressure, and temperature) and ***autonomic fibres*** that innervate sweat glands, blood vessels, and arrector pili for thermoregulation.

Hair and nails

The skin surface is covered with ***hair***, except for the mucous membranes, palms, and soles. Hair grows from follicles, which are modified epidermal structures deeply rooted in the dermis. The lower portion of the hair follicle (bulb) surrounds a richly innervated and vascularized dermal papilla. Hair grows from the stem cells in the bulb, and if this area is damaged by an inflammatory process or trauma, it leads to permanent hair loss. Follicles go through a cycle of anagen (growth), catagen (involution), and telogen (shedding). At any one time, most hairs (>90%) will be in the anagen phase, which is usually 3 to 5 years for scalp hair. Miniaturization of coarse terminal scalp hairs with age leads to hair thinning, or baldness.

Nails are modified epidermal structures that contain specialized tough keratin proteins. It takes about 6 months to grow a fingernail and 1 year to grow a toenail.

Subcutis

The subcutaneous layer consists of lobules of adipose tissue separated by fibrous septa with sparse blood vessels and nerves. This layer provides insulation and cushioning against trauma.

Further reading

Dermoscopy https://dermnetnz.org/cme/dermoscopy-course

INVESTIGATION OF SKIN DISEASE

The most common dermatology investigations used in dermatology are microbiological: microscopy and culture of microbial swabs, as well as fungal tests (mycology) of skin, hair, or nail clippings (**Box 45.6**). Histology is also carried out, when skin lesions are removed by shaving, curetting, or excision, depending on their depth and the suspected diagnosis. This can be done under local anaesthetic as an outpatient procedure. Immunofluorescence (IMF) tests on skin and serum are indicated in patients with suspected immunobullous diseases.

PRINCIPLES OF DERMATOLOGICAL THERAPY

Topical therapy

The skin is easily accessible to topical therapy, and this is an effective and generally safe way of treating many rashes (and some low-grade malignancies) with a reduced risk of systemic adverse effects compared with oral therapy. Attention to cosmetic acceptability, clear instructions about how and when to apply (**Box 45.7**), duration of treatment, and anticipated effects (including how quickly the condition should settle and the likely side-effects) optimize patient adherence. It is helpful to specify quantities in simple terms, such as a 'pea-sized' portion to treat half a face or the 'fingertip unit' for topical corticosteroids in eczema management.

Box 45.6 Common dermatological investigations

- Microbial swabs: bacterial and fungal (impetigo, folliculitis, candida); viral swab (herpes simplex virus, varicella zoster virus)
- Mycology (fungal studies) of plucked hair or scalp brushings (tinea capitis)
- Mycology of nail clippings (onychomycosis)
- Mycology of skin scrapings (tinea, pityriasis versicolor)
- Skin biopsy (histology, immunohistochemistry, direct immunofluorescence, culture)
- Blood tests (routine bloods, inflammatory markers, autoantibodies, indirect immunofluorescence)
- Urinalysis (glucose, red cells) and urine microscopy (red cells, casts)
- Patch tests for allergic contact dermatitis

Box 45.7 How to use topical treatment

Topical treatment instructions	Examples of instructions	Example of treatment type
What to apply	Ointment on trunk and limbs Lotion on scalp Shampoo on body for 1 minute	Emollient Topical steroid Antifungal shampoo
Where to apply	On the affected skin On the affected zone From the neck down On the whole skin	Topical steroid Acne treatment Scabies treatment Emollients
How much to apply	A fingertip unit covers an area the size of a palm Until the skin is shiny	Topical steroid Emollients
How often to apply	Once daily Two to four times daily As often as needed	Topical steroid Emollients Anti-itch creams
When to apply	Last thing at night After bathing When going outside	Topical steroid Emollients Sunblock
Specific instructions	Apply in direction of hair growth 30 minutes before going to bed Keep away from dogs Leave on for 2 hours Avoid touching mucosal sites	Ointments Benzoyl peroxide 5-Fluorouracil cream Topical psoriasis treatment Capsaicin cream

Topical medicaments consist of an ***active ingredient***, in a ***vehicle*** or ***base***, and ***excipients*** such as ***preservatives*** or ***emulsifiers*** to maintain stability and efficacy. The following formulations exist:

- ***Creams (oil-in-water emulsions)*** are light and easily absorbed. They usually contain preservatives, such as parabens (hydroxybenzoates), which can sometimes cause allergic contact dermatitis (see later). Aqueous cream is a popular and inexpensive soap substitute (but should not be used as a 'leave-on' emollient).
- ***Ointments*** are greasy preparations based in vehicles such as polyethylene glycol (water soluble) or petrolatum (insoluble). They feel sticky and are useful for treating dry, flaky rashes. Some have an emulsifying wax added to allow them to be used as soap substitutes.
- ***Lotions*** contain water or alcohol that evaporates after application, giving a cooling effect. They are useful for weeping skin conditions and on hairy sites (e.g., the scalp). Alcohol-based lotions should be avoided on broken skin because they sting.
- ***Gels*** are semi-solid preparations of high-molecular-weight polymers. They are non-greasy and liquefy on contact with the skin. They are used to treat oily or hairy areas (e.g., scalp).
- ***Pastes*** are thick, adherent ointments containing a powder (e.g., Lassar paste). They are rarely used now because of poor cosmetic acceptability.

Adverse effects of topical therapies

- ***Systemic absorption*** may occur when large areas of inflamed skin are treated, especially with occlusive bandages or dressings. Neonates are particularly susceptible because of their relatively high body surface area and immature skin barrier, and topical steroids can cause adrenal suppression in this age group.
- ***Contact allergy*** may develop to the drug (e.g., hydrocortisone, neomycin), vehicle (e.g., lanolin, cetearyl alcohol), or excipients (e.g., chlorocresol, parabens). It should be considered in any patient with an unresponsive dermatitis or new dermatitis after topical treatment. Ointments contain fewer excipients and so are less likely to cause contact allergy.
- ***Folliculitis*** can occur because of blockage of hair follicles, especially with ointments in hot weather and on sweaty, occluded skin (bandages, tight clothing). It can also occur as a side-effect of topical steroids.
- ***Corticosteroids*** can cause a range of adverse effects but have a very good safety profile when used correctly.

Phototherapy

Phototherapy with natural sunlight has been used to treat people with skin diseases since ancient times. Modern phototherapy uses specific wavelengths in the UV spectrum for anti-inflammatory or pigment-inducing effects. The conditions that are most commonly treated with phototherapy are psoriasis, eczema, polymorphic light eruption, cutaneous T-cell lymphoma, and vitiligo. Narrow-band UVB (311 nm/TL01) was developed specifically to treat patients with psoriasis and is the treatment of choice in most cases. UVA therapy is given with systemic or topical psoralens (psoralen + UVA = PUVA) to enhance its photobiological effects. Commercial tanning beds predominantly deliver low doses of UVA and have limited effectiveness in treating skin disease but carry a risk of inducing melanoma and other skin cancers.

Narrow-band UVB therapy is given two or three times a week for 20 to 25 treatments. It has superseded broadband UVB because it is more effective. PUVA is usually given twice a week for 20 to 25 treatments. If the psoralen is taken orally, UV-protective glasses must be worn for the day of treatment to protect the retinas. PUVA is used less widely nowadays because it is associated with an increased risk of skin cancer, especially SCC. The maximum recommended lifetime dose of PUVA is 200 treatments.

Unaffected regions of skin or high-risk areas such as the scrotum should be covered during phototherapy. To avoid adverse effects, patients should be carefully selected and their skin cancer risk assessed. Phototherapy is generally avoided in those who have already had a skin cancer. People with very fair skin and freckles are at increased risk of burning, and the UV dosage has to be carefully adjusted and escalated according to skin type or based on phototesting to look for the lowest dose that causes just perceptible redness (minimal erythema dose).

Systemic therapy

A wide range of systemic drugs are used to treat skin disease. Some are used off-licence, especially for uncommon diseases where no licensed option exists. The risks and benefits should be carefully considered and discussed with the patient so they can share in treatment decisions. Selected examples of the types of systemic drugs include:

- Antibiotic, antiviral, and antifungal therapies
- Antiandrogens (e.g., cyproterone acetate, spironolactone, minoxidil)
- Anti-inflammatories (antihistamines, antimalarials, colchicine, corticosteroids, dapsone, Janus kinase (JAK) inhibitors, methotrexate, phosphodiesterase inhibitors, thalidomide)
- Anti-scabetics, e.g., ivermectin or anti-helminths (e.g., albendazole)
- Immunosuppressants (azathioprine, ciclosporin, cyclophosphamide, mycophenolate mofetil)
- Retinoids (e.g., acitretin for psoriasis, alitretinoin for hand dermatitis, isotretinoin for acne, bexarotene for cutaneous T-cell lymphoma)
- Monoclonal antibodies (for psoriasis, eczema, urticaria, prurigo, pemphigus).
- Anticancer therapies including small molecule targeted therapy (e.g., vismodegib for basal cell carcinoma [BCC]; dabrafenib and trametinib for melanoma) and immunotherapies (e.g., checkpoint inhibitor therapy for melanoma and Merkel cell carcinoma)

DERMATOLOGICAL EMERGENCIES

Although most skin diseases are dealt with in an outpatient setting, occasionally, they can be life-threatening and require intensive inpatient treatment. Dermatological emergencies are covered in the acute medicine chapter (see Chapter 11) and include:

- ***Erythroderma*** (see Chapter 11, 'Erythroderma and loss of skin function ['skin failure']' section), describing extensive skin inflammation (affecting most or all of the body surface). Patients can become seriously unwell as a result of transdermal fluid losses, alterations in peripheral vascular resistance, thermal dysregulation, and secondary infection. Causes include primary skin diseases such as eczema or psoriasis, allergic reactions, and malignancies.
- ***Necrotizing fasciitis*** (see Chapter 11, 'Necrotizing fasciitis' section), a severe form of bacterial soft tissue infection, often associated with gas formation, extensive tissue destruction, and a high mortality rate.
- ***Severe drug reactions*** (see Chapter 11, 'Severe cutaneous adverse drug reactions' section), including ***Stevens-Johnson syndrome*** (SJS), ***toxic epidermal necrolysis*** (TEN) and ***drug reaction with eosinophilia and systemic symptoms*** (DRESS syndrome). Milder drug eruptions are discussed toward the end of this chapter.

COMMON RASHES

Acne and related disorders

Acne

Acne is one of the most common skin conditions globally, affecting more than 85% of adolescents and often persisting into adulthood. In females, a late-onset chronic variant starting in the 20s is increasingly recognized. Acne can have a profound effect on self-esteem and mood, leading to anxiety, depression, and even suicidal ideation. This may have little correlation with apparent disease severity.

Lesions arise in the pilosebaceous follicle, which becomes blocked because of abnormal keratinization and increased production of altered sebum, leading to comedones (open comedones are 'blackheads', and closed comedones are 'whiteheads'). Microcomedones (microscopic comedones) can be found in biopsies of normal-looking acne-prone skin. Blockage of the pilosebaceous duct is followed by alteration in its microbiome, especially involving *Cutibacterium acnes* (formerly *Propionibacterium acnes*), leading to activation of innate immunity (Fig. 45.8A). This causes an inflammatory response with influx of neutrophils that release elastase, causing connective tissue damage. Scarring may follow with profound long-term psychological consequences. Post-inflammatory hyperpigmentation (see 'Post-inflammatory hyperpigmentation' section) is also a common sequela in darker-skinned people.

Clinical features

Acne affects the face and upper torso, where the sebaceous glands are dense and affected skin is usually greasy (seborrhoea). Lesions are classified as:

- ***Non-inflammatory:*** open comedones (blackheads) or closed comedones (whiteheads)
- ***Inflammatory:*** papules, pustules (Fig. 45.8B), nodules, and cysts
- ***Scars:*** raised (hypertrophic) or depressed or pitted (box, rolling, and ice-pick)
- ***Post-inflammatory hyperpigmentation*** (especially in patients with skin of colour (Fig. 45.8C)

Uncommon acne variants are shown in Box 45.8.

Severe acne may be associated with various autoinflammatory syndromes characterized by fever and aseptic inflammation (Box 45.9).

Management

The diagnosis of acne is usually straightforward. Treatment should target the primary pathogenic lesion: the ***microcomedone*** and associated inflammatory response. The choice depends on whether lesions are predominantly non-inflamed or inflamed, disease severity or scarring, response to previous treatment, and degree of psychological upset (Box 45.10).

- Topical retinoids (adapalene and tretinoin), azelaic acid, salicylic acid, and benzoyl peroxide (BPO) are ***keratolytic*** and help unblock pores but also cause dry, flaky skin. This can be managed with a non-comedogenic, oil-free moisturizer.
- Topical antibiotics including clindamycin are also used and topical fixed combinations are usually recommended (e.g., adapalene/benzoyl peroxide, tretinoin/clindamycin, benzoyl peroxide/clindamycin). Treatment usually takes 6 to 8 weeks to have a noticeable positive effect.
- ***Oral tetracyclines*** and ***erythromycin*** are licensed for acne, but their prolonged use may encourage the emergence of resistant bacteria. They should not be prescribed for the long term or

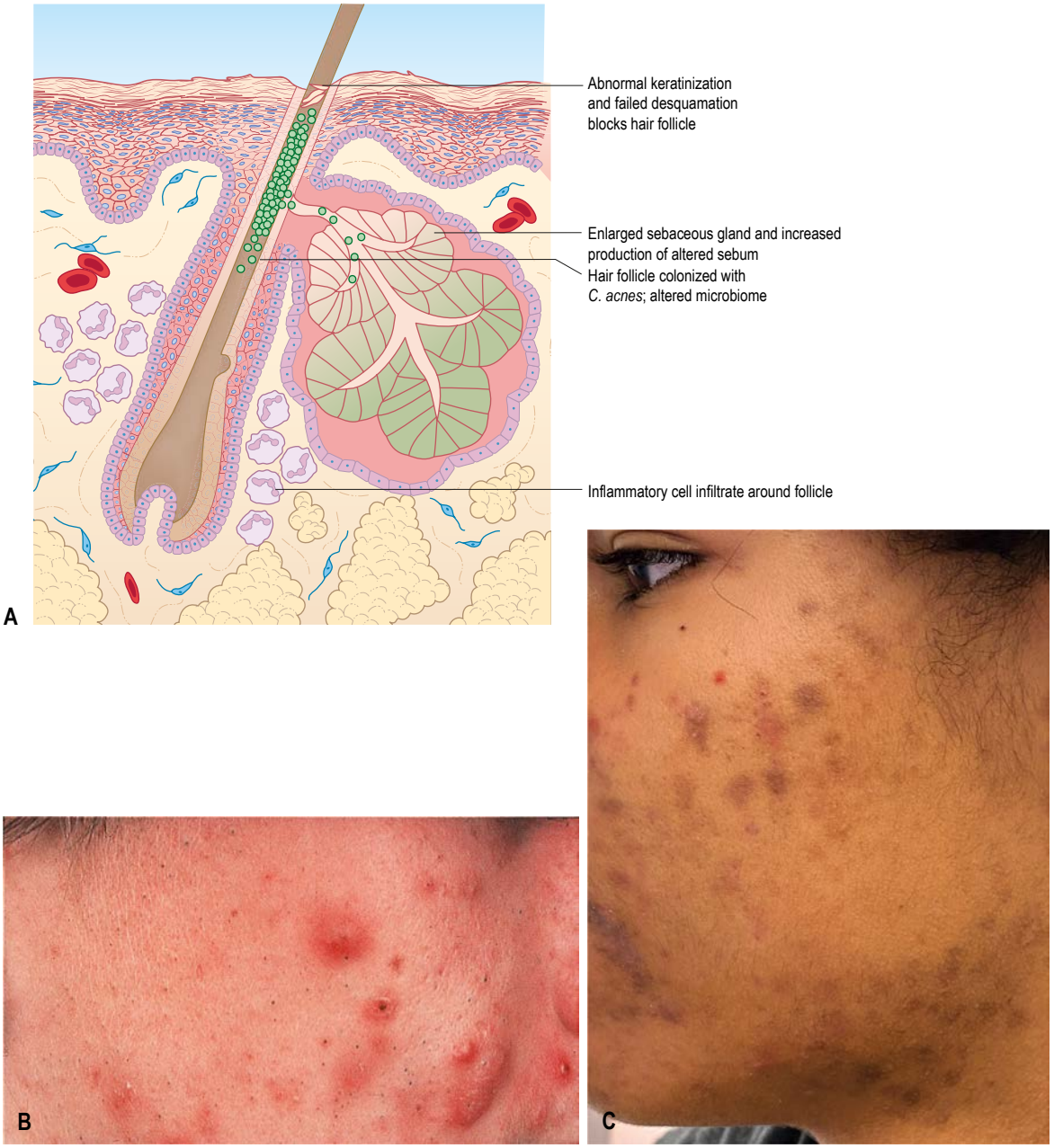


Fig. 45.8 Acne. (A) Pilosebaceous unit, which becomes blocked in acne. **(B)** Comedones and inflammatory lesions in facial acne. **(C)** Post-inflammatory hyperpigmentation secondary to acne. Such pigmentary changes are common in people with darker skin tones. (*A, From Eichenfield LF, Frieden IJ, Mathes EF et al.* Neonatal and Infant Dermatology, *3rd edn. Elsevier Inc.; 2015; Fig. 2.1. Courtesy of Randall Hayes.*)

without non-antibiotic topical therapy because the latter reduces the emergence of resistant bacteria (Box 45.11).

- In females, the ***combined oral contraceptive pill*** and a formulation containing additional ***cyproterone acetate*** (a mild antiandrogen) may be of benefit because of a reduction in sebum excretion. Low-dose spironolactone is another useful option. Progesterone-only pills ('mini-pills') may aggravate acne.
- ***Oral isotretinoin*** is a highly effective treatment licensed for severe inflammatory acne, especially with scarring. It is also used in milder recalcitrant cases associated with severe psychological upset. It causes dryness of the mucous membranes and has a range of other potential adverse effects. A major concern is its potent teratogenicity, so females of childbearing age must use highly effective contraception and follow a specific pregnancy prevention programme. Oral isotretinoin has been associated with depression and rare cases of suicide, so mood should be carefully monitored. There is now a programme in the UK whereby all patients requiring treatment with isotretinoin need documented monitoring of the psychosocial co-morbidities associated with acne, and for patients younger than 18 years of age, two signatories are required for prescriptions.
- High-intensity light treatment (blue light, intense pulsed light) has also been reported to lessen inflammatory lesions, and ablative laser therapy and physical treatment, such as dermabrasion,

Box 45.8 Clinical variants of acne

Infantile acne
- Typically starts between 3 and 12 months of age
- Triggered by maternal androgens

Acne excoriée
- Mainly affects young females who pick mild acne spots, leaving prominent excoriations
- There is usually underlying psychological upset

Drug-induced acneiform eruptions
- May be caused by corticosteroids, vitamin B_{12}, isoniazid, and epidermal growth factor receptor inhibitors
- Comedones are absent

Pomade acne
- Caused by oily scalp cosmetics
- Typically affects the forehead

Oil acne
- Caused by prolonged (e.g., occupational) contact with oils
- Usually affects the legs

Chloracne
- Caused by exposure to toxins (e.g., dioxin)
- Widespread comedones and large, straw-coloured, non-inflamed cysts

Acne conglobata
- A severe form of cystic acne with abscesses and interconnecting sinuses
- Usually affects young males

Acne fulminans
- Associated fever, weight loss, and musculoskeletal pain
- Systemic steroids and oral isotretinoin are required

Follicular occlusion triad
- Severe cystic acne with cellulitis of the scalp and hidradenitis suppurativa

Box 45.9 Acne-associated autoinflammatory syndromes

- SAPHO syndrome: synovitis, acne, pustulosis, hyperostosis, osteitis
- PAPA: pyoderma gangrenosum, acne, pyogenic arthritis
- PASH: pyoderma gangrenosum, acne, suppurative hidradenitis (hidradenitis suppurativa)
- PAPASH: pyoderma gangrenosum, psoriatic arthritis, suppurative hidradenitis (hidradenitis suppurativa)
- PASS: pyoderma gangrenosum, acne, seronegative spondyloarthritis

Box 45.10 Treatment of acne

Mild or physiological acne
- Step 1: Provide oil-free skin care for acne-prone skin, including salicylic acid washes (keratolytic).

Comedonal acne
- Step 2: Add topical retinoid or azelaic acid (keratolytic).

Inflammatory acne
- Step 3: Give topical retinoid + benzoyl peroxide.
  - Benzoyl peroxide ± topical or oral antibiotic (antimicrobial).

Severe acne
- Step 4: Consider oral isotretinoin.
  - Anti-androgen therapy (cyproterone acetate, spironolactone) may be useful in females with moderate acne and seborrhoea.

Box 45.11 Prescribing antibiotics for acne

- Avoid long-term prescription (>6 months) of oral antibiotics
- Do not use topical antibiotic monotherapy (co-prescribe with topical retinoid or benzoyl peroxide)
- Avoid prescribing dissimilar oral and topical antibiotics

may improve the appearance of scars. However, a key goal of acne treatment is to *prevent* scar formation.
- Treatment of scarring and post-inflammatory pigmentation disorders may involve topical retinoids, dermabrasion, 'peels', laser treatments, and sometimes aesthetic surgical interventions.

Acne itself is associated with significant psychological disease, including body dysmorphic disorder, suicidal ideation, changes in self-perception, and skin picking (acne excoriee). However, for most people with acne, emotional wellbeing improves as their acne clears.

Hidradenitis suppurativa ('acne inversa')

Hidradenitis suppurativa is a chronic inflammatory disorder that affects the apocrine pilosebaceous follicles of the axillae, the inguinal area, and around the breasts ('apocrine acne'). It is characterized by recurrent abscesses, draining sinuses, and scarring and is associated with the metabolic syndrome, obesity, and smoking. Patients are often disabled by pain and malodorous discharging lesions; there is no definitive treatment. Bacterial biofilms within the occluded follicles may explain the disappointing results of antibiotic therapy.

Treatment options include oral tetracycline, combined rifampicin and clindamycin, acitretin, adalimumab (anti-TNF), and IL-17 blockers (secukinumab). Surgery (usually marsupialization of affected areas, or excision and grafting) may be required to drain abscesses and to excise the affected areas of skin.

Rosacea

Rosacea (Fig. 45.9) is a common inflammatory rash predominantly affecting the central face. It usually starts in mid-adult life and is more common in fair-skinned people. The cause is unknown, but the pathogenesis probably involves vasodilation of cutaneous facial microvasculature (which may become fixed telangiectasia, i.e., non-blanching on pressure) with or without chronic sun damage and an increased innate immune response triggered by the commensal mite *Demodex folliculorum* and leading to synthesis of pro-angiogenic and pro-inflammatory mediators.

Clinical features

The main features are diffuse erythema and inflammatory papules and pustules. Flushing and telangiectasia may be early features and can be aggravated by hot drinks, alcohol, sunlight, and changes in temperature *(erythemato-telangiectatic rosacea)*. This may be

Fig. 45.9 Rosacea. Papules and pustules on background erythema. There are no comedones.

followed by papules and pustules *(papulopustular rosacea)*, but unlike with acne, there are no comedones. Ocular involvement can lead to blepharitis, keratitis, and conjunctivitis. Later complications include sebaceous gland and soft tissue overgrowth *(phymatous rosacea)*, especially of the nose in males, causing rhinophyma and facial lymphoedema. Use of topical steroids can exacerbate or trigger rosacea.

Management

Treatment is suppressive rather than curative. Avoidance of agents that make the skin flush (alcohol, saunas, steam rooms, capsaicin [the 'hot' component of chilli]-containing foods) together with topical metronidazole, azelaic acid, or ivermectin may be used for long-term control, with intermittent courses of oral tetracyclines. Subantimicrobial doses of doxycycline are effective because of an anti-inflammatory effect.

Flushing and erythema do not respond to antibacterial drugs, and treatment options include cosmetic concealers, topical brimonidine, and vascular laser therapy. Rhinophyma can be debulked surgically or with an ablative laser (e.g., CO_2 laser). Lasers may be useful for fixed telangiectasia or erythrotelangiectatic rosacea.

Perioral dermatitis

Perioral dermatitis is a distinctive rosacea-like rash that typically affects the nasolabial fold and area around the mouth in young females in a 'diamond' distribution. It is characterized by small inflammatory papules and pustules with overlying scaling. Topical steroids may trigger the condition in susceptible individuals and are often continued inadvertently because the rash flares when they are discontinued. Treatment is with an oral tetracycline, as for rosacea, and with weaning off topical steroids.

Other rarer causes of facial rashes

These include autoimmune rheumatic diseases and lymphocytic and granulomatous infiltrates.

Further reading

Ingram JR, Collier F, Brown D et al. British Association of Dermatologists guidelines for the management of hidradenitis suppurativa (acne inversa) 2018; *Br J Dermatol*. 2018; 180:1009–1017.
Xu J, Mavranezouli I, Kuznetsov L, Stephen Murphy M, Healy E; Guideline Committee. Management of acne vulgaris: Summary of NICE guidance. *BMJ*. 2021; 374:n1800.
Zaenglein AL. Acne vulgaris. *N Engl J Med*. 2018; 379:1343–1352.

Eczema and dermatitis

The terms 'eczema' and 'dermatitis' are used interchangeably to describe an extremely common type of inflammatory rash. In the developed world, eczema may affect about 10% of the population at any one time, with up to 40% experiencing an episode of eczema during their lifetime.

Eczema is classified as ***constitutional*** or ***endogenous*** and ***contact*** or ***exogenous*** (Box 45.12). It can also be classified according to duration (***acute, subacute,*** or ***chronic***). The term 'eczema' is used to refer to a number of separate clinical conditions with overlapping clinical features, including:

- Atopic eczema
- Seborrheic eczema
- Venous eczema
- Discoid eczema
- Contact of irritant eczema

Box 45.12 Classification of eczema

Endogenous
- Atopic eczema
- Seborrhoeic eczema
- Venous ('gravitational') eczema
- Discoid eczema
- Asteatotic eczema
- Chronic hand or foot eczema
- Lichen simplex or nodular prurigo

Exogenous
- Irritant contact eczema
- Allergic contact eczema

- Chronic forms of eczema, including lichen simplex and nodular prurigo.

The cardinal features of inflammation are manifest in all forms of eczema, including

- Erythema
- Swelling
- Itch or soreness
- Loss of normal skin barrier function

In acute eczema, oedema is prominent, and tiny vesicles or larger bullae may be seen within inflamed skin. Skin swelling subsides to leave dry, flaky skin. Scratching leads to excoriations with serosanguinous exudate. Repeated scratching or rubbing causes thickened skin with the prominent lines ***(lichenification)*** of chronic eczema. The thickened skin is not supple and cracks easily, predisposing to fissures, especially on the hands and feet. Secondary bacterial infection may occur and cause crusts, papules, and pustules.

The histology of eczema correlates with the clinical features: whereas in acute eczema, the keratinocytes are swollen with increased intercellular fluid (spongiosis), in chronic disease, there is little oedema but prominent thickening of the epidermis (acanthosis) and scaling (hyperkeratosis). Inflammatory cells are present around the upper dermal vessels in all patterns.

Atopic eczema

This type of constitutional eczema is extremely common, affecting 5% to more than 20% of children worldwide (numbers vary widely between countries and ethnicities). It usually starts before the age of 5 years, and the disease course is chronic and relapsing with periods or flare and remission over many years. The majority (~80%) of children with early-onset atopic eczema spontaneously improve and 'clear' the disease before the teenage years. Recurrence may occur in adulthood, especially as a localized hand eczema. Onset in later childhood or adulthood is often associated with a more chronic relapsing course. It is often associated with other atopic diseases.

Aetiology

The pathogenesis of atopic dermatitis represents an interplay between genetic factors, Th2 cell–skewed immune dysregulation, epidermal barrier dysfunction, environmental triggers, an altered skin microbiome, and neuroimmune interaction. It is debated which mechanisms are the main drivers and which combinations of mechanisms accounts for the variable disease phenotypes.

A positive family history of atopic disease is often present: there is 90% concordance in monozygotic twins but only 20% in dizygotic twins. If one parent has atopic disease, the risk of a child developing eczema is about 20% to 30%. If both parents have atopic eczema, the risk is more than 50%.

Activation of Th2 CD4 lymphocytes in the skin drives atopic eczema, with increased IL-4, IL-15, and IL-13 production that lead to raised levels of immunoglobulin E (IgE) and IL-31, a key mediator of itch. Scratching leads to further epidermal damage. The later chronic inflammatory phase is more complex and involves Th1, Th17/IL-23, Th22 cells, and the JAK/STAT pathway. Th2 immune cell activation impairs epidermal antimicrobial peptides in the skin, leading to a partial immunodeficiency and predisposition to cutaneous infection, particularly with *Staphylococcus* spp. and herpes simplex virus (HSV).

An altered skin microbiome has been demonstrated in patients with eczema with a loss of diversity and a predominance of *Staphylococcus* spp. in acute and chronic eczematous skin. It is currently unknown if the microbial changes predispose to an eczema flare or are secondary.

Epidermal skin barrier dysfunction is a further key finding in atopic eczema. Multiple factors have been shown to contribute, including a reduction in filaggrin production, an imbalance in stratum corneum proteases, an abnormality in tight junction function, and microbial colonialization. Filaggrin deficiency is of major importance, and Th2 cytokines have been shown to reduce its production in the skin. Loss of function variants in the filaggrin gene are the strongest genetic predictor of atopic dermatitis with an associated three- to fourfold increase of developing the disease.

Chronic itch is the leading symptom of atopic dermatitis, but it is unknown if it drives inflammation. Itch is in part mediated by Th2 cytokines IL-4, IL-13, TSLP, and IL-31 blockade, which have been used therapeutically in the management of eczema and asthma and itch.

Exacerbating factors

Exposure to soap and detergent skin cleansers dries and damages the epidermis and promotes inflammation. In some patients, repeated infections with *Staphylococcus aureus* induce flares, which may be mediated by superantigen exotoxins that cause widespread T-cell activation. Teething, dribbling, and lip licking may trigger facial eczema in infancy and childhood. Severe anxiety or stress may exacerbate eczema in some individuals. Pet dander can aggravate eczema. The role of house dust mites and diet is less clear-cut.

Immediate or type 1 hypersensitivity to food allergens (cow's milk, egg, soya, wheat, fish, and nuts) is common in young children with severe atopic eczema. Ingestion of food causes perioral and sometimes generalized urticaria (hives) and gastrointestinal (GI) symptoms (reflux, vomiting, and diarrhoea). Rarely, there may be a delayed hypersensitivity to foods such as cow's milk that manifests with an eczema flare after 12 to 24 hours. This cannot be investigated with prick tests or blood tests, and the mechanism is unclear. The diagnosis is made by taking a careful dietary history followed by oral food challenge tests.

Clinical features

Atopic eczema has a variable clinical presentation. The characteristic sites involved are the flexures of the elbows, knees, ankles, and wrists (Fig. 45.10) and around the neck. In infants, eczema often starts on the cheeks, while impacting the flexures in childhood. In adulthood, the hands and feet, nipples, and lateral eyebrows are more likely to be affected. Clinical features of acute or chronic eczema (see earlier) may predominate, and the pattern can change with time. Pruritus is always present and is usually severe, leading to excoriations, lichenification, and dyspigmentation.

In darker skin tones, the erythema of acute eczema may be difficult to appreciate. The distribution may be different, with an 'inverse pattern' predominantly affecting the extensor surfaces, and follicles can be prominent. Pigmentary changes (hyper- and hypopigmentation) may follow the inflammatory phase of eczema and are often

Fig. 45.10 Atopic eczema. Infantile atopic dermatitis on the extensor arms. **(A)** Chronic diffuse involvement of the extensor arms with scaling and hyperpigmentation. Note the follicular prominence on the trunk. **(B)** Nummular lesions on the extensor aspect of the arm and nipple eczema. (*B, Courtesy of Julie V. Schaffer, MD.*)

slow to resolve. Hypopigmented patches over the cheeks termed ***pityriasis alba*** are common.

In some atopic individuals, the skin of the upper arms and thighs feel rough because of follicular hyperkeratosis (keratosis pilaris). Prominent skin creases on the palms (hyperlinear palms) and dry, 'fish-like' scaling of the skin of the lower legs (ichthyosis vulgaris) point to underlying deficiencies in filaggrin. Centrofacial pallor with characteristic nasal sparing is a typical sign of atopic eczema.

Secondary infection

Secondary bacterial infection with *S. aureus* or streptococci is common, especially in moist flexural areas such as the neck and groin. Pustules, oedema, and golden-crusted inflamed papules strongly suggest secondary infection.

Cutaneous viral infections (e.g., viral warts and molluscum) may be more widespread in atopic eczema. HSV can cause extensive infection of the skin, termed ***'eczema herpeticum'***. This appears as multiple small blisters or painful, punched-out crusted papules (Fig. 45.11) that are associated with malaise and pyrexia. Urgent systemic antiviral therapy is indicated to reduce the risk of disseminated life-threatening infection. An extensive form of hand, foot, and mouth disease may also occur in children with eczema, with blisters and crusted lesions extending over large areas of the limbs and face ('eczema coxsackium'). Ocular complications of atopic eczema include conjunctival irritation and, less commonly, keratoconjunctivitis and cataract.

Fig. 45.11 Eczema herpeticum. Grouped punched-out monomorphic erosions with haemorrhagic crusts on the arm **(A)** and posterior neck **(B)**. Vesicles are rarely evident. (*B, Courtesy of Julie V. Schaffer, MD.*)

Investigations

Atopic eczema is diagnosed on the history and clinical features. About 80% of patients also have laboratory features of atopy (raised total IgE and allergen-specific IgE and mild eosinophilia).

Management

General measures include a patient-centered approach avoiding common or proven allergens and irritants (especially bar and liquid soap). Clothing should be made of soft, 'breathable' fabric such as cotton (see Box 45.13). Manipulating the diet (e.g., following a dairy-free or egg-free diet) may help a minority of infants with moderate to severe eczema. However, this should be done only under expert guidance to ensure adequate intake of nutrients such as calcium. Dermatology nurse-led education of people living with eczema and their families provides psychological support and improves adherence to treatment.

Topical therapies

Topical therapies are sufficient to control atopic eczema in most people. The key components are moisturizers or emollients and topical anti-inflammatories such as corticosteroids and calcineurin inhibitors.

Fragrance-free moisturizers should be used liberally on all dry skin areas and reapplied as often as needed. Some can also be used as soap substitutes. Moisturizers with or without topical steroids can be applied under occlusive damp tubular gauze bandages ('wet wraps'). These can improve efficacy but are time consuming to apply. Paste bandages containing ichthammol can also be useful for resistant or lichenified eczema of the limbs.

Topical corticosteroids remain the bedrock of eczema treatment, and for most patients, they provide reasonable disease control. They are applied once a day to inflamed skin. Topical corticosteroids are classified into four groups, according to potency (Box 45.14). An appropriate potency should be selected according to the body site, surface area, and patient age to minimize the risk of adverse effects (Box 45.15) and allow safe, intermittent, long-term treatment. 'Steroid phobia' is common among patients or their carers and should be appropriately addressed to improve adherence. ***Topical steroid withdrawal syndrome*** is increasingly recognized as a rare rebound flare associated with prolonged use of potent steroids characterized by a burning and stinging sensation, skin pain, and flushing. Potent steroids should be used for short periods only and weaned down to milder steroids or used intermittently as the eczema settles. Regular use of emollients may reduce steroid usage.

Box 45.13 Management of atopic eczema

- Education and explanation
- Avoidance of irritants and allergens
- Emollients
- Bath oils, soap substitutes
- Topical therapies
  - Steroids
  - Calcineurin inhibitors
- Adjunct therapies
  - Oral antibiotics
  - Sedating antihistamines
  - Bandaging
- Phototherapy
- Systemic therapy, e.g., oral prednisolone, ciclosporin, methotrexate, dupilumab, tralokinumab, lebrikizumab, JAK inihibitors (baricitinib, abrocitinib, upadacitinib)

Box 45.14 Classification of topical corticosteroids by potency

Very potent
- 0.05% clobetasol propionate
- 0.3% diflucortolone valerate

Potent
- 0.1% betamethasone valerate
- 0.025% fluocinolone acetonide

Moderately potent
- 0.05% clobetasone butyrate
- 0.05% alclometasone dipropionate

Mild
- 2.5% hydrocortisone
- 1% hydrocortisone

Box 45.15 Adverse effects of topical steroids

- Cutaneous atrophy and telangiectasia
- Striae
- Topical steroid withdrawal syndrome
- Steroid-induced rosacea, perioral dermatitis, and folliculitis
- Tinea incognito (a tinea infection modified by a topical steroid applied inappropriately)
- Ocular adverse effects (cataract, glaucoma)
- Adrenal suppression (from systemic absorption with long-term use of potent steroids)

Mild steroids rarely cause skin atrophy, but this is a significant risk with potent and super-potent steroids, especially if applied under occlusion or with a keratolytic such as salicylic acid. Mild to moderate steroids are used to treat eczema on the face and flexures. Potent and super potent steroids are usually required to treat palmoplantar eczema because of the thick stratum corneum at this site. The 'fingertip unit' can be used to guide topical steroid dosage.

Topical calcineurin inhibitors (TCIs) (tacrolimus and pimecrolimus) are licensed for the treatment of atopic eczema in adults and children older than the age of 2 years. They do not cause skin atrophy and are therefore useful for delicate areas such as the face and eyelids. They are less effective on thick skin because of poor penetration. The main adverse effect is a burning or prickling sensation that usually improves with continued treatment. Initial concerns that the immunosuppressant effects of TCIs would increase the risk of skin cancer have not proved well-founded, but patients are advised to photoprotect.

New topical treatments, including the topical PDE4 (phosphodiesterase isozyme 4) inhibitor crisaborole and the JAK1/2 inhibitor ruxolitinib, have shown efficacy in the treatment of patients with eczema but are not currently licensed in the UK.

Antibiotics and antiseptics

Antibiotics are indicated if there are clinical signs of infection and are usually given orally for 7 to 10 days. High-dose flucloxacillin is the treatment of choice for patients with staphylococcal and streptococcal skin infections. Erythromycin is an alternative for those with penicillin allergy. Topical antiseptics are often used as emollient washes or bath additives or in the form of 'bleach baths'. They may cause irritation.

Sedating antihistamines

These may be useful at before bedtime, but because histamine is not a key mediator in the itch of eczema, they are often of limited benefit.

Second-line therapy and new drugs

These are used in severe, unresponsive eczema, especially if it is interfering with an individual's life (e.g., growth, sleeping, schoolwork, or job). ***UV phototherapy*** and ***systemic immunosuppressants*** (ciclosporin, azathioprine, methotrexate) can be helpful. Short, tapering courses of oral prednisolone usually provide rapid disease control, but relapse is common on withdrawal. The risks and benefits of long-term systemic therapy require careful evaluation and are chosen based on shared decision-making.

Recent years have seen a breakthrough in the treatment of patients with severe eczema through the development of new ***therapies that target type 2 immunity***. The monoclonal antibody dupilumab blocks anti-IL4Ra, the shared receptor of IL-4 and IL-13. It is the first approved biologic shown to be an effective and safe treatment for atopic eczema in adults and children. It is given by subcutaneous injection every 2 weeks. The suggestion that IL-13 plays a major role in disease pathogenesis led to the development of tralokinumab, the first anti–IL-13 monoclonal antibody that shows effective disease control with a lower incidence of conjunctivitis than reported for dupilumab.

In addition, three ***small molecule inhibitors of JAK*** have been licensed for the treatment of patients with atopic eczema in the UK. Abrocitinib (JAK1 selective) and upadacitinib (JAK1 selective) are approved in adults and children aged 12 year and older. Baricitinib (JAK1 and 2 selective) is approved for the use in adults only. They have been shown to be a rapid and highly effective treatment and have the advantage of being administered orally. However, data from safety trials in patient with rheumatoid arthritis showed a potential increased risk of serious infections, major cardiovascular events, cancer, thrombosis, and death, but these have not been noted in clinical trials for atopic eczema, and long-term data are awaited.

Prevention

Maintenance regimens using topical steroids or calcineurin inhibitors twice weekly have been shown to reduce the frequency of eczematous flares. Studies are exploring if early treatment with targeted therapy can change the overall disease course.

Seborrhoeic eczema

Clinical features

Seborrhoeic eczema is extremely common in adults and in its mildest form is often overlooked. As the name implies, it affects greasy areas of skin because of its high density of sebaceous glands. On the face, it presents with greasy yellow-looking scale and erythema around the nose and nasolabial folds (Fig. 45.12), medial eyebrows, scalp, and retroarticular folds. Itch is variable. Dandruff is thought to be a mild variant of scalp seborrhoeic dermatitis. More severe scalp involvement can look like psoriasis (sebopsoriasis). The presternal area may be affected, particularly in males, as well as upper back, axillae, and anogenital area. A generalized form of seborrhoeic eczema presents as erythroderma in older adults. Children are not affected until puberty. Cradle cap and a non-itchy napkin dermatitis may represent infantile variants of seborrhoeic eczema.

Aetiology

The precise cause of seborrheic eczema is not known. It is thought that a lipophilic commensal yeast, *Malassezia* spp., triggers the inflammatory skin changes of seborrhoeic dermatitis in susceptible individuals. Host immunity is also involved, and seborrhoeic dermatitis is one of the earliest skin manifestations of human

Fig. 45.12 Seborrhoeic dermatitis. Note the hypopigmented plaques over seborrhoeic sites. (*From James WD, et al.* Andrews' Diseases of the Skin, *12th edn. Philadelphia: Elsevier; 2016.*)

Fig. 45.13 Chronic atopic dermatitis. (A) Hand eczema. **(B)** Thick eczematous plaques with lichenification and excoriation on the dorsal hands and wrists. (*B, Courtesy of Julie V. Schaffer, MD.*)

immunodeficiency virus (HIV) infection (see Chapter 44). Its prevalence is also increased in those with Parkinson's disease and other neurological disorders.

Management

Seborrhoeic dermatitis usually runs a chronic relapsing course, and treatment is suppressive rather than curative. Stress and cold weather are frequent triggers. Topical azole antifungal creams, with short-term use of mild- to moderate-potency steroids according to body site, are usually helpful. TCIs are an effective option but are unlicensed for this form of dermatitis. Ketoconazole-containing shampoo is useful to treat scalp involvement and as maintenance therapy.

Venous eczema (varicose eczema, stasis, gravitational eczema)

Venous eczema usually affects older adults and those with varicose veins or a history of venous thrombosis. The inner calf is involved, and there are usually coexistent signs of venous hypertension, including haemosiderin deposition, lipodermatosclerosis, and varicose ulceration. The eczematous changes range from mild erythema and scaling to an acute exudative inflammatory rash. It may be complicated by allergic contact dermatitis to medicated creams or bandages, and patch tests should be done if there is an inadequate response to treatment.

Management

This includes use of bland (perfume-free) emollients such as a 50:50 liquid paraffin:white soft paraffin mix and short-term prescription of a moderately potent topical steroid. Underlying venous hypertension should be managed with compression hosiery or vascular surgery. Caution should be exercised over prolonged used of paraffin-containing emollients because they pose a fire hazard by impregnating clothes and bedding, making them highly flammable.

Discoid eczema (nummular eczema)

Discoid eczema is characterized by well-demarcated, inflamed scaly patches, sometimes with tiny vesicles. It usually affects the limbs and torso and is intensely itchy, which helps differentiate it from psoriasis. Lesions may be secondarily infected with *S. aureus*. The cause is unknown. Potent topical steroids are usually required to clear individual lesions. Management of severe cases follows general atopic dermatitis treatment protocols as described earlier.

Asteatotic eczema (winter eczema, eczema craquelé)

This form of eczema often affects older people in the wintertime and can be intensely pruritic. It involves the lower legs, lower back, and other areas that have few sebaceous glands. These areas are particularly vulnerable to the drying effects of soap and water. The rash resembles 'crazy paving' with dry scales and inflamed fissures. Frequent use of a bland moisturizer and soap substitute is usually all that is needed.

Hand eczema

Hand eczema (Fig. 45.13) is a common problem that has a significant impact on quality of life. It often runs a chronic relapsing course for

many years, with a reported lifetime prevalence of up to 15%. Different clinical patterns are recognized, and the morphology may be mixed and can vary with time in an affected individual.

Aetiology

The underlying causes are:

- ***Contact dermatitis or eczema*** caused by an external harsh substance ***(irritant)*** or allergy-provoking substance ***(allergen)***
- ***Endogenous dermatitis or eczema*** in which no external factors can be identified. There may also be involvement of the feet.

Hand eczema often has a complex and mixed aetiology because of an inherent endogenous eczema tendency and an additional irritant or allergic component.

Mild ***irritant hand eczema*** typically causes dry, sore skin on the dorsal hands and finger webs. It is extremely common in cold, dry weather and in people who wash their hands frequently, either at home (especially when caring for young children) or in occupations such as catering, healthcare, and hairdressing. Hand eczema is a common finding in people with ***atopic eczema***, and the hand may be the only site involved in adulthood. ***Vesicular hand eczema (pompholyx)*** is characterized by multiple intensely itchy blisters, especially on the lateral fingers and palms, and can similarly affect the feet (Fig. 45.14). ***Hyperkeratotic eczema*** is characterized by dry, scaly plaques and fissures on the palms and soles and shares similarities with psoriasis at these sites. Localized eczema of the finger pulps is called ***pulpitis***.

Severe chronic hand eczema may be associated with increased sickness leave and job loss. It is a considerable financial and societal burden because it impairs the ability to work and to carry out the tasks of daily living. ***Patch testing*** should be considered in all patients with chronic hand eczema to identify relevant contact allergens because it is often not possible to distinguish which patients have an allergic contact dermatitis from the clinical appearances. A careful history of occupation and recreational activities may indicate the need to test with extra allergens beyond the standard battery of common allergens (e.g., specific plants in a gardener or gloves in a healthcare provider).

Management of chronic hand eczema includes use of emollients and moderate to potent topical corticosteroids, minimizing contact with irritants, and avoiding exposure to allergens (Box 45.16).

Fig. 45.14 Pompholyx eczema. (A) Acute pompholyx-like palmar vesicles. **(B and C)** Crusted papules from dried up pompholyx-like vesicles. (*Courtesy of Dr A. Bewley, London.*)

Box 45.16 Common contact allergens

- ***Fragrance*** ('parfum'): in fine fragrances, toiletries, moisturizers, household products
- ***Rubber chemicals:*** in household and examination gloves and footwear
- ***Metals:*** in jewellery; buckles; spectacle frames; and gadgets, including mobile phones
- ***Chemical hair dye:*** in permanent and semi-permanent colorants
- ***Preservative chemicals:*** in cosmetics, toiletries, and household and industrial products
- ***Topical antibiotics and antiseptics:*** prescribed and over-the-counter products
- ***Other ingredients in medicated creams:*** e.g., lanolin, hydrocortisone
- ***Adhesives:*** in footwear, acrylic nails, superglue
- ***Leather and textile dyes***

This may involve wearing protective gloves and sometimes even a change of occupation. The oral retinoid alitretinoin is an effective treatment for topical steroid-refractory severe chronic hand eczema. It leads to clearance or significant improvement in more than half of all patients over a 6-month course. Like all oral retinoids, it is a teratogen, and a pregnancy prevention programme must be followed.

Allergic and irritant contact eczema

Allergic contact eczema is an example of a delayed-type hypersensitivity. The rash does not usually appear until at least 12 to 24 hours after skin contact and may even be delayed by several days, so the cause is often unsuspected. The hands, face, and neck (Fig. 45.15) are commonly involved because of their frequent contact with allergens. Other body sites that are often affected by allergic contact dermatitis include the lower leg, external auditory canal, and anogenital area (from medicaments) and the feet (from footwear). Patch testing should be considered in all patients with suspected allergic contact eczema or when this cannot be excluded.

Irritant contact eczema is a common condition that mainly affects the hands. People with a history of atopic eczema are at significantly increased risk. The most common cause is frequent contact with soap or detergents and water at work ('wet work'). Occupations with increased risk include hairdressing, catering, and healthcare.

Lichen simplex

This is a chronic form of eczema in which the skin is thickened with increased skin markings (lichenified) in response to repeated rubbing or scratching (Fig. 45.16). Common sites are the nape of the neck, lateral calves, and anogenital area. Short-term use of a potent or super-potent topical steroid and topical anti-pruritics (e.g., menthol) often helps to break the ***itch–scratch cycle***, along with habit reversal techniques. There may be underlying emotional stress, hence the alternative name, 'neurodermatitis'.

Nodular prurigo

This is a chronic, widespread, intensely itchy, nodular eruption that is perpetuated by picking and scratching. It develops on a background of chronic pruritus and is often associated with a history of atopic eczema. Scattered, firm, dome-shaped, and hyperkeratotic nodules are typically found on the upper trunk and the extensor surfaces of the limbs. The upper middle back is typically spared (butterfly sign). The diagnosis is made by exclusion of other pathologies and may require a skin biopsy. Medical causes of pruritus should be excluded (see Box 45.23). Topical corticosteroids, sedating antihistamines, and antipruritics have limited effectiveness. Narrowband UVB and methotrexate can be useful for widespread disease. Dupilumab has shown promising results in clinical trials and has been approved in some countries for this indication.

Fig. 45.15 Contact eczema secondary to perfume allergy.

Fig. 45.16 Lichen simplex chronicus. (A) Lichen simplex from chronic rubbing. **(B)** Lichen simplex chronicus of the posterior neck. (*B, Courtesy of Ronald P. Rapini, MD.*)

Chronic actinic dermatitis

This uncommon form of eczema mainly affects older males and is caused by excessive sensitivity to UV radiation in sunlight. The

rash is prominent on sun-exposed sites and worse in the summer, but patients may not be aware of their underlying photosensitivity. A clear cut-off can usually be found between clothed (unaffected) and exposed (affected) sites. Specialist assessment is needed. Monochromator phototesting across the solar spectrum can be done to confirm UVB sensitivity and other causative wavelengths.

Further reading

Eyerich K, Eyerich S, Biedermann T. The multi-modal Immune pathogenesis of Atopic Eczema. *Trends Immunol.* 2015; 36(12):788–801.
Trier A, Kim B. Insights into atopic dermatitis pathogenesis lead to newly approved systemic therapies *Br J Dermatol*. 2023; 188:698–708.

Psoriasis

Psoriasis is a common chronic inflammatory skin disease that affects 2% to 3% of the total population. Various clinical patterns are recognized, and the severity is highly variable. Psoriasis can appear at any age but has two peak ages of onset: early disease (age 20–30 years) is more common and is often associated with a positive family history, and late-onset disease peaks around 50 to 60 years.

Psoriasis is associated with an inflammatory arthritis, metabolic syndrome, cardiovascular disease, inflammatory bowel disease, and autoimmune disorders. Depression and impaired body image are common co-morbidities.

Aetiology

The condition appears to be polygenic, with epigenetic, environmental, and behavioural triggers. Forty percent of individuals with psoriasis or psoriatic arthritis have a positive family history. Twin studies show 73% concordance in monozygotic twins compared with 20% in dizygotic pairs. Multiple genetic psoriasis susceptibility loci have been identified including *PSORS1* (which accounts for 35%–50% of the heritable component).

The pathogenesis of psoriasis is complex and involves an interplay between keratinocytes, immune cells, and other skin-resident cells. Plasmacytoid dendritic cells induce the maturation of myeloid dendritic cells and the production of TNF-α, IL-12, and IL-23, which leads to the activation of the pro-inflammatory T17 axis and secretion of TNF-α, IL-17, IL-21, and IL-22 and keratinocyte activation. This causes overproduction of antimicrobial peptides, inflammatory cytokines, and chemokines, amplifying the immune response and driving abnormally rapid keratinocyte proliferation and abnormal differentiation of the epidermis. The result is thickened, inflamed skin (plaques) with thick overlying scale. Biologic therapy targeting TNF-α, IL-23, and Il-17 have revolutionized psoriasis management.

Exacerbating factors

Multiple medical and behavioral conditions have been identified as risk factors, including smoking, obesity, alcohol use, stress, UV light exposure, and trauma. Drugs including beta-blockers, lithium, and anti-malarial drugs and more recently immune checkpoint inhibitors can exacerbate psoriasis.

Infections including bacterial and viral infection may induce acute flares, and post-streptococcal flares of ***guttate psoriasis*** are well recognized.

Clinical features

Different patterns of psoriasis are recognized, and many can occur together.

Chronic plaque psoriasis is the most common variant and is characterized by pink-red, well-demarcated plaques with silver-white scales, seen especially on extensor surfaces of the knees (Fig. 45.17A) and elbows (Fig. 45.17B). The lower back, ears, and scalp are often also involved. New plaques of psoriasis occur

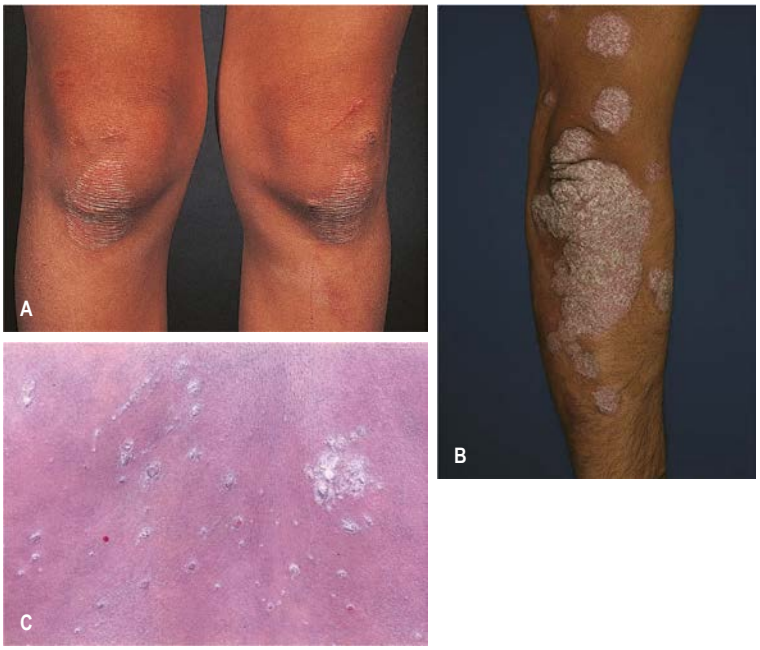


Fig. 45.17 Psoriasis. (A) Chronic plaque psoriasis of the knees. **(B)** Chronic plaque psoriasis on the elbows of an Indian patient. **(C)** Guttate psoriasis in an African patient. (*Courtesy of Dr P Matondo, Lusaka, Zambia.*)

at sites of skin trauma. The lesions can become itchy or sore and bleed easily when scales are removed. The palms and soles can develop painful fissures. The extent of involvement can vary between localized disease and almost full-body surface involvement.

Flexural psoriasis tends to affect older adults with a raised body mass index. It is characterized by well-demarcated, shiny plaques with little, if any, scaling affecting the large flexures of the groin, natal cleft, axillae, and submammary area. The rash is often misdiagnosed as *Candida* intertrigo (see '*Candida* infections' section), but it has a sharp margin and lacks satellite lesions. Anogenital involvement is common and can cause considerable psychosexual upset.

Guttate psoriasis is the term given to a 'raindrop-like' pattern of psoriasis, most commonly seen in children and young adults (**Fig. 45.17C**) with no previous history of psoriasis. The rash develops over days and consists of very small circular or oval plaques on the torso. It is usually triggered by a streptococcal throat infection. Guttate psoriasis resolves spontaneously and in up to one-third of individuals does not recur. However, the remainder experience recurrent guttate attacks or develop chronic plaque psoriasis.

Erythrodermic and pustular psoriasis are the most severe variants and are manifestations of extensive inflammatory skin disease. They may occur together and may be preceded by unstable widespread chronic plaque disease. There is often associated systemic upset with malaise, fever, and circulatory disturbance, which can be life-threatening. The pustules are sterile and can be triggered by sudden withdrawal of potent topical or systemic corticosteroids, pregnancy (impetigo herpetiformis), or infection.

Palmoplantar pustulosis is a more common localized, pustular rash on the palms and soles, and it is controversial whether it is a variant of psoriasis or a separate entity.

Acrodermatitis continua of Hallopeau with involvement of the distal digits is considered a further variant.

Associated features

Nail changes affect up to 50% of people with psoriasis (**Fig. 45.18**). Features include:

- Pitting of the nail plate (most common)
- Distal separation of the nail plate (onycholysis) and, rarely, loss of the nail plate
- Pink- (salmon patch) to yellow brown (oil-drop) discoloration
- Subungual hyperkeratosis (accumulation of scales under distal portion of the nail).

Toenail psoriasis can be difficult to distinguish from nail tinea, and the two diseases may coexist.

Fig. 45.18 Psoriasis of the nail. Yellowish brown discoloration and distal nail plate separation (onycholysis).

Psoriatic arthritis and enthesitis (see Chapter 32, 'Psoriatic arthritis' section) may affect up to 40% of individuals, although the prevalence varies widely. Joint symptoms can present before skin disease, but most individuals have scalp or nail psoriasis. People with psoriasis should be asked specifically about joint symptoms, and the Psoriasis Epidemiology Screening Tool (PEST) score can be used as a guide for rheumatological referral.

Management

There is no curative treatment, and the choice of therapy is determined on an individual basis, according to the site and severity of disease, its psychological impact, co-morbidities such as cardiovascular or liver disease, and the individual's wishes (**Fig. 45.19**). Most people with chronic plaque psoriasis have the condition for life, though it may fluctuate in severity. A supportive and empathic approach is needed because people with psoriasis often feel stigmatized by their skin disease.

Topical therapy

This is given for limited chronic plaque psoriasis, usually as a vitamin D analogue (calcipotriol, calcitriol, tacalcitol) and an emollient to reduce scaling. Moderate or potent corticosteroids may also be used in combination with vitamin D (e.g., calcipotriol with betamethasone) for short-term control. Other topical agents include coal tar and calcineurin inhibitors (tacrolimus and pimecrolimus) for facial and flexural areas. Dithranol is seldom used nowadays because it stains the skin and clothing. Salicylic acid can help remove scales but increases steroid potency when used concurrently.

Topical therapy is usually applied once or twice daily to skin plaques. Vitamin D analogues should be limited to less than 100 g per week because higher doses may cause hypercalcaemia. Topical therapy is often not practical for widespread plaque psoriasis and the next option is usually narrowband UVB phototherapy. Treatment of difficult-to-treat sites is outlined in **Box 45.17**. Topical therapy of nail disease is usually ineffective because of the difficulty of absorption of topical medication at the levels of the nail plate. Systemic medication and biologics can improve nail dystrophy.

Phototherapy

Patients with guttate psoriasis are usually treated with topical therapies, UVB phototherapy, or both. Chronic plaque psoriasis may also respond to phototherapy with UVB or PUVA. Topical PUVA

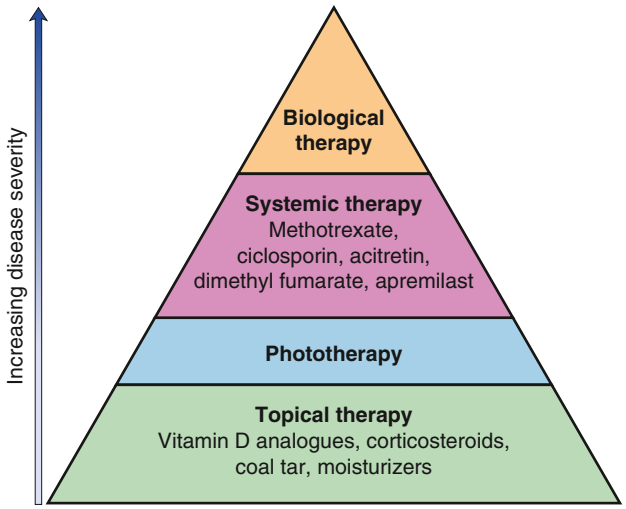


Fig. 45.19 Management of psoriasis.

Box 45.17 Difficult-to-treat sites in psoriasis

- The scalp is considered a 'difficult-to-treat' site because thick scales build up, hindering penetration of topical drugs. Scaling needs removal with keratolytic agents (e.g., silicone lotion or coconut oil and salicylic acid ointment). These can be applied overnight and washed out with a coal tar shampoo.
- Flexural and genital psoriasis requires special consideration because the skin is easily irritated and prone to secondary infection and steroid atrophy. Treatment usually consists of intermittent mild to moderate steroids used alone or in combination with antimicrobial agents. Calcitriol and 0.1% tacrolimus ointment can also be effective.
- Nail psoriasis responds poorly to topical therapy and usually requires systemic drugs.

can be helpful for palmoplantar psoriasis, including palmoplantar pustulosis.

Systemic therapy

Systemic therapy is usually indicated for widespread disease that has failed to respond to topical agents and phototherapy. It may also be needed to manage associated psoriatic arthritis and milder disease with psychological upset. All medication has potential adverse effects and toxicities, and patients should be provided with user-friendly information, such as the British Association of Dermatologists patient information leaflets.

Methotrexate is given orally or by subcutaneous injection in low dose (up to 25 mg) once a week, with folic acid on other days to reduce GI upset. It is a teratogen, so pregnancy must be avoided. Regular blood tests are needed to check liver function and monitor for bone marrow suppression. Measurement of serum procollagen 3 peptide levels and fibrosis-4 (FIB-4) and assessment by liver fibroelastography (see Chapter 23, 'Hepatic stiffness [transient elastography]' section) have evolved as routine non-invasive alternatives to liver biopsy to monitor for hepatotoxicity. Non-steroidal anti-inflammatory drugs (NSAIDs) should be avoided, especially in older adults, because they reduce renal excretion of methotrexate metabolites and may cause toxicity.

Ciclosporin is a selective immunosuppressant that inhibits T lymphocytes. It has been widely used in the acute treatment of patients with severe psoriasis and is a useful drug in unstable or erythrodermic disease because of its rapid onset of action. However, long-term use is often complicated by nephrotoxicity and hypertension and is not recommended.

Acitretin is an oral retinoid used to treat patients with psoriasis and rarer inherited scaly skin diseases (e.g., Darier disease, congenital ichthyosis). It is less effective than ciclosporin or methotrexate but lacks immunosuppressant effects and may be used in combination with UVB and PUVA. All oral retinoids are potent teratogens. Acitretin metabolites have an extremely long half-life, and pregnancy must be avoided for 3 years after completing treatment, so this drug is generally not suitable for females of childbearing potential.

Immunomodulator drugs licensed for use in patients with severe psoriasis include the phosphodiesterase-4 inhibitor ***apremilast*** and ***dimethyl fumarate***. Other drugs include the newly licensed oral selective inhibitor of tyrosine kinase 2 ***deucravacitinib***.

Biological therapy

Advances in understanding the immunopathogenesis of psoriasis have led to development of an expanding and increasingly effective group of targeted biological drugs that block pro-inflammatory cytokines or their receptors. The UK's National Institute for Health and Care Excellence restricts use of these drugs to patients with severe psoriasis (defined by a Psoriasis Area Severity Index [PASI] of >10 and DLQI of >10), in whom phototherapy and conventional systemic drug therapy have failed, are contraindicated, or are not tolerated.

Anti–TNF-α drugs (infliximab, etanercept, adalimumab, certolizumab) were the first group of biologics licensed for use in psoriasis. These have been followed by ***IL-12/23 inhibitors*** (ustekinumab), ***IL-17 inhibitors*** (secukinumab, brodalumab, ixekizumab), and ***selective IL-23 inhibitors*** (guselkumab, tildrakizumab, risankizumab). The more recent allow the majority of patients to achieve a 90% reduction in PASI score (PASI-90), and improvement can be observed within weeks. However, biologic drugs remain expensive despite the launch of several cheaper biosimilars, and they have the potential to lose efficacy over time.

Further reading

Griffiths C, Armstrong A, Gujonsson J, Barker J. Psoriasis. *Lancet.* 2021; 397(10281):1301–1315.
https://pathways.nice.org.uk/pathways/psoriasis. *NICE pathway for management of psoriasis*.

Other common papulosquamous rashes

Pityriasis rosea

Pityriasis rosea is an acute, self-limiting exanthem, which is thought to be caused by a viral infection, possibly human herpesvirus (HHV)-6 and HHV-7. It usually affects older children or young adults, and there is an increased incidence in spring and autumn. The rash is most prominent on the torso and proximal limbs and consists of circular or oval pink macules with a collarette of fine scale. It is usually preceded by a larger solitary 'herald patch' a few days prior. Lesions tend to run along dermatomal lines of the back, giving a 'Christmas tree' pattern. The rash is usually asymptomatic or mildly itchy and resolves within 4 to 8 weeks. Treatment is not usually required, but emollients and a mild steroid cream may be used for pruritus. Severe pityriasis rosea in the first trimester of pregnancy has been associated with fetal loss, and off-label use of aciclovir may be considered.

Polymorphic light eruption

Polymorphic light eruption (PLE) is a sun-induced rash is common in young females and is usually triggered by unusually intense exposure to strong sunshine on holiday. In most cases, it is mild and wrongly attributed to heat ('prickly heat'). Typically, an itchy papular rash develops on sun-exposed areas, particularly the anterior neck, shoulders, and arms, several hours to days after sun exposure and persist for several days. Papules, plaques, and vesicles may occur, hence the term 'polymorphic'. A minority of people with PLE are severely affected and develop the rash in temperate zones, starting in spring or early summer. It is thought to be caused by a delayed-type hypersensitivity to an endogenous antigen that occurs in the skin after UV exposure (photoantigen). Mild cases can be managed by avoiding the sun (with high sun protection factor [SPF] sunscreens and clothing) and gradually building up tolerance to sunlight. Short courses of oral corticosteroids can be given for prophylaxis or treatment. People who are troubled by severe PLE can be desensitized with a short course of UVB phototherapy in early spring, which enables them to tolerate subsequent sun exposure.

Lichen planus

Lichen planus (LP) is a chronic inflammatory skin disease that can affect the skin, scalp, nails, oral mucosa, anogenital area, and oesophagus.

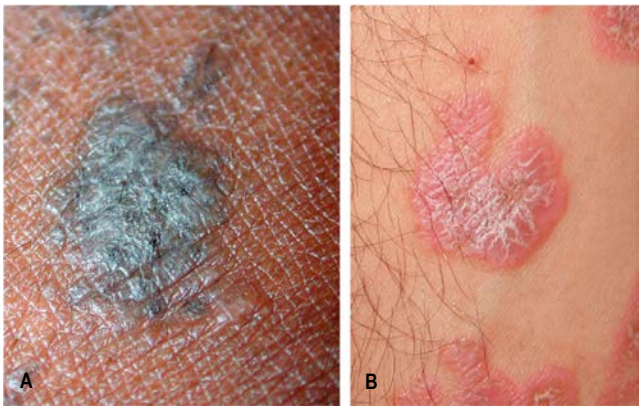


Fig. 45.20 Lichen planus. Lichen planus on pigmented skin **(A)** has a much darker appearance than on lightly pigmented skin **(B)**. Erythema and striae are less obvious. The lesion and surrounding skin become more heavily melanin-pigmented because of the chronic inflammation, producing a very different appearance.

Aetiology is thought to be a an immune-mediated mechanism involving CD8+ T cells against basal keratinocytes with a potential association with the hepatitis C virus (HCV). LP-like (lichenoid) rashes may be caused by drugs (e.g., antihypertensives, antimalarials, NSAIDs) and contact allergens. However, in most cases, there is no identifiable cause.

Clinical features

Cutaneous LP appears as clusters of intensely pruritic, flat-topped, purple-pink (violaceous), polygonal papules (Fig. 45.20) that usually affect the flexural aspect of the extremities such as the wrists, forearms, and lower legs but may be widespread. Close inspection of the papules shows fine white streaks (Wickham striae). Lesions can fuse into plaques, especially on the lower legs, where they become hyperkeratotic. After resolution, marked post-inflammatory hyperpigmentation is common, especially in people with skin of colour. LP may köbnerize, that is, localize to areas of superficial trauma. Scalp and nail involvement may lead to scarring alopecia and permanent nail damage or loss.

Mucosal LP is seen in about 50% of patients with cutaneous LP, and the oral and genital mucosae may also be affected in isolation. Involvement ranges from asymptomatic, reticulate white streaks on the buccal mucosa and lateral tongue (Wickham striae) to a severe, painful, erosive gingivitis and glossitis (erosive LP). Delayed-contact allergy to metals in amalgam fillings may cause localized LP-like changes on the oral mucosa. The anogenital mucosa may be affected by LP and lesions on the penis are often annular. A severe erosive variant can affect the vulva and vagina in females. Genital LP can result in significant scarring.

Diagnosis and treatment

Histology of LP is usually diagnostic and shows a dense infiltrate of lymphocytes along the dermo-epidermal junction (DEJ), which becomes damaged with degenerate (apoptotic) basal keratinocytes. Cutaneous LP usually resolves within 1 to 2 years, though it may relapse. Hypertrophic disease and mucosal, genital, and scalp disease are usually more persistent. Malignant change with development of SCC may rarely complicate chronic ulcerated LP.

Skin and mucosal lesions usually respond to potent or super-potent topical corticosteroids and topical tacrolimus. Widespread skin disease may require tapering courses of oral steroids, phototherapy, oral retinoids, anti-malarials, methotrexate, or systemic immunosuppressants such as azathioprine.

Granuloma annulare

Granuloma annulare usually affects children and young adults. It is often asymptomatic and characterized by clusters of small flesh-coloured or slightly erythematous papules (with no scaling) in an annular shape with a dusky centre. These typically affect the dorsal hands or feet (or both) but can be generalized in 15% of cases. The cause is unknown, but several systemic associations have been reported, including diabetes mellitus and dyslipidaemia and thyroid disease. Localized granuloma annulare is usually self-limiting and resolves after a few years, but potent topical or intralesional steroids can be tried if necessary.

Lichen sclerosus

Lichen sclerosus (LS) is a common chronic inflammatory dermatosis of the genital area. The cause is unknown, but it may be autoimmune and preceding infection or trauma, and an occluded, moist environment may act as triggers.

Clinical features

In females, the usual age of onset is before puberty or after menopause. Males may develop LS at any age. Lesions are intensely pruritic or sore, and appear as shiny, ivory-white, fissured plaques on the vulva, glans penis, distal foreskin, or penile shaft. Additional perianal involvement is common in females, leading to a 'figure-of-8' distribution. Telangiectasia may be evident, and early lesions in young females can present with haemorrhagic blisters that are occasionally mistaken for signs of sexual abuse.

The scarring and atrophy that accompany longstanding LS cause changes in the vulval structure, with loss and fusion of the labia minora and scarring of the clitoral hood. Involvement of the foreskin can cause phimosis, and urethral disease may cause strictures and impaired micturition. Profound sexual dysfunction can result from LS in both males and females.

Management

The diagnosis is usually made by the clinical appearance. Histology shows thickened (hyalinized) dermal collagen and a thin, flat (atrophic) epidermis.

Short-term courses of potent or super-potent topical corticosteroids improve the signs and symptoms of LS. TCIs may also be helpful. The condition may remit spontaneously after several years, especially in children. Unresponsive phimosis may necessitate circumcision. Longstanding anogenital LS is associated with an increased risk of SCC in affected sites.

Further reading

Lewis FM, Tatnall FM, Velangi SS et al. British Association of Dermatologists' guidelines for the management of lichen sclerosus 2018. *Br J Dermatol.* 2018; 178:839–853.

Urticaria and angio-oedema

Urticaria (hives, 'nettle rash') is common (10% of the population will have urticaria at some stage in their life) and is characterized by short-lived (<24 hours), itchy swellings or 'weals' with a surrounding redness (flare) (Fig. 45.21). The short-lived nature of individual swellings and lack of skin surface changes help distinguish urticaria from other inflammatory rashes, especially eczema.

Fig. 45.21 Urticaria.

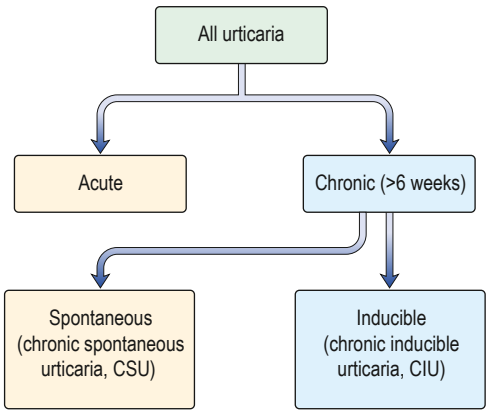


Fig. 45.22 Classification of urticaria. *CSU,* Chronic spontaneous urticaria.

Angio-oedema is a similar disorder involving deeper tissues and usually affecting the lips, tongue, and eyelids. Urticaria and angio-oedema may occur together or separately and have a range of causes. Angioedema by itself may be a feature of chronic spontaneous urticaria or may be a side-effect of angiotensin-converting enzyme (ACE) inhibitor or of an underlying hereditary angioneurotic oedema (HAE).

Aetiology and clinical features

Urticaria and angio-oedema with weals are caused by dermal mast cell degranulation and release of a range of pro-inflammatory mediators (including histamine), together with release of neuropeptides from cutaneous nerves, which together cause vasodilation and increased capillary permeability. Mast cell degranulation may be triggered by various stimuli, including ***drugs*** (opiates, aspirin), ***physical triggers*** (friction, pressure, sweating), ***allergens***, and ***autoantibodies***. Urticaria is classified as acute or chronic (>6 weeks' duration). Chronic urticaria is divided into spontaneous and inducible types (Fig. 45.22).

Acute urticaria (urticaria lasting up to 6 weeks) is often triggered by infections or (more commonly an allergic response to a foodstuff, e.g., peanuts). It may also be the presenting feature of immediate (type 1) allergy, especially in young children with atopic eczema. Common causes include food (nuts, egg, milk), drugs (penicillin), and natural rubber latex. Allergic urticaria may be localized to an area of skin (contact urticaria), such as caused by latex gloves, or take the form of a more widespread reaction that can evolve to anaphylaxis (see Chapter 11, 'Anaphylaxis' section). Measurement of allergen-specific IgE or skin prick tests and allergen challenge tests can be done to confirm an allergic cause.

Chronic inducible urticaria has a range of causes, including friction (symptomatic dermographism), pressure (delayed pressure urticaria), cold (cold urticaria), sunshine (solar urticaria), and sweating (cholinergic urticaria). Physical tests, such as lightly scratching the skin for dermographism or an ice cube test for cold urticaria, are helpful in demonstrating the cause.

Chronic spontaneous urticaria (CSU) is thought to have an underlying autoimmune basis, and in some individuals, functional autoantibodies against the high-affinity IgE receptor on mast cells and basophils or against IgE antibodies can be identified. There is also an association with autoimmune thyroid disease. Chronic urticaria is seldom caused by food allergy, and food allergy tests are not indicated.

Angio-oedema without weals is classified into hereditary and acquired variants:

- ***HAE*** is a rare autosomal dominant condition that can cause recurrent and severe swellings of the skin, upper airways, and intestinal tract. Laryngeal involvement may be life-threatening. Attacks usually start in childhood and can be spontaneous or triggered by minor trauma. Skin swellings are not itchy and usually last for several days. HAE types I and II are caused by mutations in the *SERPING1* gene that result in decreased plasma levels of functional C1 esterase inhibitor (C1NH). This allows unchecked activation of the complement cascade. Serum complement C4 can be measured as a screening test. C1NH activity is normal in type III HAE. All forms of HAE are caused by increased levels of bradykinin.
- ***Acquired angio-oedema*** may be idiopathic or associated with lymphoproliferative or autoimmune disorders. Drugs can also cause angio-oedema, especially ACE inhibitors and angiotensin II receptor antagonists. The rate is up to four times higher in Afro-Caribbean patients, and the onset may be delayed by months or years after starting therapy, so a careful drug history is essential.

Management

Acute urticaria is usually self-resolving. Allergic urticaria and angio-oedema may require emergency treatment with intramuscular adrenaline (epinephrine) and intravenous steroids (see Chapter 6, 'Allergic disease [immediate hypersensitivity]' section). ***Chronic urticaria*** is managed with nonsedating antihistamines such as loratadine 10 mg/day; the dose may be increased fourfold if needed for symptom control. Montelukast may also be helpful in those with aspirin sensitivity. NSAIDs and opiate analgesics should be avoided because they can aggravate urticaria. The biological anti-IgE drug omalizumab is licensed for antihistamine-refractory CSU.

Management of ***HAE*** used to be limited to prophylaxis with attenuated androgens (stanozolol, danazol), which were poorly tolerated, and treatment of acute attacks was with fresh-frozen plasma. Newer selective therapies include plasma-derived and recombinant C1NH, a kallikrein inhibitor (ecallantide), and a bradykinin B2 receptor antagonist (icatibant). These newer products can be used for treatment of acute attacks of HAE, and C1NH is also used for prophylaxis.

Urticarial vasculitis

If urticarial weals are painful and last for more than 24 hours, leaving bruising or a blemish, this may indicate an underlying vasculitis rather than ordinary urticaria. The patient should be thoroughly assessed with blood tests, skin biopsy, and urinalysis.

Further reading

Buss PJ, Christiansen SC. Hereditary angioedema. *N Engl J Med.* 2020; 382:1136–1148.

SKIN INFECTIONS

The skin is the largest organ in the body and is home to an enormous and diverse population of microbial organisms—the skin microbiome. The skin microbiome includes a vast array of bacteria, viruses, and fungi, most of which are commensals (permanent non-pathogenic residents) or transients (temporary residents). Considerable variation in the composition of the skin microbiome exists both across different body sites and between individuals.

Disruption of host–microbiome interactions can cause an organism to transition from commensal to pathogen, causing skin disease, and may occur as a result of alterations in environmental exposures, trauma, changes in host immunity, and host genetic variations. In addition to its role in skin infections, there has been intense interest in recent years in the role of the skin microbiome across a range of other inflammatory and malignant skin diseases.

Bacterial infections

See also Chapter 11 ('Acute skin infections' section).

Impetigo

Impetigo is a common and highly contagious skin infection affecting the superficial epidermis which is typically caused by *S. aureus* or less commonly by group A beta-haemolytic *Streptococcus pyogenes* (or sometimes by both organisms together). Although it is most common in children, impetigo can occur at any age and is transmitted by direct skin contact.

Two forms of impetigo are recognized:

- ***Nonbullous impetigo*** is most common on the face and the extremities. Infected areas initially appear as red macules (which may be more difficult to see in darker skin tones) and pustules evolving into characteristic golden (honey-coloured), crusted, and eroded plaques (Fig. 45.23). It can spread rapidly and may be itchy with associated regional lymphadenopathy, but patients are usually well.
- ***Bullous impetigo*** is caused by exfoliative toxins produced by *S. aureus*, which target intracellular adhesion molecules in the epidermal granular layer, resulting in blister formation. Lesions appear as superficial, thin-roofed blisters that rupture and ooze yellowish fluid, leaving a rim of scale, and patients may have malaise, fever, and lymphadenopathy.

The term 'impetiginization' is also used to describe secondary infection of skin lesions with similar clinical appearances to impetigo.

Fig. 45.23 Impetigo. Yellowish, crusted lesions on the chin with areas of blistering (bullous impetigo).

Predisposing factors include skin diseases such as atopic eczema, skin trauma, immunosuppression, crowded environments, and warm and humid climates. Complications include cellulitis, bacteraemia, staphylococcal scalded skin syndrome, post-streptococcal glomerulonephritis, and post-inflammatory hyper- or hypopigmentation.

Management

The diagnosis is usually based on clinical features, and a skin swab may be helpful if impetigo is recurrent or widespread. Nasal swabs can identify staphylococcal nasal carriage, which may be a source of recurrent infection.

Treatment of localized nonbullous impetigo is with topical antiseptics and topical antibiotics (fusidic acid as first line, mupirocin in cases caused by methicillin-resistant *S. aureus* [MRSA]).

Oral antibiotics are used in widespread (more than 3 lesions) nonbullous impetigo and bullous impetigo and for patients at risk of complications. Flucloxacillin is usually the first-line treatment (or erythromycin or clarithromycin in penicillin-allergic patients). Nasal staphylococcus nasal carriage should be treated in patients with recurrent infections.

Affected individuals should avoid close contact with others (e.g., school or work) until the lesions have crusted over or after at least 24 hours of treatment.

Ecthyma and ecthyma gangrenosum

Ecthyma is considered a deeper form of impetigo and is usually caused by the same bacteria. It presents as chronic, well-demarcated, deep ulcers with overlying, adherent necrotic crust and exudate. Lesions are up to 3 cm in diameter and may heal with scarring. It is most common in children, older people, and immunocompromised people and often affects the lower limbs and buttocks. Predisposing factors include poor hygiene, crowded living conditions, high temperature and humidity, malnutrition, immunocompromise, and the presence of minor skin injuries of other skin conditions. Complications include more widespread infection, scarring, and post-streptococcal glomerulonephritis. Management is with topical antiseptics, topical and systemic antibiotics, improved hygiene, and addressing predisposing factors.

Ecthyma gangrenosum is characterized by distinctive necrotic ulcers with a central thick, black-brown eschar, and surrounding redness that evolve rapidly through initial red macules and subsequent haemorrhagic pustules. It particularly affects axillae and anogenital areas. It is most frequently caused by *Pseudomonas aeruginosa* bacteraemia in an immunocompromised or neutropenic patient, with bacterial invasion of the subcutaneous vasculature causing haemorrhagic occlusive vasculitis. Other bacterial, viral, and fungal pathogens may also be implicated.

Management requires prompt diagnosis of the causal infection for which swabs, blood cultures, and skin biopsy and culture may be needed. Treatment is with the appropriate parenteral antibiotics for the causal underlying bacteraemia.

Cellulitis

Cellulitis describes bacterial infection of subcutaneous tissue and is dealt with elsewhere (see Chapter 11, 'Cellulitis and erysipelas' section). Necrotizing bacterial infection extending into deeper soft tissue structures or muscle (***necrotizing fasciitis*** or ***myonecrosis***; see Chapter 11, 'Necrotizing fasciitis' section) is a clinical emergency with a high mortality rate.

Folliculitis

Folliculitis is a common superficial infection of the hair follicle by a range of pathogens, but non-infective causes are also recognized.

- *S. aureus* is the most common cause of infective folliculitis and presents as itchy or tender small papules and pustules. Predisposing factors include friction, occlusion, humidity, diabetes, immunodeficiency, and underlying skin disease (e.g., atopic dermatitis). A deeper variant affecting hair follicles in the beard area (sycosis barbae) causes multiple papules that may coalesce to form pustule-studded plaques. The main differential diagnosis is pseudofolliculitis barbae, which is caused by ingrowth of curly hairs.
- *P. aeruginosa* folliculitis is caused by inadequately disinfected hot tubs or spas and lesions tend to cluster under swimming costumes.
- *Malassezia (pityosporum)* folliculitis caused by *Malassezia* yeast infection causes papules and pustules mainly on the trunk.
- *Extensive, itchy folliculitis* may also be a manifestation of HIV infection (see Chapter 44).
- *Non-infective folliculitis* may be caused by trauma (e.g., shaving and waxing), chemicals (e.g., coal tar), and occlusion (e.g., emollient use).

Management

Topical antiseptics, topical or oral antibiotics, or antifungals are used in bacterial and *Malassezia* folliculitis, respectively. Management should address underlying predisposing factors.

Boils (furuncles) and carbuncles

Boils (furuncles) are deeper infections of hair follicles and surrounding subcutaneous tissue and appear as painful, red, pus-containing nodules. Multiple boils may coalesce to form carbuncles, which are larger swellings with multiple pus-discharging openings and may cause systemic symptoms. The term 'skin abscess' refers to a collection of pus within the dermis and deeper skin tissue and therefore includes furuncles and carbuncles, which are usually caused by *S. aureus* infection. Factors predisposing to formation of boils include diabetes, metabolic syndrome, iron deficiency, malnutrition, and immunocompromise.

Most people with boils are otherwise healthy but may carry *S. aureus* on their skin or in their upper respiratory tract, and up to 20% of the population are staphylococcal carriers. Person-to-person spread may be facilitated by poor hygiene or overcrowding. Enhanced transmission and increased virulence is also caused by *S. aureus* strains (both MRSA and non-MRSA), which secrete an exotoxin called ***Panton–Valentine leukocidin (PVL)***. PVL *S. aureus* (PVL-SA) has been associated with outbreaks of severe recurrent furuncles, carbuncles, and skin abscesses in healthy adults.

Management

Skin and nasal swabs should be taken in severe or recurrent folliculitis, furuncles, and carbuncles to determine antibiotic sensitivities and to screen for staphylococcal carrier status and PVL-SA strains. Other predisposing causes should be addressed.

Small, isolated boils can be treated with hot compresses alone. Oral antibiotics are required for larger or more widespread infection, and intravenous antibiotics may be required for severe infections. Larger abscesses may require incision and drainage. PVL-positive strains required prolonged therapy with a combination of antibiotics such as rifampicin and clindamycin. Screening of family contacts and decolonization measures (including nasal mupirocin ointment and antiseptic washes, e.g., with chlorhexidine) are recommended as part of prevention in recurrent disease.

Erythrasma and pitted keratolysis

Erythrasma is a common superficial skin infection caused by gram positive *Corynebacterium minutissimum*. Corynebacteria are part of the normal skin microbiome but can also act as pathogens.

Fig. 45.24 Erythrasma. (A) Erythrasma of the axilla. **(B)** Pink fluorescence under a Wood lamp.

Erythrasma presents as well-defined, pink-brown plaques with fine scaling most commonly in the axillae, groin, and toe webs (Fig. 45.24), and it is often misdiagnosed as fungal infection. The diagnosis is supported by the characteristic coral-pink fluorescence of affected areas when examined with Wood lamp light (long wavelength UVA radiation), which is caused by coproporphyrin III released by the bacteria.

Corynebacteria are also implicated in ***pitted keratolysis***. This superficial skin infection appears on weight-bearing plantar skin of the feet as multiple punched-out 1- to 3-mm-diameter pits and maceration that fluoresces coral-pink with Wood lamp examination. There is usually associated malodour. The clinical features are caused by production of protease enzymes by the bacteria, which cause stratum corneum destruction. Predisposing factors include hyperhidrosis, keratoderma, wearing occlusive footwear, diabetes, and immunocompromise.

Management

Patients with erythrasma can be treated with topical antiseptics or antibiotics (fusidic acid and clindamycin) or oral antibiotics (erythromycin or tetracycline). Pitted keratolysis responds to astringents such as potassium permanganate soaks, antiperspirants, and topical imidazoles or fusidic acid combined with adequate foot hygiene.

Borrelia infection: Lyme disease

Lyme disease (see Chapter 42, 'Lyme disease' section) is caused by infection with the spirochaete *Borrelia burgdorferi* transmitted by tick bites. The main hosts for the ticks are animals, including deer, often in grass and woodlands. The skin is one of the most common organs to be affected, but infection can also involve the central nervous system (CNS), joints, heart, eyes, and liver. Depending on the infecting subspecies, *Borrelia* infection may present in the skin in two main patterns:

- ***Erythema chronicum migrans*** occurs in 70% to 80% of cases and is the most characteristic sign of Lyme disease. It appears 1 to 2 weeks after the initial tick bite as a gradually expanding, pink to purple plaque often with central clearing, which is asymptomatic, but may become itchy and associated with flu-like symptoms and lymphadenopathy. It usually resolves after 3 to 4 weeks, but if untreated, infection disseminates and joint, CNS, and cardiac involvement may occur months later.
- ***Acrodermatitis chronica atrophicans*** is the most common skin condition in the late stage of disease, occurring months to years after infection as violaceous discoloration over the dorsum of the hands, elbows, knees, and ankles followed by fibrosis and atrophy. The diagnosis is based on positive serology, consistent

histological features on skin biopsy, and positive *Borrelia* polymerase chain reaction (PCR) results. Treatment is with antibiotics, including doxycycline.

Mycobacterial infections

Leprosy (Hansen disease)

Leprosy (see Chapter 42, 'Leprosy [Hansen's disease]' section) is a chronic bacterial infection usually caused by the intracellular acid-fact bacillus *Mycobacterium leprae*. Worldwide prevalence is more than 5 million; most endemic cases are in India, Myanmar, Indonesia, Brazil, and Nigeria. Spread is probably by inhalation of airborne droplets with an incubation period of 6 months to 20 years. It typically affects the skin and peripheral superficial nerves because of its predilection for cool temperatures. The clinical features depend on the host immune response to *M. leprae*. However, in most cases, the first symptom is numbness, which may precede the appearance of skin lesions by years.

Initial skin lesions are often those of ***indeterminate leprosy***, with a few small, hypopigmented or red, scaly patches with reduced sensation. Biopsy reveals a perineural granulomatous infiltrate and scant acid-fast bacilli. It may resolve spontaneously or progress to tuberculoid or lepromatous types.

In ***tuberculoid leprosy***, there is a strong immune response and consequently fewer organisms (paucibacillary). It presents with a few hypopigmented patches or well-defined red plaques. Sensation is absent within lesions, which are also dry and hairless. Nerves may be enlarged, palpable, and tender. Biopsy shows a granulomatous infiltrate around nerves but no organisms. In ***borderline tuberculoid leprosy***, lesions are more numerous and less well defined, and anaesthesia is less prominent.

Lepromatous leprosy is a multibacillary form caused by a low or limited immune response to *M. leprae*. It presents with nasal stuffiness, discharge, and bleeding with multiple poorly defined hypopigmented and red macules with normal sensation, which progress to inflammatory plaques, nodules, and widespread skin infiltration. A characteristic 'leonine facies' is described with thickening and distortion of the forehead, nose, and earlobes and loss of the eyebrows and eyelashes (madarosis). Skin swelling and thickening and trophic ulceration of the limbs with severe disfigurement may follow. Multiple other organs may also be involved, including the eyes, liver, kidney, testes, and bones. Skin biopsy shows numerous acid-fast bacilli. In ***borderline lepromatous leprosy***, skin lesions are widespread, and symmetrical and sensation and hair growth remain normal, but there is a glove-and-stocking peripheral neuropathy with widespread peripheral nerve involvement.

Skin involvement is also a feature of ***'lepra' reactions*** during treatment. Type I lepra reactions usually occur in patients with borderline leprosy and cause increased erythema, induration, and ulceration of existing lesions and in severe reactions; nerve involvement also occurs. Type 2 reactions (erythema nodosum leprosum) occur in patients with borderline lepromatous leprosy and lepromatous leprosy and is an acute immune complex vasculitis. It causes multiple painful red nodules, sterile pustules, and ulcers and may cause systemic symptoms and affect multiple other organs.

Further clinical details, diagnosis, and management are discussed in Chapter 42 ('Leprosy [Hansen's disease]' section).

Skin manifestations of tuberculosis

Tuberculosis (TB) caused by *Mycobacterium tuberculosis* (see Chapter 20, 'Tuberculosis' section) can cause skin manifestations caused by:

- Cutaneous infection: either direct inoculation or re-inoculation of bacilli into the skin or direct extension or haematogenous spread to the skin from an underling infective focus
- Hypersensitivity reactions to underlying infection, including *tuberculides* and *erythema nodosum.*

A variety of skin lesions are caused by direct cutaneous TB infection of the skin:

- ***Primary cutaneous TB*** is caused by direct inoculation of bacilli into the skin. It usually appears 1 to 4 weeks after a penetrating injury such as tattooing or piercing, often occurs on the face, hands, or legs. An initial painless, firm red papule becomes ulcerated, and may be associated with sporotrichoid spread (along the lines of the lymphatic drainage) and local lymphadenopathy. Rarely it may follow BCG vaccination, particularly in people with STAT1 mutation or on TNF-alpha inhibitors.
- ***Lupus vulgaris*** results from re-infection, particularly on the head and neck. It presents with red–brown nodules that merge into plaques with a gelatinous 'apple jelly' appearance when pressed under a glass slide (diascopy). These heal with scarring, which may be disfiguring, and persistent lesions may predispose to cutaneous SCC (cSCC) 25 to 30 years later in up to 10% of people.
- ***TB verrucosa cutis (warty TB)*** also occurs after further inoculation in people who have previously been infected and have developed good immunity. It appears as brown, red or purple warty plaques often on the hands, feet, knees, elbows, and buttocks.
- ***Orificial TB*** is caused by autoinoculation of bacilli from internal disease, with shallow ulcers appearing at mucocutaneous junctions such as nose, mouth, and perianally.
- ***Scrofuloderma*** occurs because of invasion of the skin from an affected lymph node or bone, particularly around the jaw and neck. Firm painless lesions eventually ulcerate and heal with disfiguring scars and may also predispose to cSCC.
- ***Cutaneous miliary TB*** results from haematogenous spread of TB in patients with an active internal focus of infection. It causes multiple red papules which become ulcerated, necrotic and form abscesses and has a poor prognosis.

Cutaneous hypersensitivity reactions seen in TB infection include:

- ***Tuberculids***, which are skin lesions from which *M. tuberculosis* cannot be isolated as they represent a type III or IV hypersensitivity reaction to an underlying infection. The reaction is probably caused by haematogenous spread of mycobacterial antigens to the skin. ***Erythema induratum (Bazin disease)*** is the most common tuberculid and presents as tender red nodules and plaques on the posterior calves, particularly in females, which may ulcerate and scar.
- ***Erythema nodosum***, not strictly classified as a tuberculid, may also occur as a hypersensitivity reaction to underlying *M. tuberculosis* infection (see 'Erythema nodosum' section).

The diagnosis and treatment of patients with TB are discussed in Chapter 20 ('Tuberculosis' section).

Atypical mycobacterial infections

These are infections caused by mycobacterial species other than *M. tuberculosis* and *M. leprae*. Of more than 30 species, those most often associated with skin and soft tissue infections are *Mycobacterium marinum, Mycobacterium ulcerans* (Buruli ulcer), *Mycobacterium fortuitum,* and *Mycobacterium chelonae*. Skin infection often causes plaques, crusted nodules, and abscesses. Skin biopsy and culture are important for diagnosis.

M. marinum infection ('fish tank' or 'swimming pool granuloma') is one of the more common types and is usually acquired through recreational or occupational exposure with contaminated,

non-chlorinated water, for example, after cleaning a fish tank without gloves. The initial lesion is usually a single painless inflammatory nodule or abscess on the dorsum of the fingers or hands. Further lesions may appear in a sporotrichoid distribution and are more widespread in immunocompromised patients. The diagnosis is usually made from the history and clinical features. A skin biopsy shows granulomatous inflammation, and *M. marinum* may be identified by prolonged culture at low temperature or PCR of skin lesions.

Treatment is with combination oral antibiotics (e.g., clarithromycin and ciprofloxacin) for 8 to 12 weeks, according to microbiological guidance.

Viral infections

Viral exanthems

The most common cutaneous manifestation of viral infections is a widespread maculopapular rash or exanthem that predominantly affects the torso and proximal limbs and is usually associated with systemic symptoms. It is probably caused by deposition of antibody–viral antigen immune complexes within dermal blood vessels. Many viruses cause exanthems, including echovirus, measles, HHV-6, Epstein–Barr virus, and erythema infectiosum (fifth disease) (see Chapter 42, 'Maculopapular viral rashes' section). The rash generally resolves spontaneously in 7 to 10 days.

Herpes simplex virus

Herpes simplex is a common viral infection that causes localized blistering. HSV-1 is mainly associated with orofacial infections and HSV-2 with anogenital infections, but both viruses can affect any mucocutaneous site. HSV remains latent in spinal dorsal root nerves between recurrences. It is transmitted by direct or indirect contact with active HSV and less commonly by asymptomatic shedding in saliva or genital secretions.

Most people are infected with HSV type 1 in childhood. Primary infection is often subclinical but may cause clusters of painful blisters on the face (**Fig. 45.25**) or an ulcerative gingivostomatitis associated with fever and lymphadenopathy. Minor trauma may facilitate inoculation at other sites, such as the cheek and neck in people who play contact sports or the fingers in healthcare workers (herpetic whitlow). After a primary infection, some individuals experience recurrent attacks of HSV at approximately the same site. This usually consists of smaller, umbilicated blisters and 'punched-out' ulcers on a red base that are closely grouped and usually preceded by burning or pain. They heal without scarring, although there may be post-inflammatory hypo- or hyperpigmentation at the site, which often are more apparent in darker skin tones. Factors that may trigger recurrences include illnesses such as upper respiratory tract infections ('cold sores'), UV exposure, menstruation, and emotional stress.

HSV-2 infection (see Chapter 42, 'Herpes simplex virus infection' section) is usually sexually acquired and associated with anogenital ulceration (although HSV-1 causes genital infection in 30% of cases and HSV-2 also causes lesions at non-genital sites such as the buttock and lower back). Infection can result in significant pain and difficulty passing urine, particularly during a primary infection. Recurrences are more frequent with HSV-2, and in immunocompromised patients, recurrences of both HSV-1 and 2 may be more frequent and episodes more severe with chronic ulceration. Other complications of HSV infection include corneal ulceration, eczema herpeticum, and erythema multiforme.

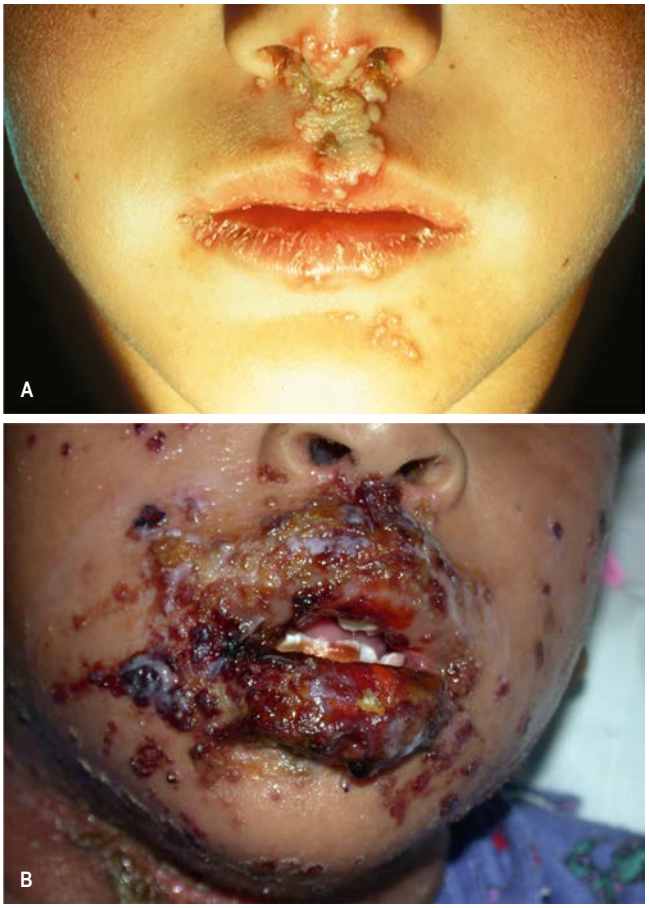


Fig. 45.25 Herpes simplex infection. (A) Labelled primary herpes simplex infection. **(B)** More extensive primary herpes simplex infection with probable bacterial superinfection.

Management

Oral antiviral therapy (e.g., aciclovir, valaciclovir) may be given for patients with severe primary HSV. Topical aciclovir started as soon as burning or discomfort is felt may shorten an episode of recurrence. However, if recurrence is more severe, oral antivirals are required as 'attack-initiated' treatment. Patients with frequent recurrences can be treated with prophylactic low-dose oral therapy. Eczema herpeticum, neonatal infection, and infection in immunocompromised patients require intravenous antiviral therapy with aciclovir or with foscarnet if resistance to aciclovir develops.

Varicella zoster virus

Varicella zoster virus (VZV, HHV-3) causes the common childhood infection varicella (chickenpox; see Chapter 42, 'Varicella [chickenpox] and herpes zoster [shingles]' section) which may reactivate later in life as herpes zoster.

Herpes zoster (shingles)

VZV remains latent in dorsal root ganglia after primary varicella infection. Herpes zoster results from reactivation of latent VZV infection, usually in adulthood. Triggers include infection, injury, radiotherapy, and contact with varicella or herpes zoster. It may be preceded by a prodromal phase of itch, tingling, or pain and may be associated with fever, headache, and feeling unwell. Within 1 to 3 days, a painful, unilateral, blistering eruption appears in the area and is in a characteristic dermatomal distribution (**Fig. 45.26**; see also **Fig. 42.10**). New lesions occur in crops over several days as initial

Fig. 45.26 Herpes zoster in an African male. (*Courtesy of Dr P Matondo, Lusaka, Zambia.*)

red papules, which then blister and may become purulent before crusting. Thoracic, cervical, and lumbosacral dermatomes are most often affected, and ophthalmic dermatomal involvement increases with age. Involvement of dermatomes affecting the mouth, ears, and genital areas can also occur. The rash lasts 2 to 4 weeks and is usually more severe in older adults or immunocompromised patients with involvement of more than one dermatome.

Complications include ocular disease (if the ophthalmic nerve is involved); motor neuropathy (e.g., facial nerve palsy in Ramsay-Hunt syndrome); and disseminated disease involving internal organs (e.g., encephalitis), which is more common in the immunocompromised patients. Severe blistering may leave post-inflammatory pigmentary changes and scarring. The risk of severe, persistent pain or burning sensation more than 1 month after the onset of herpes zoster (post-herpetic neuralgia) increases with age and is more common with facial infection.

Management

Oral antiviral therapy should be started within 72 hours of the onset of the rash and continued for 7 to 10 days to reduce pain, severity, and viral shedding. Parenteral therapy is indicated in immunocompromised patients. General measures include adequate analgesia and antibiotics if secondary bacterial infection is present. Treatment of patients with post-herpetic neuralgia includes topical capsaicin, tricyclic antidepressants, gabapentin, and pregabalin.

Recombinant and live-attenuated shingles vaccines are available to boost immunity against VZV in older adults (in the UK, adults older than 65 years of age) and people who are severely immunocompromised to reduce the risk of developing herpes zoster. Only recombinant vaccines are used in immunocompromised patients because of the risk of infection from live-attenuated vaccines.

Human papillomavirus

There are more than 200 known human papillomavirus (HPV), types and they have specific tropisms for the skin and mucosae. HPV is part of the normal skin microbiome, and infection may be asymptomatic. Viral warts caused by HPV infection are common, benign hyperproliferative mucocutaneous lesions. Persistent mucocutaneous HPV infection is occasionally associated with malignant transformation, particularly in anogenital sites (cervical cancer and some penile, vulval, and anal cancers), in the oropharynx (including some types of tonsillar cancer), and causing some types of cSCCs (including periungual cSCC). Transmission is by direct contact and is facilitated by trauma. Severe mucocutaneous HPV infection may be a feature of immunocompromise.

- ***Common warts*** (verruca vulgaris) are single or multiple papules with a hyperkeratotic, warty (papillomatous) surface, most often seen on the hands and feet and in periungual sites may be painful and cause nail dystrophy (**Fig. 45.27A**).
- ***Plantar warts*** (verruca plantaris) occur on the soles of the feet as endophytic lesions with a papillomatous surface. Small black dots (thrombosed capillaries) may be visible and cause pinpoint bleeding if the skin is pared down, distinguishing them from callosities (**Fig. 45.27B**). Plantar warts may be tender if they affect pressure points and can affect walking.
- ***Mosaic warts*** are clusters of multiple, confluent, more superficial and less tender plantar warts.
- ***Filiform warts*** are warts with elongated, 'digitate' fronds that occur most frequently on the face (**Fig. 45.27C**). Children and adolescents are usually affected. They may also affect the beard area in males and spread by shaving.
- ***Plane warts*** are less common and caused by specific HPV subtypes. They are small, flat-topped papules and may be skin-coloured or lightly pigmented; they are best seen with side-on lighting. They are usually multiple and are frequently found on the face or the backs of the hands and may exhibit the Köbner phenomenon (**Fig. 45.27D**).
- ***Anogenital warts*** are described in Chapter 43 ('Human papillomavirus [anogenital warts]' section).

Epidermodysplasia verruciformis is a genodermatosis (a hereditary condition causing a particular skin phenotype) that causes a predisposition to widespread cutaneous infection with beta HPV types. It is characterized by disseminated hypo- and hyperpigmented plane warts, pityriasis versicolor-like lesions, and red plaques. Progression to cSCC on UV exposed sites may occur. Similar features may be seen in acquired immunocompromise such as HIV infection.

Management

The diagnosis is usually clinical and can be assisted by dermoscopy. Skin biopsy is only required if lesions such as SCC cannot be excluded clinically. In most cases, spontaneous resolution occurs over 12 to 24 months, and treatment is not required. However, warts may persist for months or years, particularly in immunocompromised patients such as solid organ transplant recipients, and can cause significant discomfort and psychosocial effects. There is no available treatment that kills HPV infection, and most available treatments work by stimulating an immune response to the virus. Regular use of a topical keratolytic agents (e.g., 10%–26% salicylic acid) after removal of hyperkeratotic skin may hasten resolution and is the mainstay of treatment. Liquid nitrogen cryotherapy can also be effective but is painful and can cause dyspigmentation, and repeated cycles are usually required. Other destructive treatments include surgery and laser therapy, which may also cause pain and scarring. Treatment of patients with anogenital warts is discussed in Chapter 43 ('Human papillomavirus [anogenital warts]' section).

Molluscum contagiosum

Molluscum contagiosum is a common childhood skin infection caused by a pox virus. Lesions appear as small (1–3 mm), translucent, firm skin-coloured papules with a central indentation (umbilicated) (**Fig. 45.28**). They may have surrounding eczematous changes. They contain soft, white, keratinous matter that can be extruded by squeezing. They often occur in clusters, particularly in flexures; may spread by autoinoculation; exhibit the Köbner phenomenon; and can affect any body site. They are more common in children with atopic dermatitis. Occasionally, larger ('giant') lesions of up to 1 cm in diameter may occur, and widespread infection should raise the possibility of immunocompromise, especially HIV infection. Lesions usually appear in crops over 6 to 12 months. They may become

Fig. 45.27 Viral warts. (A) Common wart (verruca vulgaris). **(B)** Mosaic plantar warts. **(C)** Filiform wart. **(D)** Plane warts: multiple, skin-colored papules. Note pseudo-Koebner phenomenon.The *black dots* represent thrombosed capillaries within the wart. (*B, From Bolognia JL, Jorizzo JL, Schaffer JV.* Dermatology, *3rd edn. St. Louis: Saunders; 2012, with permission.*)

inflamed and crusted before resolving spontaneously and may heal with a pitted scar.

The diagnosis is usually based on typical clinical and dermatoscopic appearances. There is no specific antiviral treatment, and treatment is not usually required, particularly in children. Destructive therapies such as cryotherapy or curettage may be helpful. Potassium hydroxide 5% lotion can be used in younger children. Genital mollusca are usually sexually transmitted in adults.

Orf

Orf is a zoonotic pox virus infection acquired by direct inoculation from infected lambs and goats. Those most often affected include farm workers and children exposed, for example, on city farms. Lesions appear as solitary 1- to 2-cm red-blue papules, nodules, and plaques on the fingers, hands, and forearms with inflamed borders. They evolve into blisters or pustules and may be larger and atypical in immunocompromised people. They may be complicated by lymphangitis and lymphadenopathy. Lesions resolve spontaneously within 4 to 6 weeks, conferring immunity, and although subsequent infection may occur, it is usually less severe. Infection is diagnosed on swab, skin biopsy, or PCR of vesicular fluid. Occasionally, orf is complicated by secondary bacterial infection and erythema multiforme.

Mpox (formerly monkeypox)

Mpox is caused by the MPX orthopox virus. It produces a pustular rash with systemic symptoms similar to, but milder than, smallpox (see Chapter 42, 'Mpox [formerly known as monkeypox]' section). It is endemic in Central and West Africa with increasing outbreaks since 2010 possibly caused by waning of cross-protection from cessation of smallpox vaccination or possible environmental factors changing the interface between humans and the Mpox animal reservoir. In May 2022, a global outbreak was reported, particularly affecting Europe and disproportionately affecting males who have sex with males (MSMs). Spread is particularly associated with close person-to-person or sexual contact.

A prodrome with fever, headache, fatigue, sore throat, myalgia, and lymphadenopathy is followed by the appearance of multiple macules and papules, which evolve into vesicles and pustules. These become umbilicated or centrally eroded and are most prominent on the face but also hands and feet. After 2 to 4 weeks, the

Fig. 45.28 Molluscum contagiosum. Inoculation around the eye, a typical presentation in children.

lesions become crusted and may leave atrophic or pitted scarring. In the 2022 outbreak, lesions were more often localized to the anogenital area, were fewer in number than in previous outbreaks, and were characterized by pseudopustules. These are firm papules with a central ulceration that simulate pustules but are composed of inflammatory keratinocytic debris and not pus. Anorectal and pharyngeal involvement is also common and may be complicated by pain and lymphoedema. Lesions may heal with scarring and can be more severe in immunocompromised people, including those living with HIV.

The diagnosis and treatment of patients with Mpox are discussed in Chapter 42 ('Mpox [formerly known as monkeypox]' section).

SARS-CoV-2 infection and COVID-19

COVID-19 disease is discussed in detail in Chapter 42 ('Coronavirus infections' section).

Mucocutaneous manifestations are experienced by approximately 6% to 8% of those with COVID-19. A spectrum of presentations has been described spanning early viral exanthems to later cutaneous vasculopathies (**Box 45.18**) and in around 20% may be the only sign of infection. Maculopapular or morbilliform, urticarial, and papulovesicular eruptions and vaso-occlusive presentations are seen in older age groups and associated with more severe disease.

- ***Urticarial and morbilliform rashes*** are common, occur early in COVID-19 infection, and are generally associated with less severe disease. They may be red in lighter skin tones but grey or brown in darker skin tones.
- ***Papulovesicular and vesiculobullous rashes*** may be generalized or localized and can simulate varicella but tend to be more truncal and less itchy.
- ***Chilblain-like lesions (pseudo-pernio; 'Covid toes' or 'Covid fingers'):*** Perniosis (chilblains) is an inflammatory disorder in which pink-red-violaceous macules, papules, plaques, or nodules (with less pronounced redness in darker skin tones) present at sites of cold exposure, usually the toes, fingers, and other acral sites such as the ear. They may blister and ulcerate and are associated with pain or a burning sensation. Chilblain-like lesions are seen early in COVID-19, particularly in children and young adults who have otherwise asymptomatic or mild disease. They are thought to be caused by vascular damage by the SARS-CoV-2 virus.
- ***Livedo reticularis*** occurs concurrently or late relative to COVID-19 symptoms and consists of symmetrical, annular, violaceous patches involving trunk and limbs caused by partial or intermittent blood flow reduction to the skin resulting from SARS-CoV-2–associated endothelial cell damage.
- ***Vaso-occlusive skin lesions*** are rare, occur late, and are most often seen in older patients with severe COVID-19, often with systemic thromboembolic complications, and have the highest mortality rate of all COVID-associated skin manifestations.
- ***Multisystem inflammatory syndrome*** may occur in children and adults, and cutaneous manifestations include a polymorphic rash associated with redness and swelling of the hands and feet, conjunctivitis, and mucositis.

Box 45.18 Skin manifestations described in COVID-19

Category	Morphology
Inflammatory or exanthematous	• Urticarial rash • Maculopapular or morbilliform exanthem • Papulovesicular or vesiculobullous exanthem
Vasculopathic	• Chilblain-like acral pattern (COVID-19 toes and fingers) • Livedo reticularis • Purpura or petechiae • Vaso-occlusive (livedo racemosa; retiform purpura; acro-ischaemia; pressure-associated ulceration and necrosis)
Other rashes	• Multisystem inflammatory syndrome (children, adults) • Erythema multiforme–like rash • Pityriasis rosea–like rash • Herpes zoster reactivation • Neutrophilic dermatoses (Sweet syndrome) • Erythema nodosum or panniculitis • Vitiligo • Bullous pemphigoid
Hair and nails	• Alopecia (telogen or anagen effluvium; androgenic; alopecia areata) • Trichodynia • Beau's lines; onychomadesis; red-white discoloration with distal onycholysis
Mucosal	• Oral: Aphthous-like ulcers, angular cheilitis, white plaques, hyperpigmentation • Genital: acute genital ulcers

Many other mucocutaneous disorders have been described in association with COVID-19, including hair and nail disorders (examples are listed in **Box 45.18**), although definitive confirmation of a causative association can be challenging. COVID-19 vaccination has also been linked to a range of mucocutaneous disorders. In addition, exacerbation of pre-existing mucocutaneous disease by SARS-CoV-2 is described for conditions, including psoriasis, atopic dermatitis, alopecia areata, systemic lupus erythematosus (SLE), and bullous pemphigoid.

Fungal infections

Fungal skin diseases are very common human skin infections. In most cases, infection is *superficial* and limited to the stratum

corneum, which elicits no inflammation, or the deeper layers of the epidermis, including the hair and nails, where inflammation may be triggered by the fungus or its products. *Subcutaneous mycoses* include a range of infections of the subcutaneous tissues, usually after traumatic inoculation from environmental sources such as soil and vegetation. *Systemic fungal infections* may result from inhalation or are opportunistic and occasionally because of secondary spread from skin infection, particularly in immunocompromised patients.

Superficial mucocutaneous fungal infections

There are three groups of pathogenic fungi that commonly affect the keratinizing epidermis:

- Dermatophytes
- *Candida albicans*
- *Malassezia* spp. (formerly *Pityrosporum* spp.)

Dermatophyte infections

Dermatophyte (tinea) fungi invade and grow in keratin. The three main genera that affect humans are *Trichophyton*, *Microsporum*, and *Epidermophyton* spp. In the skin, they tend to form an expanding annular plaque caused by progressive peripheral growth, hence the name 'ringworm', and infect hair and nails.

Transmission may be human to human (anthropophilic), from animals (zoophilic), or from the environment (geophilic). The clinical appearances depend on the infecting organism, the body site affected, and the host reaction to infection.

Tinea infections are classified according to body site and include tinea corporis (body), tinea facei (face), tinea barbae (beard), tinea cruris (groin), tinea manuum (hand), tinea pedis (foot), tinea capitis (scalp), and tinea unguium (nails).

Tinea corporis: Tinea affecting the body usually presents with itchy, asymmetrical, scaly, inflamed plaques with central clearing and a scaly, raised border that may be red in lighter skin tones or violaceous to brown in darker skin tones. Occasionally, vesicles or pustules may be seen. Treatment with topical or oral steroids reduces inflammation and pruritus and masks clinical signs while allowing spread of the infection ***(tinea incognito)***. The rash typically flares when steroids are stopped, which often prompts further use of steroids and may lead to widespread infection.

Tinea cruris is more common in males than females and presents with an intensely itchy rash in the groin, with a scaly border that extends on to the thighs (Fig. 45.29A).

Tinea pedis (athlete's foot) (Fig.45.29B) is extremely common in adults and is often confined to the toe webs, where the skin looks white, macerated, and fissured. It often starts between the fourth and fifth toes and may extend more widely on to the soles and lateral borders of the feet, causing dryness, scaling, and erythema ('moccasin feet'), and the toenails are also often affected (see later). Annular lesions are rarely seen on palmoplantar skin. Tinea pedis frequently flares in hot weather, causing pustules or blisters, and this can be misdiagnosed as eczema, including pompholyx eczema.

Fig. 45.29 Dermatophyte infections. (A) Tinea cruris. **(B)** Tinea capitis. **(C)** Tinea pedis. **(D)** Onychomycosis: dermatophyte infection of the nail with white, crumbling dystrophy. (*BC, From Callen et al, 2000.*)

Infection may also spread to the palm ***(tinea manuum)***, especially in manual workers, and causes redness and scaling. ***Tinea manuum*** is often unilateral rather than bilateral and is usually associated with bilateral involvement of the feet.

Tinea capitis (Fig. 45.29C) is the most common dermatophyte infection in young children, especially those of Black African origin, in whom hair and scalp seem more susceptible to fungal invasion. Adults are rarely affected. Dermatophyte infection may be restricted to the hair shaft (endothrix) or spread over the hair surface (ectothrix), where it fluoresces under Wood lamp (UV light). Scalp ringworm is spread by close contact, particularly in schools and households, and by sharing of brushes or clippers.

The majority of UK cases are caused by *Trichophyton tonsurans* (which does not fluoresce under Wood light). Migration has led to changing patterns of fungal infections in Europe (e.g., *Trichophyton tonsurans* from Central America, *Trichophyton violaceum* from India and Pakistan).

The clinical appearance of scalp ringworm varies from mild diffuse scaling with no hair loss (similar to seborrhoeic dermatitis) to the more typical appearance of bald, scaly patches with broken hairs. An increased host response causes pustules and an inflammatory exudate, and certain types of tinea can trigger a severe inflammatory reaction with a swollen purulent mass or 'kerion'. This may be mistaken for a bacterial infection and inappropriately treated with antibiotics alone or even attempted incision and drainage.

Extensive infection is occasionally accompanied by a widespread papulopustular rash most prominent on the trunk. This so-called 'id reaction', which probably relates to the host immune response to the fungus. It resolves when the infection is treated.

Tinea unguium and onychomycosis: 'Onychomycosis' is a broad term for fungal nail infection. Tinea unguium refers to a dermatophyte infection of the fingernails or toenails (Fig. 45.29D). *Trichophyton rubrum* is the most common pathogen.

Tinea toenail infection is a common finding in older adults and is usually asymptomatic. Fingernail infection is less common. Affected nails are dystrophic, thick (subungual hyperkeratosis), and discoloured (white, yellow, or beige). Infection usually starts at the distal or lateral nail edges and then spreads proximally. The whole nail plate may be destroyed in advanced disease. The differential diagnoses include nail psoriasis and traumatic nail dystrophy (which may coexist with fungal infection).

Diagnosis and management

Asymmetrical scaly rashes should be investigated for fungal infection by mycology of skin scrapings. Multiple sites may be affected, and the skin, hair, and nails should be examined. The diagnosis should be confirmed with skin scrapings taken from the active margin or across the affected area. Clippings should be taken from the proximal crumbly, white area of affected nails, and plucked hairs or scrapings for diagnosis of scalp tinea. Specimens are transported in folded paper to keep them dry and prevent bacterial contamination. Microscopy and mycological culture confirm the diagnosis and identify the organism. Molecular diagnostic techniques, including PCR analysis, are gaining in popularity because they can offer high sensitivity and rapid results.

Localized tinea of the body or flexures is treated with an antifungal cream (clotrimazole, miconazole, or terbinafine for 1–2 weeks). Nystatin is ineffective. More widespread infection, tinea pedis, tinea manuum, and tinea capitis require oral antifungal therapy with itraconazole or terbinafine for 1 to 2 months.

Toenail tinea infection requires prolonged oral antifungal therapy. Oral terbinafine for 3 to 6 months can clear up to 80% of cases, but relapses are common. Itraconazole may be given as a continuous or pulsed regimen. Griseofulvin is a less effective drug but is still used for childhood tinea capitis.

Antifungal drug resistance of dermatophyte fungi to both topical and systemic medication is rising globally and should be suspected in extensive, treatment-resistant dermatophyte infection. This appears to be most common with terbinafine because of mutation in the squalene epoxidase gene; resistance has been particularly recognized in the Indian subcontinent and is becoming increasingly prevalent in Europe. Resistance to griseofulvin and triazole antifungals (e.g., itraconazole, fluconazole, voriconazole) is also becoming more common.

Candida infections

Candida is a common species of yeast that causes a variety of skin infections. *Candida albicans* (see Chapter 42, 'Candidiasis' section) is a commensal in the GI tract and vagina. It can overgrow on occluded moist skin, causing nappy rash, intertrigo of the large flexures in obese people (Fig. 45.30), and vulvovaginal candidiasis or 'thrush'. The glans penis may be affected in uncircumcised males (candidal balanitis). Other species of *Candida* may be isolated from vaginal infections and immunocompromised hosts.

Candidal intertrigo causes irritation and soreness; affected areas are glazed and inflamed, with a ragged, peeling edge that may contain a few small pustules. Multiple red papules (satellite lesions) may extend beyond the affected red border.

Candida may also affect the moist interdigital clefts of the toes and mimic tinea pedis. In people who frequently immerse their hands in water (e.g., cleaners, caterers), *Candida* may cause chronic infection of the nail folds (paronychia) and nail infection (candida onychomycosis).

Candida can also infect the oral mucosa, especially after broad-spectrum antibiotics and in immunocompromised people and some denture wearers. Affected mucosal surfaces are inflamed with superficial white or creamy pseudomembranous plaques, which can easily be scraped away. The diagnosis can be confirmed with microscopy and culture of a specimen taken with a routine microbial swab.

Fig. 45.30 Intertrigo with satellite lesions typical of candidiasis.

Management

Predisposing factors should be treated. Diabetes should be excluded. Topical therapy (azoles or nystatin) is usually adequate, except for nail infection, and may take the form of creams, pessaries, lozenges, or powder, depending on body site. Recurrent vulvovaginal candidiasis (see Chapter 43, 'Candidiasis' section) is a common condition in females and may require oral azole antifungal therapy.

Malassezia infection

This lipophilic yeast family (formerly called *Pityrosporum*; is part of the normal skin microbiome. Colonization is prominent at areas rich in sebaceous glands: the scalp, large flexures, and the upper trunk. More than a dozen different species of *Malassezia* have been identified, and there is increasing evidence of their causal role in three common dermatoses:

- Pityriasis (or tinea) versicolor
- Seborrhoeic eczema
- *Pityrosporum (Malassezia)* folliculitis.

Pityriasis (tinea) versicolor is a common condition in young adults, especially in warm climates. It causes widespread pink-beige slightly scaly macules on the torso. Variable pigmentation is characteristic, with pale areas developing after tanning and in individuals with darker skin tones because the yeast impairs melanin synthesis. The diagnosis is usually made clinically but can be confirmed by microscopy of skin scrapings, which shows spherical yeast and short pseudohyphae ('meatballs and spaghetti'). Treatment is with topical azoles or oral itraconazole for resistant cases. Anti-dandruff shampoos containing selenium sulphide or ketoconazole may also be used as a body wash. The pigment changes take months to resolve, and recurrences are common.

Pityrosporum folliculitis is also common in young males and causes small, itchy, monomorphic papules and pustules on the upper back, shoulders, and face and may simulate acne. Topical and oral azole antifungals can be effective, but the condition often relapses quickly.

Subcutaneous mycoses

Subcutaneous mycoses (see Chapter 45, 'Subcutaneous fungal infections' section) are a rare group of localized infections of the skin and subcutaneous tissues that follow traumatic implantation of the fungal agent, often a soil saprophyte. The causative organisms vary geographically and disease examples include sporotrichosis (worldwide), chromoblastomycosis and mycotic mycetoma (tropical and subtropical). Sporotrichosis, or 'rose gardener's disease', usually causes a slowly growing, inflamed nodule at the site of skin inoculation, with new lesions developing along lymphatics and blood vessels (sporotrichoid spread). Treatment is with oral itraconazole.

Parasites and infestations

Scabies

Scabies is an ectoparasite infestation with the mite *Sarcoptes scabiei*. There are approximately 450 million cases of scabies in the world each year, and at any one time, approximately 130 million cases are untreated, representing a huge global burden of disease. It affects people of all races and socioeconomic groups although is more common in resource-poor communities and countries, especially sub-Saharan Africa. Risk factors include social overcrowding, poor hygiene, malnutrition, and immunocompromise. It is most common in children and young adults but can affect any age group. Although scabies is considered a trivial condition, infectious complications include neonatal septicaemia, post-streptococcal glomerulonephritis, and rheumatic fever.

Scabies is spread by close or prolonged physical contact, such as within families, households, and care homes and by sexual contact. The mite survives only a few hours of leaving the host and so spread by fomites (e.g., clothing, bed linen, towels) is uncommon. Sites of predilection for the initial infestation in which the mite burrows into the stratum corneum and lays eggs are the finger webs, palms, soles, flexor wrists, elbows, axillae, male genitalia, and around the nipples and umbilicus. Fine linear or curved burrows a few millimetres long may be seen, though they are not always visible, and dermoscopy may be helpful. Small red papules, vesicles, and occasionally pustules occur anywhere on the body but rarely on the face, except in neonates (**Fig. 45.31A**).

After 4 to 6 weeks, a secondary widespread, intensely itchy, excoriated, and often eczematous rash develops that is triggered

Fig. 45.31 Scabies. (A) Itchy papules and pustules centred on the web spaces of the hand. **(B)** Crusted scabies in a patient with chronic myeloid leukaemia.

Box 45.19 Treatment of scabies

- Treat all skin below the neck, including the genitalia, palms and soles, and under the nails. Treat the head and neck regions in infants (up to 2 years of age).
- Treat all close contacts at the same time *even if asymptomatic.*
- Reapply scabicide to the hands if they are washed during the treatment period.
- Wash or clean recently worn clothes (preferably at 60°C) to avoid re-infection.
- Repeat the application after 1 week.

by a TH1 hypersensitivity response to mite antigens. Pruritus is usually worse at night and disturbs sleep. With chronic infection, skin-coloured, red-brown, or violaceous nodules can, for example, around the axillae and in the male genital area. The rash may be complicated by secondary bacterial infection with pustular lesions, blistering, and impetigo and globally is one of the most common causes of acute glomerulonephritis. Sleep disturbance and psychological distress may be significant.

Diagnosis and management

Scabies is common, and a high index of suspicion is required in patients presenting with new and widespread itchy rashes. The diagnosis is usually based on clinical features (including dermoscopic identification of burrows or mites) and risk factors, particularly contact history. It can be confirmed by demonstrating the mites or eggs or on microscopy of potassium hydroxide–treated skin scrapings from the tip of a burrow, although this is time consuming.

Treatment is based on overnight application of a topical scabicide. For the treatment to be successful, the measures listed in **Box 45.19** should be followed. Five percent permethrin cream or lotion is first-line therapy, and malathion can be used if permethrin is unavailable; 25% benzyl benzoate is use occasionally, but it is an irritant and should not be used in children. Two doses of oral ivermectin, 2 weeks apart, is effective if topical treatment has failed and in communal outbreaks (e.g., residential homes). However, it is unlicensed for this indication. Resistance to topical treatment has been increasingly recognized in recent years, and treatment with ivermectin may be required in this situation. Pruritus takes several weeks to settle after treatment and can be managed with antihistamines, cooling creams such as 2% menthol, and topical steroids for eczematous reactions.

Crusted (Norwegian) scabies

Crusted scabies is a variant in which there is uncontrolled proliferation of mites, and thousands to millions rather than usual 10 to 20 mites are present (**Fig. 45.31B**). Risk factors include increased age, immunocompromise, reduced peripheral sensation (e.g., leprosy), dementia, Down syndrome, and institutional accommodation. Pruritus may be mild or absent. Individuals are highly infectious and may be the source of outbreaks if the diagnosis is delayed. Hyperkeratotic crusted lesions characteristically affect the hands and feet. There may be more widespread inflammatory rash with scaly, crusted plaques resembling infected eczema or psoriasis. Meticulous barrier nursing is needed to protect staff against infestation. Hyperkeratotic scales should be removed with a keratolytic agent before application of a scabicide. Oral ivermectin is often required for effective eradication of all mites.

Lice

Lice are ectoparasites that feed on human blood and cause three patterns of infestation in humans.

Head lice (pediculosis capitis) is the most frequent form of infestation by lice and is extremely common in children, particularly females. Head lice spread by close head-to-head contact, and risk factors include sharing hairbrushes, beds, and clothing. Lice attach to hairs and inject anticoagulant saliva to suck up scalp blood, and they can survive a few days away from the scalp. Infestation causes pruritus, leading to excoriations and papules, particularly around the hairline of the neck and ears. The diagnosis is made by identifying transparent or whitish hatched egg cases ('nits'), which are firmly adherent to the hair shaft, or unhatched brown eggs, which are harder to detect. Adult lice are most easily detected between the teeth of a fine comb and can also be seen on dermoscopy. Complications include dermatitis, impetigo, and lymphadenopathy.

Management aims at eradication, but re-infestations are common. Lotions containing topical insecticides, such as malathion, carbaryl, permethrin, and phenothrin, are now used less widely because of increasing resistance and low cure rates. They are being supplemented or replaced by physical approaches such as application of 4% dimethicone or isopropyl myristate (suffocating agents that kill adults and nymphs) on two occasions 1 week apart and fine combing of wet hair after applying conditioner to remove young lice before they mature. This should be repeated until all eggs have hatched.

Body lice (pediculosis corporis) is less common than head lice, and infestation is associated with poverty and personal neglect, particularly if individuals are unable to wash and wear the same clothes for extended periods. Transmission is by direct contact or by sharing of infested clothing or bedding and is more common in crowded living conditions. The body louse is 2 to 4 mm in size and does not live on the skin but moves there to feed. Lice and eggs are therefore not usually seen on the skin but can be found on close inspection of clothing, particularly seams. Infestation presents with itchy papules and excoriations. Complications include secondary bacterial infections and spread of diseases such as typhus.

Management consists of malathion or permethrin application if infestation is widespread, but environmental measures are the main treatment approach and include high temperature washing and drying or ironing of clothing and bed linen. Topical insecticide spraying of clothing can also be considered.

Pubic lice (phthiriasis pubis) are often referred to as 'crabs' because of the appearance of their physical appearance. They are transmitted by direct contact, usually sexual (see Chapter 43, 'Pediculosis pubis' section). Infestation presents with itching, especially at night. Lice can be seen near the base of the hair with eggs somewhat farther up the shaft. They are most often found in pubic hair, but eyebrows, eyelashes, beard, chest, and thigh hair may also be affected. Live lice can be seen on examination or dermoscopy, and pinpoint bleeding spots associated with louse bites may also be evident.

Management is as for head lice, with use of insecticides; all sexual contacts, even if asymptomatic, should be treated and patients screened for other sexually transmitted diseases.

Papular urticaria (persistent insect bite reactions)

Papulovesicular reactions may develop to bites from airborne insects (e.g., mosquitoes) or fleas *(Cheyletiella)* from an infested animal (particularly dogs and cats). Papular urticaria presents as itchy, urticated papules that may blister and are often grouped in clusters or lines and can leave post-inflammatory hyperpigmentation. Legs and other uncovered areas are most frequently affected sites.

It probably represents an immunological reaction to bites, and some individuals react more severely than others. It is more common in children. Complications include secondary bacterial infection.

The diagnosis is usually based on clinical features, but skin biopsy may occasionally be required. Insect repellents and appropriate clothing help reduce bites from airborne insects. Topical steroids and antihistamines may reduce the size and severity of lesions and associated itching. Animals infested with fleas usually scratch vigorously, leading to scaly, thickened skin. Insect eggs can lie dormant in soft furnishings (e.g., carpets) for many months, hatching when disturbed by vibrations. Anti-flea treatment of the animal and furnishings is therefore required.

Bed bug infestation

'Bed bugs' are small, flat, brown-black, parasitic insects 5 to 15 mm in length from the Cimicidae family (including *Cimex lectulatius* and *Cimex hemipterus*). They are an increasing problem worldwide, particularly in urban populations. They do not live on human skin but emerge from the seams of bedding and bedframes at night when attracted to the warmth and carbon dioxide emissions of a sleeping human body, and they feed on human blood one to two times a week. They are attracted to warm environments; can travel in furniture, luggage, and clothing; and can live up to 12 months without feeding.

Bed bug bite reactions occur in only a subgroup of people as intensely itchy red papules on exposed areas, including the face and neck, and often appear in a linear pattern. They may be more violaceous or brown in people with darker skin tones. A punctum may be seen at the site of a bite, and vesicles or bullae, pustules, urticaria, petechiae, and post-inflammatory hyperpigmentation may occur. Complications include secondary bacterial infection and significant psychological distress.

Eradication is the key to management (see **Box 45.20**). Attempts should be made to identify the presence of bed begs in bedding, mattresses, and so on. Expert professional pest control is usually required, and resistance to pesticides is being increasingly reported. If treatment is required for bed bug bites, topical steroids and oral antihistamines may be helpful.

Tropical dermatoses

Skin diseases that feature among the 'neglected tropical diseases' defined by the WHO include scabies, leprosy, leishmaniasis, dracunculiasis (guinea worm), lymphatic filariasis, onchocerciasis, mycobacterium ulcerans (Buruli ulcer), mycetoma, tungiasis, and yaws. These are endemic in low-income countries and constitute a huge physical and financial burden on societies. These are covered in more detail in Chapter 42.

Box 45.20 Eradication and prevention of bed bugs

- Use of covers to seal mattresses (for at least 12 months).
- Removal of bed bugs with vacuum cleaners incorporating high-efficiency particulate air (HEPA) filters
- Heat and cold treatment: washing, tumble drying, or use of steam cleaners that reach 60°C will usually kill bed bugs, as will temperatures of −20°C.
- Check second-hand furniture for the possible presence of bed bugs before use
- Care to minimize exposure of luggage to bedding, carpeting, and so on while travelling
- Professional pest control companies are often required to eradicate infestations

Travellers who have visited tropical or subtropical countries may also be affected by skin diseases, especially cutaneous leishmaniasis, cutaneous larva migrans, and myiasis. Rashes may also be a feature of systemic tropical infections such as ***dengue, schistosomiasis,*** and ***rickettsial diseases***.

Leishmaniasis (see Chapter 42, 'Leishmaniasis' section) is caused by a protozoon, *Leishmania*, and is acquired from the bite of sandfly vectors. It affects approximately 12 million people worldwide. Cutaneous, mucocutaneous, and visceral disease (kala azar) may occur, and clinical features depend on the infecting organism (which varies geographically—American or 'New World' vs non-American or 'Old World') and host response. ***Cutaneous leishmaniasis*** is the most common form and initially presents as small red papules that enlarge and form a chronic ulcer ('oriental sore'), usually on non-exposed areas such as the face and limbs. The ulcer heals slowly over many months with scarring.

Cutaneous larva migrans is caused by direct contact with the larvae of one of the many types of hookworms, which may be present in faeces from dogs, cats, and other animals. It is usually acquired by walking or lying on sandy beaches. Larvae penetrate the skin—often the feet, hands, knees, and buttocks—and cause an intensely itchy serpiginous track, which migrates as the larva burrows within the epidermis (**Fig. 45.32**). It usually resolves without treatment within 8 weeks. Treatment includes topical thiabendazole, cryotherapy, or laser to destroy the larvae.

Myiasis is an infestation of the skin with developing fly larvae (maggots). Species that can penetrate intact skin to cause boil-like lesions include botflies and tumbu flies.

Human immunodeficiency virus and the skin

HIV infection is associated with a range of infective, inflammatory, and malignant skin diseases. As in the case of other immunocompromised patients, such as organ transplant recipients, people living with haematologic malignancy, and people receiving long-term immunosuppressive medication for immune-mediated inflammatory conditions, these skin disorders in people living with HIV (PLWH) may be severe, atypical,

Fig. 45.32 Cutaneous larva migrans. (*Habif T.* Clinical Dermatology, *6th edn. Oxford: Elsevier; 2016.*)

and recalcitrant (see Chapter 44). Advances in highly active antiretroviral therapy (ART) have dramatically reduced the prevalence of these diseases in recent years in countries with early access to therapy. However, awareness must be maintained because they may be a presenting feature of undiagnosed HIV infection.

Seroconversion rash

This non-specific maculopapular exanthem affects up to 75% of individuals a few weeks after becoming infected with HIV. There may be associated fever, malaise, myalgia, lymphadenopathy, and mouth ulceration (or oral candidiasis). Symptoms usually resolve within a few weeks and are often dismissed as 'flu' and the diagnosis missed. Patients are highly infectious during this time because of very high viral loads.

Cutaneous infections

Susceptibility to bacterial, viral, fungal, and parasitic infections, including opportunistic infections, is increased because of HIV-induced immune deficiency, and infections are often more severe and atypical. Skin biopsy and culture can play an important role in accurately diagnosing some infections and should be considered in any immunocompromised person presenting with atypical rash or skin lesions.

- *Bacterial infections:* Staphylococcal and streptococcal infections may be more severe; atypical mycobacterial infections are more common as are syphilis and bacillary angiomatosis (caused by *Bartonella* spp.)
- *Viral infections:* Many viral infections are seen that are typically more severe or widespread than in immunocompetent patients, including:
  - Molluscum contagiosum, especially on the face, with lesions that are large (>1 cm in diameter) and extensive (a pattern rarely seen in immunocompetent adults)
  - HSV (with increased frequency of recurrences and chronic ulceration)
  - Herpes zoster (severe shingles may be a presenting sign of HIV infection, and disseminated infection is more common)
  - HPV (causing common and anogenital warts and acquired epidermodysplasia verruciformis)
  - HHV-8 (always consider HIV infection in people presenting with Kaposi's sarcoma [KS])
  - EBV (associated with oral hairy leukoplakia; see later)
  - Cytomegalovirus (a cause of anogenital ulceration).
- *Fungal infections:* dermatophyte and candidal infections may occur early and recalcitrant and recurrent oropharyngeal candidiasis is a particular problem with low CD4 counts. Systemic fungal infections with skin manifestations include cryptococcus, aspergillus, and histoplasma (see Chapter 42, 'Systemic fungal infections' section).
- *Parasitic:* leishmaniasis (see Chapter 42, 'Leishmaniasis' section), strongyloides (see Chapter 42, 'Strongyloidiasis' section).

Inflammatory dermatoses

Severe, extensive seborrhoeic eczema is common and may be the presenting sign of HIV infection. Psoriasis is often more severe, may be atypical, and may be treatment resistant in PLWH. Management may require close liaison between dermatologist and HIV physicians, especially in patients with low CD4 counts (<200/mm^3), because of the risk of further immunocompromise if immunosuppressive systemic therapy (e.g., ciclosporin) is required or to prevent harmful drug interactions with ART.

Other inflammatory dermatoses reported to be overrepresented in PLWH include nodular prurigo and granuloma annulare.

Drug rashes

Cutaneous adverse drug reactions are significantly more common in PLWH. Reactions to co-trimoxazole and dapsone (used in *Pneumocystis* prophylaxis) and anti-retroviral drugs appear to be particularly frequent. Drug rashes may be severe (especially with nevirapine and efavirenz), resulting in erythroderma or toxic epidermal necrolysis (TEN; see Chapter 11 'Stevens–Johnson syndrome and toxic epidermal necrolysis' section). Other unusual conditions include nail or mucosal pigmentation from zidovudine, paronychia from indinavir, and lipodystrophy of the face and abdomen or buttocks with protease inhibitors.

Cutaneous malignancy

The risk of malignancy is increased in PLWH because of impaired immune surveillance, and those associated with oncogenic viruses are particularly increased.

- KS (caused by HHV-8 infection; see Chapter 44, 'Kaposi's sarcoma' section) was one of the earliest recognized manifestations of HIV infection and led to the description of the acquired immune deficiency syndrome (AIDS). KS is much more common in MSMs than in other groups of PLWH.
- Cutaneous B- and T-cell lymphomas (EBV) are also more prevalent in PLWH.
- Anogenital premalignancies caused by HPV infection (prostatic [PIN], vulval [VIN], anal [AIN], or cervical intraepithelial neoplasia [CIN]) are more frequent, and the risk of malignant transformation is high and should be assessed by screening.
- Merkel cell carcinoma (caused by Merkel cell polyomavirus or UV exposure) is increased more than 10- to 15-fold, and outcomes are worse.
- Cutaneous keratinocyte cancers (SCC and BCC) are increased 2- to 5-fold and are frequently multiple. Melanoma risk is increased 1.5- to 3-fold.

HIV-specific dermatoses

- ***Itchy folliculitis and HIV folliculitis*** (papular pruritic eruption and eosinophilic folliculitis) are commonly found in association with declining CD4 counts. It presents with intensely itchy papules and excoriations around hair follicles and occurs most commonly over the upper trunk and upper arms (**Fig. 45.33**).

Fig. 45.33 Itchy folliculitis of human immunodeficiency virus infection on the upper trunk and proximal limbs.

- ***Oral hairy leucoplakia*** presents with fixed, white plaques with vertical ridges on the sides of the tongue. Unlike candidiasis, the lesions cannot be scraped off. It was first recognized in HIV disease and associated with low CD4 counts but can rarely occur in other forms of immunosuppression and is thought to be caused by co-infection with EBV.

Immune reconstitution inflammatory syndrome

Several infections can be unmasked or can paradoxically worsen 2 to 4 months after initiation of ART as CD4 lymphocyte counts recover—the immune reconstitution inflammatory syndrome (IRIS). This is most frequently described for anogenital HSV infection, viral warts, tinea folliculitis, and molluscum contagiosum. IRIS affects up to 25% of patients on ART and is thought to reflect the response of a recovering immune system against pathogens that were already present. IRIS may also be responsible for an increase in inflammatory rashes, such as folliculitis and subacute lupus erythematosus.

Further reading

Skin related neglected tropical diseases. https://www.who.int/teams/control-of-neglected-tropical-diseases/interventions/strategies/skin-ntds
Thornhill JP, Barkati S, Walmsley S et al. Monkeypox virus infection in humans across 16 countries—April–June 2045. N Engl J Med. 2022; 387:679–691.
Visconti A, Bataille V, Rossi N et al. Diagnostic value of cutaneous manifestation of SARS-CoV-2 infection. *Br J Dermatol.* 2021; 184:880–887.

SKIN SIGNS OF SYSTEMIC DISEASE

Most skin diseases are isolated conditions, but some are associated with underlying systemic disease and may even be the presenting feature.

Multisystem inflammatory disorders

Erythema nodosum

Erythema nodosum has multiple disease associations (**Box 45.21**). It presents as painful or tender, dusky bruise-like swellings over the shins, which fade over several weeks (**Fig. 45.34**). It is most common in young females. There may be associated arthralgia, malaise, and fever. Inflammation occurs in the deeper dermis and the subcutaneous fat (panniculitis) and can be confirmed on skin biopsy.

Management includes symptomatic relief with NSAIDs, light compression socks and rest. Any underlying cause should be treated. It usually resolves spontaneously, but persistent cases can be treated with anti-inflammatory drugs (dapsone, colchicine) or prednisolone.

Box 45.21 Causes of erythema nodosum

- Streptococcal infection[a]
- Drugs (e.g., sulphonamides, oral contraceptive pill)
- Sarcoidosis[a]
- Idiopathic[a]
- Bacterial gastroenteritis, e.g., *Salmonella*, *Shigella*, *Yersinia* spp.
- Fungal infection (histoplasmosis, blastomycosis)
- Tuberculosis
- Leprosy
- Inflammatory bowel disease
- *Chlamydia* infection

[a]Common causes.

Erythema multiforme

Erythema multiforme is an acute self-limiting rash characterized by distinctive 'target lesions' on the distal limbs, palms, and soles (acral sites). These are concentric rings of erythema with a dusky centre, which can blister. There is variable inflammation or ulceration of the mouth, lips, and other mucosal sites. In erythema multiforme major, mucosal ulceration is severe, and there may be fever and systemic upset (**Fig. 45.35**). Infections, specifically HSV, are the most common triggers. Fewer than 10% of cases are caused by drugs (e.g., anticonvulsants). In 50% of cases, there is no identifiable cause. *Mycoplasma pneumoniae* infection may trigger severe erythema multiforme–like mucosal ulceration with minimal skin changes: *M. pneumoniae*–associated mucositis (MPAM).

Management is supportive, with analgesia and parenteral fluids until dysphagia has resolved. MPAM requires antimicrobial

Fig. 45.34 Erythema nodosum in a patient with human immunodeficiency virus taking co-trimoxazole.

Fig. 45.35 Erythema multiforme major. (A) Target lesions of the palm. **(B)** Mucosal involvement around the mouth.

therapy. Infection can be confirmed by HSV or *M. pneumoniae* serology.

The clinical features of erythema multiforme and SJS are shown in **Box 45.22** and are discussed in more detail in Chapter 11 ('Stevens–Johnson syndrome and toxic epidermal necrolysis' section). The role of systemic steroids in this condition remains controversial.

Pyoderma gangrenosum

Pyoderma gangrenosum (PG) is a rare ulcerating skin disease of unknown cause, often associated with systemic inflammatory conditions, especially inflammatory bowel disease. It may also be associated with inflammatory arthritis and haematological malignancies. It is classified as a neutrophilic disease and shares clinicopathological similarities with autoinflammatory diseases in which there is an infection-like host response in the absence of a microbial trigger.

The most common presentation is with painful inflammatory nodules that rapidly turn into enlarging purulent ulcers (pyoderma) with purple undermined edges. PG may occur spontaneously or postoperatively in a surgical wound or stoma, when it can be misdiagnosed as a wound infection. Skin lesions are rich in neutrophils but histological features are non-specific, so the diagnosis is made on clinical features.

Super-potent topical corticosteroids may help in very mild cases. Systemic therapy with moderate- to high-dose systemic corticosteroids, ciclosporin, or both is usually needed. Anti-TNF therapy, such as infliximab or adalimumab, may be used for refractory cases. The clinical course is unpredictable and unrelated to underlying inflammatory bowel disease activity. PG ulcers heal with mesh-like (cribriform) scars (**Fig. 45.36**).

Sweet syndrome (acute febrile neutrophilic dermatosis)

Sweet syndrome is a rare disorder characterized by fever; malaise; and an acute eruption of tender, red-purple plaques and nodules on the head, neck, and upper body (**Fig. 45.37**). As in PG, the skin lesions contain abundant neutrophils (neutrophilic dermatosis). There is usually no underlying illness, but the syndrome may be associated with haematological malignancy and can be induced by drugs, including granulocyte colony-stimulating factor. It is usually treated with systemic corticosteroids or antineutrophil drugs (dapsone, colchicine).

Behçet syndrome

Behçet syndrome (see Chapter 32, 'Behçet's disease' section) is an inflammatory disorder characterized by recurrent oral ulcers and eye lesions, genital ulcers, or skin lesions. Skin changes include erythema nodosum, papulopustular acne-like lesions, thrombophlebitis, and skin pathergy (prolonged inflammation or pustulation at the site of needle-prick or other minor skin injuries). The syndrome is probably a vasculitis affecting blood vessels of all sizes, and many organs can be involved. Patients with milder disease and skin lesions may respond well to colchicine or dapsone.

Sarcoidosis

Sarcoidosis (see Chapter 20, 'Sarcoidosis' section) is a multisystem granulomatous disorder of unknown aetiology. Skin involvement occurs in 20% to 30% of cases and usually presents with red-brown papules, nodules, and plaques. The nose may be extensively involved ***(lupus pernio)***. Lesions may hypo- or hyperpigmented, especially in people with darker skin tones. Erythema nodosum is sometimes seen in acute sarcoidosis. Although sarcoidosis may be localized to the skin, all patients should be investigated for systemic involvement (see Chapter 20, 'Sarcoidosis' section).

Patients with cutaneous sarcoid respond poorly to topical therapy, but super-potent or intralesional corticosteroids or

Fig. 45.36 Bullous pyoderma gangrenosum. This patient had ulcerative colitis. (*Habif T.* Clinical Dermatology, *6th edn. Oxford: Elsevier, Inc.; 2016.*)

Box 45.22 Clinical spectra of erythema multiforme, Stevens–Johnson syndrome, and toxic epidermal necrolysis

Diagnosis	Skin lesions	Mucosal lesions	Other signs and symptoms
Erythema multiforme (EM)	Three-ring target lesions, often on the hands and feet	EM major only	Recent infection (herpes simplex virus, *Mycoplasma* spp.)
Stevens-Johnson syndrome (SJS)	Scattered macules or blisters scattered over the face, trunk, and proximal limbs (<10% body surface area) Occasional two-ring target lesion	Always	Fever Skin tenderness and pain Recent drug exposure
Toxic epidermal necrolysis (TEN)	As for SJS but >30% body surface area involved	Always	As for SJS Respiratory, gastrointestinal, and genitourinary lesions Hypotension Decreased consciousness
SJS–TEN overlap	As for SJS (10%–30% body surface area involved)	Always	

Fig. 45.37 Sweet syndrome. Lesions are tender erythematous papules and plaques that can also be pseudovesicular or pustular. Lesions tend to be multiple but isolated lesions are seen. (*From Habif TP, Dinulos JGH, Chapman MS et al.* Skin Disease: Diagnosis and Treatment. *Elsevier Inc.; 2018; Fig. 11.33.*)

topical tacrolimus may help. Systemic therapy with antimalarials and methotrexate may also be required for those with extensive disease.

Further reading

M Cem Mat, Aysegul Sevim, Izzet Fresko, Yalcin Tuzun. Becet's disease as a systematic disease. *Clin Dermatol.* 2014; 32(3):435–442.

Pruritus

The pathophysiology of pruritus (itch) is complex and incompletely understood. It may be caused by peripheral mechanisms (as in skin disease), central or neuropathic mechanisms (as in multiple sclerosis), neurogenic mechanisms (as in cholestasis or μ-opioid receptor stimulation), or psychogenic mechanisms (e.g., parasitophobia). Evidence suggests that low stimulation of unmyelinated peripheral C-fibres is associated with the sensation of itch; high stimulation produces pain. Histamine, tachykinins (e.g., substance P), gastrin-releasing peptide, cathepsin, proteases, opioids, nerve growth factors, prostaglandins, and cytokines (e.g., IL-31) may also play a role. In the CNS, μ-opioid receptors may regulate the perception and intensity of itch.

Pruritus is a characteristic feature of many inflammatory rashes, especially atopic eczema, scabies, and LP. It also occurs with asteatotic eczema and cholinergic urticaria, in which the rash may be subtle and overlooked. Pruritus is common in older adults and usually caused by skin dryness (xerosis). ***Generalized pruritus*** without a rash can be caused by a number of medical conditions (**Box 45.23**) or drug reactions. ***Localized pruritus*** may be caused by neurological dysfunction such as in brachioradial pruritus or notalgia paraesthetica.

Management involves use of soap substitutes and symptomatic measures (as for asteatotic eczema). Phototherapy, low-dose amitriptyline, or gabapentin may help in intractable cases. If a pruritic skin disease is not evident, patients should be fully investigated for underlying systemic disease.

Fig. 45.38 Necrobiosis lipoidica on the shin in a patient with type 1 diabetes mellitus.

Box 45.23 Medical conditions associated with pruritus

- Iron-deficiency anaemia
- Malignancy (especially lymphoma)
- Diabetes mellitus
- Chronic kidney disease
- Chronic liver disease (especially primary biliary cholangitis)
- Thyroid disease
- Human immunodeficiency virus infection
- Polycythaemia vera

Endocrine disease

Thyroid disease

Individuals with ***hypothyroidism*** have cold, dry skin and brittle hair with diffuse thinning and loss of the outer third of the eyebrows. There may be a jaundice-like appearance caused by carotene excess with firm, gelatinous (myxoedematous) skin (see also Chapter 34, 'Hypothyroidism' section).

Hyperthyroidism may be associated with hyperhidrosis, onycholysis, and diffuse alopecia. The extrathyroid manifestations of Graves' disease include dermopathy (pretibial myxoedema), which classically presents with non-pitting infiltration; plaques on the shins; and in severe cases, thyroid acropachy, characterized by clubbing and soft tissue swelling of the fingers and toes with periosteal reaction of the bones.

Diabetes mellitus

Patients with diabetes mellitus (see Chapter 35) can have a range of skin manifestations and complications, mediated mainly by hyperglycaemia and hyperinsulinaemia. They include:

- Fungal infection (e.g., candidiasis)
- Bacterial infections (e.g., recurrent boils)

Fig. 45.39 Acanthosis nigricans in the axilla of an individual with a high body mass index.

- Arterial and neuropathic ulcers
- Poor or delayed healing
- Xanthomas

Specific dermatoses of diabetes include:

- ***Necrobiosis lipoidica*** (a patch of spreading erythema over the shin, which becomes yellowish and atrophic in the centre and may ulcerate; Fig. 45.38)
- Diffuse ***granuloma annulare***
- Diabetic stiff skin or ***cheiroarthropathy*** (tight, waxy skin over the fingers with limitation of joint movement owing to thickened collagen).

Cushing syndrome

Skin manifestations caused by high cortisol include hirsutism; central obesity; a full, round face ('moon face'); and dorsocervical fat pad ('buffalo hump'). Florid purple striae distensae (or 'stretch marks') skin atrophy, telangiectasia, and increased susceptibility to folliculitis and candidiasis may occur. Hyperpigmentation also occurs in cases with increased circulating adrenocorticotropic hormone (ACTH; ectopic ACTH syndromes or Cushing disease; see Chapter 34, 'Cushing's syndrome' section).

Striae distensae are caused by ruptures in dermal collagen and elastic tissue. They are a common finding over the abdomen, hips, and breasts in normal female puberty and pregnancy and in obesity and over the lumbosacral area in adolescent males. The cause is unclear, but they are associated with rapid growth and genetic susceptibility. There is no effective treatment, but they fade from pink-purple to pale silvery lines with time. Striae are also seen in those with Marfan's syndrome.

Addison disease

Generalized hyperpigmentation occurs secondary to raised circulating ACTH levels (see Chapter 34, 'Addison's disease' section). This may be prominent in the palmar creases, oral mucosa, lips, large flexures, genitals, and nail beds.

Acanthosis nigricans

Acanthosis nigricans presents with thickened, hyperpigmented, velvet-textured skin around the large flexures (Fig. 45.39). It can appear warty when advanced. The most common and mildest form is associated with obesity and insulin resistance. Late-onset, severe disease with mucosal involvement is usually a paraneoplastic phenomenon caused by underlying malignancy, especially GI tumours. Acanthosis nigricans may also be triggered by niacin therapy. If an underlying cause is present, it should be treated. Some improvement may be achieved with topical tretinoin or a vitamin D analogue such as calcipotriol.

Flushing

Facial flushing in response to emotional stimuli (blushing) is a normal physiological response that can cause embarrassment and may be associated with social phobia. Pathological causes include carcinoid syndrome, other neuroendocrine tumours, and drugs. Treatment of idiopathic flushing includes cosmetic camouflage, beta-blockers, and clonidine. Cognitive therapy and selective serotonin reuptake inhibitors may help with associated depression and anxiety.

Metabolic disease

Hyperlipidaemias

Patients with hyperlipidaemia (see Chapter 19) may present with xanthomas, which are abnormal lipid deposits in the skin. A fasting lipid profile should be checked in all patients, though it may be normal in those with the most common type, ***xanthelasma***, which presents with asymptomatic yellow plaques around the eyes. There are a number of other clinical variants of ***xanthoma***, including:

- ***Tuberous xanthoma*** (firm orange-yellow nodules and plaques on extensor surfaces)
- ***Tendon xanthoma*** (firm subcutaneous swellings attached to tendons)
- ***Plane xanthoma*** (orange-yellow macules often affecting palmar creases)
- ***Eruptive xanthoma*** (numerous small, yellowish papules, commonly on the buttocks).

Porphyria cutanea tarda

Porphyria cutanea tarda (PCT; see Chapter 36, 'Porphyrias' section) is a rare metabolic disorder associated with liver disease; it is usually precipitated by alcohol excess or HCV infection, and 20% of patients have underlying haemochromatosis (see Chapter 23, 'Hereditary haemochromatosis' section). PCT presents with sun-induced blisters, fragile skin, scarring and milia (especially on the dorsal hands), and hypertrichosis.

Management is with repeated venesection to reduce iron overload, alcohol avoidance, and low-dose antimalarial drugs. HCV infection should be treated.

Haematological disease

Anaemia (especially iron-deficient), iron overload, polycythaemia vera, and lymphoproliferative diseases such as lymphoma can present with generalized pruritus. Those with iron deficiency can also present with hair loss, glossitis, angular cheilitis, and brittle or spoon-shaped nails (koilonychia).

Amyloidosis

Amyloid may be deposited in the skin in localized cutaneous forms, including macular amyloid, lichen amyloid, nodular amyloid, and familial primary localized cutaneous amyloidosis, and is most commonly seen in South American, Asian, and Middle Eastern individuals. It may also present in association with underlying systemic

disease. About 50% of patients with systemic amyloid have skin involvement, usually in the context of an underlying monoclonal plasma cell disorder.

Macular amyloid presents with hyperpigmented plaques of 'rippled' grey-tan streaks. The upper back, chest, and upper arms are most involved. Lichen amyloid presents as skin-coloured or hyperpigmented papules that coalesce into plaques often seen on the shins. Amyloid deposits in the oral cavity present with macroglossia and mucosal papules. Deposition in and around blood vessels causes fragility with petechiae and purpura, especially in the periocular skin and large flexures. Multidisciplinary treatment of the underlying disease is required.

Chronic liver disease

Chronic liver disease may present with jaundice, palmar erythema, spider naevi, white nails (leukonychia), onycholysis, hyperpigmentation, and xerosis. Pruritus is common in cholestatic disease, including cholestasis of pregnancy (see Chapter 48).

Chronic kidney disease

End-stage kidney disease (ESKD; see Chapter 26) is commonly associated with intractable pruritus, which can lead to excoriations and nodular prurigo. Treatment involves careful management of serum phosphate levels; optimization of dialysis; and symptomatic therapy, including emollient creams and gabapentin. The selective kappa opioid receptor agonist difelikefalin has been shown to reduce itch in patients receiving haemodialysis.

Reactive perforating collagenosis is seen in up to 10% of patients with ESKD and diabetes and presents with umbilicated papules and nodules with central crusting; skin biopsy shows transepidermal elimination of collagen from the dermis. Pallor, hyperpigmentation, and ecchymoses also occur. It may rarely be associated with non-inflammatory, PCT-like blistering and fragility ***(pseudo-porphyria)***.

Calciphylaxis is a rare, life-threatening disorder caused by vascular calcification and occlusion, generally seen in patients with a long history of chronic kidney disease with suboptimal management of hyperparathyroidism. It is characterized by extremely painful, ulcerating plaques, and the associated mortality rate is high and usually caused by secondary infection (see Chapter 26, 'Calciphylaxis' section).

Nephrogenic systemic fibrosis is a severe, scleroderma-like skin disease that affects people with severe renal impairment after exposure to older gadolinium-containing magnetic resonance imaging (MRI) contrast agents. No cases have been reported in the past 15 years as newer agents have been used.

Kidney transplant recipients are at high risk of skin infections (including atypical and opportunistic infections) and skin cancers (particularly SCCs) because of immunosuppression.

Further reading

Millington GW, Collins A, Lovell CR et al. British Association of Dermatologists' guidelines for the investigation and management of generalized pruritus in adults without an underlying dermatosis 2018. *Br J Dermatol.* 2018; 178:34–60.

Autoimmune rheumatic disease

Dermatomyositis

The rash of dermatomyositis (see Chapter 32, 'Polymyositis and dermatomyositis' section) often includes prominent, periocular lilac discoloration, likened to that of the heliotrope flowering plant (heliotrope rash), and periocular oedema. The rash is usually in a photosensitive, cape-like distribution, and may be pruritic and associated with poikiloderma (reticulate pigmentation, atrophy, and telangiectasia). Purple nodules or plaques appear over the knuckles (Gottron papules; see **Fig. 32.38**) and macular pink-violaceus erythema over the extensor surfaces (Gottron sign). In individuals with darker skin tones, the rash may be hyperpigmented. A scaly scalp with associated pruritus and alopecia may be present. The nail folds and c uticles are ragged with dilated capillaries. Skin involvement only occur in amyopathic dermatomyositis on extensor surfaces. Nail folds and cuticles are ragged with dilated capillaries.

Investigations include serum muscle enzymes, myositis-specific antibodies, MRI scan, muscle biopsy, and electromyography. Skin biopsy changes are non-specific. Myositis-specific antibody tests predict the clinical disease phenotype and include:

- Anti–Mi-2 associated with mainly skin presentation and mild myositis
- Anti-SAE associated with severe skin presentation
- Anti–MDA-5 associated with interstitial lung disease and skin ulceration
- Anti–TIF1-gamma associated with malignancy-associated disease
- Anti-NXP2 associated with pronounced muscle symptoms and calcinosis.

Treatment of patients with skin disease includes photoprotection, topical corticosteroids, TCIs, and the antimalarial drug hydroxychloroquine. Patients with myositis usually require high-dose corticosteroids and steroid-sparing immunosuppressant drugs (methotrexate, tacrolimus, mycophenolate mofetil, azathioprine); intravenous immunoglobulin and biological treatments (including rituximab, anti-TNF, and anti-IL6) may also be used. This is discussed further in Chapter 32 ('Polymyositis and dermatomyositis' section).

Scleroderma

The term ***scleroderma*** refers to a thickening of the skin caused by fibrosis from abnormal dermal collagen, but the term is not a diagnostic entity in itself. Systemic sclerosis (SSc) and morphea both show sclerodermatous changes but are distinct conditions:

- ***SSc*** (systemic scleroderma) has cutaneous and systemic features and is discussed in Chapter 32 ('Systemic sclerosis [scleroderma]' section). Skin changes may occur as ***diffuse cutaneous SSc*** or ***limited cutaneous SSc*****.** The latter was previously referred to as CREST variant (calcinosis, Raynaud phenomenon, esophageal dysmotility, sclerodactyly, telangiectasia), and skin changes may be the presenting feature.
- ***Morphea*** (localized cutaneous scleroderma) is confined to the skin or subcutis and usually presents in children or young adults. It is more common in females, and the cause is unknown. Lesions usually appear on the trunk as bluish-red thickened plaques with shiny white indurated or atrophic centres (**Fig. 45.40**). A severe linear variant in childhood is associated with atrophy of underlying deep tissues and can lead to unequal limb growth or scarring alopecia.

Rarely, scleroderma-like skin changes may occur in patients with chronic *Borrelia* infection (acrodermatitis chronica atrophicans), chronic graft-versus-host disease, and eosinophilic myalgia syndrome (caused by tryptophan therapy).

Patients with skin involvement may respond to potent topical steroids and calcineurin inhibitors, phototherapy, antimalarial drugs, oral (or pulsed intravenous) corticosteroids, methotrexate, and

Fig. 45.40 Morphea. Hyperpigmentation and scarring on the trunk.

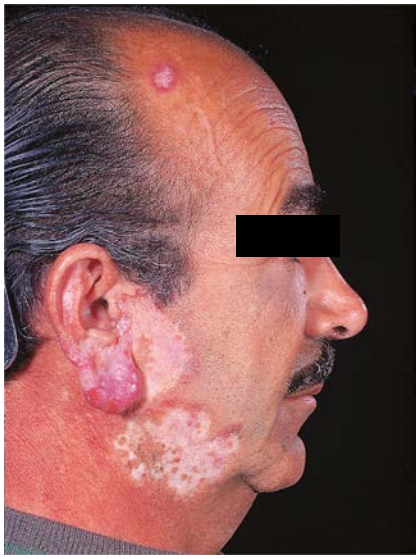


Fig. 45.41 Chronic discoid lupus erythematosus. Scaling, atrophy, and hypopigmentation.

mycophenolate mofetil. Additional systemic treatments (including cyclophosphamide, rituximab, intravenous immunoglobulin, anti–IL-6 drugs) may be indicated for those with severe subcutaneous disease to minimize scarring.

Lupus erythematosus

Lupus is a chronic multisystem autoimmune inflammatory disease (see Chapter 32, 'Systemic lupus erythematosus' section). Skin involvement falls into three broad categories:

- Chronic discoid lupus erythematosus (CDLE)
- Subacute cutaneous lupus erythematosus (SCLE)
- SLE

The cause is unknown, but inherited factors, UV exposure, and drugs may be relevant. All forms of lupus are more common in females.

Chronic discoid lupus erythematosus

CDLE is characterized by chronic red, scaly, atrophic plaques with telangiectasia, especially on the face and other sun-exposed sites (Fig. 45.41). Lesions are often hypopigmented, and close examination may show plugged follicles. Scalp involvement leads to patchy scarring alopecia. Oral involvement (red plaques or ulcers) occurs in 25% of cases. A minority of patients also have Raynaud phenomenon or chilblain-like lesions (chilblain lupus). About 30% of patients have positive antinuclear antibodies, and about 5% develop overt systemic disease.

Skin biopsy shows a dense infiltration of lymphocytes around the hair follicles and damage to the basal layer of the epidermis. Direct IMF may demonstrate granular deposits of IgM and C3 at the DEJ ('lupus band'). CDLE usually lasts for many years, with fluctuations in severity before burning out.

Treatment includes careful photoprotection and high-factor sunscreens, potent topical corticosteroids, occasionally intralesional steroids, and TCIs (tacrolimus, pimecrolimus). Widespread or unresponsive disease and associated arthritis may require antimalarial therapy (hydroxychloroquine, mepacrine), systemic steroids, methotrexate, oral retinoids, andante-proliferative immunosuppressants (mycophenolate mofetil, azathioprine), dapsone, intravenous immunoglobulins, thalidomide, and biologic treatments (rituximab).

Subacute cutaneous lupus erythematosus

SCLE is an uncommon non-scarring variant that presents with widespread annular, papulosquamous plaques on the upper trunk and upper limbs. It typically flares a few weeks after sun exposure. Systemic symptoms, such as arthralgia and mouth ulceration, are seen, but significant organ involvement is uncommon. It may be drug induced, and proton pump inhibitors, terbinafine, calcium channel blockers, and thiazide diuretics are known culprits. Anti-Ro and anti-La antinuclear antibodies are usually positive.

Treatment is similar to that used in DLE.

Systemic lupus erythematosus

In SLE (see Chapter 32, 'Systemic lupus erythematosus' section), organ involvement such as arthritis and nephritis usually predominates, but 80% of individuals will have cutaneous involvement. The classical 'butterfly' rash affects the central face in a rosacea-like distribution (Fig. 45.42). Palmar erythema, dilated nail-fold capillaries, splinter haemorrhages, digital infarcts, purpura, and livedo reticularis (see later) may be evident because of vascular inflammation or occlusion. CDLE and SCLE may overlap.

Treatment is described in Chapter 32 ('Systemic lupus erythematosus' section).

Disorders of blood vessels and lymphatics

Vasculitis

Vasculitis (see Chapter 32, 'Systemic inflammatory vasculitis' section) is an inflammatory disorder of blood vessels that causes endothelial damage. Cutaneous vasculitis may be an isolated problem or part of a multisystem disease with involvement of other organs. The most used classification is based on the size of blood vessel involved (see Box 32.39) and the presence or absence of antineutrophil cytoplasmic antibody.

The features of cutaneous vasculitis are haemorrhagic papules, pustules, nodules, or plaques, which may erode and ulcerate. Lesions are purpuric and do not blanch with pressure. Erythema in a broken livedo or reticulate pattern (see next section) may be seen, with areas of infarction and necrosis. Pyrexia and arthralgia are common associated features. Other clinical features depend on the underlying cause. A thorough clinical evaluation will determine the likely causes and guide further investigations. Routine blood tests and urinalysis should be carried out to exclude renal involvement.

Fig. 45.42 Systemic lupus erythematosus. Macular erythema on the cheeks ('butterfly rash').

Leucocytoclastic vasculitis is the most common cutaneous vasculitis and involves small vessels. It usually affects the lower legs as a symmetrical palpable purpura and is rarely associated with systemic involvement. It can be caused by drugs (15%), infection (15%), inflammatory disease (10%), or malignant disease (<5%), but often no cause is found (55%–60%). Histology of an early lesion confirms blood vessel inflammation (rather than primary occlusion) and indicates the size of the blood vessel involved, as well as the presence of additional diagnostic features such as granulomas. Treatment is with analgesia, support stockings, and oral prednisolone or dapsone to control inflammation and heal ulceration.

Livedo reticularis

The term ***livedo reticularis*** describes a net-like (reticulate), purple-red discoloration of the skin. It usually affects the lower limbs and is more noticeable in the cold but does not disappear on rewarming (unlike physiological livedo). The colour change is because of sluggish blood flow in dermal venules. This may be caused by abnormalities of the blood (high viscosity, coagulopathy, cryoglobulinaemia) or vessel wall (vasculitis such as polyarteritis nodosa or mural calcification). If the network pattern is broken and irregular ***(livedo racemosa)***, underlying systemic disease is likely. ***Livedoid vasculopathy*** is characterized by broken livedo, painful leg ulcers and *atrophie blanche* (see later). It may be associated with antiphospholipid syndrome and SLE.

Leg ulcers

Leg ulcers are common and can have many causes (**Box 45.24**). Venous ulcers are the most common type in high-income countries (**Fig. 45.43**).

Venous ulcers

Venous ulcers are the result of sustained venous hypertension in the superficial veins, usually caused by incompetent valves in the deep or perforating veins or previous deep vein thrombosis. The increased pressure causes extravasation of fibrinogen through the capillary walls, giving rise to perivascular fibrin deposition, which leads to poor oxygenation of the surrounding skin.

Venous ulcers are common in later life, affecting 1% of the population older than the age of 70 years. They are often chronic or recurrent and create a significant burden on healthcare resources.

? Box 45.24 Causes of leg ulceration

- Venous hypertension
- Arterial disease (e.g., atherosclerosis)
- Neuropathic disease (e.g., diabetes, leprosy)
- Neoplastic disease (e.g., squamous or basal cell carcinoma)
- Vasculitis (e.g., rheumatoid arthritis, systemic lupus erythematosus, pyoderma gangrenosum)
- Infection (e.g., ecthyma, tuberculosis, deep mycoses, Buruli ulcer, syphilis, yaws)
- Haematological disease (e.g., sickle cell disease, spherocytosis)
- Drugs (e.g., hydroxycarbamide)
- Others (e.g., necrobiosis lipoidica, trauma, artefact)

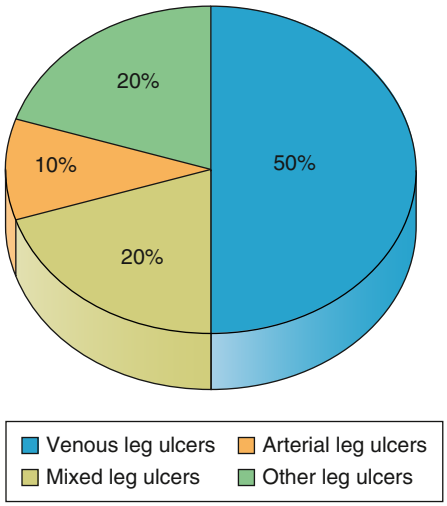


Fig. 45.43 Types of leg ulcers. (*From Griffiths C, Barker J, Bleiker T et al. eds.* Rook's Textbook of Dermatology, *9th edn. Chichester: John Wiley & Sons; 2016.*)

Fig. 45.44 Venous leg ulcer.

Ulceration is most common on the lower leg above the medial malleolus (**Fig. 45.44**). Dilated foot, leg, and thigh veins are usually evident. Other signs include:

- Oedema of the lower legs
- Venous (stasis or gravitational) eczema
- Hyperpigmentation caused by haemosiderin deposition
- Lipodermatosclerosis (fibrosing panniculitis of subcutaneous tissue), a combination of subcutaneous induration, reddish-brown pigmentation, and inflammation

- Atrophie blanche (white atrophy), which includes angular, ivory-white, depressed atrophic plaques with prominent red dots caused by dilated capillaries.

Management is with graduated multilayer (two- to four-layer) compression bandaging and leg elevation to improve venous return. Doppler studies (ankle:brachial pressure index) should be performed before compression to exclude significant arterial disease. Treatment can be delivered effectively in the community by appropriately trained healthcare professionals. Non-adherent dressings are used to keep the ulcer moist. Alternative dressings (alginate, hydrogel, hydrocolloid) may be helpful for heavy exudate and slough.

With adequate compression, up to 80% of ulcers can be healed within 6 months. Slower healing rates occur in patients with poor mobility, bilateral disease, and very large ulcers that have been present for more than 6 months. Diuretics are sometimes indicated to reduce oedema. Microbial swabs and antibiotics are indicated only when there are clinical signs of infection or cellulitis.

Painful venous leg ulcers require adequate analgesia. Split-thickness skin grafting can aid healing in resistant cases. Support stockings (individually fitted) should be worn for life after healing to lessen recurrence. Underlying venous disease should be assessed. Modern vascular surgical interventions may accelerate healing and reduce the risk of ulcer recurrence.

Arterial ulcers

Arterial ulcers present as punched-out, painful lesions on the legs or feet in people with underlying peripheral vascular disease, particularly smokers. The affected leg is usually cold and pale with loss of hair. Peripheral pulses are absent or weak. Doppler ultrasound studies can confirm underlying arterial disease.

Management depends on keeping the ulcer clean and covered, adequate analgesia, and vascular surgical intervention if appropriate. Compression bandaging must be avoided.

Neuropathic ulcers

These ulcers usually occur over pressure areas of the feet, such as the metatarsal heads, owing to repeated trauma. They are usually associated with diabetes and peripheral neuropathy. In some countries, leprosy is a common cause (see 'Leprosy [Hansen disease]' section).

Management involves keeping the ulcer clean and reducing pressure or trauma to the affected area. People with diabetes need meticulous foot care and correctly fitting shoes (see Chapter 35, 'Management of diabetes-related foot problems' section).

Pressure ulcers (decubitus ulcers, bedsores)

Pressure ulcers or sores are a severe manifestation of pressure-induced skin and soft tissue injury. They affect older, immobile, unconscious, and paralysed patients and are caused by skin and soft tissue ischaemia from sustained pressure over a bony prominence, most commonly the heel and sacrum.

Healthy individuals feel the discomfort of continued pressure, even during sleep, and relieve pressure by changing body position.

There are numerous risk factors for the development of pressure ulcers (**Box 45.25**). The majority arise in hospital, with approximately 70% appearing in the first 2 weeks of hospitalization and 70% occurring in patients with orthopaedic disorders, especially those receiving traction.

Pressure injuries may be graded and range from non-blanchable erythema of intact skin (stage 1) to deep ulcers extending to the bone (stage 4). Deep ulcers are associated with a high mortality rate. The early sign of red-blue skin discoloration can lead rapidly to ulcers within hours. Leaving patients on hard emergency room trolleys or sitting them in chairs for prolonged periods must be avoided.

Box 45.25 Risk factors for the development of pressure ulcers

Prolonged immobility
- Paraplegia
- Arthritis
- Severe physical disease
- Apathy
- Surgery and postoperative states
- Plaster casts
- Treatment on the intensive care unit

Decreased sensation
- Coma
- Neurological disease
- Diabetes mellitus
- Drug-induced sleep

Vascular disease
- Atherosclerosis
- Diabetes mellitus
- Scleroderma
- Vasculitis

Poor nutrition
- Anaemia
- Hypoalbuminaemia
- Vitamin C or zinc deficiency

Box 45.26 Management of pressure ulcers

- Ensure bed rest with pillows and fleeces to keep pressure off bony areas (e.g., sacrum and heels) and prevent friction.
- Provide special pressure-relieving mattresses and beds.
- Maintain regular turning and avoid pressure on hips.
- Provide air-filled cushions for patients using wheelchairs.
- Provide adequate nutrition.
- Apply non-irritant occlusive moist dressings (e.g., hydrocolloid).
- Prescribe adequate analgesia (e.g., may need opiates).
- Consider plastic surgery (debridement and grafting in selected cases).
- Treat the underlying condition.

Management

See **Box 45.26**.

Prevention

A risk assessment should be undertaken on newly admitted patients. The best-known risk assessment tools are the Norton Scale and Waterlow Pressure Ulcer Risk Assessment, which yield a numerical score (**Box 45.27**). Specialist tissue viability nurses help assess risk, deliver care, and train other clinical staff.

Lymphatic disease

Lymphoedema

Impaired lymphatic fluid drainage leads to chronic non-pitting oedema with impaired tissue immunity and increased fat deposition. It most commonly affects the legs and tends to progress with age. It is associated with an increased risk of cellulitis, which often becomes recurrent. The subcutaneous tissues become thickened and indurated, and the overlying skin develops a warty texture. Affected limbs may become grossly enlarged.

Box 45.27 Pressure ulcer risk assessment tools

Norton Scale for pressure ulcers[a]

Physical		Neurology		Activity		Mobility		Incontinence	
4	Good	4	Alert	4	Ambulant	4	Full	4	None
3	Fair	3	Apathetic	3	Walks with help	3	Slightly	3	Occasionally incontinent
2	Poor	2	Confused	2	Not bound	2	Limited	2	Usually incontinent
1	Very poor	1	Stupor	1	Bedfast	1	Very limited; immobile	1	Doubly incontinent

Waterlow Pressure Ulcer Risk Assessment[b]

Build/weight for height		Visual skin type		Continence		Mobility		Sex/Age		Appetite	
Average	0	Healthy	0	Complete	0	Fully mobile	0	Male	1	Average	0
Above average	2	Tissue paper	1	Occasionally incontinent	1	Restricted or difficult	1	Female	2	Poor	1
Below average	3	Dry	1	Catheter or incontinent of faeces	2	Restless or fidgety	2	14–18	1	Anorectic	2
		Oedematous	1	Doubly incontinent	3	Apathetic	3	50–64	2		
		Clammy	1			Inert/ traction	4	65–75	3		
		Discoloured	2					75–80	4		
		Broken or spot	3					81+	5		

Special risk factors		Assessment value	
1. Poor nutrition, e.g., terminal cachexia	8	At risk	10
2. Sensory deprivation, e.g., diabetes, paraplegia, cerebrovascular disease	6	High risk	15
3. High-dose anti-inflammatory or steroid in use	3	Very high risk	20
4. Smoking 10+ per day	1		
5. Orthopaedic surgery or fracture below waist	3		

[a]Low scores carry a high risk.
[b]See assessment value, bottom right of table.

Lymphoedema can be primary (and present early in life) and caused by an inherited abnormality of lymphatic vessels (e.g., Milroy disease) or can be secondary and caused by obstruction of lymphatic vessels (e.g., filarial infection or malignant disease).

Management is with compression stockings and physical massage. Causes of high central venous pressure, low albumin, and local inflammation should be treated. Long-term antibiotics are indicated for patients with recurrent cellulitis (see Chapter 11, 'Cellulitis and erysipelas' section) because each episode of cellulitis causes further damage to the lymph vessels. Surgery should be avoided.

Further reading

National Institute for Health and Care Excellence. NICE clinical guideline NG152: Leg ulcer infection: antimicrobial prescribing. NICE 2020.
National Institute for Health and Care Excellence. *Clinical Knowledge Summary: Leg Ulcer—Venous*. NICE 2019; https://cks.nice.org.uk/leg-ulcer-venous.
National Institute for Health and Care Excellence. *NICE Clinical Guideline 179: Pressure Ulcers: Prevention and Management*. NICE 2014; https://www.nice.org.uk/guidance/CG179.

Systemic malignant disease

Certain skin disorders may be a non-metastatic manifestation of an underlying malignancy; these are the ***paraneoplastic dermatoses*** (Box 45.28). Rarely, internal malignancies can metastasize to the skin, where they usually present as papules or nodules that may ulcerate.

Genetic disease

Neurofibromatosis type 1

Type 1 neurofibromatosis (von Recklinghausen disease) is an autosomal dominant condition caused by mutations in the *NF1* gene on chromosome 17. It often presents in childhood with more than five ***café-au-lait*** macules (light brown, well-demarcated, round or oval macules >5 mm in diameter) and axillary freckling. Learning difficulties and skeletal dysplasia can occur. Lisch nodules in the iris may be seen on slit-lamp examination. Neurofibromas appear in adulthood as fleshy skin polyps and smooth, soft dermal tumours, and can be extensive. Several endocrine disorders are rarely associated, including

Box 45.28 Non-metastatic cutaneous manifestations of underlying malignancy

Dermatosis	Tumour
Dermatomyositis	Lung, gastrointestinal tract, genitourinary tract
Acanthosis nigricans	Gastrointestinal tract, lung, liver
Paget disease of the nipple (localized eczema-like changes on nipple and areola)	Ductal breast carcinoma
Erythroderma	Lymphoma, leukaemia
Tylosis (thickened palms/soles)	Oesophageal carcinoma
Tripe palms (velvet palms)	Pulmonary, gastric
Ichthyosis (dry flaking of skin)	Lymphoma
Sign of Lesar-Trelat (sudden eruption of multiple seborrheic keratoses)	Lymphoproliferative, gastrointestinal
Erythema gyratum repens (concentric rings of erythema, which change rapidly)	Lung, breast
Necrolytic migratory erythema (burning, geographic and spreading annular areas of erythema)	Glucagonoma
Paraneoplastic pemphigus	Lymphoproliferative
Pyoderma gangrenosum	Myeloproliferative
Acute febrile neutrophilic dermatosis (Sweet syndrome)	Lymphoproliferative

phaeochromocytoma, acromegaly, and Addison disease (see Chapter 34, 'Addison's disease' section).

Tuberous sclerosis

Tuberous sclerosis complex (TSC) is an autosomal dominant multisystem disorder characterized by a range of benign hamartomas in multiple organs. It is caused by mutations in either *TSC1* gene on chromosome 9q, which produces the protein hamartin (10%–30%), or TSC2 on chromosome 16p, which produces tuberin. The main clinical features are learning and behavioural difficulties, epilepsy, and skin lesions:

- Angiofibromas: pink or red papules around the nose and cheeks
- Periungual fibromas: nodules arising from the nail bed
- Shagreen patches: connective tissue naevi present as flesh-coloured plaques on the trunk
- Ash-leaf hypopigmentation: pale macules best seen with a Wood UV lamp
- Forehead plaque: indurated, flesh-coloured plaques
- *Café-au-lait* macules

Hamartomas can also arise in the heart, lung, kidney, retina, and CNS. Inhibition of the mammalian target of rapamycin (mTOR) signalling pathway with drugs such as rapamycin may help to shrink TSC-associated tumours.

Ehlers–Danlos syndrome

Ehlers–Danlos syndrome (EDS; see Chapter 32, 'Hypermobile Ehlers–Danlos syndrome and hypermobility spectrum disorders' section) is a group of inherited disorders of connective tissue that affect the structure and function of the skin, eyes, ligaments, joints, blood vessels, and internal organs. At least 13 subtypes of EDS are recognized with different molecular genetic defects, inheritance, and clinical features. The underlying genetic defects involve abnormalities of collagen and other extracellular matrix molecules, affecting connective tissue synthesis and structure. These lead to:

Box 45.29 Differential diagnosis of photosensitive rashes

Photoexacerbated or provoked rashes

Autoimmune rheumatic disorders
- Lupus erythematosus
- Dermatomyositis

Metabolic disease
- Porphyrias (see Chapter 36, 'Porphyrias' section)
- Pellagra (see Chapter 25, 'Pellagra' section)

Drugs
- Thiazides
- Phenothiazines
- Tetracyclines
- Amiodarone

Plant phototoxins
- Phytophotodermatitis (photosensitivity induced by contact of the skin with certain plants, e.g., celery, hogweed, rue, lime, fig tree)

Skin disease
- Rarely, atopic eczema, psoriasis, lichen planus (these usually improve in sunlight)

Idiopathic photodermatoses
- Polymorphic light eruption
- Chronic actinic dermatitis
- Solar urticaria

- Fragile, hyperelastic and lax skin that heals with tissue paper scars
- Fragile bloods vessels with easy bruising
- Hypermobile (unstable and hyperextensible) joints.

Pseudoxanthoma elasticum

Pseudoxanthoma elasticum is a group of rare autosomal recessive diseases caused by mutations in the *ABCC6* gene encoding a protein involved in inhibition of mineralization. It causes abnormal calcification and consequent fragmentation of elastic fibres, particularly in the reticular dermis of skin, tunica or media intima of blood vessels, and Bruch membrane of the choroid. Skin changes include laxity and wrinkling and yellow papules and plaques (resembling 'plucked chicken' skin), and the changes are most evident in the flexures and lateral neck. Non-cutaneous features include angioid streaks on the retina, recurrent GI bleeding, vascular occlusion, and aneurysms.

PHOTOSENSITIVITY DISORDERS

Photosensitivity disorders are characterized by abnormal sensitivity to UV radiation and sometimes the visible range of the solar spectrum. Several skin diseases are primarily caused by sunlight ***(primary or idiopathic photodermatoses)*** including polymorphic light eruption, chronic actinic dermatitis, and solar urticaria. A larger number are exacerbated by sun exposure ***(photo-aggravated photodermatoses)***, including some connective tissue diseases (dermatomyositis and cutaneous lupus). The rashes of pellagra (see Chapter 25, 'Pellagra' section) and the porphyrias (see Chapter 36, 'Porphyrias' section) and certain drug hypersensitivities may also be triggered by UV exposure (**Box 45.29**).

BLISTERING AND BULLOUS SKIN DISEASE

There are many causes of skin blisters, the most common being ***infection*** (HSV, chickenpox, impetigo, cellulitis), ***insect bites*** (especially on the lower legs), and ***trauma*** (burns and friction). Blistering

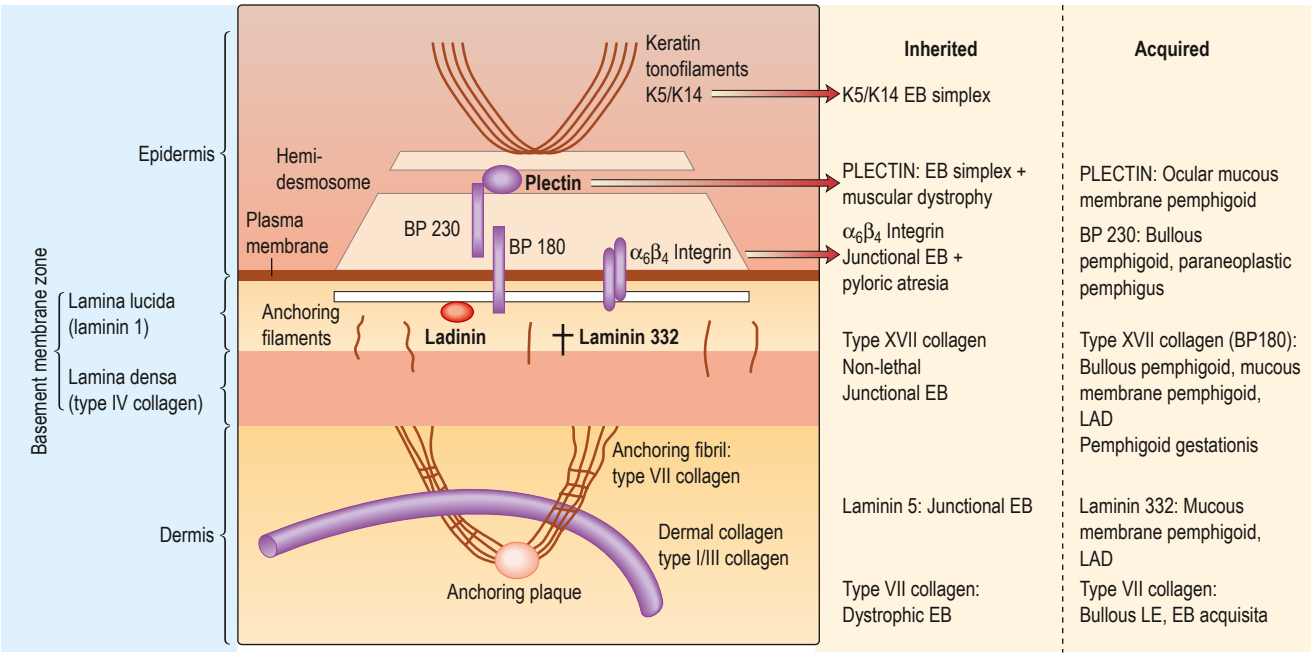


Fig. 45.45 Section of the basement membrane zone. The structural sites of damage in bullous disorders. *EB,* Epidermolysis bullosa; *K,* keratin; *LAD,* linear IgA disease; *LE,* lupus erythematosus.

may also be a feature of eczema, especially on the hands and feet (pompholyx).

Very rarely, blistering is from skin fragility caused by genetically determined abnormalities in the structural proteins of the DEJ and basal keratinocytes ***(mechanobullous diseases)*** or their destruction by antibodies in autoimmune ***(immunobullous)*** disease. The level at which fragility and blistering occur varies according to disease and affects the clinical presentation (Fig. 45.45). It can be demonstrated by light and electron microscopy and IMF studies. Autoantibody deposits (at the basement zone or between keratinocytes) can be identified by direct IMF of affected (perilesional) skin; indirect IMF identifies circulating autoantibodies in serum.

Immunobullous diseases

Pemphigus

This is a disorder affecting the upper (epidermal) layer of skin, characterized by fragile, flaccid blisters that burst easily.

- ***Pemphigus vulgaris*** is a rare and potentially life-threatening autoimmune blistering disease. It is more common in Ashkenazi Jews and people from the Indian subcontinent. It may affect all ages but usually starts in adulthood. Pathogenic IgG4 antibodies against desmosomal proteins (desmoglein 1 and 3) lead to loss of keratinocyte adhesion in the skin and mucous membranes, causing superficial (intraepidermal) blisters, erosions, and mucosal ulcers.

  Oral involvement is common and is the presenting sign in up to 50% of patients. Skin lesions appear as flaccid blisters that rupture easily, leaving weeping and crusted erosions. Blisters can be extended with gentle lateral pressure or rubbing of the skin (Nikolsky sign).
- ***Pemphigus foliaceus*** is a rarer variant, in which the target antigen is desmoglein 1. This antigen is not expressed in the oral mucosa, which is therefore unaffected. Desmoglein 1 is predominantly expressed in the upper epidermis, which detaches, leaving crusts and erosions rather than blisters. It usually affects the seborrhoeic areas before becoming more widespread. Endemic pemphigus foliaceus (fogo selvagem) occurs in rural areas of Brazil and is thought to be triggered by an infectious agent carried by biting insects. A pemphigus foliaceus–like rash can be induced by drugs (e.g., captopril, penicillamine).
- ***Paraneoplastic pemphigus*** is a rare malignancy-associated disorder characterized by painful, hemorrhagic oral erosions and a polymorphic inflammatory rash. It is usually associated with an underlying lymphoproliferative neoplasm.

Investigations

Direct IMF of perilesional skin in pemphigus shows intercellular IgG deposition. Circulating anti-epidermal antibodies can also be detected, and their titre correlates with disease activity. Histology of affected skin shows intraepidermal cleavage above the basal cell layer with separation of individual keratinocytes (acantholysis) in pemphigus vulgaris. The split occurs higher in the epidermis in pemphigus foliaceus.

Management

Topical and systemic corticosteroids are usually effective in controlling pemphigus followed by longer-term, steroid-sparing immunosuppressive drugs, including mycophenolate mofetil, azathioprine, methotrexate, cyclophosphamide, and rituximab (anti-CD20 monoclonal antibody).

Bullous pemphigoid

Bullous pemphigoid is more common than pemphigus in Western Europe and usually affects adults older than 60 years of age. Blisters form in the dermis (deeper than in pemphigus); are usually itchy; and present as large, tense, serous, or haemorrhagic blisters (bullae) with inflamed (urticated) bases. In some cases, intensely itchy, urticated plaques are the presenting feature and precede blistering. Blistering may be widespread, affecting the trunk and limbs, especially the lower abdomen and thighs (Fig. 45.46) but in some cases remain localized.

Fig. 45.46 Bullous pemphigoid: bullous presentation. (A) Typical presentation with multiple tense bullae arising on normal and erythematous skin. Several of the bullae have ruptured, leaving circular erosions. **(B)** Multiple tense bullae that contain serous fluid arising within an area of erythema in addition to haemorrhagic crusts.

Mucosal ulceration is not prominent, except in the rare variant ***mucous membrane pemphigoid*** (formerly called cicatricial pemphigoid). This presents with erosions and ulceration of the mucosal surfaces and scarring eye disease, with fusion of the eyelids, corneal damage, and visual loss. ***Pemphigoid gestationis*** is a rare variant that occurs in pregnancy and the postpartum period.

Investigations

Autoantibodies against a 230- or 180-kDa hemidesmosomal protein (bullous pemphigoid antigen 1 and type 17 collagen, respectively) play a pathogenic role and cause a split through the epidermal BMZ. This can be seen on routine histology as a subepidermal split. Direct IMF shows deposition of IgG and complement at the DEJ.

Management

Oral corticosteroids (prednisolone 0.5 mg/kg per day) are usually initiated for disease control, with a steroid-sparing drug such as azathioprine, mycofenolate mofetyl, or rituximab for long-term management. These drugs require careful monitoring in older adults because of their increased risk of adverse drug effects, which contributes to the increased mortality associated with pemphigoid. Oral tetracyclines are also used for their anti-inflammatory actions. Mild localized pemphigoid can be managed by daily application of a very potent topical steroid to drained blisters. Treatment can usually be withdrawn after a few years, but the disease may relapse.

Linear IgA disease

Linear IgA disease (LAD) is a rare pemphigoid-like disorder of adults and children caused by pathogenic IgA autoantibodies that target various BMZ proteins, including type XVII collagen and laminin-332 (see **Fig. 45.45**). The cause is unknown, but LAD may be induced by drugs, especially vancomycin. Direct IMF of skin shows linear IgA deposition along the DEJ. LAD characteristically presents with circular clusters of golden blisters, described as a 'string of jewels' (**Fig. 45.47**). It usually responds to treatment with dapsone (see later).

Fig. 45.47 Linear IgA disease: typical circular clustering of blisters.

Dermatitis herpetiformis

Dermatitis herpetiformis (DH) is a cutaneous manifestation of coeliac disease (see Chapter 22, 'Coeliac disease [gluten-sensitive enteropathy]' section). Both conditions are genetically determined and triggered by dietary gluten. DH presents with an intensely itchy, eczematous rash that predominantly affects the elbows, knees, scalp, and buttocks. Small blisters may be visible but are often scratched away, leaving excoriations and crusted erosions.

DH is more common in males (whereas coeliac disease has a female predominance); it can present at any age but usually appears in early adulthood. Nearly all patients with DH have small intestinal changes of coeliac disease (villous atrophy/inflammation) but lack GI symptoms. Autoantibodies against tissue transglutaminase (TG) 2 occur in the serum and bowel in patients with DH and coeliac disease, with additional IgA antibodies against TG3 occurring in DH. The latter can be demonstrated in the skin by direct IMF, which reveals granular IgA deposits in the dermal papillae and along the BMZ. Histology of DH shows subepidermal blistering and a florid infiltrate of neutrophils causing micro-abscesses in the dermal papillae.

Management

Patients should avoid all dietary gluten. Treatment with dapsone or sulphonamides, sometimes at low doses, usually gives rapid relief from itch. If a strict gluten-free diet is followed, oral medication can usually be withdrawn after 2 years. Dapsone causes mild dose-related haemolysis, which is usually well tolerated except in those with cardiorespiratory disease. It is contraindicated in people with glucose-6-phosphate dehydrogenase deficiency because they are at risk of severe drug-induced haemolysis. Other rare adverse effects of dapsone include hepatitis, drug hypersensitivity syndrome (including DRESS; see Chapter 11, 'Drug reaction with eosinophilia and systemic symptoms/drug hypersensitivity syndrome' section), and peripheral neuropathy.

Further reading

British Association of Dermatologists guidelines for the management of pemphigus vulgaris 2017.
British Association of Dermatologists: Guideline on assessing and managing bullous pemphigoid 2013.
Collin P, Salmi TT, Hervonen K et al. Dermatitis herpetiformis: a cutaneous manifestation of coeliac disease. *Ann Med.* 2017; 49:23–31.

Mechanobullous diseases (epidermolysis bullosa)

This rare group of inherited skin diseases is caused by defective or absent structural skin proteins, leading to fragile skin that blisters with minimal trauma. Blisters and erosions usually appear at or shortly after birth. The degree of skin fragility is highly variable, ranging from mild friction blisters to severe, extensive, and life-limiting skin loss. There are three broad groups of disorders:

- ***Epidermolysis bullosa simplex*** is the mildest form and causes localized blisters on the hands and feet that heal without scarring. It is inherited as an autosomal dominant trait and caused by synthesis of abnormal cytoskeleton proteins (e.g., keratins 5 and 14) in the basal epidermis.
- ***Dystrophic epidermolysis bullosa*** is a debilitating, life-limiting disease characterized by deep erosions, blisters, and scarring. The underlying abnormality is mutations in the type 7 collagen gene that cause impaired or absent synthesis of anchoring fibrils in the deep part of the BMZ. Mucosal involvement can affect the larynx and oesophagus, causing voice loss and dysphagia. Nails are absent, and the digits may become fused by the scarring, causing mitten-like deformities. Affected individuals have a high risk of developing aggressive SCCs within chronically inflamed skin. There is no curative therapy at present, and management is supportive.
- ***Junctional epidermolysis bullosa*** is the most severe form and is characterized by a split in the lamina lucida of the BMZ caused by mutations in genes controlling synthesis of various structural proteins (laminin-332, $\alpha_6\beta_4$ integrin, type 17 collagen). It presents at birth with widespread blistering and areas of absent skin and is usually fatal in early childhood.

Investigations and management

Investigation and treatment of patients with epidermolysis bullosa should be carried out in a specialist centre. Ultrastructural skin tests and genetic testing provide a precise diagnosis and prognosis and are required for genetic counselling. Prenatal diagnosis and preimplantation diagnosis are available for those with severe forms of epidermolysis bullosa.

Treatment options have been limited to supportive care, pain management, and skin grafting. More recently, topical birch bark triterpenes for dystrophic and junctional epidermolysis bullosa and topical gene therapy for dystrophic epidermolysis bullosa have shown effectiveness in promoting wound healing. Gene therapy and bone marrow transplantation are further new approaches that are undergoing clinical trials.

SKIN TUMOURS

Benign skin tumours

Melanocytic naevi

Melanocytic naevi are the most common benign neoplasms in humans and are caused by a proliferation of melanocytes. They are usually diagnosed based on clinical and dermatoscopic appearances.

- ***Congenital melanocytic naevi*** are present at birth in 1% to 2% of newborns and may be classified by actual or predicted adult size: small naevi are smaller than 1.5 cm, medium are 1.5 to 19.9 cm, and large (giant) are more than 20 cm, including bathing trunk naevi, which cover large areas of the trunk and proximal limbs. Small solitary lesions are common and harmless. Larger lesions and multiple smaller naevi may be associated with neurological complications, including epilepsy and an increased risk of melanoma. Specific types of congenital naevi also recognized include ***hairy congenital naevus*** (which grows long, thick hairs), ***speckled lentiginous naevus*** (naevus spilus, which is a flat, tan coloured patch with darker papules), ***naevus of Ota*** (brown-blue macule around the eye or cheek), and ***Mongolian spot*** (bluish macule over the sacral and buttock areas of newborns).
- ***Acquired melanocytic naevi (moles)*** appear in childhood, adolescence, and early adult life, increasing in size and number with age. Benign naevi may arise on any location and appearances may differ by body site. They are usually round or oval, vary in size from a few millimetres to a few centimetres, and usually have even pigmentation and regular borders. Colours vary and are often pale, pink, or tan-coloured in those with lighter skin tones and darker brown or black in darker skin tones.

Initially, melanocytic naevi may appear as flat, brown macules because of proliferation of melanocytes at the DEJ ***(junctional naevus)***. With later downward growth of melanocytes into the dermis as well as the DEJ, the mole becomes raised and palpable ***(compound naevus)***. Eventually, melanocytes may completely migrate into the dermis, and the mole becomes dome shaped with progressive loss of pigment ***(intradermal naevus)***. Some melanocytic naevi may gradually involute later in life, particularly those acquired in later childhood and adulthood. They may also change during pregnancy and after sun exposure when they can increase in size, regress, or involute.

Additional types of benign acquired melanocytic naevi with characteristic clinical appearances are summarized in **Box 45.30**. Dermatoscopic classifications of moles are also used to describe their different pigment patterns (see **Box 45.2**).

Factors influencing the number of acquired melanocytic naevi present include genetic predisposition, UV exposure, and immunosuppressive medication.

Although the risk of malignant transformation of a benign melanocytic naevus is extremely low, the presence of more than 100 common naevi is associated with a sevenfold increase in the risk of melanoma. Warning signs of possible malignant transformation are summarized in **Fig. 45.1**. Most acquired melanocytic naevi do not require treatment.

- ***Atypical naevi*** are acquired melanocytic naevi that are clinically atypical or histologically dysplastic. They may occur sporadically

Box 45.30 Types of acquired melanocytic naevi

Acquired melanocytic naevus	Clinical appearance
Acral naevi	Moles on the palm or sole; often irregular in shape; more common type in darker-skinned individuals
Subungual or nail unit naevi	Usually present as a longitudinal pigmented band beneath a nail; may affect multiple nails
Blue naevus	Dome-shaped, blue-grey lesions, often located on the forehead or dorsal hands and feet, with a characteristic dermatoscopic appearance of dark blue, structureless lesion caused by a deeper proliferation of melanocytes in the mid-dermis
Spitz naevus (epithelioid cell naevus)	Dome-shaped moles rising in children and young adults that may be pink (classic Spitz) or brown (pigmented Spitz)
Reed naevus	Flat and darkly pigmented type of Spitz naevus with a 'starburst' pattern on dermoscopy
Halo naevi (Sutton naevi)	Symmetrical, round or oval, white 'halo', which is probably caused by autoimmune depigmentation; usually seen in children and young adults; may be multiple and associated with benign melanocytic naevi that gradually regress
Meyerson naevus	Melanocytic naevi affected by a halo of eczema, which appears red and scaly and may be itchy and usually resolved spontaneously

or may be familial. Clinically, atypical features include larger size (>5 mm), irregular pigmentation, and irregular or ill-defined borders. Dysplastic histological features include lentiginous proliferation of naevus cells, cytological atypia, spindle or epithelioid melanocytes, fibrosis in the dermis, inflammatory infiltrate, and increased or enlarged blood vessels. Atypical clinical features and histological dysplasia are not always correlated. They particularly affect people with lighter skin tones and may be more numerous with increased UV exposure.

Atypical naevi that lack features of melanoma, do not warrant excision on presentation, and can be photographed and reviewed at 3- to 4-month intervals to detect early features of malignant change. Patients should be advised on self-examination and photoprotection.

- ***Atypical mole syndrome (AMS)*** refers to the presence of multiple (>5) atypical and dysplastic naevi and multiple benign naevi (>50), which are often located on atypical sites (scalp, buttocks, dorsal hands and feet, irises). The lifetime risk of melanoma is increased approximately sixfold. Familial atypical multiple mole and melanoma (FAMMM) syndrome describes AMS with a family history of melanoma (one or more first- or second-degree relatives diagnosed with melanoma) and is associated with a significantly increased risk of melanoma. People with AMS and a personal history of melanoma or FAMMM usually require long-term self-monitoring and clinical surveillance, which may be supported using dermoscopy of atypical lesions and serial photographic imaging ('mole mapping'), which may be automated and may include dermoscopy.

Epidermal naevi

These are benign overgrowth of epidermal components that appear at birth or early childhood because of mosaic genetic mutations in skin cells during ectodermal development. Epidermal naevi follow the lines of Blaschko (embryological tracks taken by genetically identical developing cells).

- The most common type is a ***linear epidermal naevus***, which usually occurs on trunk and limbs, is unilateral, is a flat and tan-brown colour at birth, and becomes raised and warty with age.
- ***Organoid naevi*** ('naevus sebaceous') also involve proliferation of skin appendages (e.g., sebaceous glands, sweat glands, hair follicles). Initially, they appear as single oval or linear smooth orange-yellow patches, often in the scalp, and become more raised during puberty. In a small proportion, a secondary tumour may occur, usually in adulthood, the most common of which is BCC.
- ***Becker naevus*** is a distinct type of epidermal naevus that usually occurs in males during adolescence in any skin type and unilaterally affects the shoulders or upper trunk. It includes the epidermis, hair follicles, and melanocytes and becomes increasingly thickened, pigmented, and hairy after puberty. Rarely, epidermal naevi may be part of syndromes, which also involve eyes, nervous system, or skeleton.

Fig. 45.48 Seborrhoeic keratoses (seborrheic warts, basal cell papillomas).

Basal cell papillomas ('seborrhoeic keratoses' or 'seborrheic warts')

These are very common, benign lesions caused by hyperproliferation of basal keratinocytes. They increase in number with age (~90% of adults older than 60 years of age have one or more) and are frequently multiple. They appear anywhere on the skin other than palms and soles as papules and plaques and range from flesh-coloured to yellow, grey, light brown, very dark brown to black, or have mixed colours. They have a greasy, 'stuck-on' appearance, and the surface may rough and warty or relatively smooth and waxy. Tiny yellow-white keratin cysts may be visible (Fig. 45.48).

Dermoscopy can be very helpful in distinguishing them from other pigmented skin lesions. They are usually asymptomatic but may be itchy. Treatment is not usually required, but cryotherapy or surgery by curettage and cautery can be used. Eruptive seborrheic keratoses may very rarely be associated with an underlying malignancy (sign of Lésar-Trelat).

Dermatosis papulosa nigra is a term used to describe multiple, small, heavily pigmented basal cell papillomas that may appear with increasing age on the head and neck of some people with darker skin tones (see Fig. 45.5). ***Stucco keratoses*** are another variant that

appear as white or yellow papules on the legs. Treatment is not needed unless the lesions are causing problems, and methods used include cryotherapy, curettage and cautery, and shave biopsy.

Dermatofibroma (benign fibrous histiocytoma)

Dermatofibromas are common, benign fibrous nodules that are most frequently found on the legs and particularly in females. The cause is unclear, but there is sometimes a history of minor trauma or an insect bite. They appear as firm, smooth, pink-red or brown dermal papules or nodules (which may be darker brown or black in darker skin tones), 0.5 to 1.5 cm in diameter, often have a peripheral pigmented margin, and exhibit dimpling of overlying skin when pinched. Although usually asymptomatic, they may be itchy or tender. Histologically, they consist of proliferating fibroblasts histiocytes, blood vessels, and varying degrees of fibrosis. The diagnosis is usually straightforward clinically, and dermoscopy may be helpful. Excision is not needed unless lesions are symptomatic or there is diagnostic uncertainty, and recurrence is common if excision is incomplete.

Lipoma

These benign subcutaneous tumours consist of fat cells, are very common, and are often multiple. They usually appear in adulthood, and there is often a family history when multiple. At presentation, they are often ovoid or round, 2 to 10 cm in size and soft, rubbery, smooth, and mobile. They are most common on the trunk and limbs but can develop wherever fat is present. They are generally asymptomatic, but those with an increased number of blood vessels ***(angiolipomas)*** may be tender on palpation. Surgical treatment is rarely required.

Epidermoid cyst ('sebaceous cyst') and pilar cyst

Epidermoid cysts are benign cysts derived from occluded pilosebaceous units. 'Sebaceous' cyst is a misnomer in that they do not arise in sebaceous glands or contain sebum. They are more common in adults and in males. They are flesh-coloured or yellowish, 1 to 3 cm in diameter, and are fixed to skin but are mobile from deeper structures. They have a central punctum and contain keratin and lipid-rich, often malodorous 'cheesy' debris, which may discharge. They are most common on the trunk but can occur anywhere and may be multiple. They may rupture, causing inflammation with redness and tenderness, and can become secondarily infected. ***Pilar cysts*** are similar but occur on the scalp. Both may be multiple and familial. Uncomplicated cysts do not need treatment, but surgical excision may be considered if the cysts is particularly symptomatic, for example, if infection has occurred.

Pyogenic granuloma

Pyogenic granulomas are acquired, benign vascular proliferations of capillary vessels that present as rapidly growing, friable, red nodules that bleed easily. They may follow minor trauma and often occur on the fingers, face, lips, and oral mucosa. They may also be linked to pregnancy and certain drugs, including retinoids. They rarely resolve spontaneously, except those associated with pregnancy. Excision is the most frequently used treatment, and lesions should be sent for histology to exclude amelanotic melanoma, which is the most important differential diagnosis.

Cherry angioma (Campbell de Morgan spots)

Cherry angiomas are benign haemangiomas that appear as asymptomatic red, purple, or blue papules and may become black if thrombosed. They occur anywhere on the skin, especially on the trunk, and vary in size from pinpoint to 0.5 to 1 cm diameter and are usually multiple. They increase in number with age, and an estimated 75% of people older than 40 years of age have at least one. No treatment is required.

Keloid and hypertrophic scars

Abnormal wound healing with excessive dermal fibrosis leads to hypertrophic scars and keloid formation.

- In ***hypertrophic scars***, thickening remains confined to the borders of the original wound, is usually slightly raised and pink-red (more pigmented in darker skin tones), and occurs within weeks of an injury. They are common in all age groups and races and ages and tend to develop if there is significant tension on a wound. They may regress spontaneously.
- In contrast, in ***keloid scars***, the excessive scar formation extends beyond the original wound margins and may develop months or even years after minor injury such as minimal trauma, acne, shingles, insect bites, immunization, and occasionally spontaneously. They tend to affect young adults with darker skin tones and are more common in those of African and Chinese heritage. They are usually red-purple (or brown in darker skin tones), firm, and may be itchy and tender, and sites of predilection are areas of high skin tension, including the shoulders, upper back and chest, earlobes, and chin (Fig. 45.49).

Both hypertrophic and keloid scars can be associated with significant cosmetic disfigurement, functional problems, and adverse psychosocial impact. Treatment options include use of silicone gel or dressings, pressure garments, topical steroids, steroid-impregnated tapes, intralesional corticosteroids, cryotherapy, and intralesional 5-fluorouracil (5-FU). Occasionally, surgical excision of keloid scars may be undertaken and followed by superficial radiotherapy. Non-essential surgery should be avoided.

Dysplastic and premalignant skin lesions

Actinic (solar) keratoses

Actinic keratoses (AKs) are common premalignant skin lesions with partial-thickness epidermal dysplasia and are considered to be

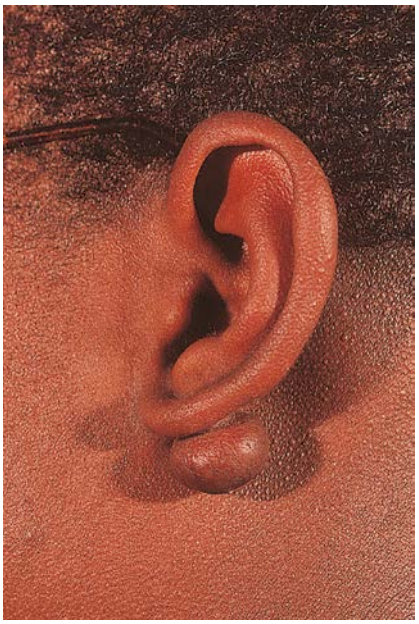


Fig. 45.49 **Keloid scar on the lobe of the ear.**

Box 45.31 Fitzpatrick skin phototype classification

Fitzpatrick classification[a]	Clinical features (skin tone, hair colour, eye colour)	Reaction to UV radiation
I	Pale white skin; light blonde or red hair; blue or green eyes	Always burns; does not tan; freckles
II	Lighter white skin; blue, green, or hazel eyes	Usually burns; tans with difficulty; freckles
III	Darker white skin; blonde or brown hair; dark blue, hazel, or brown eyes	Initially burns moderately; tans gradually to light brown; may freckle
IV	Light to moderate brown or olive skin; brown-black hair; light brown eyes	Rarely burns; tans easily to moderate brown
V	Brown skin; brown-black hair; dark brown eyes	Very rarely burns; tans easily
VI	Dark brown or black skin; brown-black hair; brown eyes	Never burns; tans very easily and deeply; never freckles

[a]This classification estimates the sunburning and tanning ability of skin to ultraviolet (UV) radiation. Although widely used, its representation of global skin colour diversity in people of colour is recognized to be limited.

precursors for cSCC. They are the result of chronic UV radiation (UVR) exposure and are most common on UV-exposed sites, particularly the ears, nose, cheeks, forehead, lips, scalp (particularly in balding males), and dorsal hands and forearms. Risk factors include lighter skin tones (particularly Fitzpatrick skin phototypes I and II; Box 45.31), a history of sunburn, immunocompromise, and a history of high UV exposure (e.g., outdoor occupation or recreational activities, residence in geographic locations with high ambient UV, increasing age). They appear as asymptomatic or tender papules or plaques, which may be white, yellow, red, or pigmented. They usually feel scaly and rough and may be palpated more easily than seen. Some become more hyperkeratotic, and cutaneous horns may also arise from AKs. AKs may be multiple and confluent and coexist with chronic photodamage (Fig. 45.50), to form areas of 'field cancerization' or 'field change'.

AKs are usually diagnosed clinically and on dermoscopy, although occasionally, a skin biopsy may be required to exclude cSCC, particularly if an AK becomes tender, enlarges, or ulcerates. Malignant transformation of an individual AK is rare, occurring in an estimated 0.1% of AKs. However, approximately 10% of patients with 10 or more AKs eventually develop cSCC. Because AKs are associated with UV damage, they are also a sign of an increased risk for other skin cancers, including BCC and melanoma.

Treatment is usually indicated for symptomatic AKs and to reduce the risk of cSCC development (although the evidence for this is limited). Strict photoprotection can reduce the numbers of AKs and cSCC. Other treatments that are often used in combination or sequentially include:

- *Lesion-directed treatments* for individual AKs, including cryotherapy, curettage and cautery, and excision
- *Field treatments* aimed at clearing multiple visible and subclinical lesions in areas of field change. These include topical treatments with different modes of action, including 5-FU (cytotoxic agent), imiquimod (immune response modifier), diclofenac (cyclo-oxygenase-2 inhibitor) and tirbanibulin (inhibitor of tubulin polymerization and protein kinase signalling). Photodynamic therapy using the topical photosensitiser 5-aminolaevulinic acid or its methyl ester followed by exposure to a source of red or blue light or daylight is another option.

Bowen disease (squamous cell carcinoma *in situ*, intraepidermal carcinoma)

This is also an cSCC premalignant lesion with full-thickness intraepidermal dysplasia histologically. It is estimated to progress to invasive cSCC in approximately 5% of cases. It is usually caused by chronic UV exposure in people with Fitzpatrick skin phototype type I or II; other risk factors include immunocompromise, arsenic ingestion (e.g., from contaminated well water is some parts of the world), human papillomavirus (HPV) infection (see later), and ionizing radiation.

Fig. 45.50 Actinic (solar) keratoses with background actinic damage ('field change').

Bowen disease lesions present as slowly enlarging, well-demarcated scaly plaques that may be pink-red, purple, or pigmented. When related to UV exposure, they typically affect the lower legs in females (Fig. 45.51A); the trunk in males and the ears, face, and hands in both sexes. They may resemble plaques of psoriasis, eczema, or tinea. They may be contiguous with AKs in areas of field cancerization or field change. The diagnosis is usually based on clinical features and supportive dermoscopy, but a skin biopsy may be required for confirmation.

Variants of Bowen disease affecting the genital mucosa are referred to as ***VIN, PIN,*** or ***AIN*** and usually result from infection with high-risk alpha HPV (e.g., HPV-16 and 18). Lesions present with nonspecific red or brown plaques (erythroplasia; Fig. 45.51B) or warty papules and plaques ***(Bowenoid papulosis)***. Anogenital disease may also be associated with periungual Bowen disease. Anogenital Bowen disease is more common in immunocompromised individuals, including PLWH. Management is discussed in Chapter 43 ('Human papillomavirus [anogenital warts]' section).

Treatment for patients with Bowen disease is similar to that for AK and includes physical destructive treatments (e.g., cryotherapy, curettage and cautery, excision, laser), topical treatments (5-FU, imiquimod), and photodynamic therapy.

Keratoacanthoma

Keratoacanthomas are rapidly growing epidermal tumours that develop as nodules with a central crater-like, keratinous plug

Fig. 45.51 **Cutaneous squamous cell carcinoma *in situ* (intra-epidermal carcinoma, Bowen's disease).** **(A)** The lower leg is a common site, and lesions are often treated non-surgically. **(B)** Erythroplasia of glans penis (penile intraepithelial neoplasia).

Fig. 45.52 **Keratoacanthoma.**

(Fig. 45.52). They typically develop on chronically sun-exposed skin in males with lighter skin tones older than the age of 60 years. They characteristically grow rapidly over 3 to 6 weeks, often reaching 2 to 3 cm in diameter, and may regress spontaneously after about 3 months, leaving a crateriform scar, although they are generally excised because they can be difficult to distinguish from well-differentiated SCC. Keratoacanthomas are usually associated with chronic UV exposure but are also more common in immunocompromised patients and occur in people with some rare genetic cancer syndromes (e.g., Ferguson-Smith syndrome and Muir Torre syndrome) and secondary to some drugs (e.g., vemurafenib and dabrafenib).

Lentigo maligna (Hutchinson melanotic freckle)

Lentigo maligna is a precursor to lentigo maligna melanoma and is a slow-growing intraepidermal *(in situ)* form of melanoma. The main risk factor is chronic UV exposure. It is more common in males aged 60 to 80 years, and patients often also have a history of AKs and keratinocyte cancers. It typically presents as a pigmented macule on chronically sun-damaged skin (usually the face) and may initially resemble a solar lentigo but grows slowly over many years, often becoming increasingly atypical (Fig. 45.55C).

Invasive lentigo maligna melanoma (see later) occurs in approximately 3% to 10% of cases. It should be suspected if thickening, increasing numbers of colours, ulceration, bleeding, and suspicious dermoscopic features develop. Skin biopsy can confirm the diagnosis, but there is a risk of sampling error, which misses an invasive focus if large lesions are biopsied.

Treatment of lentigo maligna is by excision if possible, but margins may be difficult to define clinically, and recurrence rates are approximately 20%. Confocal microscopy or excision by Mohs' micrographic surgery (see later) can help improve margin control. Because complete excision may have adverse cosmetic and functional outcomes, non-surgical approaches include topical imiquimod and radiotherapy. In some cases, particularly in very frail or older patients with large lesions and multiple co-morbidities, close observation may be a reasonable alternative.

Malignant skin tumours

Keratinocyte cancers

Keratinocyte cancers account for more than 90% of all skin cancers and are the most common human malignancies. They arise in the epidermis and include BCC and cSCC. Tumours are frequently multiple, and the AK, Bowen disease, field change, and keratinocyte cancer spectrum is often regarded as a chronic condition. 'Non-melanoma skin cancer' is a broader term that includes both keratinocyte cancers and other skin cancers, including Merkel cell carcinoma, appendageal cancers, sarcomas, and lymphomas.

Basal cell carcinoma

BCCs are the most common form of keratinocyte skin cancer and overall are two- to threefold more common than cSCC. Incidence is increasing by 5% to 10% a year. BCC probably arise from pluripotent cells in the basal epidermis or follicular structures. UV-induced DNA mutations in *PTCH* tumour suppressor gene involved in regulation of the Hedgehog intracellular signalling pathway are detected in sporadic BCCs (usually UV induced) and in Gorlin syndrome, a familial cancer syndrome characterized by the development of multiple BCCs. Risk factors for BCC include:

- Age and sex (more common with increasing age and in males)
- UVR exposure (previous episodes of sunburn; use of artificial tanning beds)
- Lighter skin tones that burn easily (Fitzpatrick phototype I and II), although they also occur in people with darker skin tones
- Previous skin cancer (BCC, cSCC, melanoma)
- Immunocompromise (e.g., risk is increased 5 to 10-fold in organ transplant recipients and PLWH or haematological malignancy)
- Genetic predisposition (family history of BCC)

Fig. 45.53 Clinical basal cell carcinoma (BCC) subtypes. (A) Nodular. **(B)** Superficial. **(C)** Morphoeic. **(D)** Nodulo-ulcerative BCC.

- Rare genodermatoses (e.g., Gorlin syndrome [naevoid basal cell carcinoma syndrome], xeroderma pigmentosum)
- Exposure to arsenic, ionizing radiation, or other drugs (including immunosuppressive agents, hydrochlorothiazide).

Clinical features

BCCs are locally invasive and typically appear as a slowly enlarging plaques or nodules that may be skin coloured, pink-red, or pigmented. They are most common on the head and neck and are frequently multiple. Several distinct clinical types are recognized:

- Nodular or nodulocystic BCC (70%): pearly or shiny plaques and nodules that bleed easily with minor trauma and fail to heal. They may have a central ulceration, making the borders appear rolled, with a pearly appearance (Fig. 45.53A), and overlying, branch-like (arborizing) telangiectasia.
- Superficial BCC is the most common type on the upper trunk and presents as a pink, red, or pigmented; scaly; irregular plaque with a thin, cord-like rolled border and microerosions (Fig. 45.53B).
- Morphoeic (infiltrative) variants present as scar-like plaques with ill-defined borders and are most prevalent on the face (Fig. 45.53C). They may have significant subclinical extension and perineural invasion; both of these features make incomplete excision and recurrence more likely with this subtype.

Although BCCs have minimal metastatic potential, they may cause extensive local tissue destruction if untreated (hence the common term 'rodent ulcer'), resulting in significant cosmetic and functional morbidity (Fig. 45.43D).

Management

The diagnosis is based on clinical and dermatoscopic appearances and a skin biopsy if the diagnosis is uncertain or non-surgical treatment is planned. Non-invasive skin imaging technologies such as optical coherence tomography (OCT) and reflectance confocal microscopy (RCM) are also increasingly being used to assist diagnosis and monitor treatment, although their availability is currently limited in most centres.

The treatment depends on the subtype, size, location, number of BCCs, and patient factors.

- The treatment of choice for most BCCs is surgical excision with histology to ensure complete removal of the tumour with adequate margins.
- *Mohs micrographic surgery* is a form of margin-controlled surgery in which tumours are excised in layers and processed so that 100% of the margins are evaluated (compared with standard 'breadloaf' histology, which evaluates a much smaller proportion of margins). Further layers are removed if excision is incomplete, ensuring that complete excision is achieved before reconstruction, while minimizing tissue removal. It is preferred for morphoeic BCCs (which are more likely to recur), BCCs in locations where tissue sparing is a priority (e.g., nose, periorbital), and recurrent BCC.
- Superficial BCCs can be managed with excision; curettage and cautery; or non-surgical treatments, including cryotherapy, photodynamic therapy, topical imiquimod cream, and topical 5-FU creams.
- Radiotherapy is an option if surgery is not appropriate or for patients who are unsuitable for surgery.
- Targeted therapy with the oral hedgehog inhibitors vismodegib and sonidegib is indicated for locally advanced and metastatic BCC not amenable to surgery or radiotherapy, but adverse effects result in a significant number of patients stopping treatment.

- Anti–programmed cell death-1 (PD-1) immunotherapy with cemiplimab has shown benefit in locally advanced and metastatic BCC when hedgehog inhibitors have been unsuccessful.

After a first BCC, approximately 30% to 50% of people develop another keratinocyte cancer within 5 years and are also at increased risk of other skin cancers. Regular self-examination and ongoing photoprotection should be advised as part of long-term management.

Cutaneous squamous cell carcinoma

cSCC is the second most common skin cancer. Like BCC, it is derived from keratinocytes but has a higher metastatic potential than BCC and can arise in pre-existing AKs or Bowen disease (cSCC in-situ). The incidence is increasing by 5% to 6% per year. UVR exposure (both solar and artificial from tanning beds) is the most important cause of cSCC, and there is a clearer correlation with chronic cumulative UV-induced skin damage compared with BCC. cSCCs have a high tumour burden of UV-induced genetic mutations in multiple genes, including *TP53*, *NOTCH1* and *NOTCH2*, *CDKN2A*, and *N-RAS*.

Risk factors for cSCC include:

- Age and sex (average age higher than for BCC; more common in males)
- UVR exposure (solar and sunbeds)
- Lighter skin tones that burn easily (Fitzpatrick phototypes I and II), but it can also occur in people with darker skin tones
- AKs, Bowen disease, and previous skin cancer (BCC, cSCC, melanoma)
- Immunocompromise (e.g., risk increased 65- to 100-fold in organ transplant recipients and 5-fold in PLWH or haematologic malignancy; cSCCs in immunocompromised people are often more aggressive with a higher rate of metastasis)
- Genetic predisposition (family history of skin cancer; specific genetic polymorphisms; rare genodermatoses, e.g., xeroderma pigmentosum, albinism, epidermodysplasia verruciformis, Muir Torre syndrome)
- Exposure to arsenic, ionizing radiation, smoking, other drugs (including immunosuppressive agents, hydrochlorthiazides, voriconazole, BRAF inhibitors, JAK inhibitors, and anti-TNF agents)
- Chronic inflammation (e.g., thermal burns, lupus vulgaris, lupus erythematosus, LP, leg ulceration)

Clinical features

cSCC mainly affects the UV-exposed sites of older people with lighter skin tones, particularly the face, ears, lips, balding scalp, dorsal hands, and lower legs. cSCCs present as enlarging, ill-defined, keratotic or warty, inflamed plaques or nodules that may ulcerate and bleed easily (Fig. 45.54). They grow more rapidly than BCC and are often painful and tender. It can also occasionally complicate areas of chronic inflammation such as ulcers (Marjolin ulcer). Metastasis risk is associated with specific high-risk features (Box 45.32) and overall is 2% to 3%, but it is higher in immunocompromised individuals. The most common site of metastasis are the lymph nodes and parotid gland (80%) but also the skin, bone, liver, lung, and brain. Most cSCC-specific deaths are caused by locoregional recurrence rather than distant metastasis.

Management

Diagnosis is based on clinical and dermatoscopic appearances and a diagnostic skin biopsy if required. Non-invasive skin imaging technologies (e.g., OCT and reflectance confocal microscopy) may also be useful when available. Staging investigations including imaging (e.g., ultrasonography, computed tomography, MRI) and biopsies of lymph nodes and other tissues may be required to assess the extent of locoregional and distant spread in high-risk tumours.

Fig. 45.54 **Squamous cell carcinoma.**

Box 45.32 High-risk features of primary cutaneous squamous cell carcinoma (cSCC)

- Tumour diameter >2 cm
- Immunosuppression
- Recurrent SCC
- SCC on ear (lips, eyelid, and scalp may also carry an increased risk)
- Histological features: tumour >4 mm thick, poor differentiation, extending into fat, perineural invasion

Patients with suspected and confirmed cSCC should be managed by a multidisciplinary skin cancer team.

- Treatment is usually complete surgical excision with a margin ranging from 4 to 10 mm depending on the risk status of the cSCC.
- *Mohs' micrographic surgery* may be used for high-risk cSCC in locations and when tissue sparing is important.
- Radiotherapy may be used as a primary treatment if surgery is not possible or unsuitable. It is also used as adjuvant treatment postoperatively for some high-risk tumours.
- Anti–PD-1 inhibitor immunotherapy (e.g., cemiplimab, pembrolizumab) is used for locally advanced and metastatic disease not amenable to surgery or radiotherapy and has significantly superior response rates to conventional chemotherapy and targeted epidermal growth factor receptor (EGFR) inhibitor therapy.

As for BCC, the risk of further cSCC is increased after a first cSCC, and patients should be advised on self-examination and photoprotection. Treatment of AKs and field change with sunscreen and 5-FU reduce cSCC risk in randomized controlled trials. Oral nicotinamide (a form of vitamin B3) reduces the risk of further cSCC in immunocompetent patients, and the oral retinoid acitretin reduces cSCC risk in immunocompromised patients.

Melanoma

Melanoma is a malignancy of melanocytes and most commonly arises in the skin, but it can also develop in non-cutaneous sites

Box 45.33 Risk factors for melanoma

- Increasing age
- Lighter skin tones (Fitzpatrick skin phototype I or II)
- High ultraviolet exposure (including sunbed use) and history of sunburn
- Multiple (more than five) atypical melanocytic naevi
- Large congenital melanocytic naevi (>20 cm in diameter)
- Family history of melanoma: mutations described in CDK4 (20%–40% of familial melanoma), *MC1R, MITF, TERT, ACD, BAP1, POT1,* and *TERF2IP*
- Immunocompromise (e.g., risk increased two- to fivefold in organ transplant recipients)
- Immunosuppressive medication, e.g., azathioprine
- Genodermatoses, e.g., xeroderma pigmentosum, albinism

(e.g., ocular, mucosal, meningeal). Although it represents only 4% of all skin cancers, melanoma is responsible for 80% of skin cancer-related deaths. The incidence increases with age, but it also disproportionately affects young people and in Australia is the most common cancer in people aged 15 to 29 years. The increasing incidence in recent years is thought to be attributable to more extensive recreational sun exposure, especially intermittent intense exposure (sunbathing in particular); sunburn in childhood; and use of artificial tanning beds, particularly in children and young adults. About 75% of cutaneous melanomas arise *de novo* from normal skin, the remainder arising from a pre-existing acquired or congenital melanocytic naevi. Risk factors for melanoma are summarized in **Box 45.33**.

Multiple oncogenes and tumour suppressor proteins have been implicated in the pathogenesis of melanoma and specific genotypes correlate with clinical phenotypes and have therapeutic implications. For example, activating mutations in the BRAF V600E protein kinase is the most common mutation in melanoma and is a therapeutic target for which treatments were first approved in 2011 (see Chapter 41, 'Malignant melanoma' section). The tumour mutational burden of cutaneous melanoma is also high, and as with cSCC and BCC, this is because of the highly mutagenic properties of UV radiation. This high mutational burden makes melanoma a good candidate for response to immunotherapy, and melanoma was the first cancer for which immune checkpoint inhibitor immunotherapy was approved in 2012 (see Chapter 6, 'T lymphocyte development and activiation' section).

Classification and clinical features

The conventional morphologic classification of melanoma recognizes four main types (**Fig. 45.55**):

- ***Lentigo maligna melanoma*** (10%–15%) is an invasive tumour that develops within pre-existing lentigo maligna. It is usually apparent as a new nodule.
- ***Superficial spreading melanoma*** (55%–60%) is a large, flat, irregularly pigmented lesion that grows laterally before vertical invasion develops.
- ***Nodular melanoma*** (10%–15%) is the most aggressive type. It presents as a rapidly growing pigmented nodule, which bleeds or ulcerates. Rarely, it is amelanotic (non-pigmented) and can mimic pyogenic granuloma.
- ***Acral lentiginous melanomas*** (<5%) arise as pigmented lesions on the palm or sole or under the nail and usually present late. They may not be related to sun exposure and are the most common skin type in people with skin of colour.

The WHO also now classifies melanoma by associated cumulative solar damage, anatomic location, epidemiology, and genotype (**Box 45.34**).

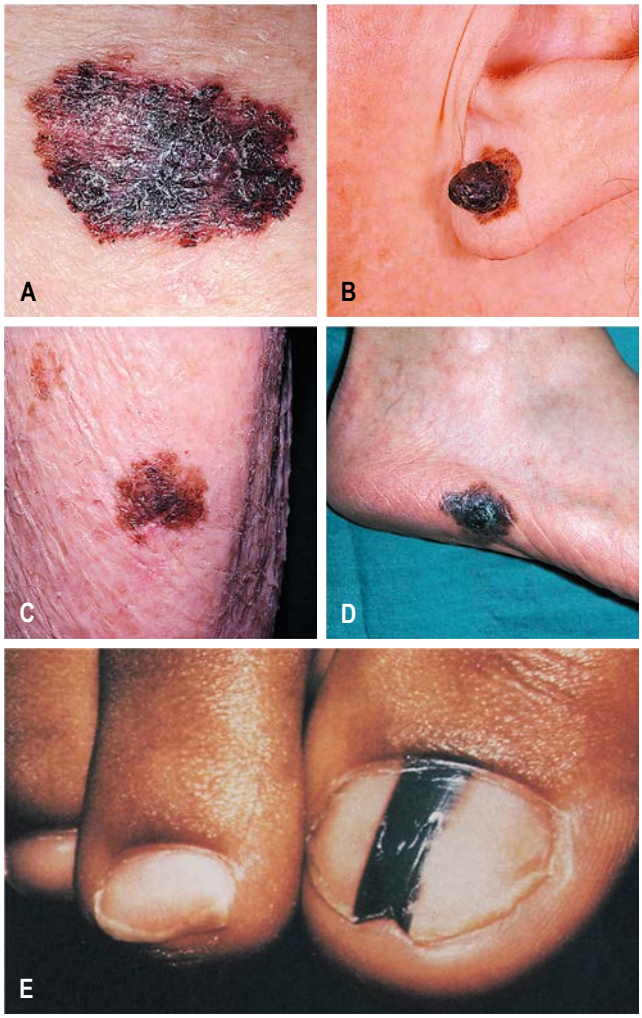


Fig. 45.55 Types of cutaneous melanoma. (A) Superficial spreading melanoma. **(B)** Nodular melanoma arising in a benign compound nevus. **(C)** Lentigo maligna, which may progress to lentigo maligna melanoma. **(D)** Acral lentiginous melanoma on the sole of the foot. **(E)** Subungual acral lentiginous melanoma. (*A–D: From Callen JP, et al:* Color atlas of dermatology, *ed 2, Philadelphia, 2000, Saunders. D, Courtesy Dr. Neil A. Fenske, Tampa, FL. E: From Habif, T. P., Campbell, J. L., Chapman, M. S., et al. [2011].* Skin disease: Diagnosis and treatment *[3rd ed.]. St Louis: Saunders.*)

Diagnosis

The diagnosis of melanoma is not always easy because many other pigmented cutaneous lesions may simulate melanoma, but a history of change, the clinical signs listed in **Box 45.35**, and recognition of the 'ugly duckling' sign (i.e., a mole that stands out as significantly different from the others) may all help distinguish melanoma from benign melanocytic naevi. ***Dermatoscopic examination*** is a key aid to clinical diagnosis, and dermoscopy is an essential skill for clinicians who assess pigmented lesions. Specific dermoscopic features of melanoma include atypical pigment network, multiple colours (five or six), brown-black dots or globules, an irregular vascular pattern, blue-white veil, and depigmentation. Other skin imaging techniques such as ***reflectance confocal microscopy (RCM)*** may be helpful in supporting the diagnosis.

Teledermatology is playing and increasingly important role in the diagnostic pathway for melanoma and other skin cancers. Artificial

Box 45.34 Classification of melanoma

Relationship with cumulative sun damage (CSD)	Melanoma type (WHO, 2018)	Examples of genetic changes
Melanomas typically associated with CSD	Cutaneous, low CSD (e.g., superficial spreading melanoma on trunk)	*BRAF* V600E mutations
	Cutaneous high CSD, (e.g., lentigo maligna on face)	Mutually exclusive mutations in *NRAS*, *NF1*, non-V600E *BRAF*
	Desmoplastic	Inactivating mutations in *NF1*; activating mutations in MAPK pathway
Melanomas not consistently associated with CSD	Acral	Amplifications of *c-KIT*, *TERT*, *CCND1*; *c-KIT* mutations
	Uveal	*GNAQ*, *GNA11*
	Mucosal	Copy number and structural variations
	Spitz	*HRAS*, kinase fusions
	Arising in congenital naevus	*NRAS*
	Arising in blue naevus	*GNAQ*, *GNA11*

Modified from WHO Classification, 2018.

Box 45.35 Clinical criteria for the diagnosis of melanoma

ABCDE criteria (USA)

- **A**symmetry of mole
- **B**order irregularity
- **C**olour variegation
- **D**iameter >6 mm
- **E**levation

Glasgow seven-point checklist

Major criteria

- Change in size
- Change in shape
- Change in colour

Minor criteria

- Diameter >6 mm
- Inflammation
- Oozing or bleeding
- Mild itch or altered sensation

Box 45.36 Visible and ultraviolet radiation

Type	Wavelength (nm)	Properties
Visible radiation (visible light)	400–700	Visible light
Ultraviolet A	400–315	Long wave, black light, not absorbed by the ozone layer
Ultraviolet B	315–280	Medium wave, mostly absorbed by the ozone layer
Ultraviolet C	280–100	Short wave, germicidal, completely absorbed by the ozone layer and atmosphere

intelligence–based diagnostic approaches are also being widely investigated and are likely to play an increasingly important supportive role in the future. However, currently, the definitive diagnosis is made on biopsy and histological confirmation.

Management

In the UK, all people with suspected melanoma should be referred urgently (2-week wait pathway) to the local multidisciplinary skin cancer team. Suspicious lesions should be completely and promptly excised with a narrow (2-mm) margin for review by a dermatopathologist experienced in melanoma diagnosis. Further management of melanoma is discussed in Chapter 41 ('Malignant melanoma' section). In people with skin of colour, identification of melanoma may be more challenging: melanomas may be more advanced at presentation, acral lentiginous melanomas are more common as are non-cutaneous melanomas, and outcomes tend to be worse.

Ultraviolet exposure and photoprotection

UV radiation is the most important carcinogen in both keratinocyte cancers and melanoma. The solar spectrum at the earth's surface includes visible light and medium- and long-wave UV (i.e., UVB and UVA). Short-wavelength UVC is highly carcinogenic but is filtered out by the atmosphere (**Box 45.36**). UVB and UVA are also potentially mutagenic and carcinogenic, especially in fair-skinned people. They cause sunburn (predominantly UVB) and premature ageing or photodamage (predominantly UVA). UV radiation also has immunosuppressive effects that enhance skin cancer development.

Additional adverse effects of UV radiation include cooperation in the direct carcinogenic effects of some systemic medications such as azathioprine, ciclosporin, voriconazole, and hydrochlorthiazides, the mechanisms responsible include drug-associated photosensitivity to UV and phototoxicity.

The wavelength of visible light is longer (400–750 mn), and it can trigger or exacerbate photodermatoses including solar urticaria and porphyria cutanea tarda (see Chapter 36, 'Porphyrias' section).

Sunscreens and photoprotection

Although sunscreens protect against UVA and UVB radiation, they are no substitute for use of protective clothing (including sunhats and sunglasses). Restriction of exposure, especially in young children, avoidance of artificial tanning beds and the 'Slip, slop, slap, seek, slide' approach should be advised (**Box 45.37**). The widely available WHO UV Index provides information on intensity of UV on a daily basis for specific geographic locations at different times of day (**Box 45.38**).

Box 45.37 Recommended approaches to photoprotection (SunSmart Australia)

Slip: Slip on protective clothing.
Slop: Slop on some sunscreen.
Slap: Slap on a sunhat.
Seek: Seek out some shade.
Slide: Slide on some sunglasses.

Box 45.38 World Health Organization ultraviolet (UV) index

UV index 1–2: No protection needed; safe to be outside
UV index 3–7: Protection needed: seek shade around midday hours; wear protective clothing
UV index 9–11: Extra protection: avoid being outside around midday hours; seek shade; wear protective clothing

Sunscreens work by using ***chemical sunscreens*** (organic filters or 'absorbers') that absorb or filter UV radiation (e.g., benzophenones, cinnamates, salicylates) or ***physical (mineral) sunscreens*** that reflect it (e.g., zinc, titanium dioxide). New sunscreen chemicals have been developed to give better protection, and the particle size of physical sunscreens can be reduced (micronized or ultrafine 'nanoparticles') to improve their cosmetic acceptability. Modern sunscreens are formulated to provide broad-spectrum protection against UVA and UVB. The SPF is a measure of UVB protection and the degree to which exposure can be prolonged before burning. The star rating (one to five stars) is an indication of UVA protection. Generally, an SPF30+ with four or five UVA stars is recommended (particularly for people with a history of skin cancer). However, in many cases, sunscreens are not applied in adequate amounts and are estimated on average to provide one-third of the SPF as labelled. UV-absorbing chemicals may occasionally cause allergic contact dermatitis and, in rare instances, photoallergic contact dermatitis (in which the sunscreen becomes allergenic with UV exposure).

Photoprotection and vitamin D deficiency

Sunlight on the skin is the source of approximately 90% of our vitamin D (which is difficult to obtain in food, being present only in a limited range of food, e.g., oily fish and some fortified butters and cereals). Individuals who do not have photosensitivity or a history of skin cancer benefit from short-term sun exposure (without burning) to maintain cutaneous vitamin D production. This is particularly necessary for people with darker skin colours (particularly Fitzpatrick phototypes V and VI) living in temperate climates, who are at risk of vitamin D deficiency. In those requiring photoprotection, oral vitamin D supplementation may be required. Advice about sun protection therefore needs to consider skin type.

Cutaneous lymphomas: primary cutaneous T-cell lymphoma (mycosis fungoides)

Cutaneous lymphomas are rare. Two main classes affect the skin: cutaneous T-cell lymphoma (CTCL) and cutaneous B-cell lymphoma.

CTCL (mycosis fungoides) accounts for 50% of primary cutaneous lymphomas. It is a rare form of non-Hodgkin lymphoma and usually follows an indolent course. It is more common in males, and its incidence increases with age with median age of onset of 55 to 60 years (although this can be younger in Black and Asian patients). It presents with pruritic, scaly patches, which typically start on the buttocks and thighs and can resemble eczema, psoriasis, or fungal infection; asymmetry and atrophy (thin, wrinkled skin) are useful clues (Fig. 45.56). Hypopigmented variants are seen in people with skin of colour and in children. Patches of mycosis fungoides may evolve into plaques, nodules, and red-violaceous tumours, which may ulcerate and become necrotic and may be complicated by secondary bacterial infection. Extracutaneous spread to the lymph nodes and internal organs occurs in 10%, and transformation into large cell lymphoma is a rare complication. The 5-year survival rate is 87%.

Fig. 45.56 Patch-stage mycosis fungoides.

The diagnosis may often require multiple skin biopsies over time to confirm the characteristic invasion of the epidermis by atypical T lymphocytes (exocytosis). T-cell receptor gene rearrangement studies show that the infiltrate is clonal.

Sézary syndrome (see Chapter 39, 'Mycosis fungoides and Sézary's syndrome' section) is an erythrodermic variant of CTCL with peripheral lymphadenopathy and atypical malignant Sézary cells in the blood. The onset is often rapid and characterized by severe itch and diffuse eczematous changes. Palmoplantar keratoderma, ectropion, nail dystrophy, and alopecia may also occur. It is seen mostly in older males and is more common in people with skin of colour who are often younger than White patients. The 5-year survival rate is less than 30%.

Treatment

Mycosis fungoides and Sézary syndrome usually run chronic relapsing courses. Treatment choices depend on disease stage and extent. Patch and plaque stage mycosis fungoides usually responds to topical treatment with potent corticosteroids, nitrogen mustard, or retinoids (bexarotene gel). Phototherapy can also be helpful. Treatment options for patients with advanced disease include localized radiotherapy or electron-beam therapy, oral bexarotene, denileukin diftitox, histone deacetylase inhibitors, chemotherapy, and brentuximab vedotin (anti-CD30 monoclonal antibody) for CD30+ disease.

Patients with Sézary syndrome are treated with extracorporeal photopheresis (in which leucocytes are mixed with psoralen

and irradiated with UVA *ex vivo*), total skin electron-beam therapy, methotrexate, bexarotene, or monoclonal antibody therapy (alemtuzumab and mogalizumab). The only potential cure is allogeneic stem cell transplantation.

Kaposi sarcoma

This tumour of vascular and lymphatic endothelium is caused by infection with human herpes virus-8 (HHV-8), also known as KS herpes virus. The incidence of KS is associated with the seroprevalence of HHV-8 globally: 30% to 50% of the population in sub-Saharan Africa is seropositive compared with 10% to 20% in the Mediterranean, Middle East, and Caribbean and fewer than 5% in Northern Europe and the USA. Transmission may be vertical, sexual, or via blood products and saliva.

Four types of KS are recognized:

- ***Classic KS*** occurs in older males, especially people of Mediterranean, Middle or Eastern European, and Jewish descent. It presents as slow-growing macules, plaques, or nodules on the foot and lower limb.
- ***Endemic KS*** occurs in sub-Saharan Africa in children and young adults and has more widespread skin and lymph node involvement with lymphoedema as a prominent feature.
- ***HIV-associated KS*** affects PLWH, most commonly MSMs in whom sexual transmission of HHV-8 may occur. Cutaneous lesions may be widespread (**Fig. 45.57**) with additional involvement of the oral cavity, bowel, and lungs. Antiretroviral therapy (see Chapter 44) has significantly reduced the incidence of KS.
- ***Iatrogenic KS*** affects people receiving immunosuppressive medication, particularly organ transplant recipients who were born in countries with a high seroprevalence of HHV-8 and in whom KS usually develops because of reactivation of latent virus.

Clinical features

KS usual presents with skin involvement as red to purplish macules, patches, papules, plaques, and nodules, which may appear dark brown-black in darker skin tones (see **Fig. 45.57**). These are usually asymptomatic initially but can enlarge, ulcerate, bleed, and become painful. Oral and nasopharyngeal mucosal involvement is common. KS lesions frequently affects the legs and feet in classic and endemic types and are associated with lymphoedema, and the latter may sometimes be the presenting symptom before the appearance of cutaneous lesions. Progression may result in involvement of lymph nodes, GI tract, and lungs.

Skin biopsy confirms the diagnosis, with slit-like spaces containing red cells, which are formed by atypical spindle cell proliferation of endothelial cells and are HHV-8 positive on immunohistochemistry. HHV-8 DNA and antibodies may also be detected in the blood.

Treatment

In localized disease, individual lesions may require treatment if they are symptomatic. Options include surgery, radiotherapy and cryotherapy. In PLWH, starting antiretroviral therapy will often result in regression of lesions. In immunosuppressed patients, reduction of immunosuppressive drugs or—in the case of organ transplant recipients—switching from calcineurin inhibitor to mTOR inhibitor immunosuppressive agents may also result in regression.

In advanced KS, treatment is with radiotherapy, chemotherapy (e.g., liposomal doxorubicin, paclitaxel), or immunotherapy (e.g., interferon-alpha), and there are ongoing clinical trials with immune checkpoint inhibitor immunotherapy.

Fig. 45.57 Kaposi's sarcoma. (A) Hyperpigmented macules and plaques. **(B)** Violaceous plaques and nodules.

Other rare skin tumours

Almost any structure in the skin can be affected by benign and malignant tumours. Rarer tumours that may be included in the differential diagnosis of the more common skin cancers include:

- ***Merkel cell carcinoma***, an aggressive neuroendocrine skin cancer that is much less common than melanoma but has a higher rate of metastasis and a 50% 5-year survival rate. It presents as an asymptomatic, enlarging, solitary, red-purple nodule (and may initially simulate other skin cancers such as cSCC but grows more rapidly; **Fig. 45.58A**). It is more common with increasing age, immunocompromise, and high UV exposure. Merkel cell polyomavirus infection is present in approximately 80% of cases. Treatment of the primary tumour usually includes combined surgery and radiotherapy. The anti–programmed death-ligand 1 checkpoint inhibitor avelumab is used in patients with metastatic disease.
- ***Appendageal tumours*** derive from epidermal appendages (hair follicles, sweat glands) and may be benign or malignant. Clinical appearances vary widely, but they often present as solitary nodules that may ulcerate (e.g., eccrine porocarcinoma). They are only infrequently diagnosed before excision.

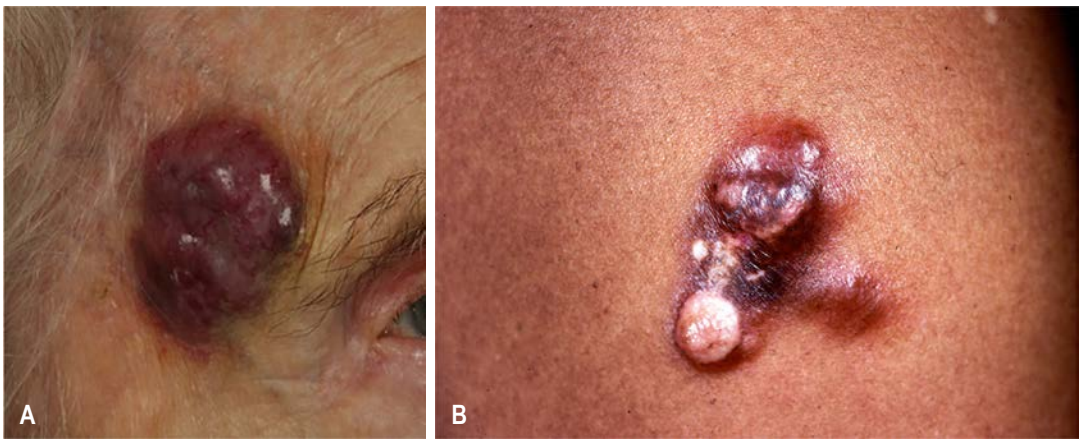


Fig. 45.58 **(A)** Merkel cell carcinoma: rapidly growing, large purple nodule of the face. **(B)** Dermatofibrosarcoma protuberans on the trunk.

- ***Dermatofibrosarcoma protuberans (DFSP)*** is a rare sarcoma arising in the dermis that presents as a rubbery, firm plaque or nodule and may be pigmented in darker skin tones in which it is also more common (Fig. 45.58B). It is most often located on the shoulder and chest areas and may simulate a hypertrophic or keloid scar. It grows slowly and often recurs after surgery but rarely metastasizes. Treatment includes surgery and radiotherapy, and in advanced disease, the tyrosine kinase inhibitor imatinib has a 50% response rate.
- ***Cutaneous angiosarcomas*** are rare and aggressive tumours of endothelial cells that occur most frequently on the head and neck of older people. They may initially present as bruise-like patches but rapidly evolve into multifocal, ulcerated, bleeding plaques and nodules. Treatment is usually with surgery or radiotherapy, but the prognosis is poor.

Further reading

Keohane SG, Botting J, Budny PG et al. British Association of Dermatologists guidelines for the management of people with cutaneous squamous cell carcinoma 2020. *Br J Dermatol.* 2021; 184(3):401–414.
Nasr I, McGrath EJ, Harwood CA et al. British Association of Dermatologists guidelines for the management of adults with basal cell carcinoma 2021. *Br J Dermatol.* 2021; 185(5):899–920.
https://melanomafocus.org/about-melanoma/.
http://www.sunsmart.org.uk. Information from Cancer Research UK on sun, UV and cancer.

DISORDERS OF PIGMENTATION

Hypopigmentation

Vitiligo

Vitiligo is a chronic depigmenting skin disorder that affects 1% to 2% of the world population; about one-third of all cases start in childhood. It is caused by autoimmune T cell–mediated destruction of melanocytes and is associated with other autoimmune disease, especially that of the thyroid. Vitiligo presents with asymptomatic, symmetrical, well-demarcated macules of complete pigment loss and typically affects the face, hands and feet, genitalia, and bony prominences (Fig. 45.59). Pigment may be lost from overlying hair, which is a poor prognostic sign because the stem cells for melanocytes (the source of repigmentation) are located in the hair follicles. Lesions are not inflamed.

Fig. 45.59 **Vitiligo of the hands.** Areas of depigmentation.

Management

Treatment is often unsatisfactory (because repigmentation is slow to occur) and has no impact on the long-term outcome. Potent topical corticosteroids, topical tacrolimus, and UV phototherapy may stimulate repigmentation, and this usually starts around hair follicles, giving a speckled appearance. Repigmentation is less likely in established lesions and acral sites because they lack hair follicles. Sunscreens should be used to prevent burning, although, paradoxically, there is no increased risk of skin cancer in vitiligo-affected skin (in contrast to oculocutaneous albinism; see later), possibly the result of increased immunosurveillance caused by autoimmunity. Patients may benefit from psychological support and skin camouflage advice.

Post-inflammatory hypopigmentation

Pale patches may be left after resolution of an inflammatory condition, especially eczema and psoriasis. They are more noticeable in people with darker skin tones or who are tanned and can be mistaken for vitiligo, but lesions are *hypo*pigmented rather than *de*pigmented. Pigmentation recovers over several months if the underlying condition has been treated. It can coexist with post-inflammatory hyperpigmentation (see later). ***Pityriasis alba*** is a common hypopigmented variant of atopic eczema seen on the cheeks and arms of

dark-skinned children, in whom there is little inflammation but prominent pallor. Hypopigmented patches are also common in ***pityriasis versicolor*** and may occur in ***mycosis fungoides***.

Leprosy

Tuberculoid leprosy and indeterminate leprosy (see Chapter 42, 'Leprosy [Hansen's disease]' section) can present with anaesthetic hypopigmented patches and should always be considered in people from endemic regions. Lesions may also show hair loss and decreased sweating.

Oculocutaneous albinism

This rare group of autosomal recessive disorders is caused by reduced or absent pigment synthesis in the melanocytes in the skin, hair, and eyes. It can affect all ethnic groups. Individuals have pale skin, white or yellow hair, and pink irises. Ocular manifestations include photophobia, nystagmus, and a squint. Meticulous UV protection is required to prevent sunburn and reduce the risk of skin cancer (in contrast to vitiligo; see earlier).

Hyperpigmentation

Freckles (ephelides)

These extremely common pigmented macules appear in childhood after sun exposure. They fade in the winter. They are commonly associated with red hair and easy sunburning and can be a clinical sign of higher skin cancer risk.

Lentigines

These persistent pigmented macules look similar to freckles, but they do not fade in winter. They may rarely be associated with systemic syndromes (e.g., Peutz–Jeghers; see Chapter 22, 'Peutz–Jeghers syndrome' section; Noonan syndrome or LEOPARD [*l*entigines, *E*CG abnormalities, *o*cular hypertension, *p*ulmonary stenosis, *a*bnormalities of genitalia, *r*etardation of growth and *d*eafness]). Solar lentigines ('age spots', 'liver spots') are common on the dorsal hands and face and on bald scalps in older people.

Café-au-lait macules

These are common flat brown birthmarks and may occur as an isolated abnormality with no clinical significance. Multiple lesions (five or more) can be a feature of some genodermatoses such as neurofibromatosis types 1 and 2, tuberous sclerosis, ataxia telangiectasia, multiple endocrine neoplasia type 1, and McCune–Albright syndrome.

Melasma (chloasma if it occurs in pregnancy)

This is a common condition in pregnant females and in females taking hormonal contraception. Asymptomatic beige-brown patches develop on the forehead, temples, and cheeks. Topical azelaic acid, retinoic acid, or 2% to 5% hydroquinone may help reduce pigmentation. Meticulous sun protection, including protection against the high-energy visible blue part of the spectrum (using sunscreen containing iron oxide), is needed to prevent relapse.

Post-inflammatory hyperpigmentation

This phenomenon occurs in dark-skinned individuals at the sites of skin trauma or inflammatory rashes, such as acne, psoriasis, LP, and eczema. It improves slowly over many months. It can occur at the same time as post-inflammatory hypopigmentation (see earlier) but is more common.

Metabolic and endocrine effects

Generalized skin darkening can occur with chronic liver disease, especially haemochromatosis (see Chapter 23, 'Hereditary haemochromatosis' section), and endocrine disease such as Addison disease.

Urticaria pigmentosa (cutaneous mastocytosis)

This disorder is caused by a benign proliferation of cutaneous mast cells. It presents most commonly in childhood as multiple pigmented macules that become red, itchy, and urticated if they are rubbed (Darier sign), and they occasionally blister. Extensive mast cell degranulation can lead to systemic symptoms, such as wheeze; flushing; syncope; diarrhoea; and, very rarely, anaphylaxis.

This may be triggered by drugs, including aspirin and opiates, and allergens. Childhood urticaria pigmentosa usually resolves spontaneously, but adult-onset disease is often persistent. Histology of skin lesions shows increased numbers of mast cells. Most cases are caused by somatically acquired activating mutations of the *KIT* receptor that controls mast cell proliferation and apoptosis. Rarely, in adult and neonatal disease, mast cell infiltration may involve the internal organs (bone, GI tract, liver, spleen—***systemic mastocytosis***). There is a small risk of developing mast cell leukaemia. Treatment of cutaneous mastocytosis is aimed at controlling the symptoms of mast cell mediator release with antihistamines and cromoglicate and minimizing the risk of anaphylaxis.

Further reading

Ezzedine K, Eleftheriadou V, Whitton M et al. Vitiligo. *Lancet.* 2015; 386:74–84.
Hill JP, Batchelor JM. An approach to hypopigmentation. *BMJ.* 2016; 355:i6534.

NAIL DISORDERS

- ***Psoriasis***, ***fungal infection***, and ***trauma*** are the most common causes of abnormal nail growth (dystrophy).
- ***Pitting*** can be associated with psoriasis, alopecia areata, atopic eczema, and trauma.
- ***Trachyonychia*** is roughening of the nails and can be seen in inflammatory conditions and in so-called 20-nail dystrophy.
- ***Onycholysis*** (distal nail plate separation) is caused by psoriasis; thyrotoxicosis; trauma; and, rarely, a phototoxic drug reaction (e.g., with tetracyclines).
- ***Onychoschizia*** is splitting of the distal end of brittle nails. It is cause by damage from wet work, heat, or chemicals such as nail varnishes. It is made worse with vitamin D deficiency.
- ***Koilonychia*** (thin, spoon-shaped nails) can be caused by iron-deficiency anaemia.
- ***Leuconychia*** (white nails) is seen in hypoalbuminaemia (e.g., in chronic liver disease).
- ***Beau lines*** (transverse lines) are horizontal grooves in the nail caused by a temporary growth arrest associated with an acute severe illness.
- ***Yellow-nail syndrome*** is a rare disorder based on the triad of thickened, slow-growing, yellow nails; pulmonary manifestations (cough, bronchiectasis, pleural effusion); and lower limb lymphoedema.
- ***Subungual hyperkeratosis*** is thickening of the nail plate, usually caused by tinea infection, psoriasis, trauma, or a combination of these. Occasionally, it can be the presenting feature of nail bed SCC.

- ***Onychogryphosis*** is severe nail thickening and curvature (resembling a ram's horn), which is common in older adults, especially in the big toenail, where trauma from ill-fitting footwear may be relevant.
- ***Longitudinal melanonychia*** (hyperpigmented streaks) in multiple nails is a common finding in darker-skinned people. An acquired solitary pigmented streak may be caused by a melanocytic naevus and needs to be differentiated from subungual melanoma, especially if the pigmentation spreads on to the adjacent nail fold **(Hutchinson sign)** or if it causes destruction of the nail plate.
- ***Subungual haemorrhage*** is common in the great toenails after trauma (football, running downhill, tight footwear). The red-brown pigmentation in located under the nail plate and grows out with the nail over several months with clear proximal growth. It might be persistent if the trauma continues.
- ***Clubbing*** is discussed in Chapter 17 ('Examination of the cardiovascular system' section).
- ***Nail dystrophy*** is also a feature of various genodermatoses, including ***nail patellar syndrome, ectodermal dysplasias*** (abnormal hair, teeth, and nails), and ***pachyonychia congenita***.
- Benign tumours around the nail (e.g., **myxoid cyst**) can exert pressure on the nail matrix and cause a longitudinal groove in the nail.

Further reading

Lee DK, Lipner SR. Optimal diagnosis and management of common nail disorders. *Ann Med.* 2022; 54(1): 694–712.

HAIR DISORDERS

Hair loss (alopecia)

There are many causes of hair loss (alopecia) (Box 45.39). They are broadly divided into localized and diffuse patterns and into ***non-scarring alopecia*** (in which follicles are preserved) and ***scarring alopecia*** (in which an inflammatory process causes permanent destruction of follicles). Some forms of alopecia start as non-scarring but can become scarring with time. Patients may complain of increased hair shedding or thinning. Dermoscopic (trichoscopic) examination can help to identify whether follicles are still present, and histology of scalp biopsies is helpful when there is diagnostic uncertainty.

Box 45.39 Causes of alopecia

Scarring alopecia
- Discoid lupus erythematosus
- Trauma
- Traction (pulling) alopecia
- Lichen planus
- Sarcoidosis
- Dissecting cellulitis
- X-irradiation
- Idiopathic ('pseudopelade')

Non-scarring alopecia
- Androgenic alopecia (can scar in late stages)
- Telogen effluvium
- Alopecia areata
- Trichotillosis (self-induced hair-pulling)
- Tinea capitis and kerion
- Metabolic (iron deficiency, hypothyroidism)
- Drugs (e.g., heparin, isotretinoin, chemotherapy)

Male pattern and female pattern alopecia

Pattern alopecia is the most common cause of non-scarring hair loss and may be regarded as a feature of normal human ageing. It is thought to be caused by a genetically determined increased sensitivity to androgens (usually with normal blood androgen levels), and there is often a positive family history. It may present in adolescence with a receding frontal hairline followed by thinning over the vertex. Females may be similarly affected but usually later in life, with milder loss and preservation of the frontal hair margin. If acne, hirsutism, and menstrual disturbance are also present, there may be underlying polycystic ovary syndrome or an androgenic disorder.

Treatment options include topical or oral minoxidil or anti-androgens, including spironolactone (in female patients) or oral 5-alpha-reductase inhibitors such as finasteride in male patients. These can halt progression and may induce modest regrowth if used early in disease and continued indefinitely. Approximately one-third of patients do not respond to either therapy. Finasteride is well tolerated but may cause sexual adverse effects, such as loss of libido in about 1% of males; it should not be taken or handled by pregnant females because it can cause feminization of a male fetus. In stable pattern alopecia, hair transplantation may be helpful in restoring areas of hair loss using occipital hair units (which are not affected by hormonal influences).

Alopecia areata

Alopecia areata is a non-scarring hair loss disorder with a lifetime prevalence of about 2%. It is thought to be an organ-specific autoimmune disease of the hair follicle that occurs because of loss of hair follicle immune privilege. It is associated with other organ-specific autoimmune diseases and runs an unpredictable course ranging in severity from a temporary small, solitary patch of scalp hair loss ***(alopecia areata)*** to total loss of scalp hair ***(alopecia totalis)*** and rarely, scalp and body hair ***(alopecia universalis)***. Hair loss can have a major impact on body image and quality of life, with associated depression and social isolation.

Clinical features

Alopecia areata usually presents in children or young adults with patches of baldness (Fig. 45.60). There is very often psychological stress that triggers the hair loss. These areas of hair loss may regrow spontaneously and in mild disease (<25% scalp involvement), there is an 80% chance of regrowth within 2 years. However,

Fig. 45.60 Alopecia areata.

it is not unusual for regrowth to be followed by new areas of hair loss. The presence of broken hairs and ***exclamation mark hairs*** (narrow at the scalp and wider and more pigmented at the tip) at the edge of a bald area is diagnostic. The nails may be pitted or roughened. The diagnosis is usually based on clinical features and trichoscopy (dermoscopic imaging of the scalp and hair). Skin biopsy may occasionally be required if there is any diagnostic uncertainty.

Management

Treatment has no effect on the long-term disease course. Potent topical corticosteroids in children or injected steroids in adults may trigger regrowth of localized patchy hair loss. TCIs may also be helpful. PUVA phototherapy or repeated application of a contact sensitizer (contact immunotherapy) such as diphencyprone can be helpful in more extensive disease. Oral JAK inhibitors can reverse the hair loss in 60% of moderate to severe alopecia areata, but the rate of relapse is around 50% when treatment is stopped. In the UK, patients with extensive hair loss are eligible for prescribed wigs and may benefit from joining patient support groups such as Alopecia UK.

Traction alopecia

This form of scarring hair loss is caused by mechanical damage to the hair follicle from pulling the hair back into a bun or tight plaits. It can be reversible initially, but continued mechanical damage can cause permanent scarring.

Telogen effluvium

This type of diffuse, non-scarring hair loss is associated with increased hair shedding and usually presents 2 or 3 months after pregnancy, severe illness, sudden weight loss, or severe psychological stress. It occurs when a larger percentage of scalp hairs enters the telogen phase of hair shedding at the same time. Spontaneous recovery is usual after a few months, though a chronic variant is recognized, particularly if the trigger is unresolved.

Metabolic causes of hair loss

Iron deficiency may be associated with increased hair shedding and hair thinning and is a common finding in menstruating females who do not eat red meat. The exact relationship between serum iron or ferritin levels and optimum hair growth is unclear. Untreated hypothyroidism is also associated with hair loss, particularly of the outer third of the eyebrows ***(madarosis)***.

Dissecting cellulitis

This is a chronic scarring folliculitis of the scalp that predominantly affects males with afro-textured hair and may sometimes be associated with severe acne and hidradenitis suppurativa ***(follicular occlusion syndrome)***. Crusted papules and pustules with underlying diffuse swelling cause scarring on the scalp (Fig. 45.61). Prolonged courses of anti-staphylococcal oral antibiotics such as rifampicin and clindamycin may be helpful. Oral isotretinoin can help control the disease (after antibiotic treatment has been completed), but there is a high rate of recurrence after stopping it.

Increased hair growth

Hirsutism

Hirsutism refers to a male pattern of hair growth in females. Ethnic variation in hair growth must be considered because people of certain ethnicities (e.g., African, Caribbean, and Asian) commonly have more male-pattern hair growth than Northern European females. If virilizing features (deep voice, clitoromegaly, dysmenorrhoea, acne) are present, a full endocrine assessment is necessary. Hirsutism may cause severe psychological distress in some individuals.

Management involves physical methods, such as bleaching, waxing, electrolysis, intense pulsed light, and laser therapy. Oral anti-androgen therapy is occasionally helpful.

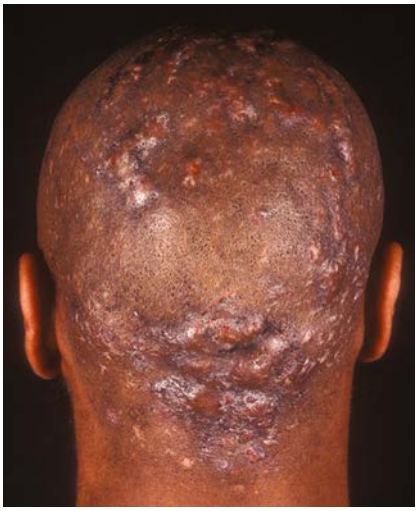


Fig. 45.61 Dissecting cellulitis of the scalp in an African-Caribbean male patient.

Hypertrichosis

Hypertrichosis refers to excessive hair growth at any site and occurs in both sexes. It can be seen in anorexia nervosa, PCT, and underlying malignancy. It is also caused by certain drugs (e.g., ciclosporin, minoxidil) and can be a side-effect of potent topical steroids

Further reading

Workman K, Piliang M. Approach to the patient with hair loss. *J Am Acad Dermatol.* 2023; 89(2S):S3–S8.

BIRTH MARKS AND NEONATAL RASHES

Various skin lesions may be present at birth or during infancy; most resolve during childhood, although ***congenital melanocytic naevi*** persist lifelong.

Port-wine stain (naevus flammeus)

Port-wine stain ('capillary haemangioma') is not a true haemangioma but an abnormal dilation of dermal capillaries. It presents at birth as a flat, red area and is commonly found on the face. It does not improve and may become thickened with age. If the lesion occurs in the distribution of the first division of the trigeminal nerve, it may be associated with ipsilateral meningeal vascular anomalies that can cause epilepsy and even hemiplegia (Sturge–Weber syndrome). Periocular lesions may be associated with glaucoma, so ophthalmological follow-up is required.

Box 45.40 Severe cutaneous adverse drug reactions

- Stevens–Johnson syndrome and toxic epidermal necrolysis
- Drug reaction with eosinophilia and systemic symptoms
- Acute generalized exanthematous pustulosis
- Erythroderma or exfoliative dermatitis
- Serum sickness–like reactions

Fig. 45.62 Morbilliform drug rash caused by penicillin allergy.

Treatment of port-wine stains is with a vascular laser, for example, a pulsed-dye laser. Facial lesions respond best, but lesions can darken after several years and require retreatment.

DRUG ERUPTIONS

Cutaneous adverse drug reactions are common and range from mild and predictable dose-related effects such as phototoxicity from doxycycline or mucocutaneous dryness from oral isotretinoin to idiosyncratic, severe, and life-threatening eruptions (severe cutaneous adverse drug reactions; **Box 45.40**). These life-threatening diagnoses are described in Chapter 11 ('Severe cutaneous adverse drug reactions' section). The diagnosis of drug eruptions can be challenging because individual drugs can cause different rashes, and these may mimic constitutional inflammatory rashes or viral rashes. The clinical features can also be mixed, and patients may be receiving multiple drugs that are potential culprits.

A high index of suspicion and a detailed drug history are essential. Underlying viral infection, especially HIV, and systemic disease, such as SLE and leukaemia, may also increase the risk of drug rashes. Allergy tests (prick tests and patch tests) have a limited role, especially in the acute setting, and drug challenge tests are time consuming and potentially risky.

Maculopapular (morbilliform) exanthems

These are the most common type of hypersensitivity rash (**Fig. 45.62**). They start on the trunk and spread to the face and limbs but spare the mucosae. They are self-limiting and usually clear within 1 to 2 weeks.

Box 45.41 Dermatoses induced or exacerbated by drugs

Dermatosis	Drugs
Acne	Androgens (e.g., danazol)
Acneiform	Corticosteroids, EGFR inhibitors
Angio-oedema	ACE inhibitors
Urticaria	Penicillin, aspirin, NSAIDs, iodine contrast media
Vasculitis	Gold, hydralazine, NSAIDs, PPIs
Fixed drug eruption	Phenolphthalein in laxatives, tetracyclines, paracetamol
Pigmentation	Minocycline (black), amiodarone (slate-grey)
Vitiligo	Anti–PD-1/PD-L1 immunotherapy
Lupus erythematosus	Minocycline, anti-TNF biologics, isoniazid, interferons, anti–PD-1/PD-L1 immunotherapy
Photosensitivity	Thiazides, quinolones, tetracyclines, diuretics, amiodarone
Pseudoporphyria	Naproxen, diuretics
Leg ulcers	Hydroxyurea
Anogenital ulcers	Nicorandil
Erythema nodosum	Sulphonamides, oral contraceptive
Erythema multiforme	Barbiturates, etravirine
Lichen planus-like (lichenoid)	Antimalarials, thiazides, statins, diuretics, anti–PD-1/PD-L1 immunotherapy
Psoriasiform	Methyldopa, gold, lithium, beta-blockers, anti–PD-1/PD-L1 immunotherapy
Toxic epidermal necrolysis	Penicillin, co-trimoxazole, carbamazepine, NSAIDs, nevirapine, efavirenz
Bullous pemphigoid	Anti–PD-1/PD-L1 immunotherapy
Pemphigus	Penicillamine, ACE inhibitors
Erythroderma	Gold, sulphonylureas, allopurinol, nevirapine, efavirenz

ACE, Angiotensin-converting enzyme; *EGFR,* epidermal growth factor receptor; *NSAID,* non-steroidal anti-inflammatory drug; *PD-1,* programmed cell death-1; *PD-L1,* programmed death-ligand 1; *PPI,* proton pump inhibitor; *TNF,* tumour necrosis factor.

Fixed drug eruptions

These inflamed patches recur at the same site each time a drug is taken. They may blister and often resolve with hyperpigmentation.

Drug-induced and drug-exacerbated dermatoses

Drugs may exacerbate pre-existing skin disease and can also trigger or induce skin disease that resembles a constitutional dermatosis (**Box 45.41**).

Severe cutaneous adverse reactions

SJS, TEN, DRESS, and acute generalized exanthematous pustulosis are discussed in Chapter 11 ('Severe cutaneous adverse drug reactions' section).

Further reading

Creamer D, Walsh SA, Dziewulski P. UK guidelines for the management of Stevens Johnson syndrome/toxic epidermal necrolysis in adults 2016. *Br J Dermatol.* 2016; 174:1194–1227.

PSYCHODERMATOLOGY

Psychodermatlogy is a newer and emerging subspecialty of dermatology. There are two main categories of disorders, as discussed next.

Primary psychiatric disease presenting to dermatology healthcare professionals

For example:

- ***Delusional infestation:*** Patients believe that they have an infestation of the skin when there is no infestation (i.e., a primarily persistent delusional disorder). Commonly, this may be associated with recreational drug usage (e.g., cocaine, cannabis, or amfetamines).
- ***Body dysmorphic disease:*** Patients have an overstated concern about a physical (usually cutaneous) change that may seem very trivial to healthcare professionals but tat for the patient is very much the focus of their attention. Patients may have social avoidance or hide their skin from others. It is usually classified as an obsessive-compulsive disorder.
- ***Dermatitis artefacta:*** Patients harm or neglect themselves in some way, often mimicking a skin disease. This is different from deliberate self-harm because patients with dermatitis artefacta do not know that they are harming themselves.

Primary dermatological disease in which there are psychosocial co-morbidities

For example:

- Changes in self-perception
- Anxiety
- Depression
- Anxiety and depression (more common than anxiety or depression in isolation)
- Suicidal ideation

Bibliography

Griffiths C, Barker J, Bleiker T et al (eds). *Rook's textbook of dermatology*, 9th edn. Chichester: John Wiley & Sons; 2016.

Significant websites

http://dermnetnz.org/ *Dermatology information and images*.
http://hardinmd.lib.uiowa.edu/dermpictures.html *Dermatology images (atlas)*.
http://www.aad.org *American Academy of Dermatology*.
http://www.bad.org.uk *British Association of Dermatologists*.
http://www.debra.org.uk *Dystrophic Epidermolysis Bullosa Association*.
http://www.eczema.org *National Eczema Society*.
http://www.evidence.nhs.uk/ *National Institute for Health and Care Excellence (NICE): clinical guidelines, systematic reviews, evidence-based synopses, image database*.
http://www.pcds.org.uk *Primary Care Dermatology Society*.
http://www.psoriasis-association.org.uk *Psoriasis Association*.
http://www.skin-camouflage.net *British Association of Skin Camouflage*.
http://www.thecochranelibrary.com *Cochrane Library: systematic reviews of treatment*.
http://www.vitiligosociety.org.uk *Vitiligo Society*.
https://www.nice.org.uk/guidance *Dermatology treatment guidelines in the UK*.
www.eadv.org
www.psychodermatology.co.uk
www.skinhealthinfo.org.uk

Questions for this chapter can be found on Elsevier eBooks+.

46 Young adult medicine

Arvind Nagra, Gary James Connett, Matthew Harmer, and Yincent Tse

Background 1631
Mortality and morbidity 1632
Mental health issues in young people 1632
Eating disorders 1633
Bullying 1633
Screen time and social media 1633
Lifestyle choices in young people 1634
Nutrition and obesity 1634
Physical activity 1635
Smoking and vaping 1635
Alcohol 1635
Recreational drugs 1635
Gambling 1636
Vaccination 1636
Sleep 1636
Long-term health conditions 1636
Shared decision-making 1638
Health literacy 1638
Other considerations when addressing the needs of young people 1638
Neurodiversity 1638
Sexual and reproductive health 1638
Menstruation 1638
Gender identity 1638

CORE SKILLS AND KNOWLEDGE

Healthcare professionals providing for the medical needs of young people need effective communication skills. Active listening, motivational interviewing techniques, and a non-judgemental approach are key to putting young people at their ease. Healthcare professionals need to create an environment in which young people feel safe talking about sensitive issues.

A broad understanding of general medicine is essential, with a more detailed knowledge of medical conditions that are more likely to occur in this age group.

Illnesses pertaining to mental health are common healthcare problem and are important co-morbidities complicating other medical conditions. Doctors treating young people in this age group need to be aware of relevant resources that are available to help meet this need in their patients.

When managing chronic conditions, doctors need to consider the way in which maturational development of the brain increases tendency to risk-taking behaviours, which can result in worse long-term outcomes. Adherence is a particular challenge. Doctors need to have strategies to ask about and address this.

Comprehensive care for young people requires collaborative working with other professionals. Psychologists and social workers have major roles in this age group. Doctors need a good understanding of the roles of the wider multiprofessional team and how best to use their skills and services.

Successful care of young people facilitates transition to adult care and the promotion of healthy lifestyle choices that can impact long-term health.

BACKGROUND

Adolescents and young people in early adult life face unique healthcare challenges. Adolescence is typically defined as the period from puberty to 18 years of age and young adulthood from age 18 to 25 years, although individuals mature over different time scales. Young people's health is an area of medicine focusing on the specific needs of these age groups. It is a time of significant physiological and psychological change, compounded by peer pressures and social, romantic, and sexual relationships (**Fig. 46.1**). There is a major reorganization of neural pathways during adolescence that continues into early adult life (**Fig. 46.2**).

A tendency to risk-taking behaviour is common and can result in poor engagement with healthcare and adherence to medical treatment, as well as misadventure. Such behaviours occur at a time of increased autonomy and independence from parents and guardians, including capacity to give informed consent for medical treatments and procedures (**Box 46.1**).

In 2019, young people made up 18% of the UK population. They are more ethnically diverse, less likely to identify as heterosexual, and more gender fluid than older age groups. Socioeconomic and ethnic disparities have a significant impact on their health outcomes. Young people from low-income families and marginalized groups often face barriers accessing healthcare, nutrition, and safe environments.

Healthcare professionals have an important role in helping young people achieve good physical and mental health outcomes. At a societal level, this might include advocating for improvements in housing, education, and employment opportunities. At a patient level, this is best achieved through multidisciplinary teams that have the skills to address the specific health needs of this age group.

Biological	Psychological	Social
• Breast development (stage 2–3 at ages 11–12) • Penis enlargement • Increase in testicular volume (10 mL at ages 13–14) • Development of pubic and axillary hair • Menarche • Growth spurt • Fat deposition • Changes in body shape • Voice maturation in males • Accretion of bone mineral density • Acne • Blushing • Increased need for sleep • Organ maturation	• Thinking changes from concrete to abstract (logic) • A sense of invincibility • Development of sexual identity and increased experimentation • Reassessment of body image • Growing educational abilities and development of vocational ideas • Learning to differentiate between laws and morality • Learning to control impulsivity • The development of passionate views	• Emotional separation from parents • Peer identification • Exploratory risk-taking behaviours (e.g., smoking, violence, sex, illicit drug use) • Romantic/sexual relationships • Vocational plans • Financial independence

Fig. 46.1 Developmental changes during adolescence. Extensive biological, psychological, and social changes occur in adolescence and early adult life.

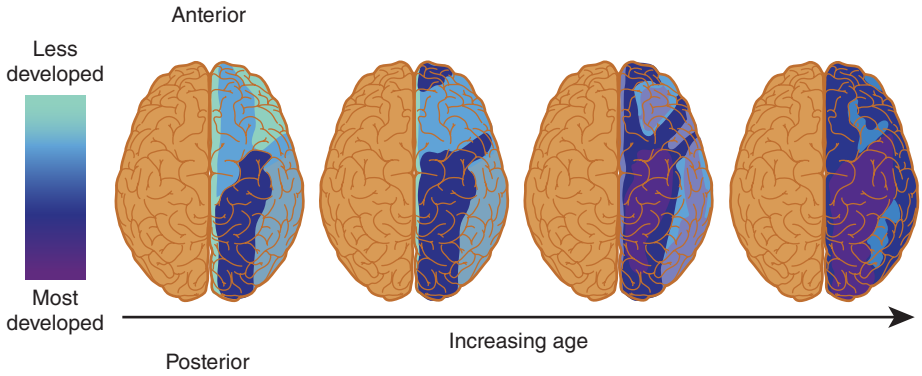


Fig. 46.2 Brain development and maturation after completion of brain growth. Although the brain stops growing by early adolescence (about 11 years old), there is ongoing maturation (from posterior to anterior) and neurological 'fine-tuning', finishing typically after the age of 25 years. The prefrontal cortex, responsible for thought and emotional regulation, is one of the last parts to mature.

Identifying and responding to the specific needs of young people can be challenging. HEeADSSS (**Box 46.2**) is an acronymic interview screening tool for use with young people attending healthcare services. It helps identify a wide range of potential psychosocial protectors and risks that may warrant further enquiry. It can be completed with a healthcare professional during consultation or used as a self-assessment tool before consultation.

Further reading

Hagell A, Wortley L, Ross L, Whitaker E. Key Data on Young People 2021: Overview and Policy Implications. 2021. https://ayph-youthhealthdata.org.uk/key-data/; and https://ayph-youthhealthdata.org.uk/wp-content/uploads/2021/11/KDYP_OverviewAndPolicyImplications.pdf.
Royal College of Paediatrics and Child Health. *State of Child Health in the UK*. https://stateofchildhealth.rcpch.ac.uk/.

MORTALITY AND MORBIDITY

Deaths among young people are fortunately rare and represent less than 1% of all-age deaths. In the UK, mortality rates fell consistently until 2015 before plateauing in all age groups, except in those aged 20 to 25 years, in which it subsequently increased. Males are at greater risk of death than females (**Fig. 46.3**). The most common cause of death is accidents. Death caused by vehicle collisions a particular concern: young people aged 10 to 14 years are at risk as pedestrians and older age groups as drivers and passengers (**Fig. 46.4**). Death by suicide is the second most common cause of death in young people, killing more than 5000 young adults in the UK in 2019. Globally, tuberculosis is a major killer of young people in low-resource settings.

MENTAL HEALTH ISSUES IN YOUNG PEOPLE

Academic pressures, social media influence, financial pressures, and a competitive job market all contribute to anxiety, depression, and stress among adolescents and young adults. Increased isolation, as occurred through SARS-CoV-2 restrictions, has disproportionately affected young people's mental health. Twelve percent of young people aged 16 to 24 years self-report feeling isolated in the past 12 months. This was higher than the 9% reported among those aged older than 85 years. Access to mental health services, school-based counselling programmes, and fostering open conversations to reduce the stigma surrounding mental health illness are supportive approaches to this problem. More recently, medical

Box 46.1 Consent and the law for young people in the UK

Consent in the context of healthcare is the process of agreeing or giving permission to have treatment. It should be achieved as part of shared decision-making. Consent can be non-verbal (e.g., attendance for a low-risk procedures such as radiography, holding out an arm for venepuncture), verbal with documentation of information given and discussion of risk (e.g., blood transfusion), or written by signing a consent form.

Young people younger than 18 years of age might make life choices that conflict with their parents or guardians. For example, they may want to access contraception or seek confidential support for substance misuse but do not want their parents or other carers to know.

The law makes provision to protect young people from making decisions that lead to serious permanent harm. This means that there are specific considerations in relation to age and consent.

Aged older than 16 years old

Gillick competency and Fraser guidelines underpin consent in young people aged younger than 16 years. If judged capable of decision-making, young people younger than 16 years old can consent to treatment, and an agreement to treatment cannot be overruled by someone with parental responsibility. Even if judged competent, refusal of treatment can be overruled by a person with parental responsibility if healthcare professionals think that treatment is in their best interests.

If a person younger than 16 years is judged incapable of deciding because of emotional maturity or learning disabilities, a person with parental responsibility can give permission on their behalf.

Aged 16 or 17 years

For anyone aged 16 years or older, the ability to consent is governed by the Mental Capacity Act 2005. The law states that young adult aged 16 years or older must give consent unless they are not able decide. Those with parental responsibility can give consent on behalf of a 16- and 17-year-olds only if they lack capacity, and an assessment of capacity should be carried out and documented.

If a competent 16- or 17-year-old patient refuses a treatment that can save them from death, severe permanent injury, or irreversible mental or physical harm, refusal can be overridden by a court.

Aged 18 years or older

Young people aged 18 years or older are adults. Parents cannot give consent even the patient cannot consent for themselves. The only exception to this is if parents have applied successfully for an Order from the Court of Protection. In an emergency, the treating team will make a decision according to the person's best interests.

Who has parental responsibility

Parental responsibility includes the legal rights, duties, powers, responsibilities, and authorities parents have in relation to their children and their children's property. A mother automatically has parental responsibility for her child from birth. A father usually has parental responsibility if he is either married to the child's mother or is listed on the birth certificate. Same-sex partners both have parental responsibility if they were civil partners at the time of artificial insemination (after 6 April 2009) but not if conception occurred after intercourse with a male. Others can apply for parental responsibility via legal routes. In safeguarding cases, the local authority may be granted parental responsibility by a court.

What should be discussed in consent

The judgement in *Montgomery v Lanarkshire Health Board* [2015] United Kingdom Supreme Court (UKSC) 11 guides what should be discussed in the consent process: The doctor is under a duty to take reasonable care to ensure that the patient is aware of any material risks involved in any recommended treatment and of any reasonable alternative or variant treatments. Whether a risk is material does not solely depend on how frequently it occurs. Understanding what matters to the patient ensures the patient understands the risk and benefits of treatment for them. For example, for a lung transplant operation, the position of the scar may be important to a young person.

Sharing information

Sharing health information with parents of competent young people younger than 18 years old must be with their explicit agreement. This is particularly relevant to letters automatically sent to a home address.

When a competent young person refuses to allow information to be shared with their parents, there should be evidence that the risks of not sharing the information have been considered. When it is in the young person's best interests to share information, there should be attempts to compromise. If this is unachievable, legal and safeguarding advice should be sought.

An example of how to discuss confidentiality with young people states: 'Everything we talk about today is confidential. That means I cannot tell other people, even your parents, unless you agree to my sharing information with them. The only exceptions would be if I thought you or someone else was at risk of serious harm. In that case, I would need to tell someone else to keep everyone as safe as possible'.

professionals have had the option of prescribing activities such as exercise, music clubs, and social get togethers. This is termed 'social prescribing' and has been shown to have multiple benefits for young people; for example, joining an organized walking group helps both mental and physical wellbeing. The effects also often last beyond the prescribed period and do not have the risks associated with prescribing medication.

Eating disorders

Anorexia nervosa, bulimia nervosa, and binge eating disorder are psychiatric diseases in which there are severe alterations of body image and eating behaviour. Anorexia and bulimia typically occur in the later teenage years and binge-eating disorders in early adult life. Treatment typically includes a combination of medical, nutritional, and psychological interventions. BEAT is a UK-based eating disorder charity with useful resources for those who are affected as well as their families and friends (https://www.beateatingdisorders.org.uk/).

Bullying

One in five secondary school-aged children experiences bullying. Males are more often involved in physical bullying and females in verbal and cyberbullying. Bullying can cause significant physical, psychological, social, and educational harm.

Those with protected characteristics are particularly susceptible; these include ethnic and religious minority groups, people with physically disabilities, LGBTQ+ individuals, travellers, and asylum seekers. Those who are very thin or obese are also at increased risk. Being bullied results in poor school attendance and reduced academic achievement. Victims are more likely to experience depression and anxiety and are at increased risk of suicide. Bullies are more likely to become known to the criminal justice system, abuse drugs and alcohol, conduct antisocial behaviour, and commit domestic violence than those who do not exhibit bullying behaviours.

Young people are particularly susceptible to cyberbullying by email, text messages, or social media. Teaching good Internet etiquette and maintaining open dialogue with young people about their Internet experiences are key.

Screen time and social media

The overwhelming majority of 12- to 15-year-olds own a smartphone. The increasing use of technology and social media has both positive and negative health effects. It provides opportunities for learning and connectivity, but excessive screen time is associated with sleep

Box 46.2 Identification of protective and risk psychosocial factors in young people. Adapted from the HEeADSSS tool, with suggested questions to facilitate use

Establish trust
- Explain the purpose of the assessment.
- Explain rights when receiving care.
- Obtain informed consent.
- Explain confidentiality but include the three safety exceptions: harming self, harming others, or someone else harming them.

HEeADSSS

Home and relationships
- Who lives at home with you?
- Do you have your own room?
- Who do you get on with best or fight with the most?
- Who do you turn to when you're feeling down?

Education and employment
- Are you in school or college? If so, which year?
- What do you like best or least at school or college?
- How are you doing at school?
- Do you have friends at school?
- How do you get along with others at school?
- What do you want to do when you finish school or college?
- Do you work? If so, how much?

Eating
- Are you worried about your weight or body shape?
- Have you noticed any change in your weight recently?
- Have you been on a diet?

Activities and hobbies
- How do you spend your spare time?
- What do you do to relax?
- What kind of physical activities do you do?

Drugs, alcohol, and tobacco
- Does anyone smoke at home?
- Have you been offered cigarettes?
- Do you smoke? If so, how many cigarettes or vapes per day?
- Have you tried or been offered alcohol? How much and how often?
- Have you tried cannabis? How much and how often?
- Have you tried any other drugs, for example, ecstasy and cocaine?

Sex and relationships
- How do you identify: as a boy, girl, other, or not sure?
- Are you seeing anyone?
- Are they male or female, and how do they identify?
- Do you have any intimate relationships? How are you managing this?
- Have you ever had sex?
- What contraception do you use?

Self-harm, depression, and self-image
- How is life going generally?
- What do you do when you feel stressed?
- Do you ever feel sad and tearful?
- Have you ever felt so sad that you feel that life is not worth living?
- Do you think about hurting or killing yourself?
- Have you ever tried to harm yourself?

Safety and abuse
- Do you feel safe at home and at school?
- Is anyone harming you?
- Is anyone making you do things that you do not want to do?
- Have you ever felt unsafe online or when using your phone?

disturbance, cyberbullying, and a lack of physical activity. Engaging with social media represents the largest proportion of screen time use in young people. Regular use of social media is reported by 65% of 8- to 12-year-olds and 91% of 13- to 18-year-olds. Inappropriate content or content that caused worry or upset has been seen by 45% of 8- to 17-year-olds, and in half of cases, this was not reported

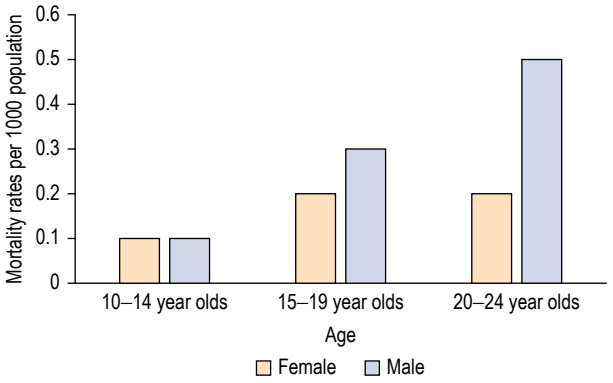


Fig. 46.3 Age- and sex-specific mortality rates in young people in England and Wales in 2019.

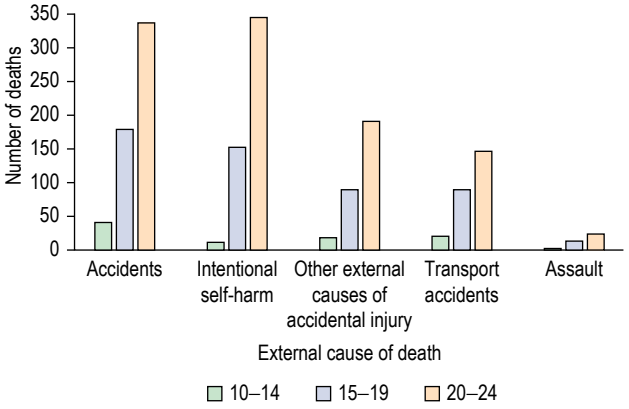


Fig. 46.4 Age-specific causes of death in young people in England and Wales in 2019.

to a responsible adult. More than 10% of 13- to 15-year-olds report receiving a sexual message within the past year.

The Royal College of Paediatrics and Child Health recommends that families should negotiate screen time limits with their children considering disruptions to physical activity, social life, and sleep. Screens should be avoided for 1 hour before a planned bedtime. The National Institute for Health and Care Excellence suggests that children should have days without any TV or have a 2-hour screen time limit.

Further reading

Children's Commissioner. *Digital Childhoods: A Survey of Children and Parents*. 2022. Accessed 21/06/2023: https://www.childrenscommissioner.gov.uk/resource/digital-childhoods-a-survey-of-children-and-parents/.
RCPCH. *The Health Impacts of Screen Time—A Guide for Clinicians and Parents*. Accessed online on 21/06/2023: https://www.rcpch.ac.uk/resources/health-impacts-screen-time-guide-clinicians-parents.
www.nspcc.org.uk/keeping-children-safe/online-safety/social-media.

LIFESTYLE CHOICES IN YOUNG PEOPLE

Achieving and maintaining a healthy lifestyle in young people is important for current health status and the long-term risk of illnesses including cancer, diabetes, and cardiovascular disease.

Nutrition and obesity

Contemporary diets are more likely to be calorie dense and nutrient poor. The average consumption of fruit and vegetables for females

aged 11 to 18 is 2.7 portions per day, with just 7% achieving the recommended five portions per day. Nearly half of all young females are not ingesting the recommended amount of iron. Males aged 11 to 18 years eat an average of 3.1 vegetable portions, but only 13% achieve five-a-day recommendations.

Obesity is an important long-term problem with many potential adverse health consequences. of The prevalence is high: 14% of young people were overweight, and 21% were obese at 17 years of age according to data from 2018 to 2019. A meta-analysis of longitudinal studies identified that obese adolescents are five times more likely to be obese adults than non-obese adolescents. Around 70% of obese adolescents will still be obese after 30 years of age. Early intervention is therefore important. Obesity disproportionately affects those living in deprived areas and is more prevalent among ethnic minority groups and young people with learning disabilities.

Physical activity

Regular exercise has long-term health benefits. Worldwide, only 22% of male and 15% of female adolescents take part in at least 60 minutes of moderate- or vigorous-intensity physical activity a day across the week. Qualitative data show that teenagers value unstructured, local, low-cost, fun, sociable opportunities and the right facilities to be active.

Smoking and vaping

Most adult smokers began in childhood with the risk of nicotine addiction occurring within 4 weeks of starting. Trends for young people smoking traditional cigarettes have progressively declined with only 2% of those aged 11 to 15 years smoking regularly in 2018.

E-cigarettes (vapes, vaporizers, and electronic nicotine delivery systems) heat nicotine-containing solutions with additional flavourings and other additives that are held in a cartridge or tank, with vapour inhaled through a mouthpiece. Since their widespread introduction in the UK in 2007, e-cigarettes can be sold to anyone older the age of 18 years in the UK and are encouraged as a less dangerous product than traditional cigarettes. It is illegal to sell these products to those younger than 18 years of age but not to give or supply them. E-cigarettes are more successful in getting people to stop smoking traditional cigarettes than other nicotine replacement therapies, but it is controversial as to whether they are more effective than discontinuation.

Although considered less harmful than traditional cigarettes, potential dangers are associated with vaping. The long-term effects of toxins in e-cigarette vapours are unknown. E-cigarettes contain nicotine, which is highly addictive. Of relevance to young people, they may have an impact on concentration ability and have toxic effects on developing brains.

Vaping by young people is rapidly increasing. In 2023 in the UK, 20.5% of young people aged 11 to 17 years had tried vaping compared with 15.8% in 2022 and 13.9% in 2020. Regular or occasional vaping occurred in 7.8% in 2023 compared with just 1.4% in 2014. Trends are similar in the USA.

When talking to young people about smoking and vaping:

- Give clear, consistent messages about the known and unknown risks of smoking and vaping. These should include health risks but also the impact on skin condition, ageing, bad breath, and cost.
- Avoid threats and ultimatums.
- Acknowledge that family significantly influences behaviour.
- Discuss early in childhood. Addiction occurs quickly and may occur before the later teen years.
- Discuss whether peers smoke or vape and empower young people to refuse cigarettes and e-cigarettes.
- Explain that most smokers quit smoking successfully without the need for any type of nicotine replacement.
- Ask young people how they might have been targeted and manipulated by tobacco companies. Such understanding can modify views and help motivate smoking cessation.

Further reading

Action on Smoking and Health (ASH). *Young People and Smoking.* 2019. www.ash.org.uk.

Alcohol

In the UK, it is illegal to sell alcohol and for an adult to purchase alcohol for those younger than 18 years old. From age 16 years, young people can drink beer, wine, or cider with a meal if accompanied by an adult. Despite these legal barriers, consumption of alcohol in young people is common. Experimentation with drinking alcohol occurs as young people explore their environment and socialize.

Alcohol exposure in young people, especially regularly or heavily, is associated with poorer neurological development and poorer scholastic achievement and increases the chance of risk-taking behaviours that may result in antisocial or violent behaviour and unprotected and unsafe sexual behaviours.

NHS data from England show that the number of young people who have ever had an alcoholic drink has fallen from around 60% in 1992 to around 40% in 2021 for school-aged children and that 6% drink alcohol at least once per week and 17% monthly or more frequently. The proportion usually drinking once a week increases with age, from 1% of 11-year-olds to 14% of 15-year-olds. Eight percent of school children report being drunk in the previous 4 weeks.

Further reading

Advice about how to talk about alcohol consumption is available on the Drinkaware website https://www.drinkaware.co.uk/advice.

Recreational drugs

In 2020, more than 280 million people worldwide aged 15 to 64 years old used illicit drugs, with use skewed to younger people. Most commonly, this was cannabis (209 million annual users) followed by opioids (61 million), amfetamines (34 million), cocaine (21 million), and Ecstasy (20 million). Use in school children in 2020 was reported as 18% (12% in the past year), although this was a decrease from 2018, when use was 24% (17% in the previous year). A similar trend was seen in 15- to 16-year-olds in Italy and the USA associated with the SARS-CoV-2 pandemic. Future monitoring will inform whether this trend of reduced use of recreational drugs has continued.

Cannabis use is now legal in many countries, including the USA, and this is associated with an increase in use. Globally, cannabis use continues to increase year on year with 16% of those aged 16 to 24 reported having used the drug in the previous year. In England and Wales, where cannabis use is not legal, use is relatively stable in all age groups and has decreased by 40% in young people since 1995.

Non-prescribed opioid-related deaths are a concern in USA and Canada with more than 100 000 deaths in 2021. More than 61 million used opioids for non-medical reasons in 2020, with a doubling of use between 2010 and 2020. Their use accounts for nearly 70% of drug-related deaths and 40% of illicit drug use treatment costs.

Drug use can have detrimental effects on physical and mental health, with significant adverse long-term consequences, including the increased likelihood of risk-taking behaviours and transmission of blood-borne infections, including human immunodeficiency virus (HIV) and hepatitis B and C. Prevention programmes, early intervention, and public awareness campaigns are essential in curbing substance abuse and supporting those living with addiction.

Gambling

Young people are vulnerable to gambling addiction. Risk factors in this age group include impulsivity, peer pressure, and seeking excitement or escape. The consequences of addiction include adverse effects on mental health, relationships, and financial stability. The ease of access to online gambling platforms and widespread social acceptance of gambling exacerbate risks. Regulation of the gambling industry including reliably imposed age restrictions and the provision of support services can help mitigate against negative effects.

Vaccination

Young people are recommended to receive human papillomavirus and meningitis (Men ACWY) vaccinations as well as booster vaccinations against tetanus, diphtheria, and polio. Outbreaks of bacterial meningitis are reported around the world in young people who remain particularly vulnerable because of bacterial carriage rates and social behaviours. Despite the introduction of routine meningitis ACWY vaccination in the UK in 2015, invasive disease, with a mortality rate of 6%, kills up to 12 people in England every year.

Sleep

Regular and adequate amounts of good-quality sleep are needed for optimal growth, cognitive function, and physical and mental wellbeing. Almost a third of males and half of females aged 15 years report disturbed sleep with more than a third reporting that this interferes with their schooling. Up to 7% of teenagers develop a particular insomnia called delayed sleep phase syndrome. This is because they produce the pineal hormone melatonin later at night than other age groups. This causes them to have difficulties getting to sleep and waking in the morning. Chronotherapy treatment aims to reset circadian biorhythms with interventions including avoiding bright light in the evening and being exposed to bright light in the morning.

Establishing a good sleep routine in early adult life can help prevent diseases in later life, including hypertension, diabetes, obesity, and depression. Good sleep hygiene is key to achieving this. Some of the key components of good sleep hygiene include having a regular sleep routine and avoiding heavy meals, caffeine, and light emitted by screens before bedtime. Encouraging young people to establish a routine that works best for them is most likely to be successful. Referral for more specialist assessment and treatment might be needed for young people with severe sleep-related disorders.

Further reading

NHS Digital. *Smoking, Drinking and Drug Use among Young People in England*, 2021. Published September 2022. Accessed online at: https://digital.nhs.uk/data-and-information/publications/statistical/smoking-drinking-and-drug-use-among--young-people-in-england.

LONG-TERM HEALTH CONDITIONS

Up to one-third of 16- to 25-year-olds self-report having a health condition (Fig. 46.5). The most reported long-term health conditions in 11- to 16-year-old young people is mental health illness followed by acne, migraine, allergy, and asthma. Healthcare professionals have an important role in providing support, education, and resources to help young people meet the challenges of long-term conditions and to facilitate informed decision-making by young people about their health and wellbeing.

Several studies report an increase in morbidity and mortality when young people with long-term health conditions move from paediatric to adult healthcare services. Young people with long-term conditions are twice as likely to have mental health problems (30% vs 15%) and more likely to be unemployed (25% vs 2%) compared with healthy peers.

Approximately one-third to half (30%–43%) of young people with long-term health conditions fail to attend adult healthcare services 2 years after transfer from paediatric services. Non-attendance is associated with reduced adherence to treatment and worsening disease control. For example, fewer than 30% of those aged younger than 25 years with type 1 diabetes have

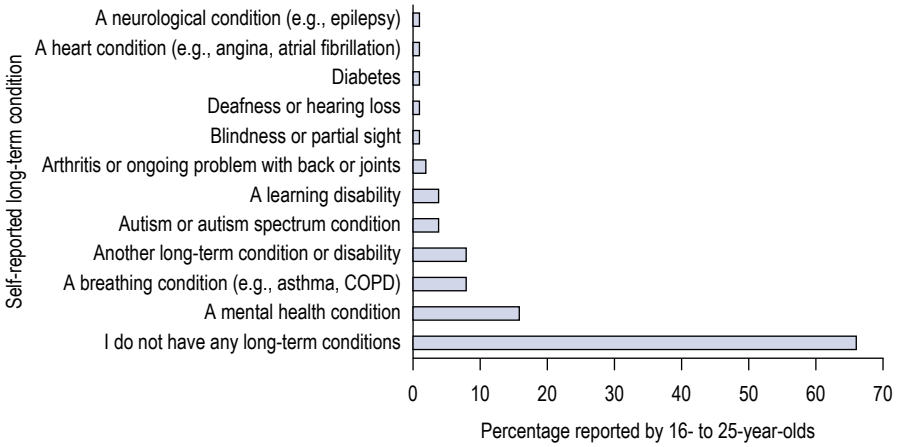


Fig. 46.5 Prevalence of self-reported long-term conditions by 16- to 25-year-olds in England and Wales in 2021. *COPD,* Chronic obstructive pulmonary disease.

an HbA_{1c} within the target range, rates of diabetes-related hospital admissions are increased, and almost one-third live with depression or eating disorders or are distressed about their condition.

The move to adult care can be challenging for many reasons (**Box 46.3**). Despite the availability of resources and evidence highlighting the risks of poorly managed transition to adult care, many young people and their families report unsatisfactory experiences. Several reports and guidelines have sought to mitigate the risks associated with transition from paediatric to adult services. Central to these efforts is the concept of 'transition'. The process of transition aims to empower young people and gradually equip them with the necessary skills and knowledge to confidently manage their healthcare. Transition needs to be structured and holistic and to address young people's medical, psychosocial, and vocational issues.

Features of transitional healthcare that are associated with improved outcomes include:

- A prolonged period of preparation, ideally starting before 13 to 14 years of age
- Accurate and comprehensive referral documentation
- Joint clinics between adult and paediatric specialist teams and a visit to the adult service before transfer of care
- Appropriate parent involvement that facilitates gradual relinquishing responsibility for treatment
- Promotion of young people's knowledge of and confidence in managing their health conditions and wider health (health self-efficacy).

Various transitional programmes have been developed that focus on empowering young people by teaching them self-advocacy and involving them in shared decision-making. One such programme is the 'Ready Steady Go' transition programme (www.readysteadygo.net). Through completion of a series of structured questionnaires, young people's knowledge of their conditions, treatment, the healthcare team, healthy lifestyle practices, sexual health, educational and vocational issues, and psychosocial concerns are assessed and addressed. Specific questions prompt discussion about potentially difficult topics such as sex and mental health (**Fig. 46.6**).

Effective transition programmes improve long-term outcomes. For example, in patients with type 1 diabetes, an effective transition increases attendance in adult services, reduces hospital admissions, and improves diabetic control as evidenced by HbA_{1c} concentrations.

i Box 46.3 Challenges in transition from paediatric to adult healthcare services

- The increased responsibilities of adult healthcare lead to apprehension, anxiety, and disengagement.
- Reluctance from parents to relinquish control over healthcare management hinders development of independence.
- Saying goodbye to trusted paediatric healthcare providers leads to feelings of loss and anxiety about building new connections with an adult team.
- Changes in care protocols between paediatric and adult services cause confusion, concern, and problems in accessing necessary services.
- Young people may worry that new healthcare providers will not fully understand their medical history and specific needs.
- Fear of being judged or stigmatized in adult healthcare settings deters young people from seeking care and support.
- Emotional stress and anxiety make it hard to cope with change.
- A poorly planned transition process results in gaps in medical care.

Further reading

Harmer MJ, Everitt LH, Parker L, Davis N, Connett G, and Nagra A. Structured transition is associated with improved outcomes in diabetes. *Pract Diab*. 2022; 39:18–22a.

Nagra A, McGinnity PM, Davis N, et al. Implementing transition: Ready Steady Go. *Arch Dis Child Educ Pract Ed* 2015; 100:313–320. Implementing transition: Ready Steady Go | ADC Education & Practice Edition (bmj.com).

National Confidential Enquiry into Patient Outcome and Death (NCEPOD). Transition from child into adult healthcare. *The Inbetweeners* 2023 www.ncepod.org.uk/2023transition.html.

Transition from children's to adults' services for young people using health or social care services 2016. www.nice.org.uk/guidance/ng43.

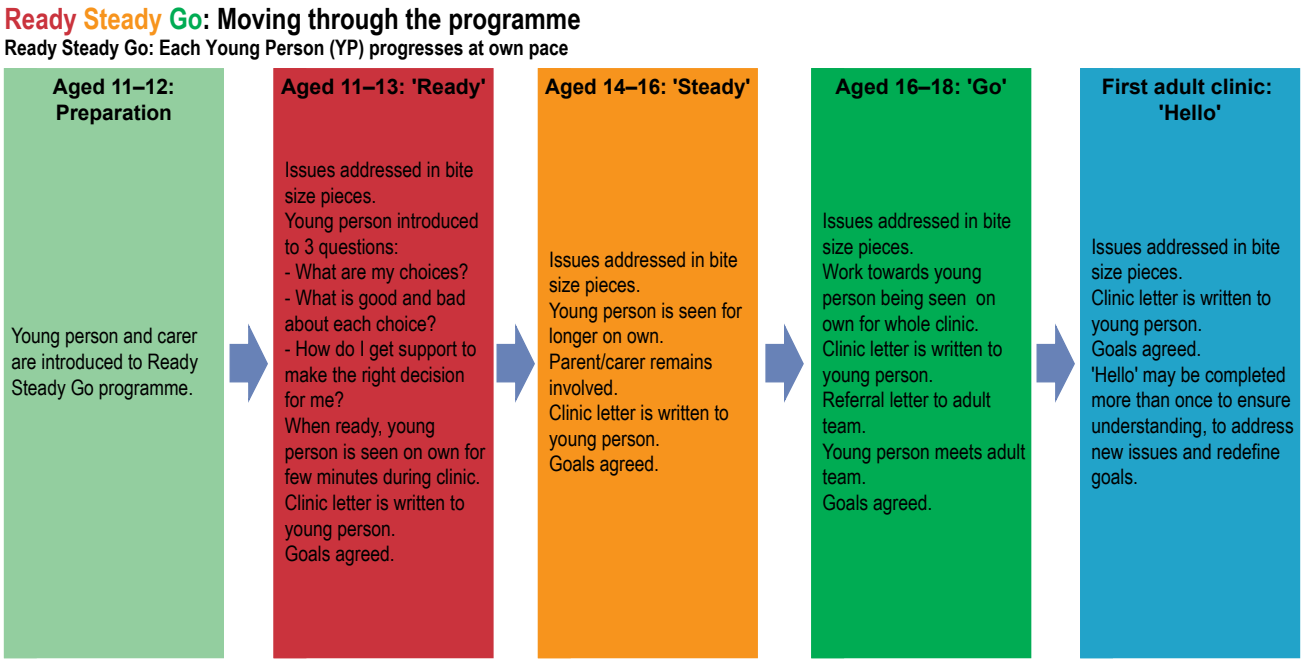


Fig. 46.6 Ready Steady Go: A step-by-step program outlining key stages of transition to adult care, providing clear guidance and support at each phase.

SHARED DECISION-MAKING

Shared decision-making is the direct involvement of young people in decisions about their healthcare. It is a key part of empowering young people to be actively engaged in managing their healthcare and improves health outcomes and patient satisfaction. It requires consideration of safeguarding, consent and parental involvement in care. The UK NHS constitution, the American Academy of Pediatrics, and the Australian Commission on Safety and Quality in Healthcare all promote shared decision-making in young people. Despite evidence for shared decision-making, fewer than half of young people and their parents feel involved in decisions about healthcare (which are similar to data from adult services).

'Ask 3 Questions' is a shared decision-making tool that provides a structured framework for discussions and resource allocation using a patient-centred approach. This approach empowers patients to become active participants in their healthcare journey and leads to improved patient satisfaction and health outcomes.

HEALTH LITERACY

For young people to be empowered, active participants in their own healthcare, they need to be able to obtain and understand information about their conditions and lifestyle choices. This skill is termed 'health literacy'. Better health literacy can be realized through improving young people's self-confidence and motivation to access information, as well as ensuring that the information provided is engaging, relevant, and appropriate to their needs and ability to understand.

In the UK, the average adult reading level is comparable to that of a 9-year-old. About 42% of English adults struggle to comprehend and use written health information, which rises to 61% when numeracy skills are necessary. These rates are similar worldwide. To address the gap in accessible health information, co-production between healthcare professionals and patients is essential. Involving young people and their parents or other carers in designing resources and services empowers patients and tailors' healthcare to young people's needs.

Further reading

NICE. *Shared Decision Making*; 2021. www.nice.org.uk/guidance/ng197.

OTHER CONSIDERATIONS WHEN ADDRESSING THE NEEDS OF YOUNG PEOPLE

Neurodiversity

Brain maturational changes during adolescence reveals variation in brain function and neurodevelopment. Neurodiversity, including autistic spectrum disorders, attention deficit hyperactivity disorder, and dyslexia, may become apparent. Health services have an important role in the diagnosis of those with neurodiversity and signposting to resources such as speech and language, occupational, and behavioural therapists. Access to education and employment support and technologies to facilitate communication may also benefit young people.

Sexual and reproductive health

Sexually transmitted infections (STIs) and unintended pregnancy are preventable. Teenage pregnancy is associated with adverse health and social outcomes, including lower education achievement, deprivation, and higher infant mortality rates, compared with older mothers and birthing people. The highest rates of STIs, HIV, and unintended pregnancy are among adolescents and young adults. Timely sex education, including information about consent, and access to youth-friendly, confidential reproductive health services providing contraception, STI testing, and treatment, are necessary measures toward achieving good sexual health in young people. Non-judgemental, accessible support is required in pregnancy, and many obstetric centres have specialist medical and midwifery teams supporting pregnancies in young people. About 40% to 60% of conceptions among females younger than 25 years of age living in the UK result in termination.

Menstruation

Key concerns in relation to menstruation in young people include ensuring the provision of accurate, timely, and age-appropriate information about menstrual hygiene and the emotional and physical changes that occur. Educational materials should promote positive societal attitudes and challenge cultural taboos when present. Access to affordable hygiene products can be a problem for those in low-income households.

Further reading

Menstrual Hygiene Day: Making Menstruation a Normal Fact of Life by 2030. Menstrual hygiene day Global (menstrualhygieneday.org).

Gender identity

Young people with gender identity concerns have specific needs. These include access to inclusive healthcare and the provision of policies and legislation to help ensure equal rights and opportunities to realize their potential. Careful consideration needs to be given to the needs of transgender and non-binary young people who menstruate and the way in which hormonal changes during the menstrual cycle can impact mood and exacerbate pre-existing mental health conditions. Young transmen should have access to safe and effective contraception and cervical screening programmes.

Further reading

World Health Organization. *Adolescent and Young Adult Health*. www.who.int/news-room/fact-sheets/detail/adolescents-health-risks-and-solutions.

Questions for this chapter can be found on Elsevier eBooks+.

47 Sex and gender in medicine

Eoin Dinneen, Michael Dinneen, J. Michael Dixon, Elizabeth Egbase, Leighton John Seal, Sophie Michelle Strong, and Rebecca Swan

Breast disease 1640
History 1640
Examination 1640
Investigation 1640
Breast cancer 1640
Congenital abnormalities of the breast 1641
Aberrations of development and involution 1641
Benign neoplasms and proliferations 1642
Other benign conditions of the breast 1643
Mastalgia 1643
Nipple discharge 1644
Breast infection 1644

Female health 1646

History 1646
Examination 1646
The life course approach 1649
Female adolescence 1649
Heavy menstrual bleeding 1649
Pre-pregnancy counselling 1650
Miscarriage and stillbirth 1651
Pregnancy 1652
The postnatal visit 1652
Contraception 1653
Termination of pregnancy 1653
Menopause 1655
Female urinary incontinence 1657
Uterine fibroids (leiomyomata) 1657
Gynaecological malignancy 1658
Domestic abuse and violence 1659
Female genital mutilation 1660

Male health 1660

History 1660
Examination 1660
Benign conditions of the prostate 1662
Prostate cancer 1664
Diseases of the penis 1666
Testicular and spermatic cord disease 1667
Loss of libido and erectile dysfunction 1669
The ageing male 1669

Transgender health 1670

The science of gender 1670
Epidemiology of gender incongruence 1670
Aetiology of gender incongruence 1670
Assessment of gender incongruence 1671
Diagnosis of gender incongruence 1673
Management of gender incongruence 1673
Treatment of non-binary people 1679

CORE SKILLS AND KNOWLEDGE

More than 90% of patients presenting to breast clinics have normal breasts or benign breast disease. Benign breast disease includes congenital abnormalities, aberrations of normal development and involution (ANDI), and other benign conditions. Timely, accurate diagnosis is key.

Approximately 1 in 20 females aged 30 to 49 years consults annually with menstrual disorders, 1 in 6 couples experiences subfertility, and 1 in 3 females will have a termination, with considerable global variance. Most gynaecological issues are managed in primary care. Complex gynaecological conditions, red-flag symptoms suggestive of possible malignancy, and symptoms that fail to respond to initial treatments require referral.

Conditions of the human body unique to the male sex include disorders of the male urogenital tract including the prostate, penis, and testes. Compared with females, males have lower health literacy and are less likely to use available health resources. Few countries have addressed the burden of male ill-health through the adoption of male-centred health strategies.

The development of a person's gender identity involves the interplay of psychological, social, and physical factors. Biology, socialization, and cognition contribute to the formation of gender identity. Gender diverse people face discrimination, leading to barriers in accessing healthcare.

Key learning objectives in sex and gender medicine include:

- Establishing gender identity, preferred name, and pronouns for all patients
- History taking in patients presenting with breast, gynaecological, or lower urinary tract symptoms (LUTS)
- Examinations of breast, genitalia, and prostate that are sensitive, efficient, and maintain dignity, with a chaperone present
- Knowledge of benign breast conditions, the menstrual cycle, and factors that impact physiology
- Understanding the benefits, risks, and potential side-effects of available contraceptive methods
- Recognizing red-flag symptoms suggestive of gynaecological and urogenital tract cancers
- Recognizing and managing urological emergencies, including testicular torsion, acute urinary retention, paraphimosis, and priapism
- The assessment of gender, physical health, and mental health in gender diverse patients
- Understanding the diagnostic criteria for gender incongruence and gender dysphoria
- Understanding that assessment and management of gender incongruence and dysphoria requires a multidisciplinary team, including psychological, social, and physical health
- Recognizing that gender diverse people face discrimination and barriers to accessing healthcare.

BREAST DISEASE

History

The primary focus is to assess whether breast cancer could be present. A history of presenting symptoms, risk factors for breast cancer, family history, and current medication is needed. The duration and onset of symptoms can be helpful: cysts can appear overnight, but cancers usually grow slowly. The growth of breast tissue in males (gynaecomastia) is usually self-limiting but can rarely be a sign of underlying systemic disease or drug toxicity.

Examination

Breast inspection should take place in good light with the patient sitting with their arms in three positions: by their side, above their head, and pressed on their hips to contract pectoralis major. Skin dimpling in the lower part of the breast may only be evident with the arms elevated or when the pectoral muscles are contracted. Contraction of the pectoral muscles allows assessment of deep fixation of palpable lumps.

Breast palpation is best performed with the patient lying flat with the arms above or behind the head. The fingertips are used to palpate all areas of the breast. Abnormalities should be assessed for contour, texture, and size. Patients with breast pain should be examined lying on each side, allowing the breast tissue to fall away from the chest wall to differentiate between true breast pain (mastalgia) and chest wall pain. Axillary and supraclavicular lymph nodes should be palpated. Palpable nodes may be palpable in up to 30% of patients without clinically significant disease. Benign causes of enlarged axillary nodes include infection, postvaccination response, and implant silicone leak causing silicone granuloma.

Investigation

Most breast masses or concerning areas of asymmetric nodularity should have *triple assessment* consisting of clinical examination, imaging, and core biopsy. Bilateral two-view (oblique and cranio-caudal) mammography is performed in patients older than 40 years of age with palpable abnormalities or nipple discharge. In patients younger than 40 years, the breasts are relatively radio-dense, and ultrasound scan is preferred. Ultrasonography can characterize benign and malignant features: whereas breast cysts are well demarcated, cancers usually have indistinct outlines. Magnetic resonance imaging (MRI) may be indicated in benign breast disease and is the optimum imaging method when the patient has breast implants. Ultrasound-guided core biopsy is the standard procedure to determine the cause of suspicious or indeterminate lesions or suspicious nodes in the axilla. Fine-needle aspiration cytology is no longer used to assess breast masses, although it may be useful in the assessment of axillary or supraclavicular lymphadenopathy. Large-bore vacuum-assisted biopsy can be used to completely remove lesions such as fibroadenomas and small papillomas. Open biopsy should be used only after imaging and at least one attempt at core biopsy.

Breast cancer

Although most patients presenting to breast clinics have normal breasts or benign breast disease, breast cancer should be excluded in patients presenting with breast symptoms.

Breast cancer may present as:

- **Breast lump or nodularity:** A breast lump is discrete and feels different from the surrounding breast tissue; it has definable borders and is measurable. Nodularity can be bilateral and symmetrical, or localized; it can fluctuate with the menstrual cycle. Localized nodularity is a common presentation of invasive lobular cancer.
- **Skin dimpling or change in contour** are present in up to 25% of symptomatic patients with breast cancer.
- **Nipple changes:** Nipple inversion occurs in both breast cancer and inflammatory breast conditions. Nipple inversion associated with breast cancer may be eccentric with the nipple being distorted. Slit-like central retraction of the nipple (inversion without distortion) is characteristic of benign disease.

Box 47.1 lists breast symptoms that require referral to exclude breast cancer.

Breast masses require triple assessment:

1. Examination
2. Imaging (mammography or ultrasonography)
3. Tissue sample (fine-needle aspiration or core biopsy)

Areas of asymmetrical nodularity require clinical and radiological assessment. A cause for nodularity needs to be identified; if no cause is found and the nodularity is localized, then core biopsy is indicated.

i Box 47.1 Breast symptoms requiring assessment for malignancy

Breast lump
- New discrete lump
- New lump within pre-existing nodularity
- Asymmetrical nodularity after the age of 35 years
- Asymmetrical nodularity that persists after menstruation (review after next menstrual cycle, within 4 weeks)
- Abscess or breast inflammation that does not settle rapidly after one course of antibiotics
- Palpable axillary mass

Pain
- If associated with a lump
- Intractable pain that interferes with a patient's lifestyle or sleep that has failed to respond to reassurance, supportive measures, and anti-inflammatory drugs
- Unilateral persistent breast pain (distinguished from chest wall pain)

Nipple discharge
- In females aged over 50 years
- In females 50 years or younger with any of:
  - Blood-stained discharge
  - Spontaneous single-duct discharge
  - Bilateral discharge sufficient to stain clothes

Nipple retraction or distortion

Nipple eczema

Change in skin contour

Family history
- Genetic testing offered if *BRCA1* and *BRCA2* mutation carrier possibility is 10% or more using available risk assessment tools

Box 47.2 Aberrations of normal breast development and involution

Age (years)	Normal process	Aberration
Younger than 25	Breast development: Stromal or lobular	Fibroadenoma Juvenile hypertrophy (macromastia)
25–40	Cyclical activity	Cyclical mastalgia or cyclical nodularity (diffuse or local)
35–55	Involution: Lobular Stromal Ductal	 Macrocysts Sclerosing lesions Duct ectasia

Adapted from Dixon JM, Thomas J. Congenital problems and aberrations of normal development and involution. In: ABC of Breast Diseases, *4th edn. Blackwell Publishing; 2012.*

Each component of the assessment—clinical examination (E), mammography (R) and ultrasound (U), cytology (C), and core biopsy (B)—is graded:

1. Normal (C1 = acellular)
2. Benign
3. Indeterminate or probably benign (B3 = atypical)
4. Probably malignant (B4 = suspicious)
5. Malignant (B5a = *in situ* carcinoma; B5b = invasive cancer)

Any mass, localized nodularity, or skin change considered clinically E3 or above requires biopsy even in the presence of normal or benign imaging. In the presence of a definite lesion, a normal core biopsy (B1) may be unrepresentative, and further biopsies should be considered by the multidisciplinary team (MDT). In any lesions considered suspicious of malignancy should be marked in the centre of the lesion at the time of biopsy, including suspicious lymph nodes. Management of breast cancer is detailed in Chapter 41.

Congenital abnormalities of the breast

Supernumerary nipple has an incidence of approximately 0.2% to 2.5%, most found inferior to the breast. Supernumerary breasts are less common with an incidence of 0.1% to 1%, most commonly in the lower axilla. Some degree of breast asymmetry is usual, and the left breast is more commonly larger than the right. Absence or hypoplasia of one breast can be associated with defects in one or both pectoral muscles. True unilateral absence of breast tissue is associated with absence or hypoplasia of the pectoral muscles in around 90% of patients. In contrast, 90% of patients with pectoral muscle defects have normal breasts. Tubular breasts have a narrow base and resemble a tubular hourglass structure. Pectus excavatum and scoliosis can result in symmetrical breasts appearing asymmetrical. Correction of asymmetry and tubular breasts involves tissue expansion, autologous fat grafting, or implants.

Aberrations of development and involution

The breast undergoes regular changes in relation to the menstrual cycle passing through phases of development, cyclical activity, and involution. ANDI is an established framework that refers to common benign breast conditions that occur during one specific period and are considered variations of a normal process, not caused by pathology (Box 47.2).

Box 47.3 Management of fibroadenoma

Conservative: If imaging and pathology are concordant with fibroadenoma, the patient can be reassured and discharged.

Excision (surgical excision or vacuum-assisted excision):

- If pathology shows a cellular fibroepithelial lesion that is difficult to confidently distinguish from a phyllodes tumour, open surgical excision is recommended.
- Rapidly growing lesion.
- Large fibroadenoma relative to the breast size or superficial fibroadenoma with significant distortion of the breast.
- Discomfort and pain are relative indications for surgical excision. Pain may not improve after excision.

Fibroadenoma

Fibroadenomas are the most common causes of palpable breast masses before 20 years of age, accounting for 60% of masses. They account for 13% of palpable symptomatic breast masses. Clinically, they are mobile, discrete, rubbery masses. An ultrasound scan can differentiate a fibroadenoma from a cancer. Core biopsy of the lesion should be performed if patients are aged 25 years or older or if there are indeterminate clinical or radiological features. Management is reassurance or excision (Box 47.3). Juvenile fibroadenomas occur in adolescent females and can undergo rapid growth.

Tubular and lactating adenoma

A tubular adenoma is a benign circumscribed lesion managed in a similar way to fibroadenoma. Lactating adenomas occur during pregnancy, usually regress after cessation of breast-feeding, and are managed conservatively.

Juvenile hypertrophy

During puberty or lactation, uncontrolled overgrowth of stromal and ductal breast tissue can occur with no measurable endocrine abnormality. Reduction mammoplasty should be considered within a careful shared decision-making pathway and can improve quality of life.

Pain and nodularity

Cyclical pain and nodularity are common and regarded as physiological, not pathological. Pain that is severe or prolonged is regarded as an aberration of cyclical change. Nodularity is the most common cause of referral to a breast clinic. Imaging reveals normal breast tissue. Focal, asymmetric nodularity that persists after menstruation requires triple assessment. Repeat review may be required because breast cancer frequently presents as localized asymmetrical nodularity in young females.

Palpable breast cysts

Palpable breast cysts are a common reason for presentation to a breast clinic, affecting 7% of females. They are distended and involuted lobules most commonly in perimenopausal females and less common after the menopause. Presentation is usually as a smooth, mobile, discrete breast mass, which can be firm. Box 47.4 summarizes the investigation and management of palpable breast

Box 47.4 Investigation and management of breast cysts

Investigation

Mammography: in females older than 40 years

Ultrasonography: distinguishes solid and cystic lesions; allows image-guided aspiration

- Simple cyst: smooth outline, posterior enhancement, absent internal echoes
- Complex cyst: thickened or irregular wall, internal echoes or thin septations, absent posterior enhancement

Core biopsy:

- If projections in the cyst wall to exclude intraductal papilloma or carcinoma
- Indeterminate clinical or radiological findings
- Persistence of a palpable mass after aspiration
- Rapid and persistent refilling of a cyst

Management

Simple, asymptomatic: reassure

Large, symptomatic, or complex characteristics:

Aspirate to dryness and send only blood-stained fluid for cytology

Re-examine after aspiration to ensure resolution; perform further imaging and biopsy if the mass persists

Recurrent: repeat aspiration; perform further investigation including imaging and biopsy if there is rapid and persistent refilling

cysts. Although there is a minimally increased risk of breast cancer in females with cysts, especially at younger ages, this does not warrant routine follow-up. Complex cysts are rarely malignant, but aspiration or repeat review and imaging is required.

Duct ectasia

The major subareolar ducts dilate and shorten with age. By age 70 years, 40% of females have duct ectasia. Presentation is with symmetrical, slit-like nipple retraction, cheesy nipple discharge, or a palpable mass that may be hard or doughy. Surgery may be indicated if discharge is troublesome or the patient wishes the nipple to be everted.

Sclerosing lesions

Sclerosing lesions are aberrations of stromal involution and can mimic cancer on mammography, causing diagnostic challenges during breast screening. There are three types: sclerosing adenosis, radial scars, and complex sclerosing lesions (CSLs). Sclerosing adenosis is benign and does not increase the risk of breast cancer. Radial scars and CSLs are associated with malignancy in up to 20% of cases, usually ductal carcinoma in situ (DCIS) or low-grade invasive cancers. Vacuum-assisted large-bore core biopsy usually provides sufficient tissue to exclude malignancy and avoid surgical excision.

Benign neoplasms and proliferations

Epithelial hyperplasia

Epithelial hyperplasia is characterized by an increase in the number of cells lining the terminal duct lobular unit. It is graded as mild, moderate, or florid (severe). If the cells display cellular atypia, the condition is called atypical hyperplasia. There are two types of atypical proliferation of the terminal duct lobular unit:

- **Lobular intraepithelial neoplasia (LIN):** characterized by loss of E-cadherin expression; includes atypical lobular hyperplasia and lobular carcinoma in situ

Fig. 47.1 Large ulcerated borderline phyllodes tumour.

- **Atypical ductal hyperplasia (ADH)** or atypical intraductal epithelial hyperplasia: express E-cadherin

ADH is the only benign breast condition associated with an increased risk of breast cancer. ADH carries a four- to fivefold increased risk of breast cancer, and the risk is doubled with a positive family history of breast cancer. When ADH is diagnosed on core biopsy, large-volume vacuum-assisted core biopsy or excisional biopsy is recommended because DCIS (up to 30%) or invasive carcinoma (9%) can be present in adjacent breast tissue. Tamoxifen 20 mg/day reduces the cancer risk associated with ADH by more than 80%, with more recent data showing that 5 mg/day is also effective, with limited toxicity.

Duct papilloma

Duct papillomas can cause nipple discharge that is often bloody. They are commonly detected through breast screening. Although the epithelial component can range from benign to *in situ* carcinoma, they usually have low malignant potential. Single-duct or total-duct excision is used to establish the diagnosis and stop troublesome nipple discharge. Core biopsy cannot always confidently differentiate between benign and malignant papillomas; therefore, a vacuum-assisted excision or surgery is recommended. Given the low malignant potential of these lesions, small papillomas without atypia can be safely observed.

Phyllodes tumours

These are rare fibroepithelial neoplasms that range from benign to borderline to malignant. The majority are benign in nature. They are commonly seen in patients aged 35 to 55 years. The aetiology is unknown. They feel like large fibroadenomas. They can grow rapidly and rarely cause marked distortion, cutaneous venous engorgement, and ulceration (Fig. 47.1). The diagnosis is with core biopsy. Excision is required to remove benign phyllodes with a clear margin. A wide margin excision is required in malignant lesions. Local recurrence is seen in fewer than 10% of benign phyllodes, and the rate of metastasis is 10% to 20% in malignant lesions.

Box 47.5 Benign nipple conditions

Condition	Features	Investigation	Management
Nipple adenoma	Non-discrete, benign, palpable growth in the nipple papilla Most common at ages 40–50 years Bloody nipple discharge is characteristic Cause nipple mass, discoloration, erosion, or contour change	Punch or core biopsy	Complete surgical excision
Eczema of the nipple	Usually affects areola first	Punch or core biopsy to differentiate from Paget's disease	Remove allergen; use hypoallergenic washing agents Topical steroid
Nipple Raynaud's disease	Vasospasm causes nipple to turn white, then blue, then red Patients usually have Raynaud's disease symptoms elsewhere		Avoid cold Low-dose nifedipine

Lipomas

Lipomas are common, soft, lobulated, radiolucent lesions. They can be confused with pseudolipomas, which are soft masses around a breast cancer caused by indrawing of fat surrounding a spiculated carcinoma. Ultrasound examination establishes the diagnosis. Biopsy or excision is only required if there has been rapid growth.

Granular cell tumours

Granular cell tumours are benign neoplasms that arise from Schwann cells of the peripheral nerve sheath. Although largely benign, malignant lesions have been described.

They can be challenging to differentiate clinically and radiologically from breast carcinoma because of their fibrous consistency, fixation to the pectoral fascia, and skin retraction. Core biopsy confirms the diagnosis. Granular cell tumours should be excised with a narrow clear margin to prevent recurrence.

Other benign conditions of the breast

Nipple conditions

Benign breast conditions of the nipple are summarized in **Box 47.5**.

Haematomas

Breast haematoma occurs most commonly after trauma such as road traffic accidents or surgery. Breast haematoma may occur spontaneously in patients on anticoagulant therapy. Examination reveals bruising or a palpable collection in a post-surgical patient. Management is supportive, including reassurance, educating patients on anticoagulant therapy, and repeat review if the haematoma does not resolve or the patient has any further concerns.

Fat necrosis

Fat necrosis can produce a mass that can feel like a breast carcinoma. Causes include:

- Trauma to the breast
- Surgery: Mammaplasty, lipofilling, and breast reconstruction with lower abdominal flaps can cause areas of fat necrosis.

Management is supportive with reassurance of the benign nature of the palpable breast abnormality. Lipomodelling may be used for significant distortion.

Diabetic mastopathy

This is a form of sclerosis associated with longstanding type 1 diabetes. It can present as one or more hard breast masses that are indistinguishable from malignancy. Histology reveals sclerosis. There is no increased risk of malignancy, and the treatment is reassurance.

Mondor's disease

Mondor's disease is thrombosis of the superficial veins in the skin of the breast, most commonly involving at least one of three venous channels:

- The thoracoepigastric vein
- The lateral thoracic vein
- The superior epigastric vein

It presents as a painful, thickened palpable cord, which may be erythematous. It is self-limiting. Topical non-steroidal anti-inflammatory drugs (NSAIDs) can be used.

Hamartomas

Hamartomas are common benign breast lesions composed glandular epithelium, adipose tissue, and fibrous tissue in variable proportions. They are usually asymptomatic but may present as a soft, smooth, mobile lump indistinguishable from fibroadenoma. Core biopsy confirms the diagnosis and excision is not required.

Mastalgia

Mastalgia is pain originating from within the breast. This needs to be distinguished from chest wall pain, which is the commonest cause of pain presenting to breast clinics (**Boxes 47.6 and 47.7**).

Few patients with breast pain require treatment. Reassurance that there is no serious underlying pathology is the mainstay of treatment. Tamoxifen 10 mg in the luteal phase of the menstrual cycle can be used for cyclical pain, and selective serotonin reuptake inhibitors can be used for premenopausal mastalgia. The fruit extract agnus castus has described benefits but is not widely used. Chest wall pain may be relieved by wearing a soft, supporting bra 24 hours a day. NSAIDs, providing there are no contraindications, can also be effective. Depot prednisolone combined with long-acting

Box 47.6 Breast pain overview

Breast pain alone or painful lumpiness is common and accounts for approximately 50% of new breast clinic referrals.
Breast pain is rarely a symptom of breast cancer.
Pain in the breast can arise from the breast or the underlying chest wall.
Careful examination can differentiate chest wall pain from true breast pain.
Medical treatment rarely needed.

Box 47.7 Breast pain

Clinical features

Most patients with breast pain who reach the breast clinic have pain arising from the chest wall. It is rare for breast carcinoma to present with breast pain.

True breast pain

- Worse before and relieved after menstruation; aetiology unknown
- Exacerbated by fluctuating hormone concentrations (perimenopause) and exogenous hormones (hormone replacement therapy, contraceptive pill); more common with combined oral contraceptive pill use than progesterone methods of contraception but seen in both

Chest wall pain

- Unilateral
- Exacerbated by activity
- Reproduced by direct pressure over a specific area of chest wall

Examination

Sitting up and lying on each side to allow breast tissue to fall away from chest wall

Investigation

Mammography: in females older than 40 years of age to exclude malignancy. The positive predictive value of mammography for malignancy in breast pain is very low (0.17 for age 40–49 years vs 2.8 after 70 years)
Ultrasonography: in all patients with localized area of pain
Triple assessment: in all patients with a palpable lump

local anaesthetic can be effective for patients with localized chest wall pain.

Nipple discharge

Nipple discharge accounts for 5% of referrals to breast clinics. Up to 20% is caused by *in situ* or malignant disease. Serosanguinous or bloody discharge from a single duct raises the suspicion of an underlying malignancy, although it is rare for an invasive cancer to cause nipple discharge in the absence of a palpable lump. The aim is to differentiate physiological causes from ductal pathology. Causes of nipple discharge include:

- Physiological discharge: opalescent, multiple duct discharge
- Galactorrhoea: bilateral, copious, pale milky, multiple duct discharge. This can be caused by prolactinoma or medications that increase prolactin concentrations.
- Duct ectasia: a yellow, thick discharge, usually at older ages
- Papilloma, epithelial hyperplasia, and DCIS: can all present with serosanguinous or bloody discharge from a single duct
- Invasive carcinoma: serosanguinous or bloody discharge from a single duct, often with a palpable lump. The clinical features, investigation, and management of nipple discharge are shown in **Box 47.8**.

Box 47.8 Nipple discharge

Key clinical assessment

Does it arise from a single duct or multiple ducts?
Is it bloodstained?
Induced by squeezing the nipple or spontaneous?
Affecting one or both breasts?
Frequency, colour, and consistency

Examination

Breast examination for lump
Firm digital pressure to allow assessment of site and character of discharge

Investigation

Test discharge for blood by inspection and urinalysis strip testing
Mammogram in patients older than 40 years of age
Ultrasonography of ducts close to the nipple can sometimes identify papillomas and malignant lesions

Management

Spontaneous single-duct discharge requires surgery if it occurs on more than one occasion and has at least one of the following features:

- Blood stained or contains moderate or greater amounts of blood on testing
- Persistent (occurs at least twice per week)
- Associated with a suspicious mass

Microdochectomy, in which a single duct is removed, is considered for single-duct discharge in younger females. This should not damage the surrounding ducts and allows lactation. Total-duct excision is an option for females of non-childbearing age and is less likely to miss underlying malignancy than microdochectomy. Multiple-duct spontaneous discharge is treated by total-duct excision. It is very rare for duct surgery to be indicated in males.

Breast infection

Breast infection most commonly affects females between the ages of 18 and 50 years and can be divided into lactational and non-lactational infection. As with all breast pathology, breast infection is rare in males. Cellulitis in male breast is uncommon but can be seen in the neonatal and pubertal periods. Specific breast infections are summarized in **Box 47.9**. The four guiding principles in treating patients with breast infection are:

1. Provide early appropriate antibiotics to reduce the risk of abscess development.
2. Refer to hospital for infections that do not settle within 48 hours of appropriate antibiotic treatment.
3. Perform ultrasonography if abscess is suspected to confirm a collection before drainage.
4. Exclude breast cancer if there is a solid lesion on ultrasound or if infection fails to resolve with adequate antibiotic treatment.

Peripheral non-lactating breast abscesses

Most of these have no underlying cause. They can be associated with diabetes, rheumatoid arthritis, steroid treatment, granulomatous lobular mastitis, and trauma. If the skin overlying the abscess is not thinned, management is by aspiration and

Box 47.9 Breast infections

Neonatal infection

Presentation: rare, occurring in first few weeks of life when breast bud is enlarged

Organism: usually *Staphylococcus aureus*; occasionally *Escherichia coli*

Management: antibiotics or small peripheral incision of abscess to avoid damage to the breast bud

Lactating infection

Presentation: most commonly seen during first 6 weeks of breast-feeding; history of nipple or skin abrasion causing local oedema and reduced milk flow

Organism: usually *S. aureus*

Management: antibiotic therapy according to sensitivity (e.g., flucloxacillin, erythromycin, co-amoxiclav); avoid tetracyclines and chloramphenicol when possible because of adverse neonatal effects; promote milk drainage; application of gel packs or cold cabbage leaves to the breast

Periareolar infection

Presentation: most common in young females; more common in smokers; periareolar inflammation or abscess with mass, pain, nipple retraction, or discharge; organism: usually anaerobic organisms

Management: antibiotics including anaerobic cover (e.g., co-amoxiclav, erythromycin plus metronidazole); recurrent episodes may be caused by persistence of diseased ducts, and excision may be considered, weighed with risk of further infection and nipple necrosis, particularly in smokers

irrigation with local anaesthesia along the aspiration track and to wash out the abscess until fluid is clear. If the skin overlying the abscess is thinned or necrotic, then pus is drained through a small skin incision through the thinned skin under local anaesthesia without packing. Drainage under general anaesthesia is rarely required. Patients should be reviewed every 2 to 3 days and aspiration or irrigation continued until no further pus is present.

Granulomatous lobular mastitis

This is a rare condition that presents most frequently in young parous females as an irregular mass indistinguishable from a carcinoma or as multiple or recurrent breast abscesses. Pathological features include non-caseating granulomas and microabscesses confined to a breast lobule. The role of specific organisms in the aetiology of this condition is unclear, but an association with *Corynebacterium* spp. is described in some studies. The diagnosis is confirmed on core biopsy. Management has a limited evidence base, and approaches include:

- Treat episodes of infection with antibiotics, aspiration, or drainage and ensure close surveillance.
- Avoid large incisions because of risks of poor wound healing and persistent discharge.
- Steroids and immunosuppressants have been used with varying efficacy in small series.

Mammary duct fistula

This describes an abnormal communication between the skin and a major subareolar breast duct. It occurs as a complication of periductal mastitis after incision and drainage, spontaneous discharge of a peri-areolar inflammatory mass, or core biopsy of a peri-areolar inflammatory mass. Treatment is by opening (fistulotomy) or excising (fistulectomy) the fistula, but recurrence is common.

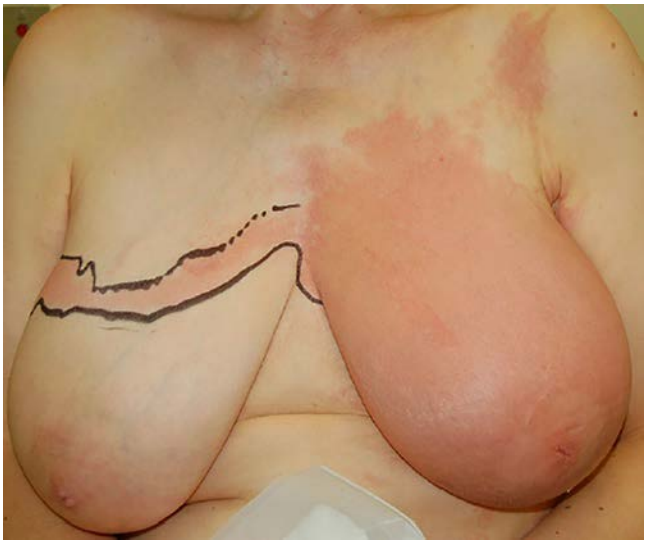


Fig. 47.2 Left breast cellulitis with lymphangitis across the right breast.

Skin-associated infection

Primary infection of the skin of the breast can present as cellulitis or an abscess and most commonly affects the lower half of the breast. Radiotherapy disrupts the blood and lymphatic flow of the breast and increases the risk of skin-associated infection and cellulitis (Fig. 47.2). When inflammation occurs after radiotherapy, it is important to distinguish lymphoedema from cellulitis. Cellulitis of the male breast is rare but is seen in the neonatal and pubertal periods. Beta-haemolytic streptococci and *Staphylococcus aureus*, including methicillin-resistant *S. aureus*, are the most common pathogens. Treatment is with antibiotics according to organism and sensitivity. Outlining the area of cellulitis allows assessment of response to treatment. Epidermoid cysts of the skin of the breast may become infected.

Those who are overweight, have large breasts, or have intertrigo are susceptible to recurrent infections in the skin under the breast. Treatment includes antibiotics if infection is present, weight reduction, and advice on measures to keep the area as clean and dry as possible.

Hidradenitis suppurativa

Hidradenitis suppurativa is a chronic disease of follicular occlusion that can affect the infra-mammary regions. Treatment of secondary infection is with appropriate antibiotics and drainage of pus if required. Excision of the affected skin is associated with a high risk of recurrence despite surgery. Patients should be supported to stop smoking.

Post-surgical infection

Post-surgical infection can occur in the early postoperative period or after wound healing. The most common organisms in the early operative period include skin flora and *S. aureus*. Perioperative prophylaxis with a single dose of a broad-spectrum antibiotic such as co-amoxiclav has

been shown to be beneficial in prevention. Delayed cellulitis may occur months to years after breast-conserving surgery and radiation therapy and should be managed as skin-associated infection as described earlier.

Further reading

Association of Breast Surgery Summary Statement, Management of Fibroadenomas [Internet]. [cited 2023 Jun 1]. Available from: https://associationofbreastsurgery.org.uk/media/64950/abs-summary-statement-fibroadenomas-v1.pdf.
Dixon JM. *ABC of Breast Diseases*. 4th edn. Blackwell Publishing Ltd. 2012.
Dixon JM, Khan LR. Treatment of breast infection. *BMJ*. 2011; 11;342:d396.
Guzik P, Gęca T, Topolewski P et al. Diabetic mastopathy. Review of diagnostic methods and therapeutic options. *Int J Environ Res Public Health*. 2021; 19(1).
Lawrence RA. 2—Anatomy of the breast. In: **Lawrence RA, Lawrence RM, editors.** *Breastfeeding (Ninth Edition) [Internet]*. Philadelphia: Elsevier; 2022. p. 38–57. Available from: https://www.sciencedirect.com/science/article/pii/B9780323680132368013400002X.
Park HL, Pyo YC, Kim KY et al. Recurrence rates and characteristics of phyllodes tumors diagnosed by ultrasound-guided vacuum-assisted breast biopsy (VABB). *Anticancer Res*. 2018; 38(9):5481–7.
Pinder S, Shaaban A, Deb R et al. NHS Breast Screening multidisciplinary working group guidelines for the diagnosis and management of breast lesions of uncertain malignant potential on core biopsy (B3 lesions). *Clin Radiol.* 2018 Aug;73(8):682–92.
Presazzi A, Di Giulio G, Calliada F. Breast hamartoma: ultrasound, elastosonographic, and mammographic features. mini pictorial essay. *J Ultrasound*. 2015; 18(4):373–7.
Walker S, Hyde C, Hamilton W. Risk of breast cancer in symptomatic women in primary care: a case–control study using electronic records. *Br J Gen Pract*. 2014; 64(629):e788.

FEMALE HEALTH

History

Direct but sensitive questioning about intimate issues may be required. People may not feel comfortable talking about these in front of their family or friends Ensure you have a confidential area to take a history (**Box 47.10**).

Examination

Consider the preliminaries listed in **Box 47.11**. Ensure a chaperone is available to conduct an examination. Vaginal (bimanual) and speculum examinations (**Fig. 47.3**) are needed as part of the investigation of symptoms including abnormal vaginal discharge, intermenstrual or postmenopausal bleeding, pelvic pain, and dyspareunia. These examinations may be perceived as invasive, intimate, and uncomfortable. It is essential to gain competence in examination techniques using pelvic training simulators before examining patients. Gain informed consent and empower patients to tell you to stop the examination at any point.

Box 47.10 Taking a gynaecological history

Menstruation
- Age when periods began and stopped (if postmenopausal)
- Regularity of periods: usual length of cycle and duration of bleeding
- Any significant changes in menstrual cycle
- First day of the last menstrual period
- Heavy or painful periods:
  - As perceived by the patient, including impact on quality of life
  - Semi-objective measures: days off work, frequency of sanitary product change

Pregnancy
- Number, timing, and outcomes
- Sensitive approach to adverse outcomes and difficult decisions: e.g., miscarriage, termination, stillbirth
- Avoid confusing gravida and parity jargon: a detailed chronological list is more valuable

Medical history
- Medical and surgical history
- Personal and family history of gynaecological and breast disorders
- Current and past medications, including teratogenic or fetotoxic medication, e.g., sodium valproate, warfarin

Cervical cancer risk factors
- Human papillomavirus vaccination
- Uptake of cervical screening
  - Date and outcome of most recent and previous screening
  - Ensure accessible screening for females in same-sex relationships and transgender men

Sexual history
- Sexual partners and types of intercourse in the preceding 6 months
- History of sexually transmitted and pelvic infections
- Current and past use of contraception
- Current and future fertility wishes

As appropriate
- History of non-consensual sexual activity or abuse

Consider referral for forensic medical examination
Support with referral to police
Consider need for psychological support
Female genital mutilation

Box 47.11 Preliminaries to a vaginal examination

- Ensure the environment is comfortable and private.
- Wash your hands.
- Introduce yourself (name and role) and confirm patient details.
- Ensure a chaperone is present (ideally, a female chaperone) and record their name.
- Explain the proposed procedure clearly and answer all questions.
- Obtain explicit consent; explain that the patient may change their mind at any time.
- Ask whether there is any possibility that the patient might be pregnant.
- Wear gloves.
- Provide a sheet or modesty blanket and ensure the door is closed and locked.
- Ask the patient to remove their underwear and lie on the examination couch with the heels drawn to buttocks and the knees out to sides (modified lithotomy position).

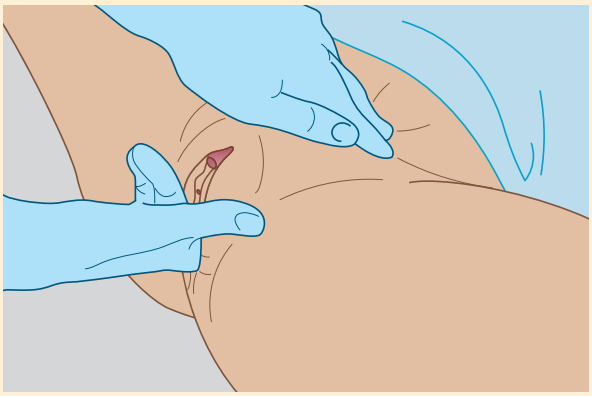
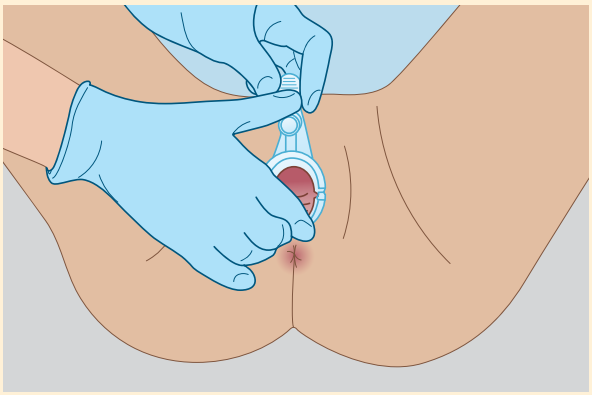
Bimanual examination	Speculum examination
	
• Inspect vulva: - Ulcers - Vaginal discharge - Masses (malignant or abscesses) • Enter vagina using two lubricated fingers • Assess vagina (including the fornices) for masses or irregularities • Assess cervix: - Masses - Os (open/closed) - Pain on palpation – 'cervical excitation' • Assess uterus: - Place second hand on lower abdomen - Use internal fingers in posterior fornix to push uterus upwards and palpate between hands - Assess size, shape and position • Assess adnexa: - Place internal fingers into left lateral fornix - Place other hand to palpate deeply in left iliac fossa - Move internal fingers towards other hand to assess for tenderness and masses - Repeat on right • Close the encounter: - Stop examination; remove gloves - Offer paper towels and encourage patient to dress - Summarize findings and next steps - Document procedure and chaperone identity	• Inspect vulva (see left) • Insert speculum: - Warn patient you will now insert speculum - Lubricate speculum (if possible, also warm metal speculums) - Part labia using thumb and first finger of non-dominant hand - Insert the speculum sideways; rotate 90 degrees once inside (so handle faces upwards) - Open speculum blades and visualize cervix - Tighten locking nut to hold blades open • Assess cervix: - Os (open/closed) - Ulcers - Masses - Discharge • Perform any necessary diagnostic procedures: - Swabs for conventional culture or nucleic acid amplification tests (NAATs) - Smear test for cytology • Remove speculum: - Loosen locking nut - Partially close blades, taking care not to trap vaginal wall between blades - Rotate 90 degrees and withdraw speculum • Close the encounter (see left) • Ensure all samples are sent for appropriate tests

Fig. 47.3 **Bimanual and speculum examination.**

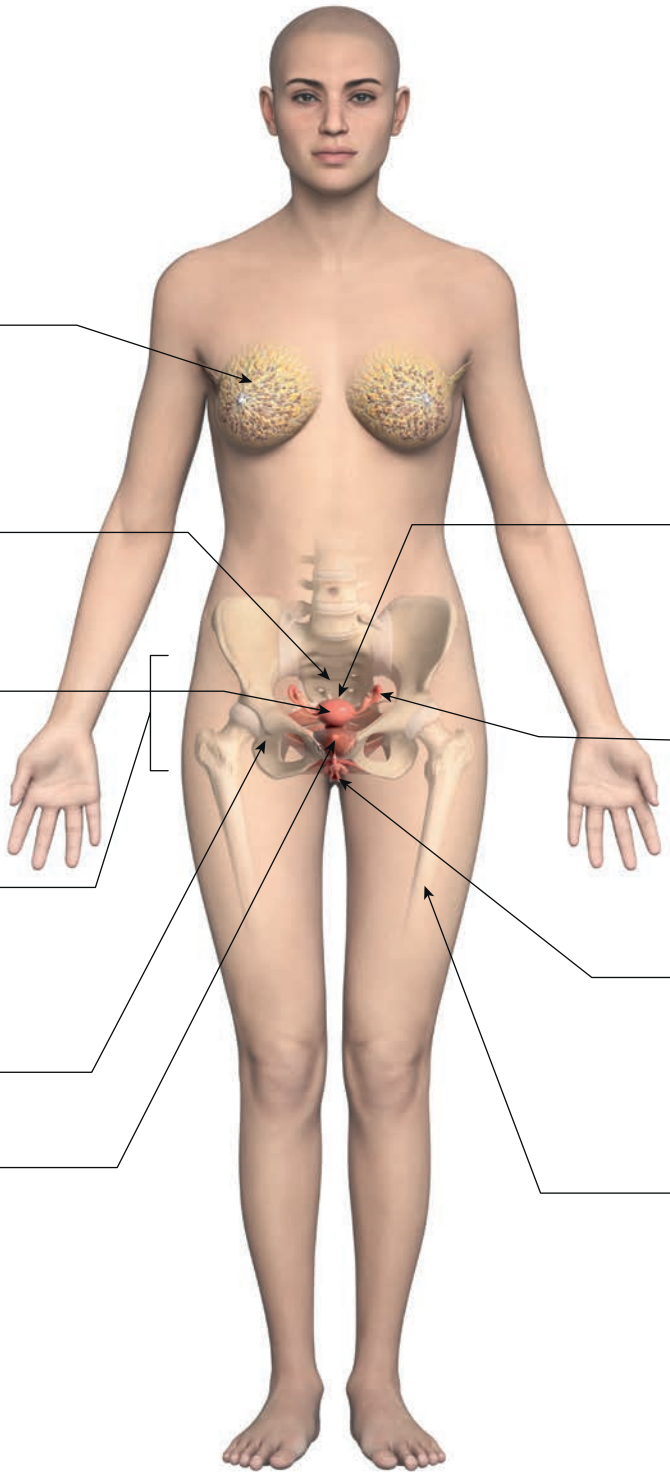
Each patient will have their own experience of menarche and the pre- and postmenopause. Issues during these life phases may result in females interacting more with health services compared with male patients. Female patients have a higher lifetime risk of mental health issues, anaemia, and osteoporosis while also being at risk of complications related to pregnancy and delivery.

Female reproductive and sexual health needs are largely predictable. Most females want to enjoy healthy sexual relationships and control their fertility to prevent unwanted pregnancies. Patients are increasingly choosing to delay conception until later in life, which will likely increase the demand for fertility services in the future. Increased life expectancy means that the postmenopausal period is longer. An understanding of the impact of societal changes on future healthcare needs is warranted. Female patients are more likely to experience sexual and domestic violence than male patients, the latter being even more common in pregnancy and the postnatal period.

Ethnic inequalities in obstetric healthcare are a concern. In the UK, patients of Black ethnicity are 3.7 times more likely to die in pregnancy or the postpartum period (34 per 100 000 giving birth) compared with patients of White ethnicities. Patients of Asian ethnicities are 1.8 times more likely to die (16 per 100 000 giving birth).

Definitions in female health include:

- ***Puberty.*** The process of physical maturation to reach sexual maturity and be capable of reproduction. The mean age of menarche is 12.8 years, but it may take several years for the menstrual cycle to establish a regular pattern. Initial cycles are typically anovulatory and unpredictable.



Breast
Mass: contour, texture, size
Pain (mastalgia)
Nipple: mass, discoloration, discharge, retraction, vasospasm
Skin: contour, infection
Congenital abnormality
Venous thrombosis
Reconstruction
Infection: abscess, cellulitis
Atrophy

Abdominal/pelvic pain
Exclude ectopic pregnancy

Uterus
Fibroid
Amenorrhea
Heavy menstrual bleeding
Intermenstrual bleeding
Post-coital bleeding
Menopause

Gynaecological malignancy
Pain
Bloating
Pelvic mass
Post-menopausal bleeding
Weight loss
HPV infection

Lymphadenopathy

Bladder
Urgency
Incontinence

General
Inequality
Violence/abuse

Pregnancy
Infertility
Miscarriage
Hypertension: pregnancy-induced hypertension/pre-eclampsia
Elevated plasma glucose: gestational diabetes
Fetal growth restriction
Preterm delivery
Stillbirth
Termination
Post-partum depression/psychosis
Prediction of future health risk

Premenstrual symptoms
Bloating
Weight gain
Breast tenderness
Headache
Mood changes/irritability

Vagina
Discharge
Dyspareunia
Dryness
Genital mutilation
Skin change: lichen sclerosus

Bone
Reduced bone mineral density

- ***Precocious puberty.*** This is defined as menses or breast development before the age of 8 years.
- ***Delayed puberty.*** This is the absence of secondary sex characteristics after 14 years. Causes may be central (e.g., anorexia, excessive exercise, chronic illness, pituitary tumour) or failure of gonadal function (e.g., Turner's syndrome, XX gonadal dysgenesis).
- ***Primary amenorrhoea.*** This is the absence of menstruation by 16 years of age.
- ***Secondary amenorrhoea.*** This is a new absence of menses for more than 6 months in a female who is not pregnant, lactating, or menopausal.
- ***Menorrhagia.*** This is menses with abnormally heavy or prolonged bleeding affecting quality of life.

- ***Premenstrual syndrome.*** Physical or emotional symptoms before a period are very common and include progesterone-induced bloating, weight gain, mastalgia (breast tenderness), headaches, depression, and irritability. In 5% to 15% of females, severe symptoms affect lifestyle and personal relationships. Treatments, including lifestyle modifications, hormonal preparations, cognitive-behavioural therapy (CBT), selective serotonin reuptake inhibitors, and surgery, are variably effective.
- ***Premature ovarian insufficiency.*** The cessation of periods before age 40 years occurs in 1% of females.
- ***Menopause.*** Menopause is a retrospective diagnosis made 1 year after the final menstrual period in the absence of hormonal suppression or pregnancy. The average age of menopause in high-income countries is 51 years.
- ***Surgical menopause.*** This is a menopausal state induced by surgical removal of both ovaries (oophorectomy).
- ***Postmenopausal bleeding.*** Most cases are caused by benign atrophic vaginitis and respond to topical oestrogens. Approximately 10% are caused by endometrial pathology, including atypical hyperplasia (precancerous changes) or endometrial cancer. Ultrasound examination is used to assess endometrial thickness. Tissue sampling is indicted for patients with endometrial thickness greater than 3 to 5 mm or recurrent bleeding episodes and in high-risk patients. Patients on tamoxifen for breast disease may have an irregular, thickened, and cystic endometrium, which requires hysteroscopy and directed tissue biopsy.

The life course approach

A life course approach examines the effects of biological, social, and behavioral exposures that occur during fetal development, childhood, adolescence, and adulthood, including pregnancy. The different stages of the life course impact health, wellbeing, and disease in later life and across generations. A life course perspective recognizes that the circumstances of pregnancy can affect the health of the offspring, early life events impact long-term outcomes, and early interventions may reduce the risk and severity of disease.

Female adolescence

Sex and relationship education became statutory in primary and secondary schools in the UK in September 2019. It provides information on:

- The constituents of a healthy relationship
- Consent, empowering young people to say no, and how to seek help
- Contraception
- Risks of sexually transmitted infections (STIs)
- Vaccination and screening programmes to prevent and detect early or pre-disease

Human papillomavirus (HPV) vaccination at age 12 to 13 years and *Chlamydia trachomatis* screening for those younger than 25 years aim to prevent cervical cancer, ectopic pregnancy, and tubal infertility. These encounters provide early opportunities to discuss the long-term benefits of a healthier lifestyle before considering pregnancy.

Early sexual health education reduces the incidence of teenage pregnancy. In the UK, the Teenage Pregnancy Strategy is a sustained, coordinated local and national action that has led to the lowest conception rate in those younger than 18 years since 1969. However, England's teenage birth rate remains higher than other Western European countries, and adolescent mothers have poor outcomes: 90% leave education and 20% become pregnant again within their teenage years.

Heavy menstrual bleeding

The menstrual cycle is shown in Fig. 47.4.

Heavy menstrual bleeding (HMB) accounts for one in five gynaecology clinic consultations and is the most common cause of iron-deficiency anaemia in females. Methods to quantify menstrual blood loss are inaccurate; clinical assessment uses the patient's perception of blood loss and impact on quality of life. The passage of large blood clots, severe pain, or flooding that prevents a patient leaving home require investigation and management.

Detailed menstrual and medical histories are required, with special consideration for symptoms of anaemia. Abdominal and pelvic examination is carried out to exclude pelvic masses, visualize the cervix, perform a smear (HPV test) if needed, and take swabs if infection is suspected. Investigations to consider are:

- ***Full blood count and iron studies*** to assess need for iron replacement and, in extreme cases, blood transfusion.
- ***Haematological opinion*** if history is suggestive of a bleeding disorder.
- ***Ultrasound scan*** if a pelvic mass is palpated or the bleeding pattern suggests endometrial pathology.
- ***Endometrial biopsy*** for irregular or intermenstrual bleeding and when drug therapy fails (Fig. 47.5).

Ultrasound and haematological investigations are indicated in young females presenting with menorrhagia from their first period.

The most appropriate treatment for HMB considers patient preference, risks and benefits, fertility and contraceptive requirements, and contraindications to medical or surgical therapies.

Medical treatments

- ***Levonorgestrel intrauterine system*** (Mirena coil). This has revolutionized treatment of patients with HMB. It leads to a 95% reduction in blood loss (30% of females are amenorrhoeic within 1 year of insertion), is an effective contraceptive, improves dysmenorrhoea, and has few side-effects. It should be considered in most females with HMB as an alternative to surgical treatment. Some experience irregular bleeding and persistent spotting for 3 to 9 months after insertion.
- ***Tranexamic acid.*** This is an antifibrinolytic associated with a 50% reduction in HMB.
- ***Mefenamic acid.*** This NSAID is taken during menses and is associated with a 25% reduction in bleeding.
- ***Combined oral contraceptive pill.*** This reduces bleeding and provides effective contraception, but side-effects exist, including an increased risk of venous thromboembolism (VTE).
- ***Oral progestogens.*** Taken on days 6 to 26 of the menstrual cycle, these regulate the bleeding pattern but are not contraceptives, and breakthrough bleeding can occur.
- ***Ulipristal acetate.*** This can be added to other strategies in females with fibroids larger than 3 cm. Severe liver injury is rare.
- ***Gonadotrophin-releasing hormone (GnRH) agonist*** drugs lead to desensitization of the gonadotroph GnRH receptor, resulting in the suppression of luteinizing hormone (LH) follicle-stimulating hormone (FSH) and secretion and amenorrhoea. They are suitable only for short-term use because they produce a hypoestrogenic state predisposing to osteoporosis.

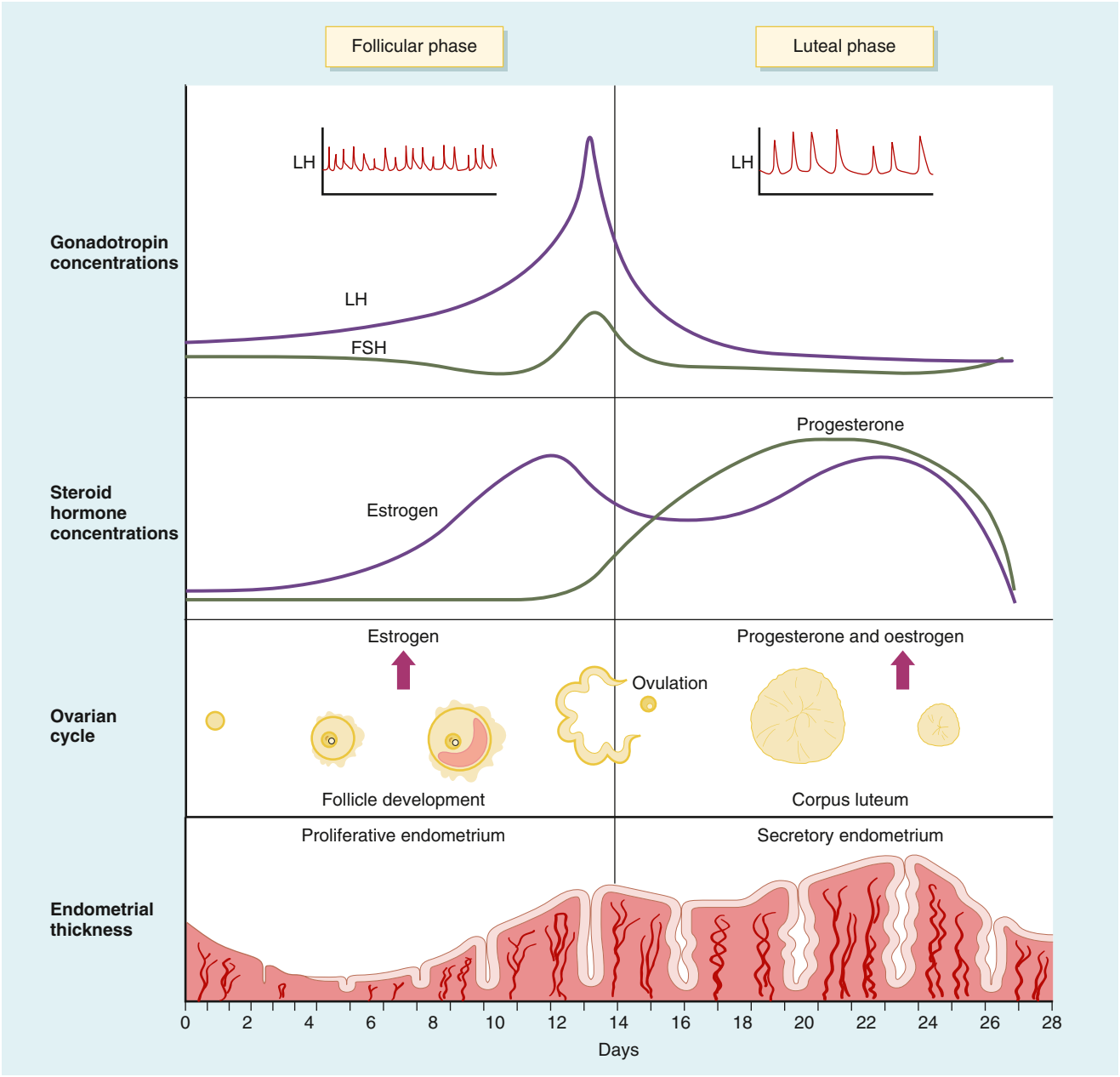


Fig. 47.4 **The menstrual cycle: luteinizing hormone (LH), follicle-stimulating hormone (FSH), oestrogen, and progesterone concentrations during the follicular and luteal phases of the menstrual cycle, with associated endometrial thicknesses and ovulation.**

Surgical treatments

Surgical treatments for patients with HMB are usually offered when medical treatments have failed and patients have completed their families.

- ***Endometrial ablation.*** This destroys the endometrial lining to a depth that prevents cyclical regeneration. Second-generation microwave and thermal ablation techniques are undertaken as day-case procedures.
- ***Hysterectomy.*** Removal of the uterus via an abdominal, vaginal, or laparoscopic method depends on the size of the uterine body, the degree of laxity of the pelvic muscular floor, and surgical expertise. Vaginal or laparoscopic surgical procedures usually mean shorter hospital stays and faster return to normal activities, but a large fibroid uterus may require an open incision.

Pre-pregnancy counselling

Forty percent of pregnancies in the UK are unplanned, and opportunities to provide preconceptual advice on diet, weight, exercise, smoking and alcohol, vaccination, and folic acid intake are often missed before pregnancy. Pre-pregnancy counselling allows a review of general health and medication and ensures that a management plan for pregnancy is in place.

Most maternal deaths in the UK occur in females with medical or mental health co-morbidity. Between 2018 and 2020, 60% of deaths in pregnancy or within the first postpartum year were in females with pre-existing medical problems, and 37% had pre-existing mental health problems. More than 25% of females' deaths were in those living with obesity (body mass index [BMI] ≥30 kg/m^2).

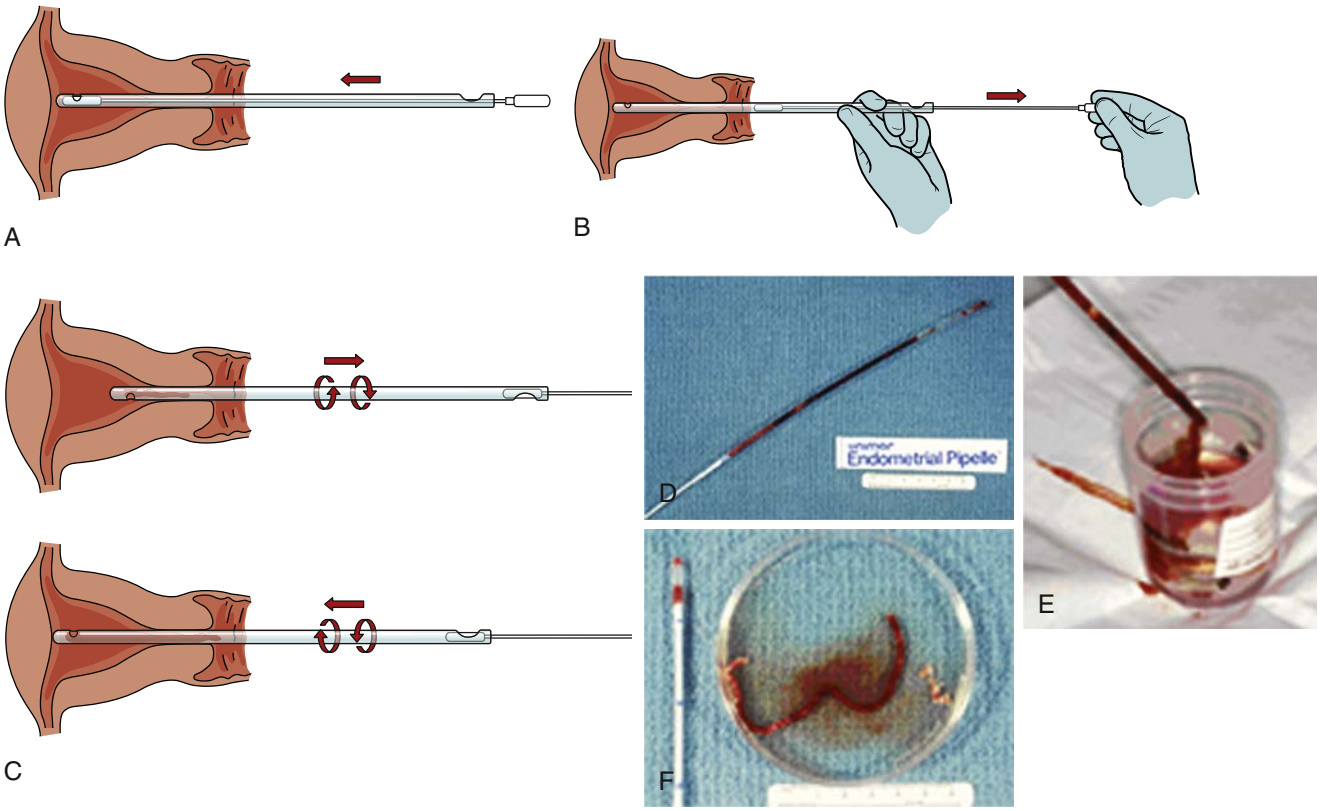


Fig. 47.5 Pipelle endometrial biopsy. The Pipelle is passed through the cervix into the endometrial cavity **(A)**. The plunger is then drawn back to create a small vacuum **(B)**. The Pipelle sampler is then moved through 360 degrees against the endometrium **(C)** to obtain an adequate sample **(D)**. The sample is put into formalin **(E)** and sent for histopathological analysis **(F)**.

Around 20% of females in the UK remain childless, either by choice or otherwise. A smaller but growing number pursue assisted conception, including *in vitro* fertilization, to overcome an age-related decline in fertility, of which the British Fertility Society has highlighted that few females are aware. Older females are more likely to have medical co-morbidities and an increased risk of adverse pregnancy outcomes. Access to fertility treatment in the UK is recommended, but local referral criteria and available funding are a source of inequality.

Further reading

MBRRACE-UK Mother and Babies: Saving Lives, Improving Mothers care 2018–20. 2022. Available at: www.npeu.ox.ac.uk.

Miscarriage and stillbirth

Miscarriage is a pregnancy that ends spontaneously before the fetus reaches a viable gestation. The term includes all pregnancy losses between conception and 24 weeks' gestation. Sporadic miscarriage is the most common pregnancy complication; most cases are thought to be caused by random genetic abnormalities in the embryo. The incidence of clinically recognized miscarriage is thought to be 15% to 20% of all pregnancies, but the total number of very early conceptions lost may be as high as 50%. Sensitive home pregnancy testing kits have identified many biochemical pregnancies that fail to implant successfully and resolve spontaneously as a delayed or heavy menstrual period. Miscarriage incidence decreases with advancing gestation, with rates of second trimester loss between 1% and 2%.

Maternal age at conception and reproductive history are independent risk factors for miscarriage. The risk of fetal loss rises steeply after 35 years of age, reaching 75% by 45 years presumed to be because of poor-quality oocytes leading to chromosomally abnormal conceptions. Delayed childbearing not only leads to increased rates of miscarriage, but also ectopic pregnancy, and later pregnancy complications, including stillbirth. Less common factors contributing to early pregnancy loss are uterine abnormalities, infection, endocrine disorders, and exposure to teratogenic drugs and chemicals.

Different types of miscarriage are classified by clinical presentation and the results of investigations (Box 47.12). Vaginal bleeding in early pregnancy is common and does not always indicate miscarriage (Box 47.13). A detailed clinical history is needed. Clinical assessment includes:

- ***Speculum examination*** assesses the cervix and excludes other lower genital tract pathology (e.g., cervical polyp, ectropion). A closed cervix in the context of miscarriage means either a 'complete miscarriage' or a 'missed miscarriage' has occurred in which the pregnancy has stopped growing but the patient has not yet had pain or bleeding. An open cervix represents either a 'threatened miscarriage' if the pregnancy remains in the uterus or an 'incomplete miscarriage' often with products of conception seen in the cervical os.
- ***Transvaginal ultrasonography*** determines pregnancy location, gestation, and viability. It offers valuable reassurance if fetal heart activity is seen.
- **Serial serum beta-human chorionic gonadotrophin (β-hCG) concentrations** are measured when a conceptus cannot be visualized and the site of pregnancy is therefore unknown.

Box 47.12 Classification of miscarriage

Type of miscarriage	Uterus ultrasonography	Clinical presentation
Biochemical pregnancy	Empty uterus	PV bleeding; 'late period' after positive pregnancy test result
Threatened miscarriage	Intrauterine pregnancy	PV bleeding and pain; cervical os closed
Inevitable miscarriage	Intrauterine pregnancy	PV bleeding and pain; cervical os open
Incomplete miscarriage	Retained products of conception	PV bleeding and pain; cervical os open ± products in cervical os
Complete miscarriage	Uterus empty; no retained products	Pain and PV bleeding resolved; cervical os closed
Missed miscarriage	Fetal pole >7 mm (or gestation sac 25 mm) and no fetal heartbeat	PV bleeding and pain

PV, Per vaginam.

Box 47.13 Causes of vaginal bleeding in early pregnancy

- Threatened, inevitable, or incomplete miscarriage
- Bleeding from a viable pregnancy
- Ectopic pregnancy
- Lower genital tract pathology, e.g., cervical polyp or ectropion
- Trophoblastic disease

Bleeding and pain can be severe when products of conception distend the cervical canal. Prompt removal is required using sponge forceps. Severe pain localized to the iliac fossa or peritonism suggests a ruptured ectopic pregnancy, and urgent laparoscopic assessment and management may be needed. However, many ectopic gestations resolve spontaneously by being resorbed or miscarried from the tube into the peritoneal cavity.

Management of miscarriage depends on clinical presentation and patient choice. First-line management of miscarriage in the UK is expectant for 7 to 14 days. Prostaglandins (oral or vaginal misoprostol) induce uterine contractions and expulsion and are cheap and effective. A follow-up urinary pregnancy test or pelvic scan is needed to confirm miscarriage is complete. Patient preference, risk of haemorrhage, previous adverse events, and evidence of infection may warrant surgical management (evacuation of the retained products of conception [ERPC]). ERPC can be performed by suction under general anaesthesia or by manual vacuum aspiration (MVA) under local anaesthesia. Associated morbidity is uncommon but includes uterine perforation, cervical trauma, intrauterine adhesion formation, and postoperative pelvic infection.

Recurrent miscarriage

Recurrent miscarriage is defined as three or more consecutive miscarriages. It affects 1% to 2% of females trying to conceive and is associated with significant psychological morbidity. Investigation considers genetic (parental translocations), anatomical, infective, endocrine, immune, and thrombophilic causes. Most cases remain unexplained, which conveys a good prognosis for future pregnancy outcomes.

Antiphospholipid syndrome is present in up to 20% of recurrent miscarriages. Subsequent pregnancies are managed with a low-dose aspirin, sometimes with addition of low-molecular-weight heparin, achieving a live birth rate of more than 75%.

Females with a history of recurrent miscarriage have an increased risk of pre-eclampsia, fetal growth restriction, and preterm delivery when they achieve a successful pregnancy, presumed to be caused by factors related to the quality and depth of early implantation. There is also an increased risk of stroke at an earlier age than females with no miscarriage history.

Miscarriages can lead to anxiety, depression, denial, anger, blame, severed relationships, and a sense of loss and inadequacy. Compassionate counselling services and access to support are needed.

Stillbirth

Various definitions exist that distinguish stillbirth from miscarriage, generally based on gestational age (varying from 20–28 weeks). The World Health Organization (WHO) defines stillbirth as a pregnancy loss with a fetal weight of 1000 g or more, an assumed equivalent of 28 weeks' gestation. Worldwide, there were 2.6 million third-trimester stillbirths in 2015. Despite reductions in maternal and child mortality rates, the number of stillbirths remains high. This includes a high number of intrapartum losses (1.3 million), many of which are preventable.

Pregnancy

Complications that arise in pregnancy and confer a risk of future disease include:

- ***Pre-eclampsia*** and ***pregnancy-induced hypertension*** resolve after delivery but are associated with an increased risk of ischaemic heart, cerebrovascular disease, and chronic kidney disease later in life.
- Approximately 50% of those with ***gestational diabetes*** will develop type 2 diabetes within 10 years. Annual screening for diabetes should take place after pregnancy, along with inclusion in a diabetes risk register and diabetes prevention programme in primary care. Children born of pregnancies affected by diabetes are more likely to develop diabetes later in life.
- ***Teenage pregnancies*** and ***preterm deliveries*** repeat across generations.
- Perinatal ***mental health disorders*** affect one in five females and are the leading direct cause of maternal death in the first postpartum year. The 2022 Mothers and Babies: Reducing Risk through Audits and Confidential Enquiries across the UK (MBRRACE-UK) report into maternal mortality in 2018 to 2020 showed death by suicide during or up to 6 weeks after the end of pregnancy was three times higher compared with 2017 to 2019. Psychiatric and cardiovascular disorders are now responsible for the same number of maternal deaths in the UK.

Pregnancy should be viewed as the healthcare opportunity of two lifetimes, offering significant potential for population health gain and an opportunity to tackle and interrupt cross-generational transmission of ill-health.

The postnatal visit

The postnatal visit is an ideal time to plan future healthcare needs for both parent and baby. It can positively impact lifestyle, including factors such as obesity, exercise, dental health, contraception, smoking,

and drug misuse. A review of the recent pregnancy will provide valuable information on the risk of future morbidity and offers an opportunity to prevent and screen for diabetes, cardiovascular disease (CVD), and recurrent pregnancy complications. This visit also provides a pivotal opportunity to assess mental health, including the postnatal blues, postnatal depression, and puerperal psychosis.

Contraception

Preventing unplanned pregnancy is essential for sexual, physical, and mental wellbeing. Those who lack access to contraceptive methods are at increased risk of morbidity and mortality because of obstetric complications and the consequences of illegal abortion. More than 30% of maternal deaths and 10% of child deaths can be prevented by ensuring a 2-year interval between pregnancies.

Efficacy of contraception methods depends on mechanism of action and ease of use (**Box 47.14**). Failure rates of different contraceptive methods are reported according to 'perfect use' and 'typical use'. 'Perfect use' represents correct and consistent use. However, 'typical use' should be presumed because it considers human error (e.g., split condom, forgotten contraceptive pill). Each contraceptive option has advantages and disadvantages. The UK Medical Eligibility Criteria for Contraceptive Use (UK-MEC) criteria provide evidence-based guidance on the safe use of contraception in those with underlying medical conditions. Side-effects must be judged against the hazards of pregnancy, especially if other contraceptive methods are not practicable or acceptable.

Hormonal contraception

Hormonal contraception includes pills, transdermal patches, vaginal rings, subdermal implants, and intrauterine devices (IUDs). They can be either a combination preparation (oestrogen and progesterone) or progesterone alone.

- ***Combined oestrogen and progesterone contraceptives*** prevent ovulation, make the endometrium hostile to implantation, and alter cervical mucus. Oestrogen suppression by gonadotrophins prevents follicular development, ovulation, and luteinization. The side-effects of these preparations are shown in **Box 47.15**. Most serious side-effects are rare with oestrogen doses of 20 to 30 mcg. Thromboembolism may be slightly more common with newer third-generation pills containing desogestrel and gestodene (30 in 100 000 patient-years) compared with second-generation pills containing levonorgestrel and norgestrel (15 in 100 000 patient-years) and no treatment (5 in 100 000 patient-years), although overall incidence is low. The risks of the combined pill are increased in smokers and those with obesity or cardiovascular risk factors.
- ***Progestogen-only preparations*** act on cervical mucus thickening and thin the endometrium. Some preparations (e.g., desogestrel) also inhibit ovulation. They avoid the side-effects of oestrogen and are suitable for those with risk factors for VTE and CVD. Common side-effects include irregular bleeding, simple ovarian cysts, mood swings, breast tenderness, and acne.

Long-acting reversible contraception

Long-acting reversible contraceptive methods include both progesterone-only methods (subdermal implants, injectables, and IUDs) and non-hormonal methods (copper coil). They have lower failure rates and better continuation rates than user-dependent contraception. They are more cost-effective than oral contraceptives.

- ***Subdermal implants*** are inserted in the upper arm under local anaesthetic and last 3 years. There is return of fertility after removal. Irregular bleeding is common in the initial period and sustained for some.
- ***Depot injectables*** are given intramuscularly and are effective for 12 weeks. Side-effects include irregular periods and a delay in the return of fertility of up to 6 months.
- ***IUDs*** are inserted into the uterine cavity by trained personnel. A thread is left protruding from the cervix into the vagina, allowing removal by traction.
  - Levonorgestrel-releasing intrauterine systems (e.g., Mirena) are licensed for contraception, for treatment of HMB, and as part of hormonal replacement therapy regimens. The duration of use is 3 to 5 years, depending on the device. Irregular vaginal spotting tends to resolve by 6 months after insertion but can be troublesome and is the most common reason for removal, aside from wanting to conceive. Other side-effects may include acne, breast tenderness, and mood swings, but these often regress. With use, there is transition to amenorrhoea or lighter, less painful periods, making them effective treatment for patients with pelvic pain caused by endometriosis and adenomyosis.
  - Copper IUDs are cheap and effective for up to 10 years. They are toxic to both eggs and sperm (spermicidal) before fertilization can occur. They are suitable for those in whom hormonal use is contraindicated or unwanted. Patients should be counselled about the possibility of heavier or more painful periods.

Barrier methods

Condoms, diaphragms, spermicides, and sponges require consistently correct use. They have high typical-use failure rates but do offer protection against STIs.

Emergency contraception

Emergency contraception can be used after unprotected intercourse to prevent unwanted pregnancy. The earlier the use after unprotected sexual intercourse, the greater the efficacy. Levonorgestrel needs to be taken within 72 hours and is available without prescription in the UK. Ulipristal acetate delays ovulation and can be used up to 120 hours after intercourse. It requires a prescription in the UK. The failure rate of both pills is 1% to 2%, and a repeat dose should be given if vomiting occurs within 3 hours. Copper IUD insertion within 5 days of unprotected intercourse is a highly effective emergency contraceptive and provides ongoing effective contraception for up to 10 years.

Sterilization

Male and female sterilization offers irreversible contraception, with low failure rates of 1 in 2000 and 1 in 200, respectively. Male vasectomy can be performed under local anaesthetic and carries fewer risks than tubal ligation, clipping, and salpingectomy in females.

Termination of pregnancy

Unplanned pregnancy is common, and termination of pregnancy is a frequently performed gynaecological procedure. One in three females will have a termination in their lifetime in the UK. Termination of pregnancy continues to generate ethical, cultural, and political debate, and legal frameworks governing abortion services vary considerably between jurisdictions. Globally, 50 million pregnancies end in termination, of which 20 million are illegal. Approximately 70 000 maternal deaths occur each year because of unsafe abortions.

Box 47.14 Contraceptive methods, advantages, disadvantages, and effectiveness

	Type of contraception and mode of action	Advantages	Disadvantages	Contraindications	Typical use failure rate (%)[a]
Non-hormonal	**Fertility awareness** Intercourse is restricted to non-fertile times using: 1. Body temperature (rises 0.2°–0.4°C after ovulation) 2. Cervical mucus changes 3. Menstrual cycle length	No physical side-effects No co-morbidity effect or consequence Acceptable across faiths and cultures	Relies on regular menstrual cycles, record-keeping, and commitment Can take three to six cycles to learn effectively Illness, stress, and travel may alter fertility indicators No protection against STIs	Nil	24
	Withdrawal Penile withdrawal before ejaculation ('coitus interruptus')	No physical side-effects No co-morbidity effect or consequence	Variable control of ejaculation and presence of sperm within pre-ejaculatory fluid		22
	Barrier methods Condom Femidom Diaphragm Cervical cap	Effective if used correctly No systemic side-effects Widely available Protection from STI and HIV transmission with condoms and femidoms	Must be applied before penetration Can reduce sensation Diaphragm and cap must be fitted and checked by a trained professional Placement can be incorrect Condoms can split with incorrect use	Latex allergy for latex-containing products (non-latex polyurethane and polyisoprene products are widely available) Concurrent use of spermicide (e.g., nonoxynol-9) does not increase protection from STI and may increase HIV transmission	Male condom: 18 Femidom: 21 Female diaphragm: 12
	Copper coil	No hormonal side-effects Fertility resumes after removal	Heavier or more painful periods Low risk of expulsion, uterine perforation and 'lost threads' requiring hysteroscopic retrieval	48 hours to 4 weeks postpartum Current PID	0.8
	Female sterilization (tubal ligation)	No hormonal side-effects Permanent contraception	Irreversible	Possible future fertility wishes	0.5
	Male sterilization (vasectomy)	No hormonal side-effects Permanent contraception	Delay to azoospermia, which requires confirmation before unprotected intercourse Irreversible Low risk of chronic testicular pain	Possible future fertility wishes	0.15
Hormonal	**Combined hormonal contraception** Combined oral contraceptive pill Patch Vaginal ring	Reliable if taken correctly Convenient and not intercourse related Improves dysmenorrhoea, menorrhagia, and premenstrual symptoms Reduced functional ovarian cysts Reduced incidence of ovarian, endometrial, and colon cancers	VTE Hypertension Relative increase in cardiovascular and cerebrovascular disease risk Small increase in cervical intraepithelial neoplasia and cervical cancer after prolonged use Mixed evidence regarding breast cancer risk Breakthrough bleeding, headaches, nausea, breast tenderness, mood changes Does not protect against STIs	History of VTE Atrial fibrillation Reduced effectiveness with enzyme-inducing medication Reduced effectiveness with vomiting or diarrhoea for oral methods	9
	Progesterone-only pill	Can be used if contraindication to oestrogen methods Safe use in lactation Improved dysmenorrhoea or menorrhagia	Irregular breakthrough bleeding in early use Acne, breast tenderness, headaches	Breast cancer	9
	Progesterone-containing intrauterine system (e.g., Mirena, Jaydess)	Shorter, lighter periods; amenorrhoea Fertility resumes after removal	Uncomfortable insertion Uncommon complications expulsion, uterine perforation, lost threads requiring hysteroscopic retrieval If pregnancy occurs, risk of ectopic pregnancy	48 hours to 4 weeks postpartum Current PID Breast cancer	0.2
	Subdermal implant (e.g., Nexplanon, Implanon)	Shorter, lighter periods; amenorrhoea	Breakthrough bleeding	Breast cancer	0.05
	Depot progesterone injections	Shorter, lighter periods; amenorrhoea Can use with enzyme-inducing medication	Can delay fertility for up to 1 year after stopping Acne, breast tenderness, breakthrough bleeding, headache	Breast cancer	6

[a]Percentage of females conceiving during the first year of typical use.

HIV, Human immunodeficiency virus; *PID*, pelvic inflammatory disease; *STI*, sexually transmitted infection; *VTE*, venous thromboembolism.

Box 47.15 Combined oestrogen–progesterone oral contraceptives: adverse effects

General
- Loss of libido
- Pigmentation (chloasma)
- Breast tenderness

Cardiovascular
- Increased blood pressure
- Myocardial infarction
- Stroke (migraine with aura is a risk factor)

Gastrointestinal
- Nausea and vomiting
- Gallstones
- Abnormal liver biochemistry
- Hepatic adenoma

Nervous system
- Headache
- Migraine
- Mood changes

Malignancy
- Increased risk of breast cancer (reduced risk of ovarian and endometrial cancer)

Gynaecological
- Amenorrhoea
- 'Spotting'
- Cervical erosion (ectropion)

Haematological
- Increased risk of venous thromboembolism

Endocrine or metabolic
- Conflicting data on glucose tolerance
- Dyslipidaemia

Drug interactions
- Enzyme-inducing antibiotics, e.g., rifampicin
- Carbamazepine
- Phenytoin
- St John's wort

Legalized termination in high-income countries is safe. The risk of death from an early surgical termination of pregnancy is less than 1 per 100 000, far lower than the maternal mortality rate associated with pregnancy. Of the 190 000 procedures performed annually in the UK, 95% of them are performed under Clause C of the Abortion Act, which states that pregnancy has not exceeded its 24th week and that continuance of the pregnancy would either involve more risk than termination or injury to the physical or mental health of the pregnant patient. The earlier an abortion is performed, the safer it is. Currently, 91% of abortions in the UK are carried out before 13 weeks' gestation. Only 1% of terminations are performed for a fetal abnormality. Contraception should be discussed before termination and started immediately after the procedure to avoid a further unintended pregnancy.

Early medical termination (before 9 weeks' gestation), using mifepristone for cervical preparation and misoprostol to expel the products, is complete in more than 80% of cases. Patients may be observed for 4 to 6 hours in a clinical setting after misoprostol insertion, although early medical termination at home is increasingly supported following UK legislation passed in 2018. Later medical termination requires careful counselling. Feticide with intracardiac potassium injection is performed before medical termination after 21 weeks and 6 days' gestation to ensure a live birth does not occur.

Termination of pregnancy can be undertaken surgically under local or general anaesthesia. The type of procedure depends on the gestation, availability, and patient preference. Early surgical terminations before 7 weeks' gestation can be performed by MVA syringe under local anaesthesia in suitable patients. Surgical terminations before 14 weeks' gestation are usually performed as suction terminations under general anaesthesia. Preoperative priming of the cervix with either oral or vaginal misoprostol reduces the risk of cervical trauma and haemorrhage; osmotic cervical dilators may be considered in later gestations or if there are contraindications to misoprostol. Surgical terminations after 14 weeks' gestation are done via a specialist service in the UK.

Contemporary techniques and routine screening for pelvic infection in high-risk females mean that the risk of subfertility after termination is low. Later terminations are associated with more complications, including an incomplete procedure with retained products of conception, uterine perforation, intrauterine adhesion formation (Asherman's syndrome), pelvic infection, and trauma to the genital tract.

Many females are emotionally vulnerable after termination. Feelings of guilt and regret frequently mix with relief. However, there is no evidence of increased severe psychiatric disease after termination. Psychological sequelae can be reduced with careful counselling before the procedure and postabortion support.

Further reading

British Pregnancy Advisory Service (BPAS) https://www.bpas.org/abortion-care/abortion-aftercare/.

Menopause

Menopause, or cessation of periods, occurs naturally at 45 to 55 years old. During the fifth decade of life, FSH and then LH concentrations rise, probably as follicle supply diminishes. Oestrogen concentrations fall, and the menstrual cycle becomes disrupted. Most females notice a gradual change to irregular, scanty periods, although sudden amenorrhoea or menorrhagia occur in some. Eventually, the menopausal pattern of low oestradiol concentrations with elevated LH and FSH (usually >50 and >30 U/L, respectively) is established. Premature menopause may occur because of surgery, radiotherapy to the ovaries, or ovarian disease.

In patients older than 45 years of age, the following can be diagnosed without laboratory tests:
- ***Perimenopause:*** vasomotor symptoms (hot flushes, night sweats) and irregular menstrual periods
- ***Menopause:*** no period for at least 12 months without the use of hormonal contraception (or based on symptoms in those without a uterus)

Features of oestrogen deficiency are hot flushes (which occur in most and can be disabling), vaginal dryness, and breast atrophy. There may also be loss of libido, self-esteem, non-specific aches and pains, sleep disturbance, irritability, depression, loss of concentration, and weight gain. There is a rapid loss of bone density in the 10 years after menopause, and the female premenopausal protection against ischemic heart disease and stroke disappears.

Management of menopause should include explanations with information detailing:
- The stages of menopause
- Common symptoms
- Lifestyle changes to help general health and wellbeing, such as cessation of smoking, increased aerobic exercise, and avoidance of food triggers
- The benefits and risks of treatments
- The long-term implications of menopause

Box 47.16 Risks and benefits of hormone replacement therapy (HRT)

Benefits

- Symptomatic improvement in menopausal symptoms
  - Oestrogen-deficient symptoms respond well to systemic oestrogen replacement
  - Vaginal symptoms respond to local oestrogen (e.g., oestriol vaginal creams, pessaries)
- Protection against osteoporotic fractures of the wrist, spine, and hip
- Reduced risk of colorectal cancer
- Possible improvement in general wellbeing and mood
- Improved libido
- Reduced risks of coronary heart disease and Alzheimer's disease when oestrogen is started early
- Reduced risk of type 2 diabetes mellitus

Risks

- Oestrogen only HRT used when the uterus has been removed has little or no increased risk of breast cancer
- Combined oestrogen and progestogen HRT is associated with a higher risk of breast cancer comparable to other lifestyle factors (see **Box 47.17**)
- Oral HRT tablets increase the risk of VTE; transdermal (e.g., patch, gel, spray) preparations do not increase risk of VTE
- Possible increase in ischaemic heart disease when combined HRT started after 60 years of age or with pre-existing IHD. The first 10 years after menopause is considered the cardiovascular 'window of opportunity'
- Increased stroke risk with oral HRT started after 60 years of age

IHD, Ischaemic heart disease; *VTE,* venous thromboembolism.

Box 47.17 Breast cancer risk attributed to lifestyle factors and hormone replacement therapy (HRT)

Lifestyle risk factors		Absolute excess risk per 1000 females in 5 years aged 50–59 years
Postmenopausal obesity	Overweight vs healthy weight[a] Obese vs healthy weight[a]	+4 +10
Alcohol	4–6 units/day ≥6 units/day	+8 +11
Unopposed oestrogen use for 5 years	WHI study, 2020 NICE Menopause Guidance, 2015 CGHFBC, 2019	–6 +3 +3
Combined HRT use for 5 years	WHI study, 2020 NICE Menopause Guidance, 2015 CGHFBC, 2019	+8 +9 +10

[a]Healthy weight, body mass index (BMI) <25 kg/m^2; overweight BMI, 25–29.9 kg/m^2; obese BMI ≥30 kg/m^2.

CGHFBC, Collaborative Group on Hormonal Factors in Breast Cancer; *NICE,* National Institute for Health and Care Excellence; *WHI,* Women's Health Initiative.

Treatment options include:

- Hormonal (e.g., hormone replacement therapy [HRT])
- Non-hormonal (e.g., clonidine)
- Complementary (e.g., isoflavones, black cohosh)
- Psychological (e.g., CBT)

Hormone replacement therapy

Symptomatic patients are offered HRT with the lowest effective dose. The benefits and risks are detailed in **Box 47.16**. Risks depend on whether oestrogen is used in isolation or in combination with progesterone. Individual risks and benefits also depend on the background risk of specific disease. Transdermal oestrogen preparations are associated with a low risk of VTE, but there is no evidence that different hormone preparations or routes of administration (oral, transdermal, implant) modify cardiovascular and cancer risks. The Women's Health Initiative study estimated that, over 5 years of treatment, 1 extra person in every 100 develops an illness that would not have occurred had they not taken HRT, though the association between long-term HRT and an increased risks of breast cancer, stroke, and VTE has recently been challenged. There is little difference in the risk of breast cancer incurred by lifestyle factors, including current and past use of HRT (**Box 47.17**). Ischaemic heart disease increases after menopause, but this occurs 10 years later than in males. HRT is not recommended for the prevention of postmenopausal osteoporosis in the absence of menopausal symptoms. However, patients with premature ovarian insufficiency (menopause before 40 years of age) should be offered HRT until age 50 for bone and cardiovascular protection.

The decision about HRT is individualized and based on:

- Severity of menopausal symptoms
- Individual risk of conditions that may be prevented or made more likely by HRT
- Patient choice.

Patients with a uterus require both oestrogen and progesterone. Oestrogen provides symptomatic relief, and progesterone provides endometrial protection from unopposed oestrogen, which could lead to endometrial hyperplasia or cancer.

Selective oestrogen receptor modulators (e.g., raloxifene) offer a potentially attractive combination of positive oestrogen effects on the bone and cardiovascular system, with no impact on oestrogen receptors of the uterus and breast associated with cancer incidence; however, long-term outcome studies are awaited.

Tibolone is a once-per-day oral synthetic hormone with combined oestrogenic, progestogenic, and testosterone effects branded as the first 'bleed-free' HRT. Tibolone appears to be particularly helpful for patients with endometriosis and fibroids, reducing the risk of hormonally driven disease flares.

Patients with vaginal atrophy are treated with topical oestrogens, supplemented with vaginal moisturisers and lifestyle modifications (unperfumed soaps and products, water-based lubricant before intercourse). Long-term use is often required.

Further reading

British Menopause Society: Benefits and risks of HRT before and after a Breast Cancer Diagnosis. https://thebms.org.uk/wp-content/uploads/2024/04/08-BMS-ConsensusStatement-Benefits-risks-of-HRT-before-after-a-breast-cancer-diagnosis-MARCH2024-A.pdf.

MBRRACE-UK Mother and Babies: Saving Lives, Improving Mothers Care 2018–20. 2022. Available at: www.npeu.ox.ac.uk.

National Institute for Health and Care Excellence. *NICE Clinical Guideline NG23: Menopause: Diagnosis and Treatment*. NICE last updated 2019; https://www.nice.org.uk/guidance/ng23.

National Institute for Health and Care Excellence. *NICE Clinical Guideline NG123: Urinary Incontinence and Pelvic Organ Prolapse in Women: Management*. NICE last updated 2019. https://www.nice.org.uk/guidance/ng123.

Royal College of Obstetricians and Gynaecologists (RCOG) *Better for Women: Improving the Health and Wellbeing of Girls and Women, 2019*. chrome-extension: https://www.rcog.org.uk/media/h3smwohw/better-for-women-full-report.pdf.

Female urinary incontinence

Urinary incontinence is common in females with overall population prevalence of 25% to 45% across all ages in adulthood. Prevalence increases with age; 40% of females are affected by the age of 70 years with idiopathic detrusor activity, diabetes, mobility, and medication all possible contributors.

Types of urinary incontinence (Box 47.18) include:

- Stress incontinence
- Overactive bladder and urgency incontinence
- Mixed urinary incontinence.

Urinary incontinence negatively impacts quality of life, with many too embarrassed to seek medical help, leading to underreporting and unsustainable individual-led lifestyle modifications, such as avoiding sexual intercourse or public sporting activities.

Lifestyle interventions, including weight loss, smoking cessation, caffeine reduction, and timed fluid restriction, can significantly reduce symptom burden. Physiotherapy and bladder retraining may be helpful in cases of stress and urgency, respectively. In postmenopausal females, oestrogen deficiency exaggerates urinary symptoms, and topical vaginal or systemic HRT are treatment options. Specialist investigations includes pelvic floor assessment, including exclusion of uterine prolapse, urinary tract imaging, and urodynamic evaluation when appropriate. Further management is based on the type of incontinence and the female patient's preferences and future fertility wishes (Box 47.19).

Box 47.18 Types of urinary incontinence in females

- ***Stress incontinence:*** involuntary urine leakage on effort, exertion, sneezing, or coughing
- ***Urge incontinence:*** involuntary urine leakage accompanied or immediately preceded by a sudden compelling desire to urinate that is difficult to delay; can occur with or without overactive bladder syndrome
- ***Mixed urinary incontinence:*** involuntary urine leakage associated with urgency and exertion, effort, sneezing, or coughing
- ***Other:*** overflow incontinence caused by chronic urinary retention, urogenital fistula, or urethral diverticulum

Uterine fibroids (leiomyomata)

These are the most common benign pelvic tumours in females of reproductive age (Fig. 47.6). Prevalence increases with age and is

Box 47.19 Management of overactive bladder and stress urinary incontinence

Order of treatment	Treatment	Mechanism of action	Contraindications	Side-effects
Overactive bladder				
First	**Oxybutynin**	Anticholinergic and mucosotropic	Glaucoma, ulcerative colitis, myasthenia gravis, severe liver impairment	Dry mouth, confusion
Second	**Tolterodine** **Darifenacin**	Anticholinergic and mucosotropic		
Third	**Mirabegron**	Beta-3 adrenoreceptor agonist	Atrial fibrillation	Vaginitis
Overactive bladder: nocturia				
First	**Desmopressin**	V2-agonist	Aged older than 65 years, hypertension, cardiac disease	Fluid retention, hyponatraemia
Second	**Vaginal oestrogen**	Hormone replacement		
Overactive bladder: failed medical treatment				
First	**Botulinum toxin**	Prevention of bladder contraction	Urinary retention	Urinary retention (intermittent self-catheterization needed in 10%–15%)
Second	**Presacral nerve stimulation with an implant**	Excitatory and inhibitory reflexes on the bladder		Infection or migration of the prosthesis
Third	**Posterior tibial nerve stimulation**	Impulses from tibial nerve block abnormal bladder signals		
Fourth	**Surgical reconstruction**			
Stress incontinence				
First	**Duloxetine**	Selective serotonin-noradrenaline reuptake inhibitor	Urinary hesitancy, glaucoma, liver disease	Nausea and vomiting, change in bowel habits, dry mouth, blurred vision, anorexia
Stress incontinence: failed medical treatment				
First	**Burch colposuspension**			5% risk of need for intermittent self-catheterization Risk of developing an enterocoele (14%)
Second	**Autologous rectus fascial sling or bulking agents**			5% risk of need for intermittent self-catheterization
Third	**Tension-free vaginal tape**			Chronic pelvic pain, 10%; dyspareunia, 10%; bladder injury, 4%; mesh erosion, 1%–2%

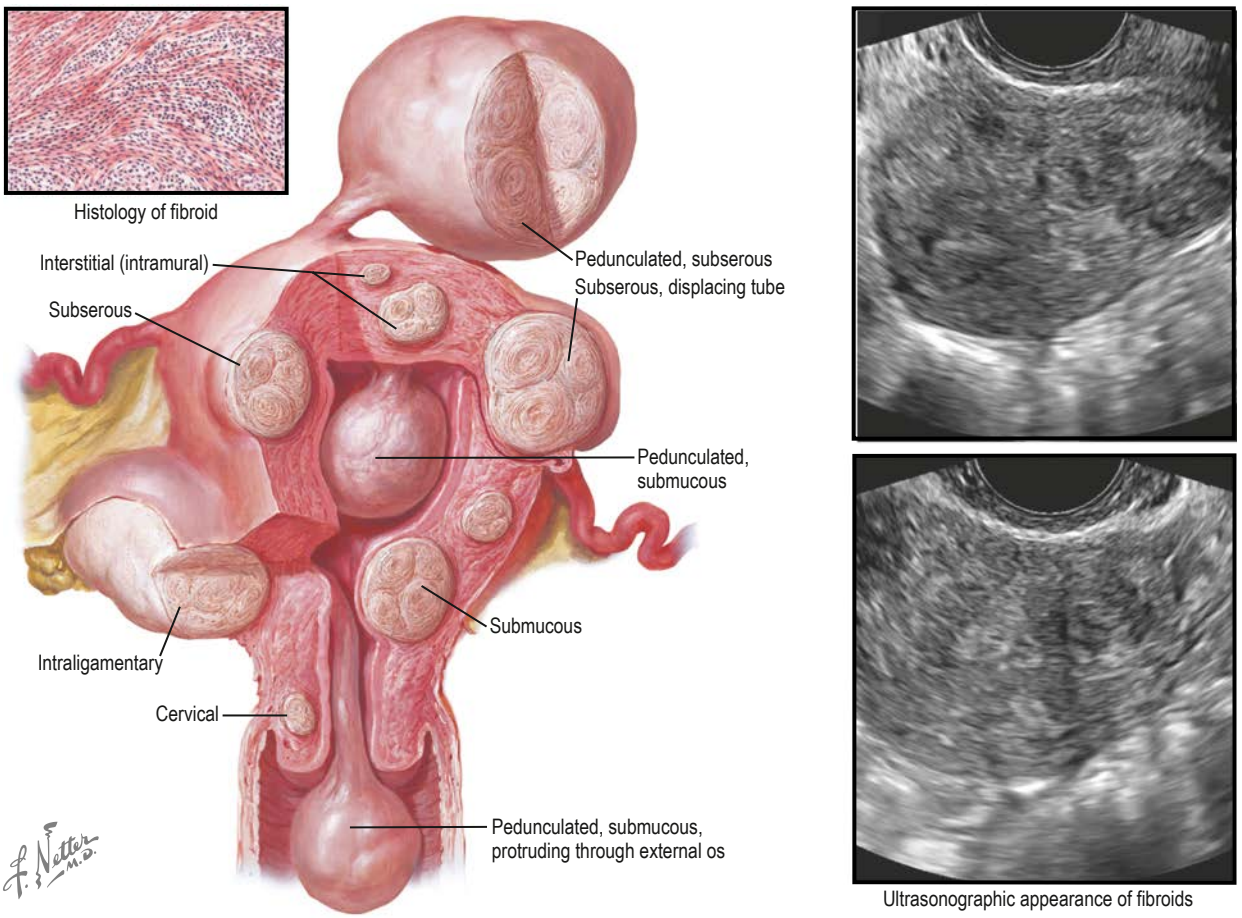


Fig. 47.6 **Uterine fibroids (leiomyomata).**

higher in females of Afro-Caribbean ethnicity. Heavy and prolonged periods, pressure symptoms, and reproductive dysfunction are common presentations. Mechanisms by which fibroids cause HMB are poorly understood, but submucosal lesions within the uterine cavity are particularly troublesome.

Treatment is with:

- Magnetic resonance-guided and -focused ultrasound uterine artery embolization (UAE). Uterine arteries are selectively catheterized via femoral approach, and microbeads are injected until flow is demonstrably reduced. There is a small risk of menopause in which the ovarian circulation arises dominantly from the uterine arteries; this should be screened for with angiography before embolization. Pregnancy after UAE carries an increased risk of abnormal placental implantation; therefore, UAE should generally only be considered if future pregnancy is not intended. Risks of sarcoma with fibroids (0.05%–0.29%) must be presented in a balanced way when considering non-hysterectomy management.
- GnRH agonists can alleviate HMB as a short-term measure, shrinking overall fibroid size and vascularity before surgery.
- Surgical excision. In laparoscopic myomectomy, fibroids are morcellated to allow removal.

Gynaecological malignancy

Gynaecological cancer is any cancer that starts in the female reproductive organs. The five main types of gynaecological cancers are cervical, endometrial, ovarian, vaginal, and vulval cancers (**Box 47.20**). Endometrial and ovarian cancers are most common. Patients with suspected cancers should be referred urgently usually via rapid-access gynaecology or colposcopy clinics, where diagnostic imaging and histology can be obtained.

Endometrial (uterine) cancer is the fourth most common cancer in females in the UK. Most cases present with postmenopausal bleeding. This is considered a 'red flag' symptom with 5% to 10% caused by endometrial cancer. It is estimated that 34% of endometrial cancer cases are preventable. Obesity is a contributing factor; increased peripheral androgen conversion to oestriol in adipose tissue leads to unopposed oestrogen action on the endometrial lining.

Ovarian cancer is the fifth most common cancer in females in the UK. Most occurs in postmenopausal females. Presentation is with late, non-specific symptoms, including abdominal pain, bloating, and weight loss. Management is with specialist surgery and chemotherapy.

Cervical cancer is the 14th most common cancer in females in the UK. Almost all (99.8%) are caused by HPV infection. National cervical cancer screening aims to detect high-risk infections. Females and males aged 12 to 13 years are offered HPV vaccination in the UK; a consequent reduction in incidence of cervical carcinoma *in situ* is already evident.

The management of gynaecological malignancies in discussed in Chapter 41.

Box 47.20 Key clinical characteristics of cervical, endometrial, and ovarian cancers

Cancer	Cervical	Endometrial	Ovarian
UK incidence (cases/year)	3200	9700	7500
5-year survival rate (%)	65	75	46
Peak age of incidence (years)	30–34	75–79	75–79
Ethnicity	Higher mortality rates in minority ethnic populations (excluding White minorities)	Higher incidence in Black ethnic groups Higher mortality rates in minority ethnic populations (excluding White minorities)	
Risk factors (percent of tumours with risk factor)	HPV (>99%) Smoking (21%) HIV	Obesity (34%) Diabetes Endometrial polyps or hyperplasia Unopposed oestrogen PCOS	Genetic (5%–15%): e.g., *BRCA1/2* Obesity (7%) Hormone replacement therapy Endometriosis Diabetes
UK screening	HPV testing for females: every 3 years for ages 25–49 years; every 5 years for ages 50–64 years	None	None; long-term follow-up data awaited
Presenting symptoms	Postcoital bleeding Intermenstrual bleeding Dyspareunia Vaginal discharge Pelvic pain	Postmenopausal bleeding Intermenstrual bleeding Vaginal discharge	Early satiety Loss of appetite Abdominal pain Abdominal bloating Urinary frequency Fatigue Unexplained weight loss Change in bowel habit
Common stage at presentation	1	1	1 in 31%; 3 in 31%
Diagnosis	Cervical biopsy	Endometrial biopsy: target biopsy at hysteroscopy, endometrial Pipelle biopsy	Ultrasonography of the pelvis; CT

CT, Computed tomography; *HIV,* human immunodeficiency virus; *HPV,* human papillomavirus; *LLETZ,* large loop excision of transformation zone; *PCOS,* polycystic ovarian syndrome.
Adapted from Cancer Research UK 2016–2018 data.

Further reading

https://bgcs.org.uk/professionals/guidelines.html. *British Gynaecological Cancer Society Guidelines*.
https://www.cancerresearchuk.org/about-cancer/type and https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type. *Cancer Research UK*.

Domestic abuse and violence

Domestic or intimate partner violence includes physical, sexual, emotional, psychological, and emotional abuse and is a significant public health problem. It negatively impacts physical, mental, sexual, and reproductive health and increases the risk of acquiring human immunodeficiency virus (HIV). The WHO estimates that 30% of females worldwide have been subjected to physical or sexual intimate partner violence or non-partner sexual violence in their lifetimes. One in four females in England and Wales experience domestic abuse or violence. Domestic violence accounts for one-third of violent crimes, costing an estimated £23 billion per annum in the UK alone. The recent SARS-CoV-2 pandemic lockdowns highlighted pre-existing domestic abuse problems with an increase in partner and family abuse. Females in poverty, females from minority ethnic groups, trafficked people, and sex workers are more likely to experience violence and abuse. The odds of being a victim of violence are twofold higher in females with physical disabilities and threefold higher in those with mental health conditions. Around 30% of domestic abuse begins or escalates during pregnancy. The prevalence of domestic violence in females requesting a termination of pregnancy is six times higher than among females attending antenatal clinics. Domestic abuse is the second leading cause of trauma during pregnancy after road traffic accidents, with multiple sites of injury likely. Victims of domestic abuse are significantly more likely to commit suicide.

Domestic violence and abuse are serious health concerns. Healthcare professionals are particularly well placed to identify abuse and intervene; 80% of survivors seek help from health services, with victims commonly presenting to emergency medicine, psychiatry, obstetrics, gynaecology, and primary care services. All healthcare workers need to be trained to recognize signs of violence and abuse, use targeted questioning, and know how to act and refer to ensure safety. Key actions to reduce violence against females include:

- ***Protection*** via an effective criminal justice system: investigation, prosecution, victim support or protection, and perpetrator programmes
- ***Provision*** to allow people to continue with their lives: effective advice and support, emergency and acute services, refuges, and safe accommodation
- ***Prevention*** of violence and addressing attitudes toward violence, including awareness campaigns, safeguarding, early identification, and education

Box 47.21 Classification of female genital mutilation

Type	Definition
I	Partial or total removal of the clitoris, the prepuce (clitoridectomy), or both
II	Partial or total removal of the clitoris and the labia minora, with or without excision of the labia majora (excision)
III	Narrowing of the vaginal orifice with creation of a covering seal by cutting and appositioning the labia minora, the labia majora, or both, with or without excision of the clitoris (infibulation)
IV	Other harmful procedures to the female genitalia for non-medical purposes, e.g., pricking, piercing,[a] incising, scraping, and cauterizing

[a]Piercing is part of World Health Organization classification, but the legal status is unclear in the UK.

World Health Organization. Eliminating female genital mutilation: an interagency statement. Geneva: World Health Organization; 2008.

Female genital mutilation

Female genital mutilation (FGM) refers to all procedures involving partial or total removal of external female genitalia or injury to the female genital organs for non-medical reasons. The different classifications of FGM are shown in **Box 47.21**. More than 200 million females have been subjected to FGM in the 30 countries in Africa, the Middle East, and Asia where FGM is concentrated. In the UK, it is estimated that around 137 000 females have undergone FGM, and 60 000 females younger than 15 years of age are at risk. FGM was made illegal in the UK by the Female Genital Mutilation Act of 2003.

FGM has wide-ranging sequelae. Immediate complications include pain, abscess formation, sepsis, tetanus, haemorrhage, urinary retention, and blood-borne virus transmission. Chronic sequalae include dyspareunia, chronic pain, dysmenorrhoea, UTI and urinary outflow obstruction, post-traumatic stress disorder, and difficulty conceiving. In pregnancy, FGM is associated with increased risks of prolonged labour, postpartum haemorrhage, perineal trauma, caesarean delivery, an increased need for neonatal resuscitation, and risks of stillbirth and early neonatal death.

Management of patients with FGM involves a specialist multidisciplinary team. De-infibulation is a minor surgical procedure that divides the scar tissue at the vaginal introitus in type 3 FGM. It can be performed in the elective clinical setting or intrapartum by a trained health professional.

Further reading

Dey T, Thakar, R. Domestic violence: An invisible pandemic. *Obstet Gynaecol,* 2022; 24:90–2.

Faculty of Sexual and Reproductive Healthcare, Royal College of Obstetricians and Gynaecologists. *The UK Medical Eligibility Criteria for Contraceptive Use 2016*. London: RCOG; Amended 2029; https://www.fsrh.org/documents/ukmec-2016/.

Knight M, Bunch K, Tuffnell D et al. (eds). *MBRRACE-UK Mother and Babies: Saving Lives, Improving Mothers Care 2018–20.* 2022. Available at: www.npeu.ox.ac.uk.

National Institute for Health and Care Excellence. *NICE Guideline 88: Heavy Menstrual Bleeding: Assessment and Management*. NICE Updated 2021; https://www.nice.org.uk/guidance/ng88.

Royal College of Obstetricians and Gynaecologists. *Female Genital Mutilation and Its Management*. Green-top Guideline no. 53. London: RCOG; 2015; https://www.rcog.org.uk/media/au0jn5of/gtg-53-fgm.pdf.

Royal College of Obstetricians and Gynaecologists. *Recurrent Miscarriage (Green-top Guideline No. 17), RCOG*; 2023; https://www.rcog.org.uk/guidance/browse-all-guidance/green-top-guidelines/recurrent-miscarriage-green-top-guideline-no-17/.

MALE HEALTH

History

An accurate, detailed history of lower urinary tract symptoms (LUTS) includes exacerbating factors and the impact on quality of life. Storage and voiding symptoms can be distinguished (**Box 47.22**). Understanding whether a patient has one or both help determine the underlying pathology and necessary treatment

- *Storage LUTS* are caused by disorders that affect storage of urine due to irritation or inflammation of the bladder lining: lower urinary tract infection (UTI), cystitis, sexually transmitted infection, bladder stones, neurological conditions (e.g., spinal cord damage), consumption of bladder irritants (e.g., caffeine), prostatitis.
- *Voiding LUTS* are caused by disorders that affect voiding. Benign and malignant prostate diseases are most common. 'Overflow' voiding is the frequent passage of small volumes of urine, which may be accompanied by lower abdominal pain, tenderness, and a palpable bladder. Involuntary urinary leakage and nighttime incontinence (nocturnal enuresis) require exclusion of high-pressure chronic urinary retention. Obstructive nephropathy and irreversible kidney damage can result if diagnosis is delayed.

Patient-reported outcome measures can be used to objectively capture and record the experience of symptoms. These remove physician bias and provide numerical data that can be helpful in medical research. Examples include the International Prostate Symptom Score (IPSS; see later) and the International Consultation on Incontinence Urinary Incontinence Short Form (ICIQ-UI SF).

Examination

Digital rectal examination

Complete examination includes inspection and palpation of external genitalia and a digital rectal examination (DRE) of the prostate and for anorectal abnormalities. Informed consent is required, and a chaperone should be offered. The patient should lie in a lateral position with the hips and knees flexed to allow inspection of the anus. Lubricant jelly should be used for insertion of the examiner's gloved finger. The finger is then rotated anteriorly to examine the prostate (**Fig. 47.7**). A fleshy or soft-feeling prostate is likely to be benign, whereas a hard

Box 47.22 Lower urinary tract symptoms

Storage symptoms

Dysuria: pain or burning on urination
Frequency: passing small volumes of urine at frequent intervals
Urgency: a sudden urge to pass urine
Urge incontinence: urgency leading to involuntary loss of urine
Nocturia: waking at night to pass urine

Voiding symptoms (usually in males)

Hesitancy: a longer than usual wait for urine flow to begin
Weak stream: a reduction in urine flow
Intermittency: urine flow that stops and starts
Terminal dribbling: weak urine flow continues after an attempt is made to stop
Straining: the need to increase abdominal pressure to urinate

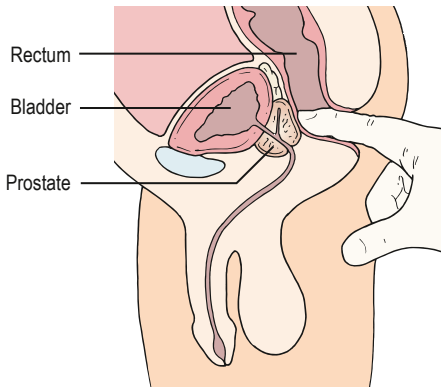


Fig. 47.7 Digital rectal examination. A gloved, lubricated index finger is inserted into the rectum to feel the prostate, which lies anteriorly. (*From Scully C.* Scully's Handbook of Medical Problems in Dentistry. *Elsevier; 2016, with permission.*)

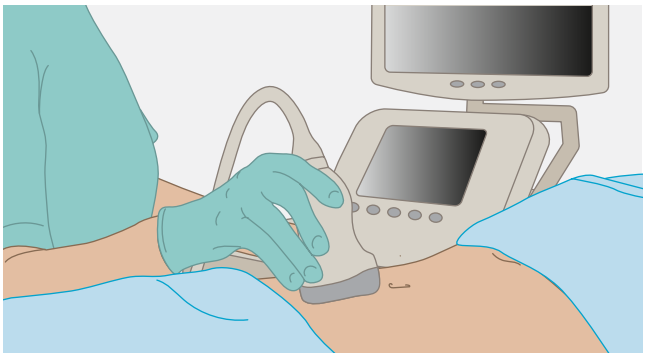


Fig. 47.8 Portable bladder scanner.

Box 47.23 Bladder scanning technique

1. Switch on the bladder scanner.
2. Apply aqueous gel to the head of the transducer.
3. Palpate the symphysis pubis and place the scanner 3 cm above this, pointing down into the pelvis.
4. Hold the transducer steady while pressing the scanning button. The bladder volume is shown on the screen.
5. Repeat the procedure three times, with slight alterations in the position of the transducer.
6. Record the highest volume in the patient's medical record.

or craggy gland is more likely to be malignant. Prostate size can be crudely estimated by measuring the 'finger widths' of the prostate and multiplying it by 10 to provide the approximate size in millilitres. Prostate size assessment by DRE can be correlated with more precise measurements from MRI and transrectal ultrasound examination while the skill is learnt. Tenderness when pressing on the prostate may indicate prostatitis.

Box 47.24 Male catheterization

1. Obtain consent.
2. Use an aseptic no-touch technique.
3. Clean the glans penis with antiseptic solution.
4. Instill lidocaine anaesthetic gel into the urethra.
5. After 5 minutes, insert an appropriately sized catheter (generally 14, 16, or 18 Fr) through the external urethral meatus. When the catheter is inserted to the hilt, urine drains out.
6. Inflate the catheter balloon with 10 mL of saline and withdraw the catheter until the balloon is snug against the bladder neck.
7. Note the volume of urine drained.
8. Replace the foreskin, if present, over the glans penis.

Fig. 47.9 Male catheterization of a mannequin. (*From O'Neill P et al, eds.* Macleod's Clinical OSCEs. *Elsevier; 2016, with permission.*)

Bladder scanning

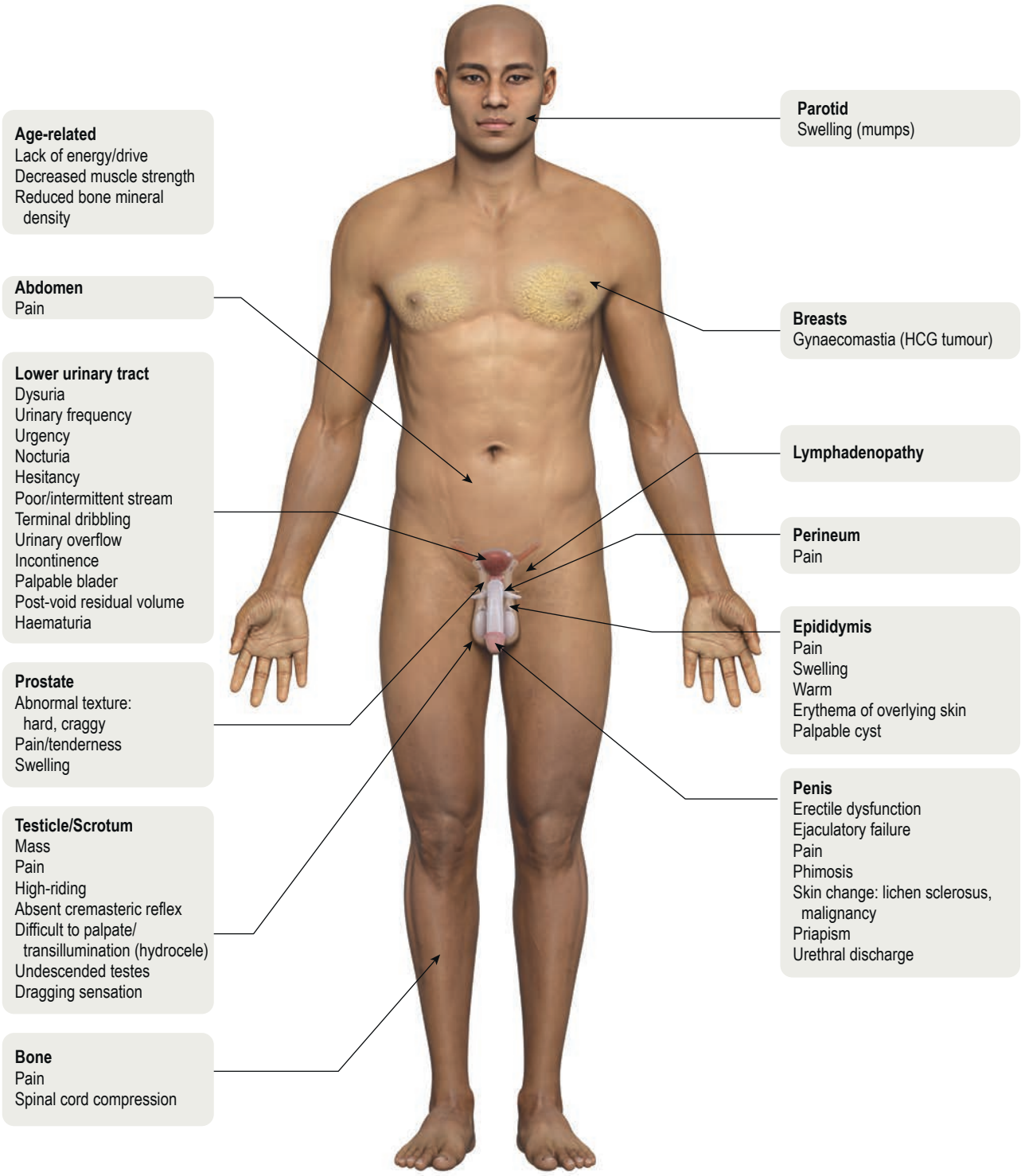
Portable bladder scanners use ultrasound technology to assess bladder volume at the bedside without a requirement for formal training (Fig. 47.8 and Box 47.23). Bladder scans can be used for:

- Diagnosis of urinary retention
- Assessment of successful voiding after catheter removal
- Assessment of complete voiding (post-void residual)

Acute urinary retention is the painful inability to void. Bladder volumes of up to and above 1 L are possible. Urgent urethral catheterization is required (Box 47.24 and Fig. 47.9). A bladder volume of more than 150 mL after attempted voiding is considered significant.

Chronic urinary retention is defined as a painless post-void residual greater than 300 mL, though volumes may be more than this.

Insertion of a urethral catheter is a key clinical skill.



Benign conditions of the prostate

The prostate is an accessory gland of the male reproductive system located under the bladder, through the centre of which passes the first part of the urethra. Alongside the seminal vesicles and other lesser accessory glands, the prostate contributes a small component of the seminal fluid, which helps mature spermatozoa through processes such as semen liquification and sperm head capacitation.

Prostatitis

Prostatitis describes inflammation of the prostate gland. It is divided into three types:

- Acute bacterial prostatitis
- Chronic bacterial prostatitis
- Chronic abacterial prostatitis (a pelvic pain syndrome)

Acute bacterial prostatitis

This condition presents with dysuria, frequency, malaise, fever, and genital region pain. The prostate is swollen and tender on examination. A fluctuant prostate indicates an abscess that warrants either transurethral or transperineal drainage. Infectious agents include Gram-negative uropathogens (chiefly *Escherichia coli*) and organisms that cause STI such as *C. trachomatis* and *Neisseria gonorrhoeae*. Antibiotic treatment is required to prevent sepsis and abscess formation.

Chronic bacterial prostatitis

This is defined as bacterial prostatitis when symptoms persist beyond 3 months. Pain, voiding dysfunction, and ejaculatory problems are the typical features. Longer courses of oral antibiotics with good prostate penetration (e.g., fluoroquinolones) should be prescribed, with awareness of side-effects, including tendinopathy.

Chronic abacterial prostatitis

Chronic abacterial prostatitis is a cause of chronic pelvic pain, the aetiology and pathophysiology of which are poorly understood. Symptoms are present for more than 3 months with no evidence of bacterial colonization in blood, urine, semen, and expressed prostate secretions captured at the time of cystoscopy and prostate massage. The condition generally occurs in males aged 30 to 50 years. Clinical features wax and wane over several years with pain being the main feature but typically include a combination of pain, urinary symptoms, and sexual dysfunction. Pain can be experienced in the penis, scrotum, perineum, or anus. Voiding difficulties are common. Dynamic obstruction to urinary flow can occur in the absence of prostate enlargement, attributed to high levels of adrenergic tone in the smooth muscle of the bladder neck and prostatic urethra. There may be erectile dysfunction or pain on ejaculation. Treatment can be difficult. Options include antibiotics (even in the absence of microbiologically proven infection), low-dose phosphodiesterase-5 (PDE-5) inhibitors (e.g., tadalafil), alpha-blockers (e.g., tamsulosin), 5-alpha-reductase inhibitors (e.g., finasteride), and NSAIDs. Exercise, dietary modification, stress management, relaxation, identification and avoidance of lifestyle triggers, psychotherapies such as CBT, and low-dose amitriptyline may all be beneficial. Chronic pain expertise may be required.

Benign enlargement of the prostate gland

Benign enlargement of the prostate is a very common condition. It occurs most often in males after the age of 60 years. In young males, the normal size of the prostate is 16 mL. Prostatomegaly (enlargement of the prostate) is defined as prostate volume of more than 30 mL. The relationship between prostate size and symptom severity is not reliable nor linear, but the frequency of males requiring intervention does increase with prostate size, particularly when prostate volumes are greater than 100 mL. Endogenous androgens have a key role in pathogenesis, with an absence of prostate enlargement in males with hypogonadism (e.g., Klinefelter syndrome; 47,XXY). The histological appearance is benign hyperplasia affecting glandular and connective tissue elements of the prostate transition zone (the proximal and upper part of the gland that surrounds and can compress the urethra). Enlargement of the prostate gland stretches and distorts the urethra, obstructing bladder outflow.

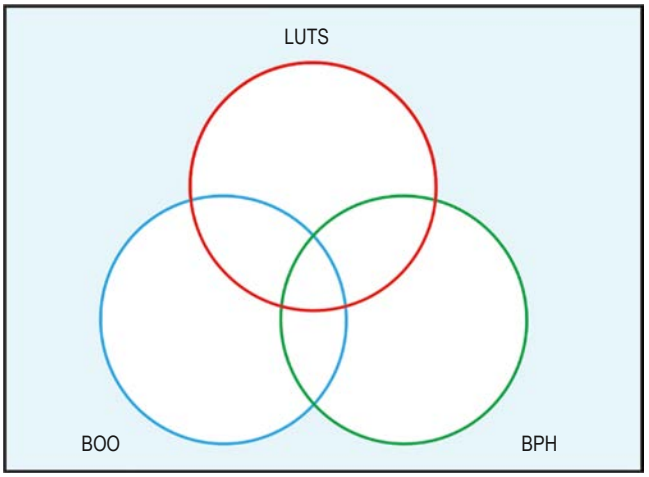


Fig. 47.10 Hald's rings. The three-ring diagram portrays benign prostatic hyperplasia (BPH) as a syndrome of three overlapping yet distinct entities: benign prostatic hyperplasia or enlargement (BPH, *green ring*), bladder outflow obstruction (BOO, *blue ring*), and lower urinary tract symptoms (LUTS, *red ring*).

Symptoms: Frequency of urination, often first noted as a single episode of nocturia, is a common early symptom. Voiding LUTS include hesitancy in initiating urination, a reduced or intermittent urine stream, and post-void dribbling. As the prostate causes bladder outflow obstruction and the pressure required to void increases, hypertrophy of the bladder detrusor muscle can contribute to secondary symptoms including frequency, urgency, and urge incontinence. Hald's rings (Fig. 47.10) provide a diagrammatic conceptual framework to show how symptoms, bladder outflow obstruction, and microscopic tissue changes within the prostate interplay. Acute retention of urine or retention with overflow incontinence may occur. Occasionally, haematuria results from rupture of prostatic veins or because of bacteriuria or bladder stone disease, which occur in conjunction with prostate enlargement. The severity of symptoms is assessed by the IPSS, based on seven questions about urinary symptoms and one question concerning quality of life (Box 47.25). The IPSS is a common patient-reported outcome measure used in the assessment and treatment of male LUTS, reducing clinician bias.

Examination: Abdominal examination for bladder enlargement is performed to exclude urinary retention. DRE is performed as an indicator of prostate size. A benign prostate characteristically feels smooth and firm.

Investigations should include:

- Urine dipstick, microscopy, and culture
- Serum prostate-specific antigen (PSA) concentration
- Kidney function
- Ultrasonography of the kidney, ureters, and bladder with prostate size estimation
- Measurement of urinary flow rate and post-void residual volume.

Treatment: In acute retention or chronic retention with overflow, the priorities are to relieve pain and/or drain the bladder to prevent damage to the kidneys. Both of these aims can be achieved by inserting a urinary catheter. If urethral catheterization is impossible, a suprapubic catheter can be placed.

Box 47.25 The International Prostate Symptom Score

Symptoms

Each question is answered on a Likert score from 0 (not at all) to 5 (almost always over the preceding month).

- Incomplete emptying: How often do you experience incomplete emptying?
- Frequency: How often do you have to urinate again within 2 hours?
- Intermittency: How often do you stop and start during urination?
- Urgency: How often have you found it difficult to postpone urination?
- Weak stream: How often have you had a weak urinary stream?
- Straining: How often have you had to push or strain to begin urination?
- Nocturia: How many times do you typically get up to urinate during the night?

Quality of Life

If you were to spend the rest of your life with your urinary condition just the way it is now, how would you feel about it? 0 (delighted) to 6 (terrible)

Score: 0 = asymptomatic, 1–7 = mildly symptomatic, 8–19 = moderately symptomatic, 20–35 = severely symptomatic.

Box 47.26 Management of benign prostatic enlargement

Mild symptoms

Lifestyle change: attention to amount and types of fluid ingested

Moderate symptoms

- Alpha-blocker (e.g., tamsulosin): relax prostate and bladder smooth muscle tone
- 5-α-reductase inhibitor (e.g., finasteride): block conversion of testosterone to dihydrotestosterone, which drives prostatic growth
- Combination therapy: e.g., alpha-blocker with an anticholinergic for concomitant overactive bladder

Surgical options

- Transurethral resection of the prostate
- Laser therapy: enucleation or vaporization
- Open prostatectomy: uncommon in high- and middle-income settings because of the availability of laser therapy

Management of benign prostatic enlargement is outlined in **Box 47.26**. The balance between benefits and risks of treatment is largely determined by the degree to which the patient is affected by their symptoms. This can be captured in the IPSS, including an evaluation of how the symptoms impact on quality of life. Typically, medical management includes alpha-blockers (e.g., tamsulosin), 5-alpha reductase inhibitors (e.g., finasteride), anti-cholinergics (e.g., solifenacin), or beta-3 sympathomimetics (e.g., mirabegron). Medical management is considered first-line treatment when symptoms are mild or moderate and when the risks of surgery are high or unacceptable to the patient.

Surgery is reserved for those who develop high-pressure chronic retention of urine requiring catheterization (which cannot be removed without surgical relief of obstruction) and for significant LUTS refractory to lifestyle measures and medication. The type of surgical intervention is based on symptoms, potential side-effects, size of the prostate, safety of general anaesthesia, bleeding risk, and use of concomitant anticoagulation (and the safety of withholding this). Although numerous minimally invasive surgical treatments are available, transurethral resection of prostate is most often performed. It requires a catheter and an overnight stay for most patients. Ejaculatory dysfunction can occur. It is more difficult to perform when the prostate volume is more than 80 mL, when preferential treatment is with enucleation (preferably endoscopically with a laser, e.g., holmium laser enucleation of the prostate), or using open surgical techniques.

Box 47.27 Age-specific prostate-specific antigen (PSA) concentrations requiring investigation for prostate cancer in symptomatic males

Age (years)	PSA threshold (μg/L)
Younger than 40	Clinical judgement
40–49	>2.5
50–59	>3.5
60–69	>4.5
70–79	>6.5
Older than 79	Clinical judgement

NICE [2015] Suspected cancer: recognition and referral. Available from www.nice.org.uk/guidance/ng12 All rights reserved. Subject to Notice of rights. NICE guidance is prepared for the National Health Service in England. All NICE guidance is subject to regular review and may be updated or withdrawn. NICE accepts no responsibility for the use of its content in this product/publication.

Prostate cancer

Prostate cancer accounts for 7% of male cancer and is the sixth most common cancer in the world. In the UK, there are 47 000 new cases per year, representing 1 in every 8 males, with 12 000 deaths per year. The risk of prostate cancer increases with age: 58% of those diagnosed with prostate cancer were 70 years or older in 2022. By the age of 80 years, 80% of males have malignant foci within the gland. Risk is increased in those with a family history of prostate (two to three times the risk for a first-degree relative) or breast cancer and in males of Black ethnicity, for whom the risk is one in four. Histologically, the cancer is almost invariably an adenocarcinoma.

Pathogenesis: Genome-wide association studies have identified more than 100 single-nucleotide polymorphisms associated with prostate cancer. The homeobox gene *HOXB13* predisposes to prostate cancer, and the *BRCA2* gene increases risk by five to seven times.

Clinical features and diagnosis: Most patients with early prostate cancer do not have any symptoms. LUTS may be due to concomitant benign enlargement. The diagnosis is suggested by an increase in serum PSA concentration or an abnormal prostate on examination. Late symptoms of metastatic disease include bone pain, malignant spinal cord compression, abdominal pain caused by hypercalcaemia, weight loss, and anaemia. In 2022, 17% of males diagnosed with prostate cancer had metastatic disease, many with radiological rather than symptomatic metastases at presentation.

PSA is a seminal fluid protein produced exclusively in the prostate and by cancerous prostate cells within the gland and when they metastasize. However, a high PSA concentration is not specific for prostate cancer. It is also associated with benign enlargement of the prostate, thus making it an imperfect screening test. PSA concentration is not currently recommended in the UK as a screening test for prostate cancer in asymptomatic patients. In males with possible symptoms of prostate cancer, PSA thresholds for referral and investigation are available (**Box 47.27**).

Multiparametric MRI (mpMRI) is a contemporary tool use to diagnose prostate cancer. Before the advent of mpMRI, random (typically transrectal) sampling of the prostate led to missed diagnosis and overdiagnosis of 'clinically insignificant' prostate cancer (i.e., indolent disease that did not require intervention). mpMRI can avoid the need for prostate biopsy while improving cancer detection rates. When biopsy is undertaken, transperineal biopsy is now more common than transrectal biopsy. This reduces infection because

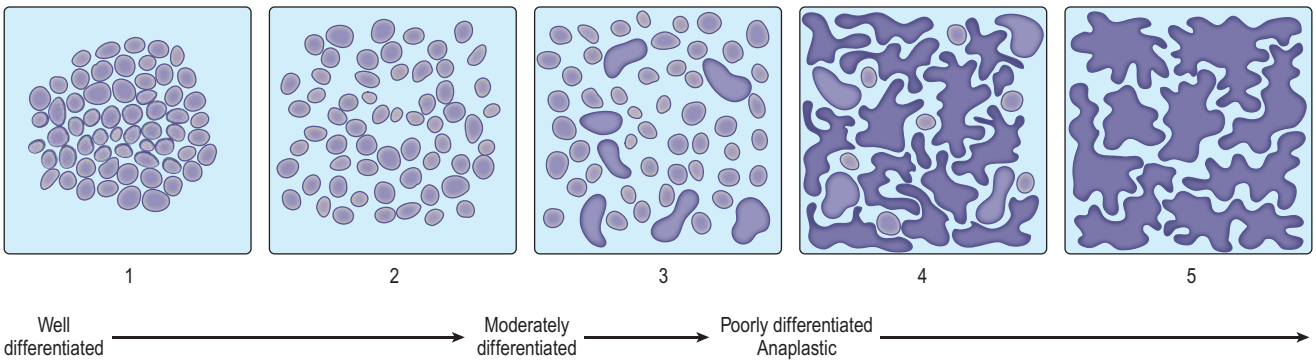


Fig. 47.11 Gleason's pattern scale. The Gleason is a grading scale that runs from 1 to 5, the number indicating the extent of differentiation of the adenocarcinoma cells of the prostate. Because of the multifocal and heterogeneous nature of prostate cancer, the Gleason score is calculated by adding together the two most prevalent grades. A Gleason score of 2 to 6 is considered well differentiated, Gleason 7 is moderately differentiated, and Gleason 8 to 10 is poorly differentiated. *(1)* Small, uniform glands. *(2)* More space (stroma) between glands. *(3)* Distinct infiltration of cells from glands at the margins. *(4)* Irregular masses of neoplastic cells with few glands. *(5)* Lack of pr occasional glands and sheets of cells. (*Redrawn after Kirby RS, Brawer MK, Denis LJ. eds.* Prostate Cancer: Fast Facts*, 3rd edn. Health Press; 2001.*)

Box 47.28 European Association of Urology risk groups for biochemical recurrence of localized and locally advanced prostate cancer after radical treatment

Risk	Low	Intermediate	High	
PSA (ng/mL)	<10	10–20	>20	Any
Biopsy grade	GS <7 (ISUP grade 1)	GS 7 (ISUP grade 2 or 3)	GS >7 (ISUP grade 4 or 5)	Any GS (any ISUP)
Clinical staging based on DRE	T1-2a: limited to less than half of one side of the gland	T2b: within half of the gland	T2c: both sides of the gland	T3-4: extraprostatic (or nodal metastases on CT scan)
Terminology	Localized			Locally advanced

CT, Computed tomography; *DRE,* digital rectal examination; *GS,* Gleason score; *ISUP,* International Society of Urological Pathologists; *PSA,* prostate-specific antigen.

the biopsy needle passes through the colonic mucosa, avoiding inoculation with rectal flora. Prostate cancer commonly metastasizes to bone. Bone metastases appear as osteosclerotic lesions on plain radiography and are also detected by isotopic bone scans.

Prognosis: Histological appearances are graded according to the combined Gleason score (**Fig. 47.11**). The International Society of Urological Pathologists grade group score is used to classify aggressivity of the histological pattern seen. This is combined with serum PSA concentration and accurate T staging of local disease with imaging or final prostatectomy specimen (if radical prostatectomy has been performed) to identify prognostic groups (**Box 47.28**). The long-term prognosis for patients with organ-confined prostate cancer is extremely good: 97% of males survive to 15 years after diagnosis. Treatment decisions therefore balance performance status and age with the predicted behaviour of the cancer (with developments in gene profiling for improved prognostication awaited).

Treatment: In the UK, treatment options are determined by multidisciplinary clinical teams and informed patients (see Chapter 41). A key priority is the avoidance of unnecessary overtreatment (the rate of which has fallen to 7% in the UK). Discussions between patients and clinicians are vital to enable appropriate treatment. Active surveillance entails regular PSA tests and usually yearly mpMRI, with repeat prostate biopsies performed if either test suggests a change that may warrant treatment.

Patients with localized disease that requires treatment are managed with curative intent. In the UK, this most commonly robot-assisted radical prostatectomy, external-beam radiotherapy, or brachytherapy with radioactive iodine implants. All achieve equivalent survival but differ in the spectrum of side-effects, including incontinence and sexual dysfunction. Radiotherapy tends to be used in patients who wish to avoid surgery. Traditionally, therapies with curative intent have targeted the whole gland. However, there is a developing body of evidence for focal therapies, comparable to a 'lumpectomy' for breast cancer. High-intensity focused ultrasound and cryotherapy can be used to ablate part of the prostate harboring a cancer. Treatment of locally advanced and metastatic disease uses endocrine therapy to achieve androgen blockade, alongside systemic chemotherapy (see Chapter 41).

Further reading

Bergengren O, Pekala K, Matsoukas et al. 2022 Update on prostate cancer epidemiology and Risk factors—A systematic review. *Eur Urol* 2023; 84(2): 191–206.

Hamdy F, Donovan J, Lane J et al on behalf of the PROTECT investigators. Fifteen-year outcomes after monitoring, surgery, or radiotherapy for prostate cancer. *NEJM* 2023; 388 (17): 1547–1558.

Kasivisvanathan V, Rannikko A, Borghi M et al on behalf of the PRECISION investigators. MRI-targeted or standard biopsy for prostate-cancer diagnosis. *NEJM* 2018; 378:1767–1777.

National Prostate Cancer Audit Annual Report 2022; https://www.npca.org.uk/content/uploads/2023/01/NPCA-Annual-Report-2022_12.01.23.pdf.

Parker C, James N, Brawley C et al on behalf of the STAMPEDE investigators. Radiotherapy to the primary tumour for newly diagnosed, metastatic prostate cancer (STAMPEDE): A randomized controlled phase 3 trial. *Lancet* 2018; 392:2353–2366.

Prostate cancer: diagnosis and management (Nice Guideline 131) published 2019 updated December 2021. https://www.nice.org.uk/guidance/ng131.

Diseases of the penis

Phimosis

Phimosis is a condition in which the contracted foreskin (prepuce) has a tight, narrow orifice and cannot be retracted over the glans of the penis (Fig. 47.12). A physiological phimosis is present at birth, but penile growth, erection, movement, and accumulation of epithelial debris (smegma) under the foreskin lead to separation, enabling the foreskin to be retracted. By the age of 17 years, only 1% of males have phimosis. Although physiological phimosis is usually asymptomatic, patients may present with ballooning of the foreskin with urination, and there may be trauma to the foreskin during coitus.

Phimosis is typically pathological in adults. Pathological phimosis secondary to infection or inflammatory conditions may cause pain, dysuria, bleeding, or pus discharge from the orifice. Pathological phimosis can be caused by recurrent balanitis (inflammation of the glans penis). In genital lichen sclerosus, hypopigmentation or discoloration of the foreskin is typically associated with pruritic plaques on the foreskin and glans. Severe genital lichen sclerosus is sometimes termed 'balanitis xerotica obliterans' because of the dry glans appearance with destructive changes to architecture. Non-surgical managements include routine attempts at gentle retraction, avoidance of irritant soaps, and topical low-potency corticosteroids. Any infection contributing to phimosis should be treated. If these measures are ineffective, the mainstay of treatment for adults is circumcision. In children, phimosis is usually physiological; the foreskin will invariably become retractile and loosen with age. Indications for circumcision in children include lichen sclerosus, recurrent balanitis, and recurrent UTIs in association with urinary tract abnormalities (e.g., vesico-ureteric reflux, posterior urethral valves, neuropathic bladder dysfunction) because circumcision reduces the frequency of UTI.

Paraphimosis

Paraphimosis is an inability to return retracted foreskin to its resting position (see Fig. 47.12). This can progress to become a urological emergency. Prolonged retraction produces a tight ring of tissue proximal to the corona of the glans. Venous congestion, oedema, and swelling of the glans can then progress to arterial occlusion and necrosis, analogous to compartment syndrome. Treatment of

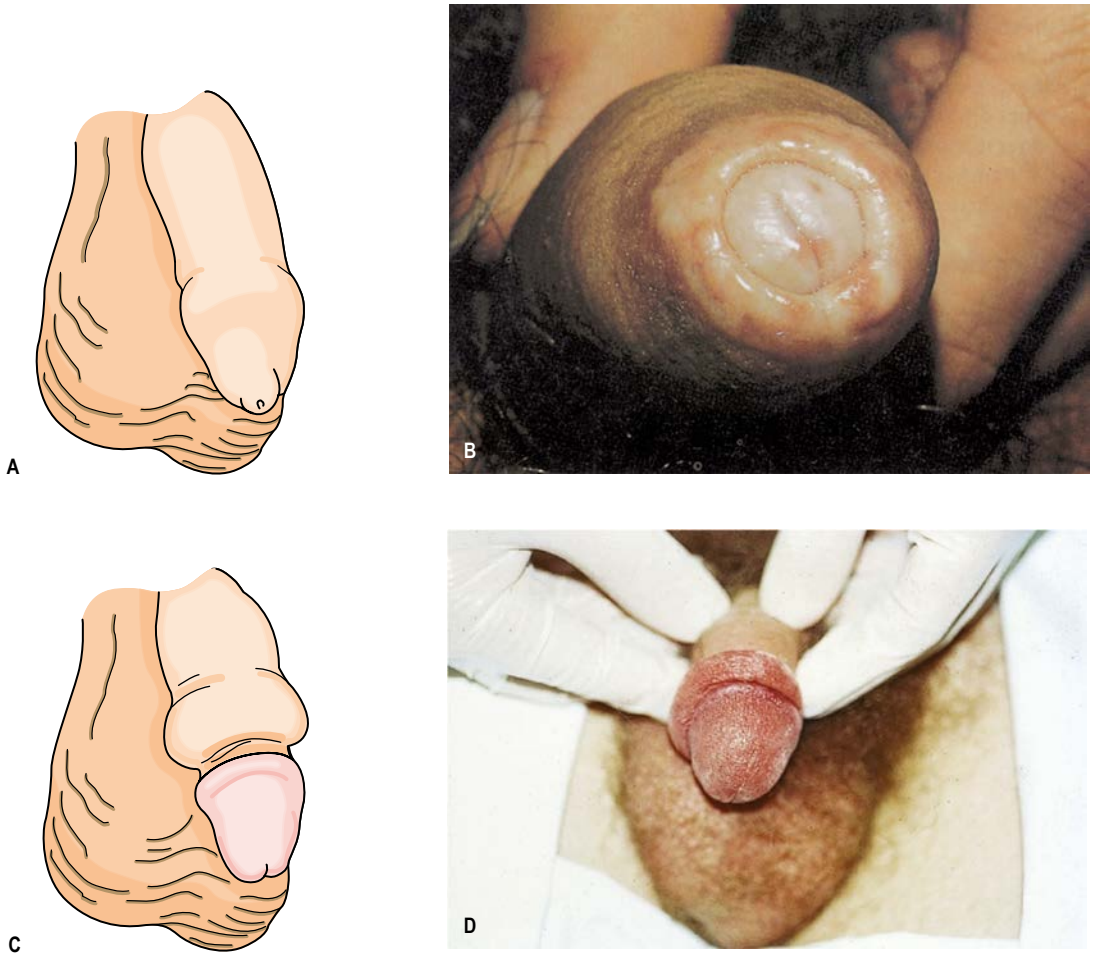


Fig. 47.12 Phimosis and paraphimosis. **(A)** Phimosis: the foreskin has a narrow opening that is not large enough to permit retraction over the glans. **(B)** Phimosis: lesions on the prepuce secondary to infection cause swelling, and retraction of the foreskin may be impossible. Circumcision is usually required. **(C)** Paraphimosis: the foreskin is retracted over the glans but cannot be reduced to its normal position. Here it has formed a constricting band around the penis. **(D)** Oedematous foreskin is the result of paraphimosis. (*From Roberts JR, Hedges JR.* Clinical Procedures in Emergency Medicine*, 5th edn. Saunders; 2009.*)

paraphimosis is with persistent manual compression of the oedematous glans with subsequent attempts to return the retracted foreskin to its normal position. If this manoeuvre fails, a dorsal slit or circumcision is definitive management.

Priapism

Priapism is a prolonged penile erection (>4 hours) maintained without sexual stimulation and persistent despite ejaculation or orgasm. Priapism is a urological emergency requiring an accurate diagnosis and urgent intervention. The two subtypes of priapism are:

- Ischaemic priapism ('low-flow'; more common). This is painful failure of blood supply to the corpora cavernosa caused by sickle cell disease, occult malignancy (priapism with no explanatory features should be investigated for cancer of unknown origin), treatment for erectile dysfunction, or recreational drugs.
- Non-ischaemic priapism ('high-flow'; rare) is caused by a fistula between an artery and the corpora cavernosa and is typically painless. It can occur after perineal or penile injury (e.g., a fall astride a bicycle).

It is important to distinguish ischaemic priapism because prompt intervention to achieve penile detumescence is required to prevent irreversible tissue damage and lifelong erectile dysfunction.

Penile cancer

More than 95% of penile malignancies are squamous cell carcinomas arising from the epithelium of the inner prepuce or the glans. Penile cancer is rare in high-income countries, with an incidence of around 1 per 100 000 in Europe and the USA. In contrast, in South America, South-east Asia, and parts of Africa, the incidence is higher with penile cancer accounting for 1% to 2% of male malignancies. Penile cancer increases with prevalence of HPV most commonly HPV subtypes 6, 16, and 18. Other risk factors include smoking, genital lichen sclerosis, multiple sexual partners, HIV infection, and exposure to ultraviolet light for treatment of psoriasis (rare). Neonatal circumcision reduces the frequency of penile cancer.

The aims of treatment are tumour removal with as much organ and functional preservation as possible to minimize the impact on quality of life. Local recurrence has little influence on long-term survival, so organ preservation strategies are justified. When penile cancer is suspected, examination for groin lymphadenopathy is important. Penile cancer metastasizes in a stepwise manner through the lymphatic system from inguinal to pelvic to distant nodes. Fewer than 5% of patients present with metastases, which have a poor prognosis. The most important prognostic indicator of survival in penile cancer is the presence and extent of nodal metastases with a 5-year survival of approximately 95% for no positive node metastases, 80% for one positive node, 65% for multiple or bilateral lymph node, and 35% for immobile pelvic nodes or when cancer breaches the capsule of the lymph node.

Because penile cancer is rare, management in the UK is centralized to a small number of higher volume referral centres with experience and expertise. When cancer is limited to the prepuce, circumcision can be curative. When the skin of the glans penis is involved, circumcision is necessary alongside treatment options to the lesion, including surgical removal (e.g., excision biopsy, glans resurfacing, glansectomy), liquid treatment (e.g., imiquimod), or superficial laser treatment (e.g., carbon dioxide laser). If the cancer has invaded the deeper structures of the penis, surgical treatment is more extensive. Radiotherapy is rarely used in the UK. Chemotherapy is reserved for cancers with proven metastatic spread.

Further reading

BAUS Section of Andrology Genitourethral Surgery, Muneer A, Brown G et al. BAUS consensus document for the management of male genital emergencies: priapism. *BJUI* 2018; 121(6):835–839.
EAU Guidelines Penile Cancer. https://uroweb.org/guidelines/penile-cancer.

Testicular and spermatic cord disease

Testicular torsion

Testicular torsion is the lateral to medial inward twisting of the spermatic cord ('torsion twists toward'). This results in strangulation of the blood supply to the testis and the epididymis and is a time-sensitive urological emergency. Testicular torsion occurs most often in the neonatal period and around puberty with peak incidence at 13 to 15 years of age, though it can infrequently occur outside of these ages. Typical presentation a short history (<12 hour) of severe, sudden-onset hemiscrotal pain, which can wake the person from sleep. Pain can radiate to the groin, loin, or epigastrium; cause nausea and vomiting; and lead to an altered, broad-based, 'cowboy' gait. On examination, the classical finding is an absent cremasteric reflex, though this is unreliable in older children and adults. The testicle may be swollen, tense, and exquisitely tender. The testicle may be high riding and in a horizontal lie. Elevation of the testicle does not ameliorate symptoms (negative Prehn's sign). Clinicians should be aware that individual patient circumstances, delays in presentation, and provision of analgesia can obfuscate symptoms and signs. The diagnosis is usually clinical. Doppler ultrasound examination cannot rule out testicular torsion and delays definitive treatment, so it should not be relied on in acute presentations. Management is immediate surgical exploration in patients presenting with less than 24 hours of pain. Delay can result in permanent ischaemic damage and potential loss of the testis. Surgical detorsion (untwisting the spermatic cord) and bilateral orchidopexy (fixation of the testicle to prevent recurrent torsion) are performed in the presence of a viable testicle. If the testicle is non-viable, orchidectomy and contralateral orchidopexy are carried out.

Further reading

Clement K, Light A, Asif A et al. A BURST-BAUS consensus document for best practice in the conduct of scrotal exploration for suspected testicular torsion: The Finding Consensus for Orchidopexy in Torsion (FIX-IT) study. *BJUI* 2022; 130: 662–670.

Epididymitis and orchitis

Epididymitis is an inflammatory condition of the epididymis and can be acute or chronic. Often, in the context of infection, inflammation involves the testis at the same time and is termed 'epididymo-orchitis'. Epididymitis is usually caused by a bacterial infection that ascends from the urethra or the bladder. In sexually active males, the infective organism is usually *N. gonorrhoeae*, *C. trachomatis*, or a coliform bacterium. In older males and children, infective organisms are more commonly urinary tract pathogens such as *E. coli*. Viral mumps orchitis can occur, and the patient's vaccination history should be determined.

Acute presentations can mimic testicular torsion though systemic features (e.g., fever) and local symptoms (e.g., preceding dysuria, urethral discharge, urinary frequency) may be helpful distinguishers. The duration of pain is usually longer than that of testicular torsion and has a gradual onset. On examination, the patient may have a fever and appear systemically unwell. The epididymis will be tender and swollen on palpation. If examination is performed

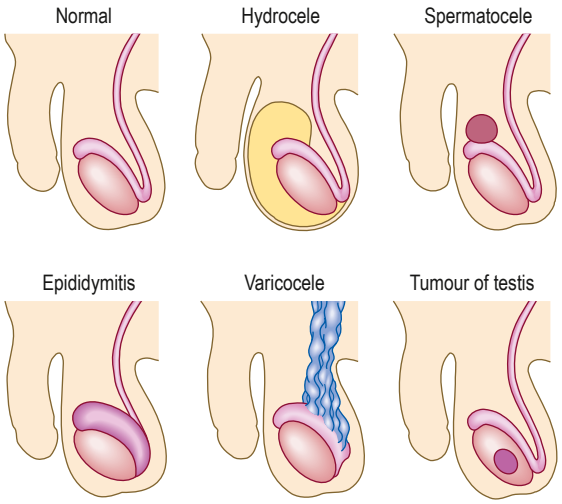


Fig. 47.13 Diseases of the testicle. *(From Innes JA, Dover AR, Fairhurst K, eds.* Macleod's Clinical Examination, *14th edn. Elsevier; 2018, with permission.)*

carefully, tenderness localized to the epididymis but not the testis can be appreciated. Prehn's sign (pain improves when the testis is raised) is classically positive, but this is not reliable for diagnosis and must be interpreted alongside all available clinical information. The hemiscrotum is often swollen and warm to the touch, and erythematous skin changes can be seen. Urinalysis may suggest infection. Examination for parotid swelling is indicated in patients who have not received mumps vaccination or if vaccination history is unclear. Testicular cancer mimics epididymo-orchitis in about 10% of cases. Examination for lymphadenopathy is warranted, which may accompany viral epididymitis.

Patients should be treated for the most likely organism according to local microbiological guidelines, and scrotal elevation is advised to reduce swelling. Advice should be provided about potential complications such as abscess formation, testicular infarction, chronic pain, chronic infection, and infertility. Those who are sexually active should be screened for STIs.

Hydrocele and epididymal cyst

A hydrocele is an abnormal collection of fluid within the tunica vaginalis of the scrotum or along the spermatic cord (Fig. 47.13). It is usually painless, but the testicle is difficult to palpate because of the tense fluid collection. It is possible to transilluminate it because the fluid is clear or straw coloured. Epididymal cysts are benign, fluid-filled cystic swellings that arise from the head of the epididymis. The fluid in the cyst is clear, so differentiation between a large epidydimal cyst and a hydrocele can be challenging for both clinician and ultrasonographer and may only become clear at the time of scrotal exploration to treat the condition. Epididymal cysts can, however, be multiloculated, unlike a hydrocele. The superior margin of a hydrocele and an epididymal cyst is palpable. The clinician being able to 'get above' the lump differentiates from an inguinal hernia. Surgical excision of a hydrocele or epididymal cyst is indicated if it is large and uncomfortable, but small and asymptomatic epidydimal cysts and hydroceles do not require intervention because they do not have an impact on long-term risk or morbidity. If the cyst contains spermatozoa, it is termed 'spermatocele' (see Fig. 47.13).

Varicocele

A varicocele is dilation of the pampiniform plexus of veins surrounding the testis and extending up into the spermatic cord (see Fig. 47.13). Varicoceles are usually small and asymptomatic. They occur more commonly on the left side because of drainage of the gonadal vein into the left renal vein. When symptomatic, patients describe a dragging sensation or dull ache in the scrotum, which classically worsens after long periods of standing, walking, or exercise. Rarely, a new varicocele can be a sign of renal malignancy caused by compression of the gonadal vein and compromised ipsilateral venous drainage. Varicoceles are more common in males with abnormal semen analysis (25%) compared with the overall population (12%). Treatment of a varicocele is therefore recommended in infertile males (unable to conceive with a fertile female partner within 12 months) with abnormal semen analysis parameters. Treatment is not undertaken for subclinical varicoceles with normal semen analysis. Treatments options are embolization or microsurgical ligation of the dilated veins via a small incision at the level of the superficial inguinal ring. Both treatments have high success rates. The selection often depends on local availability. Where available, embolization may be preferred to avoid the need for general anaesthesia.

Cryptorchidism (undescended testes)

In utero, testicular descent from the level of the kidneys into the scrotum is guided by the gubernaculum. This occurs in two phases:

- 7 to 8 weeks' gestation: descent from genital ridge to the internal inguinal ring
- 24 to 28 weeks' gestation: descent through the inguinal canal into the scrotum

Failure of descent is termed 'cryptorchidism' or 'undescended testis'. The incidence in term neonates varies between 1.0% and 4.6% and is up to 45% in preterm neonates. Patients with an intra-abdominal testis have an eightfold increased risk of testicular cancer. There are also increased risks of testicular torsion and inguinal hernia. Fertility may be reduced, although it improves if orchidopexy is performed before 2 years of age, which is routine practice in the UK.

Testicular germ cell tumours

Testicular cancer is the most common solid cancer in young males in high-income populations, representing 1% of male cancers and 5% of urological cancers. The incidence is 3 to 10 per 100 000 per year in high-income countries. Most patients present with a painless unilateral testicular or scrotal mass. Although there are no data to demonstrate regular self-examination for a mass reduces mortality rates, it seems sensible, and it is widely recommended that males examine their testes once a month. Up to 27% of patients have pain or a dragging sensation as their first symptom. A detailed history uncovers the 11% of cases with back and flank pain caused by metastases and the 7% with gynaecomastia caused by tumour secretion of hCG. Presentation mimics epididymo-orchitis in about 10% of cases (see earlier). The investigation and management of testicular cancer is detailed in Chapter 41.

Further reading

Park JS, Kim J, Elghiaty A et al. Recent global trends in testicular cancer incidence and mortality. *Medicine* 2018; 97(37):e12390.
https://uroweb.org/guideline/penile-cancer/. *EAU Guidelines for Penile Cancer*.
https://uroweb.org/guideline/paediatric-urology/#1. *EAU Guidelines for Paediatric Urology*.

Loss of libido and erectile dysfunction

Loss of libido is a loss of sexual desire. Erectile dysfunction is a complex phenomenon defined as the persistent inability to attain and maintain an erection sufficient to permit satisfactory sexual performance. Loss of libido in males and erectile dysfunction may occur together or separately and can precipitate each other. They affect physical and psychosocial health and have a significant impact on quality of life. In some instances, symptoms are attributed to low serum testosterone concentrations, and both loss of libido and erectile dysfunction occur in patients with hypogonadism. Erectile dysfunction can also occur because of inadequate arterial blood flow, and investigation for underlying cardiovascular disease may be indicated. However, in most cases, testosterone concentrations are normal with no identifiable organic cause identified. Nocturnal emissions and satisfactory morning erections make hormonal and vascular aetiologies unlikely. Psychological factors and lifestyle stresses are thought to be important in pathogenesis, and they may coexist with other causes (**Box 47.29**).

A detailed medical and sexual history should be obtained. Physical examination includes the genitourinary system and for any suspected aetiology (e.g., peripheral pulses). Investigations are determined by the history and risk factors.

Modifiable risk factors, including lifestyle (e.g., smoking, obesity), should be addressed and possible contributing medications reviewed. Treatment of underlying systemic diseases, including diabetes mellitus and hypertension, may be beneficial. If no organic disease is found or if there are psychological contributors, psychosexual counselling should be offered.

Treatments for patients with erectile dysfunction balance invasiveness and efficacy. The most common first-line therapy is an oral PDE-5 inhibitor (e.g., sildenafil, tadalafil, vardenafil, avanafil). These cause smooth muscle relaxation and an increase penile arterial blood flow, enabling an erection. The time from ingestion of medication to effectiveness varies but is generally between 15 and 60 minutes. PDE-5 inhibitors are contraindicated in patients requiring nitrate treatment (e.g., GTN [glyceryl trinitrate], isosorbide mononitrate, isosorbide dinitrate), hypotension (blood pressure <90/50 mmHg), hypertension (blood pressure >170/100 mmHg), unstable angina, angina with sexual intercourse, severe heart failure, within 6 months of a myocardial infarction, stroke, or life-threatening arrhythmia.

Vacuum erection devices provide passive engorgement of the corpora cavernosa. A constrictor ring is then placed at the base of the penis to retain blood within the corpora. This option is contraindicated in patients with bleeding disorders and in those on anticoagulation. It can be offered to those with contraindications to medical management.

Alprostadil is a synthetic prostaglandin E1 analogue that prevents cyclic adenosine monophosphate (cAMP) breakdown, leading to dilation of corpora cavernosa blood vessels via a reduction intracellular calcium. Topical alprostadil is applied to the glans and absorbed via the urethral meatus, or a pellet can also be administered into the navicular fossa of the distal urethra. Patients who do not respond to oral or topical methods can be offered intracavernous alprostadil injections after patient or partner training to ensure the correct technique.

Surgical implantation of a penile prosthesis may be considered in patients who do not respond to pharmacotherapy or who prefer a permanent solution.

? Box 47.29 Causes of erectile dysfunction

Psychogenic
Vascular: peripheral vascular disease, cardiovascular disease
Neurogenic: diabetic neuropathy, spinal cord lesions, nerve damage during radical prostatectomy
Hormonal: hypogonadism, hyperprolactinaemia
Drug induced: finasteride, cyproterone acetate

Further reading

EAU Guidelines for Erectile Dysfunction and Lack of Libido. https://uroweb.org/guidelines/sexual-and-reproductive-health/chapter/management-of-erectile-dysfunction.

The ageing male

There is no direct male equivalent to female menopause. However, there is a progressive loss of sexual function, including a reduction in morning erections and frequency of intercourse, with age. Typically, there is a gradual decrease in testicular volume and increase in sex hormone–binding globulin and gonadotrophin concentrations with ageing, but patients may have low testosterone concentrations without elevation of LH or FSH. Low testosterone concentrations increase the risk of osteoporosis with a variable increase in cardiovascular risk in some studies. It remains unclear to what extent general symptoms, including a lack of energy, drive, muscle strength, and general wellbeing, relate to hormonal changes. A recent trial of testosterone replacement showed no benefit for general symptoms, although there was an improvement in sexual function. Loss of libido and erectile dysfunction are common symptoms even when hormone concentrations are normal, and long-term outcome studies of testosterone replacement are awaited. The decision to offer testosterone replacement to an ageing male patient is based on clinical and biochemical assessment, and thorough discussion of potential risks (including prostate disease) as well as possible benefits. If serum testosterone concentration is unequivocally low (<7 pmol/L) with specific symptoms of androgen deficiency (low libido, erectile dysfunction, loss of early morning erections), replacement likely to be recommended. However, few would treat a testosterone concentration over 12 pmol/L with normal LH and FSH. A large proportion of cases have testosterone concentrations 7 to 12 pmol/L, for which the optimum diagnosis and treatment are less clear. There is an increasing move toward measuring fasting testosterone concentrations because food intake may decrease testosterone concentrations, leading to an incorrect diagnosis of androgen deficiency.

Targeted PSA testing from the age of 45 years can be considered for all male patients with risk factors for prostate cancer, including a family history of prostate or breast cancer, African or Caribbean ethnicity, or known genetic predisposition.

Further reading

Moore C, Frangou E, McCartan N et al on behalf of the Re-Imagine Study group. Prevalence of MRI lesions in males responding to a GP-led invitation for a prostate health check: a prospective cohort study. *BMJ Oncology* 2023; 2 e000057.
Vickers A, O'Brien F, Montorsi F et al. Current policies on early detection of prostate cancer create overdiagnosis and inequity with minimal benefit. *BMJ* 2023; 381:e071082.

TRANSGENDER HEALTH

In gender incongruence and gender dysphoria, a person experiences a significant discordance between their internal sense of gender and the secondary sexual characteristics they possess. People who manage their own gender diversity by changing social expression of gender rather than accessing medical interventions do not require assessment. If assessment is required, it should be carried out by a professional with specific training in this area. Assessment includes gender history and assessment of both mental and physical health.

The science of gender

The development of a person's gender identity involves an interplay of many external influences, including psychological, social, and physical factors. When combined, these factors form a person's internal sense of gender and how they express that gender. Theories of gender development can be divided into three broad categories: biological, social, and cognitive.

Biological theories emphasize the reproductive imperative for survival. In these theories, females are psychologically adapted for nurturing and display empathy, caring, and better verbal fluency. Males are adapted for hunting and so have better visuospatial skills and aggression and are more likely to take risks. There are also biological differences in neuroanatomy. The areas of the brain that are sexually dimorphic accumulate testosterone, which is aromatized to oestradiol. Oestradiol modulates neural pruning by either initiating or inhibiting apoptosis in sexually dimorphic areas of the brain. Animal models show that gender dimorphic behaviours, such as sexual behaviour, can be modified by testosterone exposure or withdrawal during critical windows in development. For example, exposing female rodents in the critical window to testosterone results in male sexual behaviour. This biological model of gender incongruence postulates that lack of testosterone during the critical window for brain masculinization leads to feminine brain development. Conversely, exposure to high testosterone concentrations could result in masculinization. However, the critical window that affects behaviour has not been established in humans.

Social learning theories view gender-specific behaviours as the product of the gender-stereotype messaging that children receive during development. Children get positive reinforcement for 'gender-appropriate' and negative reinforcement for 'gender-atypical' behaviours. This is internalized by children as the way they should perform their genders and creates their gender identities.

Cognitive theories extend the social learning construct as the child self-regulates their behaviour because they have an internal sense of gender and use it to guide their behaviours.

No single theory can explain gender development. A strong interplay between biology, socialization, and cognition forms a person's gender identity.

Sex, gender, and terminology

Terms commonly used to discuss sex and gender in medicine are described in **Box 47.30**. For gender-diverse people, the internal schema of gender is different to the biological sex registered at birth. This creates a dissonance between the internal sense of gender and the societal expectation of their gender role. This dissonance can lead to distress and is termed 'gender dysphoria'.

Box 47.30 Sex and gender terms in medicine

Sex	A collection of biological characteristics that determine whether an organism produces male or female gametes. Sex many be defined based on chromosomes (XX, XY), gonads (ovaries, testes), or secondary sexual characteristics.
Intersex	Individuals with sexual development in which anatomy, reproductive organs, hormonal patterns, and/or chromosomal patterns do not fit binary definitions of male and female.
Gender	The social, behavioural, cultural, and psychological characteristics associated with being a male or female. Gender may be used to describe an individual's perception of themselves ('gender identity') or society's expectations ('gender role').
Gender incongruence	A gender identity that is different to an individual's sex registered at birth. Although gender incongruence may be binary (e.g., an individual who identifies as a female but was registered male at birth), some individuals may identify as non-binary, gender fluid, or non-gendered.
Gender dysphoria	A sense of dissonance between internal sense of gender and an individual's secondary sexual characteristics, resulting in distress and impaired psychosocial functioning.
Gender-affirming care	An approach to medical care that aims to help an individual to live according to the gender with which they identify, supporting social gender role transition and/or the use of hormonal therapy or surgery.
Transgender	A person whose gender identity does not correspond with their registered sex at birth.
Cisgender	A person whose internal sense of gender corresponds with their registered sex at birth. In this chapter, the term is used when required to describe a comparator population of the same registered sex not undergoing gender-affirming care to assess benefits and risk.

Epidemiology of gender incongruence

The epidemiology of gender nonconforming people is difficult to establish. Estimates of prevalence vary widely depending on whether cohorts are community based or those seeking medical interventions. In healthcare systems–based studies, the estimated prevalence is 0.02% to 0.1%. In studies of self-reported transgender identity, estimates range from 0.3% to 0.5% of adults and 1.2% to 2.7% of children and adolescents.

Historically, more people registered male at birth attended gender identity services than those registered female, with a ratio of 3.6 to 1. More recent studies suggest that this ratio has now reduced to 1.4 to 1, and population-based studies suggest that there is no sex bias.

Aetiology of gender incongruence

The aetiology of gender incongruence includes neuroanatomical, genetic, hormonal, neurodevelopmental, and environmental theories.

Neuroanatomical

Despite extensive research, no clear neurological marker of gender incongruence has been found. Postmortem histological studies (e.g., third interstitial nucleus of the anterior hypothalamus, cortical

thickness patterns) and functional MRI studies (response to visuospatial tasks and verbal fluency) suggest that neuroanatomy and possibly the pattern of cerebral activation are more congruent with gender identity (experienced or felt gender) than assigned or registered gender (phenotypic gender).

Genetic

Twin studies suggest that the heritability of gender identity is comparable to that of behaviour and personality traits. The degree of heritability varies between 24% and 50% in monozygotic twins, and the incidence in dizygotic twins is half these values. This suggests that the formation of adult gender identity is caused by both genetic and environmental factors.

Hormonal

Disorders of sexual differentiation give us information about the influence of hormones on behaviour. Girls with congenital adrenal hyperplasia, who are exposed to excess androgen levels *in utero*, display more male 'gender-appropriate' play in childhood, have a decreased desire for maternity, and increased occurrence of bisexuality and homosexuality compared with unaffected siblings. In complete androgen insensitivity, a person is genetically XY, but the androgen receptor is not functional. These children display female gender identity and psychological traits. This suggests that hormone exposure influences gender identity.

Neurodevelopmental

Approximately 10% of gender-diverse people are neuroatypical. This has led to conjecture that gender diversity may be a neurodevelopmental condition, but evidence for this is limited.

Environmental

In the UK, referrals of young people to the Gender Identity Service at the Tavistock Clinic in London rose exponentially from 2014, with the biggest rises in adolescents registered female at birth. The reasons for the marked increase in prevalence is unclear with social media, childhood trauma, changes in society's acceptance of gender diversity, and socializing effects in peer groups and educational institutions hypothesized to be contributing factors. The Cass Review of the Tavistock Clinic in 2024 highlighted the complexity of many of the cases referred, in which gender dysphoria existed alongside mental health problems, neurodiversity, and adverse childhood events; children in the social care system or adopted were overrepresented.

Assessment of gender incongruence

Many people with gender diversity do not present to medical services and manage their gender diversity themselves by changing social expression of gender rather than accessing medical interventions. People in this situation do not require assessment.

It is important that any assessment process does not pathologize the patient. The World Professional Association for Transgender Health (WPATH) has published international standards of care for working with gender-diverse people (currently version 8). When assessing a person, start by establishing their gender identity, preferred name, and pronouns to maintain a respectful approach.

Assessment should identify coexisting mental health and physical or psychosocial concerns and distinguish these from gender dysphoria, gender incongruence, and gender diversity. The clinician should ascertain the patient's expectations of care: what interventions an individual is seeking to achieve an appearance consistent with their internal sense of gender. It is important to discuss the implications of physical changes on the individual's psychological, physical, and social functioning. The important elements of assessment are outlined in **Fig. 47.14**.

History

Gender history

Gender history should document when gender dissonance started and how their sense of gender identity developed. At what age did they start to express their preferred gender, for example, using clothing, makeup, or hairstyles? When did they begin to feel uncomfortable in their body and start thinking about aligning their body with their preferred gender? At what age did they tell others about their gender identity, and what impact did it have on their social relationships? Are they able to occupy the social gender role consistent with their gender identity ('socially transition'), and for how long have they been able to do this? What is the psychological impact of this? Are partner, family, and friends supportive of the person's gender identity? What is their sexual orientation, and how will gender transition affect this?

Mental health history

Gender-diverse people may have had trauma during childhood, including bullying, stigmatization, transphobia, and abuse. Depression and anxiety are extremely common in people with gender incongruence, leading to an increased incidence of deliberate self-harm and suicidal behaviour. Physical therapies can affect mood. It is important to assess mental health and co-morbid psychological conditions, especially self-harm and suicidal ideation. Does the person have any forensic history that is relevant to treatment for gender diversity, such as violence, that could be increased by testosterone treatment?

Medical history

A history of birth and childhood development, including age of puberty, aids identification of ambiguous genitalia or abnormal genital development. People who are registered male sex at birth should be asked about hypogonadism, genital abnormalities (e.g., micropenis, cryptorchidism, hypospadias), gynaecomastia at puberty, reduced facial and body hair, decreased morning erections, and infertility. People registered female sex at birth should be asked about their onset of menarche and usual menstrual pattern and for features of hyperandrogenism such as increased facial hair and body hair, increased clitoral size, oligo- or amenorrhoea, and infertility. It is important to find out if the individual has been taking hormonal treatment before assessment and what effect it had.

Hormone therapies have metabolic consequences and confer thromboembolic and hormone-dependent cancer risks. Cardiovascular status and risk are important because testosterone can increase cholesterol concentrations, and oestrogen can increase insulin resistance. Clinical history should include a personal or family history of myocardial infarction, stroke, or peripheral vascular disease and clinical features of polycystic ovarian syndrome. The use of tobacco and alcohol should be noted. People should be advised to stop smoking before starting hormone therapy and drink alcohol within safe limits. Risk factors for thromboembolism are important because hormone therapy, particularly oestrogen, can increase risk. Hormone-dependent cancers, including breast cancer, endometrial cancer, or prostate cancer, in the individual and their family should be documented.

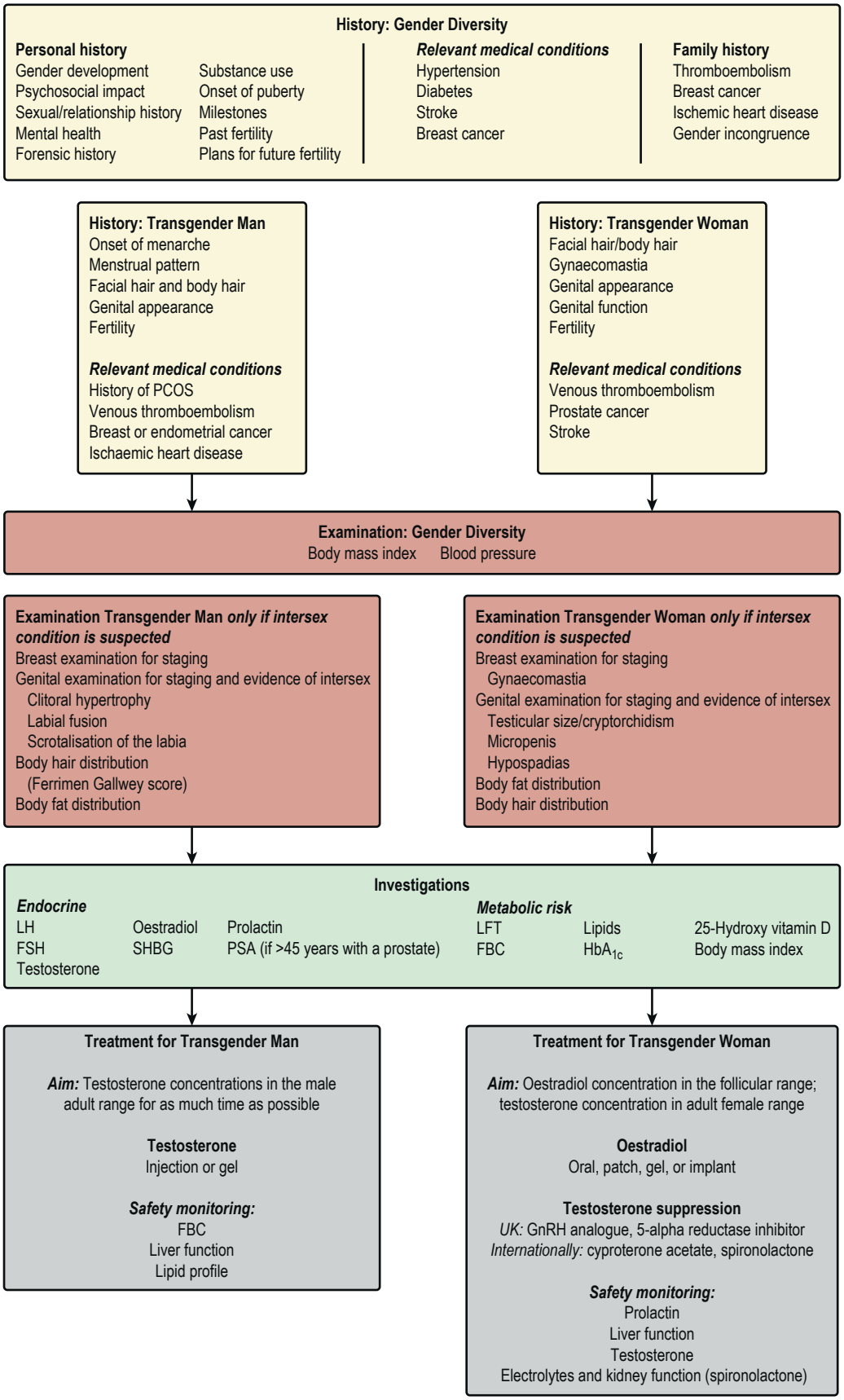


Fig. 47.14 A summary of the history, examination, investigation, and treatment of the medical management of a gender-diverse person. *FBC,* Full blood count; *FSH,* follicle-stimulating hormone; *GnRH,* gonadotrophin-releasing hormone; *HbA_{1c},* hemoglobin A_{1c}; *LH,* luteinizing hormone; *PCOS,* polycystic ovarian syndrome; *PSA,* prostate-specific antigen.

Examination

There are no specific physical features associated with gender incongruence. Clinical assessment has two purposes. First, it can identify individuals who may have a disorder of sexual differentiation who can present with gender confusion. Second, it allows a cardiovascular and metabolic risk assessment of the individual as hormone therapies can increase risk.

Blood pressure and body mass index should be measured as part of the assessment of cardiovascular risk.

Assessment of secondary sexual characteristics, including examination of external genitalia, is only required if a disorder of sexual differentiation is suspected on history or blood tests. This examination is not routine because people with gender diversity can find genital examination distressing. Any intimate examination must be done sensitively and efficiently and in a way that maintains dignity and respects the patient's sense of gender identity. It is important to gain consent for all examinations and to reassure individuals that they can decline examinations. Anatomical terms for genitalia are very gendered. Using gendered terms may cause distress to a gender-diverse person. Ask the person being examined their preferred terms for genitalia and use these words during any examination. Chaperones of an appropriate gender may make the patient more comfortable. It is rarely necessary to perform an internal examination because observation of the external genitalia is usually sufficient to determine whether there is a disorder of sexual differentiation.

The relevant clinical features on examination are outlined in Fig. 47.14.

Investigation

There are no diagnostic tests for gender incongruence or dysphoria. Investigations aim to identify if there is any significant disorder of gonadal function and to assess the metabolic status of the patient. The panel of blood tests is outlined in Fig. 47.14.

Diagnosis of gender incongruence

The clinical features of gender dysphoria and gender incongruence are outlined in Box 47.31. In both diagnoses, the clinical features are broadly similar. To make a diagnosis of gender incongruence or gender dysphoria, a person should experience a significant discordance between their internal sense of gender and the secondary sexual characteristics they possess. In both cases, the feeling of discordance must be present for a minimum of 6 months. Gender incongruence is the most widely used diagnosis in gender medicine practice and does not require dysphoria for diagnosis.

The differential diagnosis of gender incongruence includes bipolar disorder, psychosis, and dysmorphophobia. However, it is very rare for these conditions to be confused with gender incongruence because differences are usually clear, such as rapid onset in the context of a mental health crisis or unhappiness with the body's appearance without significant gender incongruence.

Management of gender incongruence

There is considerable international variation in the recognition, acceptance, and management of gender incongruence within societies. In many liberal societies, health systems have adopted an approach termed ***gender-affirming care***, which aims to support individuals to live in the gender with which they identify.

In adults, there is little high-quality evidence to guide the treatment of people experiencing gender dysphoria (e.g., controlled trials of medical and psychological interventions). Management of patients with gender incongruence is best done in the context of a multidisciplinary team that addresses psychological, social, and physical health. Assessment includes discussion around social functioning and the social gender role that the individual wishes to have.

Recent controversy has focussed on children and young people in the UK. The 2024 Cass Review into gender identity services for children and young people in England and Wales highlighted the paucity of evidence to support use of GnRH analogues and recommended against a medicalized pathway for most young people.

This chapter outlines the principles, benefits, and risks of gender-affirming care in adults. The additional potential vulnerability of young adults, particularly at the time of transfer to adult medical

Box 47.31 Diagnostic criteria for gender dysphoria and gender incongruence

	Gender dysphoria	Gender incongruence
Defined by	*Diagnostic and Statistical Manual,* fifth edition, American Psychiatric Association	International Classification of Disease-11, World Health Organization
Defined where	List of psychiatric conditions	List of sexual health conditions
Diagnostic criteria (at least two should be present for 6 months)		
Incongruence between experienced or expressed gender and primary and/or secondary sex characteristics	☑	☑[a]
A strong desire to be rid of one's primary and/or secondary sex characteristics	☑	☑
A strong desire for the primary and/or secondary sex characteristics of the other gender	☑	☑[b]
A strong desire to be of the other gender	☑	☑[b]
A strong desire to be treated as the other gender	☑	☑[b]
A strong conviction that one has the typical feelings and reactions of the other gender	☑	☑[b]
This condition causes clinically significant distress or impairment in social, occupational, or other important areas of functioning	☑	†

[a]In adolescents, a desire to prevent the development of the anticipated secondary sex characteristics.
[b]Alternative gender different from one's registered gender.

services, should be acknowledged, along with an appreciation that the developmental trajectory of young people may continue into their mid-20s.

Social gender role change

It is important to explore alternatives to physical interventions in addressing gender dissonance. A gender-diverse person may decide to change their social gender role by altering the social cues that people use to read their gender. These changes may include clothing, makeup, and hairstyle. It is also possible for a person to change their name and prefix using a legal instrument such as a deed poll to make their legal name congruent with their gender identity. In many countries, it is also possible for an individual to change their birth certificate to reflect their gender identity. This is done in the UK by provision of a gender recognition certificate.

Physical interventions

If the person is sure that physical interventions are necessary for them to have a body that is congruent with their gender identity, these can be considered. A gender specialist assesses the need for interventions including psychological counselling, facial hair removal, hormone therapy, speech and language therapy, and surgical interventions, with referral to appropriate services.

Hormone therapy

There are three aims of therapy:

- Reduce natal hormone production
- Induce secondary sexual characteristics of the desired gender
- Prevent hypogonadism after a person has undergone gonadectomy.

Hormonal therapy can be recommended by a gender specialist after a period of assessment. A gender specialist will assess capacity to consent to treatment, diagnostic criteria for gender incongruence or dysphoria, longevity of gender identity, expectations of hormone therapy, and whether hormone therapy is in the person's best interest. Before hormone therapy is commenced, psychological and physical health should be optimized.

Fertility, contraception, and sexual health

The impact of therapy on fertility should be discussed before hormone treatment. If fertility is desired in the future, referral for fertility preservation (sperm storage or ova preservation) should be offered.

Hormone therapies are not licensed as contraceptives, so gender-incongruent people on hormone therapy need additional effective contraceptive methods if necessary. Barrier contraception can be used by transgender men and women. Progesterone-based contraceptives (progesterone-only pill, implant, depot, and IUD) are suitable for transgender men and more effective at preventing pregnancy than barrier methods. It is not advisable to use oestrogen-containing methods because they may produce feminization.

It is important to discuss available vaccination programmes including HPV, hepatitis B, and hepatitis A. Transgender men should be encouraged to engage with the national cervical screening programme. It is important to counsel about HIV prevention, including pre- and post-exposure prophylaxis.

Hormone therapy for transgender women

Oestrogen therapy in transgender women initiates the sequences seen in female puberty. Hormone therapy should be individualized with doses that achieve physiological oestrogen and testosterone concentrations. The aim is to achieve an oestradiol concentration in the follicular range, increasing the dose over 6 to 12 months until the correct concentration is achieved. Doses two to five times more than standard hormone replacement for menopausal females are usually required. If oral therapy does not achieve sufficient oestradiol concentrations, gel or patch treatment is used. Ethinylestradiol and conjugated equine oestrogen are avoided because of increased cardiovascular and thromboembolic risk compared with oestradiol. To minimize cardiovascular risk, topical rather than oral oestrogen preparations are used for those older than 45 years of age.

Oestrogen monotherapy reduces testosterone to the female range in approximately 20% of transgender women. Androgen suppression is therefore usually required to achieve a testosterone concentration in the female range. There is international variation in the medications used for androgen suppression. GnRH analogues are used in the UK, cyproterone acetate is used in Europe, and spironolactone is preferred in USA. International variation is largely based on cost. GnRH analogues overstimulate the pituitary gonadotroph cells, causing down-regulation of LH and FSH secretion, which stops testicular testosterone production and spermatogenesis. The usual side-effects of menopausal hot flushes, tiredness, accelerated heart disease, and reduced bone mineralization do not occur because of simultaneous oestrogen replacement. GnRH analogues can prolong the QT interval.

Cyproterone acetate is a progesterone derivative that acts as an androgen receptor antagonist, a progestin receptor agonist (inhibiting LH and FSH), and a partial glucocorticoid agonist. Side-effects include abnormal liver function, insulin resistance, increased BMI, reduced high-density lipoprotein cholesterol, and depression. Long-term use is cautioned because it can induce the development of both meningioma and prolactinoma.

Spironolactone is a mineralocorticoid receptor antagonist, a partial androgen receptor antagonist and an oestrogen receptor agonist. Side-effects include hyperkalaemia through potassium-sparing diuretic action, gastrointestinal bleeding, depression, and abnormal liver function. Spironolactone is less effective than cyproterone acetate or GnRH analogues in reducing testosterone concentrations.

Finasteride prevents the activation of testosterone to dihydrotestosterone by blocking the enzyme 5-alpha reductase. This reduces the action of testosterone on the body and has been used to stabilize scalp hair loss and reduce body hair. Both depression and liver function disturbance have been described as side-effects; however, depression is not as prominent compared with cyproterone acetate use in transgender women.

The WPATH standards of care state that there is insufficient evidence to recommend the use of progesterone in transfeminine HRT. Although there are anecdotal data of improved wellbeing, sleep, and breast development, progesterone in combination with oestrogen increases thromboembolic and cardiovascular risk.

Monitoring in the initial phase is every 3 to 6 months, converting to every 6 months when stable, and then to annual review. Standard therapies and the monitoring required are outlined in **Boxes 47.32 and 47.33**. Other than the change from oral to topical oestrogen treatment at age 45 years, no other adjustments are made for age. After therapy has been started, it should be continued unless the person develops a significant side-effect or a condition that is an absolute contraindication to hormone therapy (e.g., oestrogen-dependent breast cancer).

Box 47.32 Oestrogen preparations used for feminization in UK practice and monitoring required

Route	Preparation	Dose	Frequency	Oestradiol concentration	Monitoring			
					Baseline tests	Therapeutic monitoring	Timing of therapeutic monitoring	Long-term monitoring
Oral	Oestradiol valerate or oestradiol hemihydrate	1–8 mg	Once a day	Midfollicular level: 400–600 pmol/L	LH FSH Testosterone Oestradiol SHBG Prolactin PSA (older than 45 years) LFT FBC Lipid profile HbA_{1c} Vitamin D BMI BP	Oestradiol Testosterone SHBG Prolactin LFT	Blood 4–6 hours after tablets	8 weeks after any change in dose On stable dose: 6 monthly for 2 years; then annually
Patch	Oestradiol	50–200 mcg twice weekly	Change patch(es) twice a week	Midfollicular level: 400–600 pmol/L			Blood 48 hours after patch application	
Topical gel	Oestradiol	0.5–1 mg (sachets) or 0.75 mg (pump actuation; maximum, 5 mg/day)	Once per day (not to breasts)	Midfollicular level: 400–600 pmol/L			Blood 4–6 hours after application and no gel on arms	
Implant	Oestradiol	50–100 mg	6–24 monthly	Trough value: 400–500 pmol/L			5 months after implant; then repeat monthly	

BMI, Body mass index; *BP,* blood pressure; *FBC,* full blood count; *FSH,* follicle-stimulating hormone; *HbA_{1c},* hemoglobin A_{1c}; *LH,* luteinizing hormone; *LFT,* liver function test; *PSA,* prostate-specific antigen; *SHBG,* sex hormone–binding globulin.

Box 47.33 Effects and risks of feminizing hormone therapy

Effect	Onset	Maximum	Risks of oestrogen	Risks of cyproterone	Risks of spironolactone	Risks of GnRH analogue	Possible risk if additional risk factors are present
Decreased libido	1–3 months	3–6 months	DVT or PE	Meningioma	Hyperkalaemia	Prolonged QT interval	Osteopenia or osteoporosis
Decreased spontaneous erections	1–3 months	3–6 months	Infertility	Prolactinoma	Polyuria or dehydration	Joint pains	Type 2 diabetes
Redistribution of body fat	3–6 months	2–3 years	Hypertriglyceridemia	Increased insulin resistance	Depression	Other symptoms if used without oestrogen: osteopenia, osteoporosis, CVD, menopausal hot flushes	
Decrease in muscle mass and strength	3–6 months	1–2 years	Weight gain	Hypertension	GI bleed	Anxiety	
Softening of skin and decreased oiliness	3–6 months	Unknown	CVD	Weight gain		Low energy	
Breast growth	3–6 months	2–3 years	Cerebrovascular disease	Depression			
Decreased testicular volume	3–6 months	2–3 years	Cholelithiasis				
Decreased terminal hair growth	6–12 months	3 years	Hypertension				
Decreased sperm production	Unknown	>3 years	Erectile dysfunction				
Male sexual dysfunction	Variable	Variable	Hyperprolactinemia				
Scalp hair	No regrowth						
Voice changes	None						

CVD, Cardiovascular disease; *DVT,* deep vein thrombosis; *GI,* gastrointestinal; *GnRH,* gonadotrophin-releasing hormone; *PE,* pulmonary embolism.

Intended beneficial effects of hormone therapy in transgender women

Skin and hair

Oestrogen treatment results in finer skin texture and a reduction in facial and body hair. This effect is maximal after 4 months of treatment. Hormone therapy, including testosterone reduction, is not adequate to control facial hair growth and local measures such as electrolysis, waxing, shaving, sugaring, or laser therapy are needed to facilitate female presentation. Androgenic alopecia does not respond to hormone therapy. Testosterone suppression can stabilize the condition and minoxidil can increase hair density in mild androgenic alopecia. Where scalp hair loss is significant a hairpiece or hair transplant may be needed.

Breast development

Breast development begins about 2 to 3 months after the start of oestrogen treatment with maximum effect after 2 years. Average breast development is 14.5 ± 1.2 cm, which equates to a B cup brassiere size. Genetic factors limit breast development even when HRT is adequate.

Body shape

Body composition changes and there is an average of 3.8 kg weight gain. This is mainly as fat centred on the hips and buttocks giving a more female fat distribution with a decrease in lean body mass.

Genital changes and fertility

Oestrogen therapy reduces libido, decreases genital size, and inhibits fertility. Erectile dysfunction will occur. For most transgender women this does not require treatment. If there is a wish to retain genital function, PDE-5 inhibitors such as sildenafil can be used.

Mood and cognition

Transgender women using oestrogen therapy generally find a positive effect on their mood. Many transgender women report an increased feeling of femininity and a calmer mood.

Cancer risk

Prostate cancer is unusual in transgender women due to the use of androgen blockade. One study estimated the odds ratio for prostate cancer in transgender women to be 0.2–0.3 compared to cisgender men.

Adverse effects of hormone therapy in transgender women

Venous thromboembolism

The risk of thromboembolism increases by two- to fivefold compared with the background rate. This risk can be reduced by stopping smoking, maintaining a normal BMI, using topical rather than oral oestrogen (especially after the age of 45 years), and using oestradiol rather than ethinylestradiol or conjugated equine oestrogen.

Hyperprolactinaemia

This is reported in 4% to 15% of transgender women treated with oestrogen. Cyproterone acetate for androgen suppression has been associated with an increased risk of prolactinoma and is avoided in UK practice.

Abnormal liver function

Oestrogen treatment causes obstructive liver dysfunction in approximately 4% of transgender women. However, it is usually transient and resolves spontaneously in 3 to 6 months. Oestrogen therapy is associated with a fivefold increased risk of gallstones.

Cardiovascular risk

Oestrogen therapy leads to an increase in triglyceride concentrations, so a healthy diet is advised. Oestrogen therapy is also associated with an increase in insulin resistance.

Osteoporosis

Standard HRT regimens maintain bone mineralization for transgender women. However, vitamin D deficiency is common in transgender people, and bone mineralization may be reduced even before hormone treatment. Good bone health is promoted with weight-bearing exercise, not smoking, and maintaining vitamin D concentrations. If vitamin D deficiency is present, supplementation is required.

Cancer risk

The odds ratio for developing invasive breast cancer in transgender women is 46.7 (95% confidence interval [CI], 27.2–75.4) compared with cisgender men but 0.3 (95% CI, 0.2–0.4) compared with cisgender women. Because breast cancer risk is lower than in cisgender women, oestrogen therapy is usually lifelong for transgender women.

Hormone therapy for transgender men

Masculinization of a transgender man is achieved by giving testosterone. Physical changes follow a male puberty pattern and take 2 to 5 years to complete. Doses of testosterone used are the same as for a hypogonadal cisgender man. Short-acting mixed testosterone ester injections are the only preparation licensed for use in transgender medicine in the UK. Newer preparations, including topical gels and longer acting injections, are also effective. The common testosterone products used, monitoring, and effects are outlined in **Boxes 47.34 and 47.35**.

There should be regular monitoring of hormone concentrations until stable adult hormone concentrations are achieved followed by annual review of blood tests, weight, height, and blood pressure. After therapy has been started, it should be continued unless the person develops a significant side-effect or a condition that is an absolute contraindication to hormone therapy (e.g., severe polycythaemia).

Intended beneficial effects of hormone therapy in transgender men

Facial and body hair

Testosterone treatment in transgender men results in an increase in body hair and a change to a masculine hair pattern with hair growth on the face, chest, abdomen, lower back, and inner thighs. The onset of this is typically 6 to 9 months after starting testosterone treatment and is complete by 48 to 56 months. The skin texture coarsens, and acne can occur. Androgenic alopecia occurs in susceptible individuals.

Body shape

Testosterone increases lean body mass and upper body strength, with a concomitant reduction in body fat mass. This results in a masculine body shape with more muscle definition, increased upper body strength, and a decrease in hip-to-waist ratio. These are the same changes that occur in cisgender men taking testosterone.

Genital changes and fertility

The first physical change usually noticed by a transgender man on testosterone treatment is clitoral enlargement. Onset is by 3 to 4 months and is largely complete by 1 year with final clitoral length of 4 to 5 cm. This amount of growth is not usually sufficient for penetrative intercourse. Menstruation usually ceases after 3 to 4 months of treatment. Ovarian morphology becomes polycystic after 6 months of testosterone exposure, but it is not known if these changes affect ovarian function. Testosterone treatment is not an effective contraceptive, and appropriate contraception should be recommended whenever it is required.

Mood and cognition

Testosterone therapy has important effects on behaviour. Assertiveness, aggression, and general drive are increased when testosterone is given to cisgender men. Transgender men report higher levels of energy, aggression, and libido. Studies show that

Box 47.34 Medications used for masculinization in UK practice and monitoring required

Route	Preparation	Dose	Frequency	Testosterone concentrations	Monitoring			
					Baseline tests	Therapeutic monitoring	Timing of therapeutic monitoring	Long-term monitoring
Monthly injection	Testosterone enanthate Testosterone cypionate	50–250 mg 50–100 mg	10–35 days 7–14 days	Trough: 8–12 nmol/L (bottom of adult male range) Peak: 25–30 nmol/L (top of adult male range)	LH[a] FSH[a] Testosterone[a] Oestradiol[a] SHBG[a] Prolactin[a] LFT FBC Lipid profile HbA_{1c} 25-Hydroxy vitamin D Weight BMI BP	Testosterone SHBG Lipid profile LFT FBC	Trough: before injection Peak: 7 days later	If dose change: • Three to four injections after a change in dose • 8 weeks after a change in gel dose On stable dose: 6 monthly for 2 years; then annually
Longer acting injection	Testosterone undecanoate	500–1000 mg	Every 6–15 weeks (has loading phase)	15–20 nmol/L (middle of adult male range)			On day of injection before injection is administered	
Gel	Testosterone	50–150 mg (sachets) or 10–150 mg (pump actuation)	Daily	15–20 nmol/L (middle of adult male range)			4–6 hours after application of gel (ensure no gel on arms)	

[a]Blood tests are taken in the follicular phase of the menstrual cycle.

BMI, Body mass index; *BP*, blood pressure; *FBC*, full blood count; *FSH*, follicle-stimulating hormone; *HbA_{1c}*, hemoglobin A_{1c}; *LH*, luteinizing hormone; *LFT*, liver function test; *SHBG*, sex hormone–binding globulin.

Box 47.35 The effects and risks of masculinizing hormone therapy

Effect	Onset (months)	Maximum (years)	Risks of testosterone	Possible risk if additional risk factors are present
Cessation of menses	1–6	N/A	Polycythemia	Type 2 diabetes
Skin oiliness or acne	1–6	1–2	Infertility	CVD
Clitoral enlargement	1–6	1–2	Acne	
Vaginal atrophy	1–6	1–2	Androgenic alopecia	
Fat redistribution	1–6	2–5	Hypertension	
Deepening of voice	6–12	1–2	Sleep apnoea	
Increased muscle mass and strength	6–12	2–5	Weight gain	
Facial and body hair growth	6–12	4–5	Decreased HDL cholesterol	
Scalp hair loss	6–12	Unknown	Increased LDL cholesterol Hypertriglyceridemia	

CVD, Cardiovascular disease; *HDL*, high-density lipoprotein; *LDL*, low-density lipoprotein; *N/A*, not applicable.

transgender men taking testosterone have improved visuospatial ability.

Voice

Testosterone causes growth of the larynx and thickening of the vocal cords, resulting in a lower vocal pitch. Usually the voice 'breaks' after 9 to 12 months of testosterone treatment, but voice change can take up to 2 years to complete.

Bone density

Most studies in transgender men show that testosterone therapy appears to maintain bone mineralization.

Cancer risk

Breast cancer risk after masculinizing chest surgery is approximately one-third of the risk in cisgender women. However, transgender men should be offered the same breast cancer screening as females if they have not had male chest reconstructive surgery.

Adverse effects of hormone therapy in transgender men

Lipid profile

Despite cisgender men having a more adverse lipid profile than cisgender women, testosterone therapy for transgender men results in no change in total cholesterol or low-density lipoprotein cholesterol concentrations. There is a minor increase in triglycerides and a minor decrease in blood high-density lipoprotein concentrations that does not translate to a significant increase in cardiovascular risk with an odds ratio for myocardial infarction of 0.3 to 1.0 compared with cisgender men.

Polycythaemia

Testosterone increases the production of red blood cells by increasing plasma erythropoietin levels. Testosterone therapy increases haematocrit by 8% to 12%, conferring a 3% to 5% risk of polycythaemia. This risk is higher with injectable than topical testosterone. Polycythaemia is managed with hydration and changing from injectable to gel therapy. If indicated, therapeutic venesection may be performed.

Abnormal liver blood tests

Increased concentrations of transaminases are seen in about 4% to 7% of transgender men. Treatment is continued provided concentrations are less than three times the upper limit of the reference range.

Cancer risk

Testosterone can be aromatized to oestradiol, which could theoretically result in endometrial hyperplasia. However, in clinical practice, reports of endometrial hyperplasia are rare, and endometrial cancer is extremely rare in transgender men. In the UK, endometrial surveillance using interval ultrasonography is recommended. Internationally, monitoring is only carried out if the patient has pelvic symptoms such as bleeding or pain. Transgender men may elect to have a hysterectomy to prevent any risk of endometrial hyperplasia. Ovarian cancer risk appears to be very low. Transgender men should be offered cervical screening as per the national screening programme for females.

Menstruation

Testosterone monotherapy may not be sufficient to suppress ovarian function. Menstrual activity can be suppressed with continuous progestin such as medroxyprogesterone acetate. If this is not adequate or tolerated, a GnRH analogue can be used.

Speech and language therapy

The larynx is not significantly affected by hormone therapy in transgender women. Speech and language therapy is needed for a transgender woman to feminize the voice. Voice therapy can help with altering the pitch, timbre, cadence, and modulation of voice, all of which are important elements in voice feminization. Communication therapy is important in feminizing vocal patterns, which vary between genders. Testosterone does lead to change to the larynx in transgender men, but speech and language therapy may augment masculinization.

Surgical Intervention

Gender-affirming surgical procedures are designed to make a person's body more congruent with their internal sense of gender identity. The assessment for surgical interventions is the same as for hormone treatment, but the person should also be stable on hormone therapy for at least 6 months. Some surgical procedures require a longer duration of hormone therapy to achieve an optimal result (e.g., metoidioplasty in which clitoral tissue hypertrophied using testosterone is used to fashion a penis). In the UK, masculinizing chest surgery and masculinizing and feminizing genital surgery are provided through NHS-funded services. Other gender-affirming surgical procedures, including breast augmentation, facial surgery (either masculinizing or feminizing), or phonosurgery (to increase voice pitch), are not currently funded by the NHS.

Masculinizing surgeries

Masculinizing chest surgery involves the removal of the breast tissue (mastectomy) and recontouring of the chest wall and nipple complex to achieve a masculine chest appearance. Chest surgery has been shown to improve quality of life and decrease gender dysphoria.

Genital surgery in which a new phallus is created is done by metoidioplasty or phalloplasty. Metoidioplasty is often a single-stage procedure. The new phallus created is not large enough for penetrative sexual function, but it usually allows the individual to urinate standing up. Phalloplasty is a multistage procedure allowing the formation of a new phallus that is large enough for sexual function, using skin from a donor site. Genital surgery may include hysterectomy and oophorectomy, rendering the person permanently hypogonadal. Testosterone therapy is usually continued during the perioperative period. Because testosterone therapy suppresses ovarian function, the dose of testosterone usually does not change after genital surgery unless oophorectomy has been performed.

Feminizing surgeries

Feminizing genital surgery usually involves orchidectomy. Although this may be performed as a standalone procedure, it is usually accompanied by penectomy, vulvoplasty, and formation of a neo clitoris. If the person desires sexual function, then vaginoplasty can also be performed. It is usual for hormone therapy to be stopped 6 weeks before genital surgery to decrease thromboembolic risk. It is restarted when the person is mobile, usually 2 weeks after surgery. Androgen suppression treatment continues during the perioperative period, and patients may experience menopausal symptoms during this time. Because most patients have their testosterone

suppressed into the female range before surgery, the dose of oestrogen does not change after surgery.

Treatment of non-binary people

Approximately 10% of people attending a gender identity clinic have a gender identity that does not sit within the gender binary of male or female. A non-binary sense of gender may include people who feel their gender identity is partially male and partially female. Other people may have a gender identity that fluctuates with time termed gender fluidity. Some people have a sense of no gender at all, termed 'agender' or 'neutroid'. For other non-binary people, their sense of gender does not have a descriptive term in the English language (e.g., the two-spirit people of native North American communities).

Cultural competence is important when working with people with a non-binary identity. The usual pronouns of he/him and she/her usually do not apply; neutral pronouns such as they/them or neo-pronouns such as ze/zir may be preferred.

There are no clearly established guidelines for the hormonal treatment of non-binary people; it is therefore best that the treatment of an individual with physical measures such as hormone or surgical interventions is done in the context of a multidisciplinary team. The assessment process should establish the ideal mix (or lack) of masculine and feminine physical features for the individual. It should also explore how fluid the gender identity is and if this is likely to change with time. The final stage of assessment is to explore whether physical changes are feasible, achievable, and safe using available hormone and surgical treatments.

Many non-binary people wish to have a partial degree of masculinization or feminization to allow them to have an androgenous gender presentation. This is achieved either by using lower doses of masculinizing or feminizing hormone therapy than those used in binary gender transition or by using full-dose therapy for a limited period to achieve partial pubertal development. For people who wish to have a neutral gender expression, removal of sex steroid production using GnRH analogues may be undertaken. Side-effects include a reduction in bone mineralization and the adverse metabolic effects of dyslipidaemia. It is important to counsel about these risks. People should have monitoring of their lipid profile and dual-energy X-ray absorptiometry (DEXA) scanning. If metabolic abnormalities or reduced bone mineralization develop, they should be treated. As with people undergoing a binary gender transition, people with a non-binary gender identity should be counselled about the impact on their fertility and offered fertility preservation before starting treatment

Further reading

Cass H. *The Cass Review: Independent Review of Gender Identity Services for Children and Young People.* 2024 Available at https://cass.independent-review.uk/.

Coleman E et al. Standards of care for the health of transgender and gender diverse people, Version 8. *Int J Transgend Health* 2022; 23:S1–S259.

Hembree WC et al. Endocrine treatment of gender-dysphoric/gender-incongruent persons: an Endocrine Society Clinical Practice Guideline. *Endocr Pract* 2017; 23:1437.

Richards C, Barrett J. *Trans and Non-Binary Gender Healthcare for Psychiatrists, Psychologists, and Other Health Professionals*. Cambridge, Cambridge University Press; 2021.

Seal LJ. Hormone treatment for transgender adults. In: **Bouman WPA and Jon A (eds.)**, *The Transgender Handbook: A Guide for Transgender People, Their Families and Professionals*. New York: Nova Publishers; 2017.

Seal LJ. Adult endocrinology. In: **Richards CPBW and Barker MJ (eds.)**, *Genderqueer and Non-Binary Genders (Critical and Applied Approaches in Sexuality, Gender and Identity).* London: Palgrave; 2018.

Questions for this chapter can be found on Elsevier eBooks+.

48 Obstetric medicine

Catherine Nelson-Piercy and Melanie Nana

Clinical skills 1681
History 1681
Examination 1681
Investigations 1681
Screening 1681
Background 1683
Hypertensive disorders 1683
Chronic or pre-existing hypertension 1683
Gestational hypertension 1683
Pre-eclampsia 1683
HELLP 1684
Liver disease 1684
Hyperemesis gravidarum 1684
Intrahepatic cholestasis of pregnancy 1684
Acute fatty liver of pregnancy 1685
Cardiac disease 1685
Thromboembolic disease 1686
Respiratory disease 1686
Asthma 1686
Pneumonia 1687
Tuberculosis 1687
Sarcoidosis 1687
Cystic fibrosis 1687
Severe restrictive lung disease 1687
Neurological disease 1687
Epilepsy 1687
Migraine 1687
Multiple sclerosis 1688
Myasthenia gravis 1688
Stroke 1688
Endocrine disease 1688
Type 1 and 2 diabetes 1688
Gestational diabetes 1689
Thyroid and parathyroid disease 1690
Pituitary disease 1690
Adrenal disease 1691
Kidney disease 1691
Urinary tract infection 1691
Chronic kidney disease 1691
Kidney transplantation 1691
Acute kidney injury 1692
Skin disorders 1692
Rheumatic disease 1692
Prescribing in pregnancy 1693

CORE SKILLS AND KNOWLEDGE

Obstetric medicine refers to the care of pregnant persons with medical problems. These include the 10% who enter pregnancy with pre-existing diseases (e.g., asthma, epilepsy, hypertension, and diabetes), as well as those with diseases that arise acutely in pregnancy (e.g., pneumonia and venous thrombosis).

Obstetric medicine is a new and growing medical specialty composed of physicians who specialize in providing care in pregnancy complicated by medical conditions and who increasingly work alongside specialist obstetricians (who have also received specialist training in obstetric medicine), specialist obstetric anaesthetists, and specialist midwives. Historically (and currently in many centres without an obstetric physician), this care was provided by general obstetricians alongside general adult physicians. Training programmes in most medical specialties require trainees to learn about the care of patients within this specialty in the context of pregnancy. General practitioners also have a crucial role because they provide continuity of care before and after pregnancies.

Special skills necessary for physicians to practice obstetric medicine include:

- The ability to take a detailed obstetric history
- An understanding of the relevance of chronic medical diseases to the physiological adaptations required for pregnancy
- An understanding of how to interpret blood results and other investigations in pregnancy
- An in-depth knowledge of the safety of radiological investigations and medications in pregnancy and breast-feeding, balanced with the risks of not investigating or treating in the same way as outside of pregnancy.

Opportunities to learn obstetric medicine may arise by completing training attachments in obstetrics and gynaecology, by attending obstetric medicine multidisciplinary team meetings or clinics if they exist, and by seeking out pregnancies under close surveillance because of coexisting medical problems. A diploma in obstetric medicine is now available through the Royal College of Physicians, with specific training posts available in England.

CLINICAL SKILLS

History

Many symptoms occur in normal pregnancy, including fatigue, nausea, vomiting, palpitations, breathlessness, gastro-oesophageal reflux, constipation, and oedema. A careful history is required to discriminate pathology from physiology. Typically, pathological symptoms arise suddenly, differ from those experienced in previous pregnancies or before pregnancy, and can be particularly severe. Even with significant experience, differentiating pathological from physiological symptoms in pregnancy can be challenging.

Examination

Physiological changes occurring in pregnancy are illustrated in the figure on the next page. They contribute to the clinical signs in normal pregnancy, including a resting tachycardia of up to 110 beats/min, relative hypotension, and oedema.

Investigations

Blood tests need interpreting based on pregnancy reference ranges, which differ from those outside pregnancy because of underlying changes in physiology. Serum creatinine falls below 77 μmol/L, there is a physiological anaemia related to haemodilution, and alkaline phosphatase is raised because of placental production (**Box 48.1**).

Screening

Healthcare systems vary as to which screening tests are carried out during pregnancy, according to local prevalence of different diseases and availability of resources. **Box 48.2** describes tests routinely performed in the UK. Clinic appointments are usually made at booking (generally 8–12 weeks' gestation) and then at 16, 25, 28, 31, 34, 36, 38, 40, 41, and 42 weeks' gestation in healthy pregnancies, with fetal ultrasonography at 12 and 20 (and in some units 36) weeks' gestation.

i Box 48.1 Reference values in pregnancy

Test	Reference interval in pregnancy	Reference interval in non-pregnant population
Haemoglobin (Hb)	105–140 g/L	115–155 g/L
Mean corpuscular volume (MCV)	80–100 fL	80–96 fL
White cell count (WCC)	6–16 × 10^9/L	4–11 × 10^9/L
Platelets	150–400 × 10^9/L	150–400 × 10^9/L
Sodium	130–140 mmol/L	135–146 mmol/L
Potassium	3.3–4.1 mmol/L	3.5–5.0 mmol/L
Urea	<4.5 mmol/L	2.5–6.7 mmol/L (8–25 mg/dL)
Creatinine	<77 μmol/L	45–90 μmol/L (0.5–1.0 mg/dL)
Bilirubin	<16 μmol/L	<17 μmol/L (0.3–1.5 mg/dL)
Aspartate aminotransferase	10–30 U/L	12–40 U/L
Alanine aminotransferase	6–32 U/L	<40 U/L
Alkaline phosphatase	30–418 U/L	39–117 U/L
Albumin	28–37 g/L	35–50 g/L

i Box 48.2 Screening tests in pregnancy

Confirmation of pregnancy
- Urine beta-human chorionic gonadotrophin (β-hCG) testing kits are highly sensitive for confirming pregnancy in the context of a missed period (and need not be repeated if a female has performed a test that has proved positive at home).
- A serum β-hCG level is not used in routine practice for diagnosing pregnancy but is used in managing certain conditions, including ectopic pregnancy and molar pregnancy.

Infectious diseases
- Human immunodeficiency virus, hepatitis B, and syphilis are screened for in early pregnancy so treatment can be instituted to reduce the risk of transmission to the fetus.

Obesity
- Body mass index is calculated from height and weight measured at the booking visit.
- Lifestyle advice to encourage healthy behaviour is given at this point and repeated throughout pregnancy.

Hypertension and pre-eclampsia
- BP measurement and urine dipstick examination are performed at each antenatal appointment.
- Pregnancies complicated by hypertension or proteinuria require close monitoring; kidney and liver function is measured.

Anaemia
- An FBC is measured at booking and at the 28-week visit.
- Haemoglobin concentrations are lower during pregnancy.
- Ferritin, folate, and vitamin B_{12} concentrations are measured if anaemia is diagnosed. Iron deficiency is common.

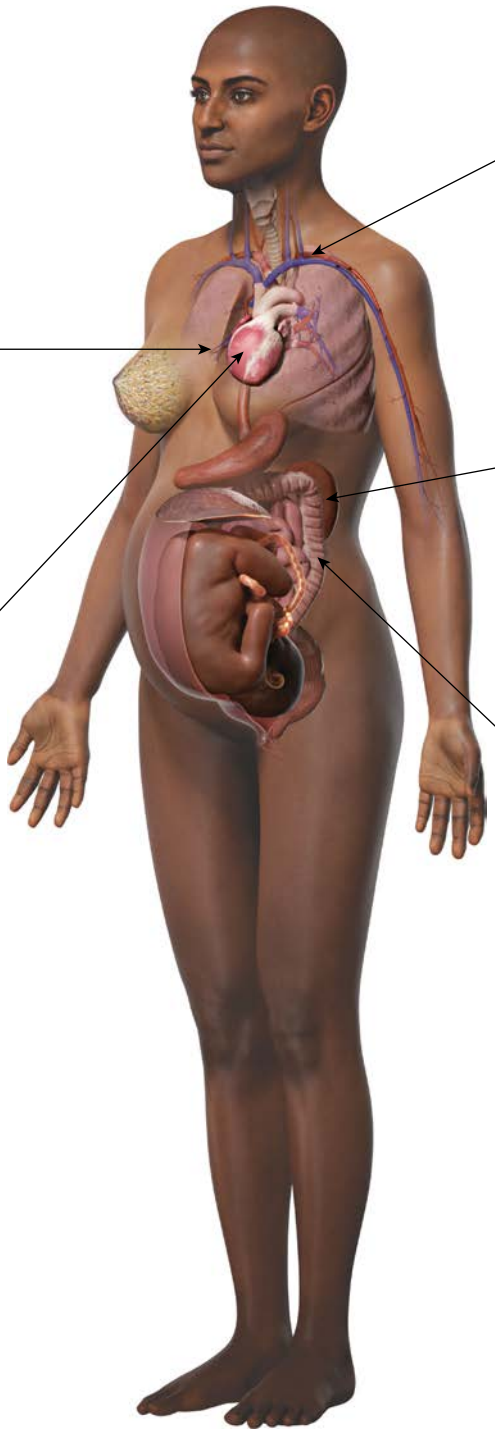
Gestational diabetes mellitus
- An oral glucose tolerance test is offered at 24 to 28 weeks' gestation to all who have risk factors, including those with previous gestational diabetes mellitus, previous high-birth-weight babies, or a family history of type 2 diabetes.

Blood group and Rhesus status
- Blood group is recorded in case there is a transfusion requirement after delivery.
- Females who are Rhesus negative are offered anti-D in the third trimester to prevent formation of Rhesus antibodies (from transfer of Rhesus-positive blood from a Rhesus-positive baby), which may cause haemolytic disease of the newborn in a subsequent pregnancy (see Chapter 38). Free fetal DNA is increasingly used to target the prophylaxis only to those with Rhesus-positive babies.

Screening for inherited diseases
- Thalassaemia and sickle cell screening is offered in all pregnancies to those in high-risk populations to identify (via partner screening if the female is a carrier) the risk of disease in the fetus, which can be confirmed by chorionic villous sampling.

Screening for fetal chromosomal abnormalities

Physiological changes in pregnancy



Respiratory
↑ Oxygen consumption (by 20%)
↑ Minute ventilation (by 40%–50%)
↑ PO_2
↓ PCO_2
↓ Forced vital capacity (third trimester)

Cardiac
↑ Cardiac output (by 40%)
↑ Stroke volume
↑ Heart rate (by 10–20 bpm)
↓ Blood pressure (↓ in first and second trimesters)
↓ Systemic vascular resistance (by 25%–30%)
↓ Serum colloid osmotic pressure (by 10%–15%)

Endocrine
Insulin resistance
↑ Prolactin (10-fold)
↑ Cortisol
↑ Renin, angiotensin, aldosterone

Haematology
↑ Plasma volume (by 50%)
↓ Haemoglobin
↓ Platelets
↑ Fibrinogen (by 50%)
↑ Clotting factors VIII, IX, X
Venous stasis (left leg > right leg)

Renal
↑ Renal blood flow (by 60%–80%)
↑ Glomerular filtration rate (by 55%)
↑ Protein excretion (up to 300 mg/24 h)
↓ Serum creatinine (max = 77 μmol/L)
Glycosuria
Physiological renal tract dilation

Gastroenterology
↓ Gut motility
↑ Alkaline phosphatase
↓ Albumin (by 25%)

General
Fatigue
Weight gain
Nausea/vomiting
Constipation
Breathlessness
Palpitations
Ankle oedema

Skin
Palmar erythema
Dry skin
Telangiectasia
Pruritus (in 20%)

BACKGROUND

The prevalence of medical conditions that complicate pregnancy is increasing, partly because:

- Females in high-income countries are delaying childbearing.
- Fifty percent of pregnant persons are now overweight or obese, resulting in more patients with diabetes, hypertension, and associated co-morbidities.
- Those with chronic medical conditions, including congenital heart disease, diabetes, chronic kidney disease (CKD), and cystic fibrosis (CF), now survive into their childbearing years because of better medical care.
- Widening access to assisted conception, including use of donor eggs, means that age is no longer a barrier to pregnancy.

Death during or soon after pregnancy in the UK is now more often caused by medical conditions than obstetric complications. Confidential enquiries show that inadequate knowledge and care are implicated in one-quarter of deaths from medical conditions in pregnancy.

The anatomical and physiological changes that occur as an adaptation to normal pregnancy result in symptoms and signs that can overlap with clinical features of disease outside pregnancy. Many symptoms and signs are benign, but clinicians need to be aware of those that warrant further investigation and may be associated underlying pathology.

Females with pre-existing disease should be made aware of the required physiological adaptation to pregnancy, and symptoms may worsen or improve, depending on the body system involved. The physiological stress of pregnancy may also result in previously subclinical disease presenting for the first time.

Infectious diseases are a major problem in pregnancy leading to increased maternal and fetal mortality, particularly in low- and middle-income countries.

HYPERTENSIVE DISORDERS

During normal pregnancy, there is a reduction in blood pressure (BP) because of a decrease in systemic vascular resistance; this is maximal by gestational weeks 18 to 20.

Chronic or pre-existing hypertension

This presents with a raised BP before 20 weeks or with treated high BP if the patient was taking medication for hypertension before pregnancy.

Management

- ***Angiotensin-converting enzyme (ACE) inhibitors, angiotensin II receptor blockers (ARBs)***, and ***thiazide diuretics*** are associated with congenital abnormalities and should be discontinued.
- ***Dietary salt*** intake should be kept low (<2 g or 100 mmol daily).
- ***Aspirin*** 150 mg should be given from 12 weeks' gestation to those at risk of pre-eclampsia (see later).
- ***Oral labetalol*** is first-line therapy during pregnancy in the UK. Alternative agents are modified-release nifedipine and methyldopa.
- BP is treated in pregnancy when it is sustained at 140/90 mmHg or more.
- After treatment, ***target BP*** is less than 135 mmHg systolic and less than 85 mmHg diastolic. Difficulty controlling BP may be an indication for delivery.

Box 48.3 Risk factors for pre-eclampsia

- Age older than 40 years
- Obesity: body mass index >30 kg/m^2
- Family history
- Primiparity
- Multiple pregnancy
- Previous pre-eclampsia
- Long birth interval
- Hydrops with a large placenta
- Hydatidiform mole
- Triploidy (particular association with pre-eclampsia of very early onset—before 24 weeks' gestation)
- Medical disorders
  - Pre-existing hypertension
  - Kidney disease
  - Pre-existing diabetes mellitus
  - Antiphospholipid syndrome
  - Autoimmune rheumatic diseases
  - Sickle cell disease

In ***breast-feeding***, ACE inhibitors (enalapril), beta-blockers, modified-release nifedipine, and amlodipine are safe. Methyldopa should be avoided because of the risk of depression.

Gestational hypertension

This is defined as BP of 140/90 mmHg or more after 20 weeks' gestation in a previously normotensive female.

Investigations and management are aimed at distinguishing gestational hypertension from pre-eclampsia and maintaining a safe maternal BP for pregnancy:

- ***Blood tests:*** full blood count (FBC), serum creatinine and electrolytes, and liver biochemistry; estimated glomerular filtration rate (eGFR) is not validated for use in pregnancy
- ***BP measurements*** once a week
- ***Urine tests*** for protein
- ***Treatment*** of BP of 140/90 mmHg or more (which may be required for up to 3 months postpartum)
- ***Admission to hospital*** if BP of severe maternal hypertension (≥160/110 mmHg)

Pre-eclampsia

Pre-eclampsia is a common direct (obstetric) cause of maternal and perinatal death and a cause of significant morbidity. The risk factors are shown in **Box 48.3**. It is a heterogeneous multisystem endothelial disorder that has widespread effects.

There is good evidence that low-dose aspirin is beneficial in preventing preterm pre-eclampsia; it is safe in pregnancy. It is given from 12 weeks' gestation and is usually stopped at 36 weeks' gestation or sooner if the patient goes into labour or develops pre-eclampsia.

Clinical features

Pre-eclampsia varies in severity, timing, progression, and order of onset of different clinical features. Symptoms include:

- Severe or persistent headache
- Visual blurring or flashing
- Nausea and vomiting
- Severe epigastric or right upper quadrant pain
- Rapidly progressive oedema of the hands, feet, and face.

Signs include new hypertension (140/90 mmHg or more) after 20 weeks' gestation, oedema, and proteinuria (urinary protein:creatinine ratio >30 mg/mmol). Hyper-reflexia, clonus, and seizures (eclampsia) may occur.

Pre-eclampsia (including eclampsia) may present antenatally or intra- or postpartum. Postpartum disease is more likely to be associated with symptoms, particularly headache. Complications and crises are shown in **Box 48.4**.

Box 48.4 Complications and crises seen in pre-eclampsia

- Eclampsia
- HELLP syndrome (*h*aemolysis, *e*levated *l*iver enzymes, *l*ow *p*latelets)
- Pulmonary oedema
- Placental abruption
- Cerebral haemorrhage
- Cortical blindness
- Disseminated intravascular coagulation
- Acute kidney injury
- Hepatic rupture
- Transient left ventricular systolic or diastolic dysfunction

Investigations

- ***FBC:*** thrombocytopenia
- ***Serum electrolytes and creatinine:*** acute kidney injury (AKI)
- ***Liver biochemistry:*** transaminitis
- ***Urine protein:creatinine ratio:*** new proteinuria
- ***Ultrasonography for fetal growth:*** to assess for placental dysfunction and inform timing of delivery
- ***Plasma placental growth factor (PlGF) concentration alone or as a ratio to soluble fms-like tyrosine kinase-1 (sFlt1:PlGF):*** novel immunoassays recommended by the National Institute for Health and Care Excellence (NICE) to diagnose or rule out pre-eclampsia after clinical assessment

Management

Close monitoring is required to ensure BP control and to assess for the development of severe disease features. This is required at least twice per week. Hospital admission may be indicated if BP control is difficult, if severe features develop, or if there is any other clinical concern. Close fetal monitoring is required because of the risks of placental insufficiency and intrauterine growth restriction.

If a patient develops eclampsia (convulsions or coma) or HELLP syndrome (see next section), she should be admitted to a critical care or high-dependency unit. Intravenous antihypertensives (labetalol, hydralazine) can be used if oral medications are ineffective or the oral route is not safe or possible. In eclampsia (or when there is risk of developing eclampsia), magnesium sulphate (4 g bolus dose followed by an infusion of 1 g/h) is used to treat and provide prophylaxis of seizures. Patients should be given steroids (intramuscular betamethasone/dexamethasone 12 mg repeated at 24 hours) for fetal lung maturation if delivery before 35 weeks' is likely within 7 days. Delivery is the only cure for pre-eclampsia, and this may be indicated for either fetal or maternal reasons. The recurrence risk of pre-eclampsia is increased with early-onset disease. Those whose pregnancy has been complicated by pre-eclampsia are significantly more likely to develop hypertension, ischaemic heart disease, cerebrovascular disease, and CKD in later life.

HELLP

HELLP is a syndrome describing ***h***aemolysis, ***e***levated ***l***iver enzymes, and ***l***ow ***p***latelets (<100 × 10^9/L) in a pregnant female. It is a form of severe pre-eclampsia, more common in multiparous females, and can occur in the second and third trimesters or postpartum. The haemolysis is caused by a microangiopathic haemolytic anaemia. Hypertension and proteinuria may be absent or mild. Clinically, patients complain of epigastric or right upper quadrant pain, nausea, vomiting, and jaundice (5%). Prompt delivery is usual and advised in all patients who are beyond 34 weeks' gestation.

Further reading

Bartsch E, Medcalf KE, Park AL et al. Clinical risk factors for pre-eclampsia determined in early pregnancy: Systemic review and meta analysis of large cohort studies. *BMJ* 2016; 353:i1753.

National Institute for Health and Care Excellence. *NICE Clinical Guideline 133: Hypertension in pregnancy: Diagnosis and management*. NICE 2019; http://www.nice.org.uk/guidance/ng133.

LIVER DISEASE

Liver function is not impaired in normal pregnancy. Pregnancy-specific liver diseases include intrahepatic cholestasis and acute fatty liver of pregnancy. In addition, liver disease from any cause can occur incidentally and coincide with pregnancy. For example, viral hepatitis accounts for 40% of all cases of jaundice during pregnancy. Pregnancy does not necessarily exacerbate established liver disease. It is less common for females with advanced liver disease to conceive.

The following effects are seen:

- Plasma and blood volumes increase during pregnancy, but the hepatic blood flow remains constant.
- The proportion of cardiac output delivered to the liver therefore falls from 35% to 29% in late pregnancy; drug metabolism can thus be affected.
- The size of the liver remains constant.
- Liver biochemistry remains unchanged, apart from a rise in serum alkaline phosphatase from the placenta (up to three to four times increased) and a decrease in total protein and albumin concentration owing to increased plasma volume.
- Triglyceride, cholesterol, caeruloplasmin, transferrin, α_1-antitrypsin, and fibrinogen concentrations are elevated because of increased hepatic synthesis.
- Pregnancy and the postpartum period confer an increased risk of venous thromboembolism (VTE). Acute Budd–Chiari syndrome can occur.

Hyperemesis gravidarum

Pathological vomiting during pregnancy can be associated with liver dysfunction and jaundice. Liver dysfunction resolves when vomiting subsides.

Intrahepatic cholestasis of pregnancy

This condition of unknown aetiology usually presents with pruritus in the third trimester. It has a familial tendency, and there is a higher prevalence in Scandinavia, Chile, and Bolivia.

Liver biochemistry shows raised serum bile acids that may or may not be associated with raised aminotransferase concentrations. Gamma-glutamyl transferase may also be slightly raised. Jaundice is rare. Liver biopsy is not indicated but would show centrilobular cholestasis. The prognosis is usually excellent for the pregnant person, but there is increased risk of fetal loss in those with serum bile acids greater than 100 μmol/L and an increased risk of preterm labour in those with serum bile acids greater than 40 μmol/L.

Treatment with ursodeoxycholic acid 1 to 2 g daily in divided doses may reduce spontaneous preterm labour but does not treat itch or improve liver biochemistry. Delivery is timed to try to avoid

Box 48.5 Pregnancy-associated risks of different cardiac diseases

Low risk
- Uncomplicated small or mild pulmonary stenosis, PDA, or MVP
- Successfully repaired ASD, VSD, or PDA
- Atrial and ventricular ectopic beats

Small increased risk of death and moderate increased risk of morbidity
- Unrepaired ASD or VSD
- Repaired tetralogy of Fallot
- Most arrhythmias

Moderate increased risk of death or severe morbidity
- Mild LV dysfunction (LVEF >45%), hypertrophic cardiomyopathy
- Repaired coarctation
- Most native or tissue valve disease (except those with extremely high risk)

Significant increased risk of death or severe morbidity
- Mechanical valves
- Systemic RV with good or mildly impaired ventricular function
- Fontan circulation
- Unrepaired cyanotic congenital heart disease

Extremely high risk of death or severe morbidity
- Pulmonary arterial hypertension
- Severe systemic ventricular dysfunction: LVEF <30% (NYHA class III or IV; see **Box 17.20**)
- Previous peripartum cardiomyopathy with residual LV impairment
- Severe mitral stenosis or symptomatic aortic stenosis
- Severe aortic (re)coarctation
- Systemic RV with moderate or severely decreased ventricular function
- Dilated aortic root
  - Marfan's syndrome >45 mm
  - Turner's syndrome with ASI >25 mm/m^2
  - Tetralogy of Fallot >50 mm
  - Ascending aorta >50 mm with bicuspid aortic valve
- Vascular Ehlers–Danlos syndrome
- Fontan circulation with any complication

ASD, Atrial septal defect; *ASI,* aortic size index; *LV,* left ventricular; *LVEF,* left ventricular ejection fraction; *MVP,* mitral valve prolapse; *NYHA,* New York Heart Association; *PDA,* patent ductus arteriosus; *RV,* right ventricle; *VSD,* ventricular septal defect.

the increased risk of fetal loss, depending on disease severity. The condition resolves after delivery. Recurrent cholestasis may occur during subsequent pregnancies or with the ingestion of oestrogen-containing oral contraceptive pills.

Acute fatty liver of pregnancy

Acute fatty liver of pregnancy (AFLP) is a rare but serious condition occurring in the third trimester and affecting 1 in 6000 to 20 000 pregnancies. It is more common in slim females with multiple gestations. Patients present with abdominal pain, vomiting, and jaundice. Polyuria and polydipsia caused by diabetes insipidus may occur because of reduced breakdown of placentally derived vasopressinase. AKI and lactic acidosis are commonly seen. In fulminant AFLP, encephalopathy, coagulopathy, and hypoglycaemia can occur. AFLP may be associated with features of pre-eclampsia.

Investigations

- FBC: neutrophilia
- Serum electrolytes and creatinine: AKI
- Plasma glucose: hypoglycaemia
- Serum uric acid: increased
- Liver biochemistry: hepatocellular damage: raised serum aminotransferases and raised bilirubin
- Coagulopathy: increased prothrombin time
- ***Liver biopsy:*** rarely performed and not necessary for diagnosis but would show fine droplets of fat (microvesicular) in the hepatocytes with little necrosis

Treatment includes delivery and supportive care as for acute liver failure (see Chapter 23). *N*-acetyl cysteine is safe in pregnancy. Early diagnosis and treatment have reduced the mortality rate to less than 20%.

Further reading

EASL Clinical Practice Guideline. *Liver Disease in Pregnancy* (published July 3rd 2023). https://easl.eu/publications/clinical-practice-guidelines/.

CARDIAC DISEASE

Cardiac disease, both congenital and acquired, is commonly encountered in pregnancy and is mostly benign. Many pregnant patients experience palpitations and breathlessness because of awareness of physiological sinus tachycardia, ventricular premature or ectopic beats, and increased minute ventilation.

Cardiac disease, predominantly acquired, is the most common cause of maternal death in the UK. The main conditions causing severe morbidity and mortality are as follows:

- ***Myocardial infarction (MI)*** in pregnancy is more often caused by coronary artery dissection or thrombus compared with non-pregnant patients. Intervention with primary angioplasty and coronary artery stenting is appropriate.
- ***Peripartum cardiomyopathy*** is defined as the development of heart failure with left ventricular ejection fraction (LVEF) less than 45% at the end of pregnancy or in the months after delivery when no other cause of heart failure is found. It is a dilated cardiomyopathy and is associated with obesity and multiparity. It is treated with conventional heart failure therapy (including thromboprophylaxis) except ACE inhibitors and ARBs are withheld until after delivery. Nearly half of patients recover to normal within 6 months, but many are found to have a pre-existing but undiagnosed dilated cardiomyopathy that became symptomatic in pregnancy.
- ***Pulmonary hypertension*** and fixed pulmonary vascular resistance are dangerous and often (10%–20%) fatal in pregnancy.

Females with significant heart disease need multidisciplinary care in a specialist centre that includes obstetricians, cardiologists, and anaesthetists with expertise in the care of heart disease in pregnancy. The levels of risk associated with pregnancy for individual conditions are summarized in **Box 48.5** based on the European Society of Cardiology's guidelines.

Pregnancy may be contraindicated where there is a high risk of death or serious morbidity (**Box 48.5**). Absolute and relative contraindications to pregnancy include a dilated aortic root of more than 4.5 cm, severe left heart obstruction from critical mitral or aortic stenosis, and severe impairment of left ventricular function. If pregnancy is contraindicated, contraceptive advice and provision is essential.

Pregnancy and prosthetic heart valves

Any strategy for anticoagulation of a mechanical heart valve in pregnancy is associated with risks to the pregnant person, fetus, or both. ***Pre-pregnancy counselling*** is essential, explaining that

Box 48.6 Risk factors for venous thromboembolism in pregnancy

- Previous venous thromboembolism
- Thrombophilia
- Active inflammatory disease, e.g., inflammatory bowel disease, systemic lupus erythematosus, pyelonephritis, nephrotic syndrome
- Age older than 35 years
- Obesity (body mass index >30 kg/m^2) before or in early pregnancy
- Parity of three or more previous deliveries
- Smoking
- Gross varicose veins (symptomatic or above the knee or with associated phlebitis, oedema, or skin changes)
- Dehydration, hyperemesis gravidarum, ovarian hyperstimulation syndrome
- Paraplegia, immobility
- Multiple pregnancy
- Current pre-eclampsia
- Caesarean section
- Prolonged labour (>24 hours)
- Mid-cavity rotational operative delivery
- Stillbirth
- Preterm birth
- Postpartum haemorrhage (>1 L or requiring transfusion)
- Surgery in pregnancy or the puerperium

warfarin is the safest option for the pregnant person but is associated with a 5% risk of warfarin embryopathy and increased risks of miscarriage, stillbirth, and fetal intracranial haemorrhage. The alternative is full anticoagulant doses of low-molecular-weight heparin (LMWH), given twice daily and titrated upward in pregnancy, based on ***regular monitoring of anti-factor Xa*** levels. This is safe for fetuses but may offer less effective anticoagulation, conferring a risk for valve thrombosis.

Further reading

Lester W, Walker N, Bhatia K, et al. British Society for Haematology guideline for anticoagulant management of pregnant individuals with mechanical heart valves. *Br J Haematol*. 2023; 00:1–14.
Task Force for the Management of Cardiovascular Diseases during Pregnancy of the European Society of Cardiology (ESC). ESC guidelines on the management of cardiovascular diseases during pregnancy. *Eur Heart J* 2018; 39:3165–3241.

THROMBOEMBOLIC DISEASE

Pulmonary embolism (PE) is the most common direct cause of death in pregnancy and the puerperium. Pregnancy and especially the puerperium are associated with a sixfold increased risk of thrombosis. Although the risks are highest after emergency caesarean section, those with risk factors (**Box 48.6**) are at risk antenatally, especially if admitted to hospital, and after vaginal delivery.

Diagnosis

The diagnosis of deep vein thrombosis (DVT) and PE must be made in those with a suggestive history. ***D-dimers*** are not useful because concentrations are often raised in normal pregnancies. DVT is much more common in the left leg (ratio, 9:1) in pregnancy because the gravid uterus compresses the left common iliac vein, which lies under the right common iliac artery. ***Ventilation/perfusion (V/Q) scans*** and computed tomography ***pulmonary angiography*** are associated with negligible radiation to the fetus and are safe, but V/Q scans may be preferred because of reduced radiation to the breast.

Management

Treatment of VTE in pregnancy involves:

- LMWHs (e.g., enoxaparin, dalteparin, tinzaparin) continued for the rest of the pregnancy and the puerperium. Higher doses are required compared to outside of pregnancy.
- Avoidance of warfarin and direct oral anticoagulants.

Decisions regarding thromboprophylaxis in pregnancy relate to a past history of VTE, the presence of thrombophilia, and other risk factors (see **Box 48.6**). Antenatal and postnatal thromboprophylaxis with LMWH should be offered to those with a history of previous VTE and in those with significant risk factors for VTE. This should begin as early in pregnancy as possible. LMWH and warfarin are safe in lactation.

Further reading

Royal College of Obstetricians and Gynaecologists. *Green-top Guideline no. 37a: Thrombosis and Embolism during Pregnancy and the Puerperium, Reducing the Risk*. RCOG 2015; http://www.rcog.org.uk/.
Royal College of Obstetricians and Gynaecologists. *Green-top Guideline no. 37b: Thrombosis and Embolism during Pregnancy and the Puerperium: Acute Management*. RCOG 2015; http://www.rcog.org.uk/.

RESPIRATORY DISEASE

Asthma

The treatment of asthma in pregnancy does not differ from that in non-pregnant patients. Pregnancy itself does not usually influence the severity of asthma. For the majority, asthma has no adverse effect on pregnancy outcomes. Key messages for asthma in pregnancy include:

- Poorly controlled severe asthma presents more of a risk to the fetus than the medication used to prevent or treat it. Risk is minimized with good control.
- Education and reassurance, ideally before pregnancy, concerning the safety of asthma medications during pregnancy are integral parts of management.
- Decreasing or stopping inhaled anti-inflammatory therapy during pregnancy can cause a deterioration in disease control.
- Corticosteroids (inhaled, oral, and intravenous) and β_2-agonists (short- and long-acting, inhaled, nebulized, and intravenous) and the leukotriene receptor antagonist montelukast are safe to use in pregnancy and while breast-feeding. Data for omalizumab and other newer biologics used for asthma and eczema suggest no adverse effects in pregnancy, with significant benefit for maintaining disease control.
- Effective control of asthma and its accompanying symptoms is a priority. An increase in the dose or frequency of inhaled steroids should be the first step if symptoms are not optimally controlled and the inhaler technique is good.

Pneumonia

The bacteria that cause pneumonia are no different in pregnancy. Chest radiographs are safe in pregnancy. Most antibiotics are safe to use in pregnancy and during lactation, though tetracycline and doxycycline should not be given. Varicella, influenza, and COVID-19 pneumonia may be fatal in pregnancy, and active steps must be taken to prevent chickenpox infection. Non-immune pregnant females exposed to varicella in pregnancy should be given varicella zoster immunoglobulin. If a pregnant female does contract chickenpox, she should be treated with aciclovir as soon as possible. All pregnant females should receive immunization against seasonal influenza and SARS-CoV-2.

Tuberculosis

Tuberculosis (TB) is common, particularly in Africa and Asia. A chest radiograph should be performed if TB is suspected, and the advice of a respiratory physician should be sought. Extrapulmonary TB is as common as pulmonary TB in pregnancy. The diagnosis must be confirmed bacteriologically, which may necessitate bronchoscopy or biopsy, which are safe in pregnancy but is often delayed because of inappropriate reluctance to perform invasive investigations in pregnancy. The neonate should be given Bacille Calmette–Guérin vaccination.

Sarcoidosis

The course of sarcoidosis is unaltered by pregnancy. Disease flare is more likely postpartum. Erythema nodosum may occur in a normal pregnancy. Serum ACE is not reliable for the diagnosis of sarcoidosis in pregnancy. Systemic steroids should be used if indicated (see Chapter 20).

Cystic fibrosis

Females with CF often reach reproductive age with adequate lung function or will have received a lung transplant and be able to consider pregnancy. Pre-pregnancy counselling, with consideration of lung function and concomitant diabetes, is required. Joint care with a CF centre should be maintained throughout pregnancy. Pregnancy outcome is related to pre-pregnancy lung function. Perinatal outcome is usually good but preterm delivery rates are high. Maternal outcome is variable and worse in the presence of pulmonary hypertension. Specialist dietary advice with additional energy supplements should be given. Chest physiotherapy can and should be continued in pregnancy, and infective exacerbations should be treated aggressively. There is an increased risk of gestational diabetes. The advent of cystic fibrosis transmembrane conductance regulator modulators has revolutionized the lives of many people with CF, and pregnancy outcomes have improved as a result. The significant benefits of these medications mean that they are continued in pregnancy in the absence of evidence of harm.

Severe restrictive lung disease

Females with severe lung disease are better able to tolerate pregnancy than females with severe cardiac insufficiency. If the forced vital capacity is over 1 L, a successful pregnancy is usually possible, but individual assessment is necessary. Immunosuppression for diffuse parenchymal lung disease should be continued in pregnancy, with the exception of mycophenolate which is teratogenic. Respiratory diseases complicated by pulmonary hypertension have a poor prognosis in pregnancy.

NEUROLOGICAL DISEASE

Epilepsy

In most, the frequency of seizures is not altered in pregnancy, provided there is *adherence* with antiepileptic drugs (AEDs). Free drug concentrations tend to fall in pregnancy, so increased doses of AEDs, particularly lamotrigine, may be required. The risk to the fetus of uncontrolled seizures is greater than the risks of continuing AED treatment; poorly controlled epilepsy and recurrent seizures are associated with adverse outcomes for both pregnant person *and* fetus, including an increased risk of sudden unexplained death in epilepsy. Monotherapy at the minimum effective dose is preferable when this offers effective disease control.

Birth defects

The overall risk of birth defects in people without epilepsy is 3%. The risk with lamotrigine and levetiracetam is not increased. The risk with carbamazepine and topiramate is 4% to 5% and with phenytoin and phenobarbital is about 6%. Sodium valproate is associated with the highest rate of serious malformations (e.g., neural tube defects) and long-term neurodevelopmental defects (e.g., reduced IQ, increased risk of autism spectrum disorder and attention-deficit disorder), and it should be stopped and substituted if necessary before conception.

Management

Counselling before conception is essential. Folic acid (5 mg/day) supplements should be taken before conception and throughout the first trimester. Antenatal screening for congenital abnormalities is necessary. Patients with epilepsy are advised to shower rather than bathe and should not be left unattended in the peripartum period. Taking AEDs is compatible with breast-feeding, though precautions regarding breast-feeding, changing the baby at floor level, and only using shallow water to wash the baby are advised.

Contraception and epilepsy

AEDs that induce hepatic enzymes (e.g., carbamazepine, phenytoin, and phenobarbital) reduce the efficacy of oral contraceptives. A combined contraceptive pill containing a higher dose of oestrogen or the progesterone-only pill provides greater contraceptive security. An intrauterine device or progesterone implant is often recommended in preference to oral contraceptives.

Migraine

Migraine can occur as a pregnancy-related phenomenon without a prior history of migraine. Those with pre-existing migraine can improve in pregnancy. Hemiplegic migraine, particularly aura without headache, can mimic transient ischaemic attacks. Sumatriptan may be used to treat migraine. Low-dose aspirin, beta-blockers, tricyclic antidepressants, and pizotifen are used for prophylaxis. Greater occipital nerve injections are safe in pregnancy. Ergotamine should be avoided in pregnancy.

Multiple sclerosis

Pregnancy has no effect on the long-term prognosis of patients with multiple sclerosis. Relapses are less likely during pregnancy but more likely in the postpartum period. Prophylactic treatments, such as interferon-beta, glatiramer, and natalizumab, may be used in pregnancy and breast-feeding. Discontinuation of natalizumab may result in flare-ups in pregnancy and is not advised, but the injections can be spaced and interrupted in late pregnancy to avoid excessive transfer to the fetus. Vitamin D supplements of at least 4000 units daily should be given to all. Those with disability may require extra help during pregnancy and while caring for their infants after delivery. There is no contraindication to epidural anaesthesia except that documentation of a pre-existing neurological deficit in the legs is necessary to avoid any postpartum exacerbation of multiple sclerosis being inappropriately attributed to a regional block.

Myasthenia gravis

The course of myasthenia gravis in pregnancy is unpredictable, but in those with stable disease, pregnancy outcomes are usually good. Postpartum exacerbations occur in 30%. Usual immunosuppressant therapy with steroids, azathioprine, intravenous immunoglobulin, and calcineurin inhibitors should be continued. Plasmapheresis is not contraindicated in pregnancy. Increased doses of long-acting anticholinesterases may be required. Many drugs should be avoided in myasthenia gravis, and an experienced obstetric anaesthetist should be consulted. Transient neonatal myasthenia gravis may develop in up to 10% of neonates because of transplacental passage of immunoglobulin G acetylcholine receptor antibodies.

Further reading

Dobson R, Dassan P, Roberts M, Giovannoni G, Nelson-Piercy C, Brex PA. UK consensus on pregnancy in multiple sclerosis: 'Association of British Neurologists' guidelines. *Pract Neurol*. 2019; 19(2):106–114.
Norwood F, Dhanjal M, Hill M, et al. Myasthenia in pregnancy: best practice guidelines from a U.K. multispecialty working group. *J Neurol Neurosurg Psychiatry.* 2014; 85(5):538–43.

Stroke

The risks of arterial ischaemic stroke, cerebral venous thrombosis, and intracranial haemorrhage are increased, particularly in the puerperium. In pregnancy, the prevalence of haemorrhagic stroke equals that of ischaemic stroke, in contrast to outside of pregnancy, when ischaemic stroke is more common (>75% of strokes). Causes of stroke in pregnancy are listed in **Box 48.7**.

? Box 48.7 Some causes of stroke in pregnancy

Ischaemic
- Cardiac causes, e.g., arterial emboli or arrhythmias
- Peripartum or dilated cardiomyopathy
- Paradoxical embolus through an atrial septal defect or patent foramen ovale
- Aortic or carotid artery dissection
- Antiphospholipid syndrome
- Sickle cell disease

Haemorrhagic
- Pre-eclampsia or eclampsia
- Ruptured arteriovenous malformation or berry aneurysm

ENDOCRINE DISEASE

Type 1 and 2 diabetes

Diabetes is one of the most common medical conditions in pregnancy, affecting up to 5% of pregnancies. The epidemiology of pre-existing diabetes in pregnancy is changing because of trends of delaying pregnancy and the increased prevalence of obesity and type 2 diabetes at reproductive age. Whereas 30 years ago, type 2 diabetes was rare in pregnancy, the latest UK national audit has indicated that type 2 diabetes is now more prevalent than type 1 diabetes in pregnancy.

Adverse effects of diabetes in pregnancy

Pregnancy complications increase with maternal hyperglycaemia. Pre-existing diabetes may make it more difficult to conceive, with higher rates of miscarriage and congenital anomalies (twice those of the background population). The risk of congenital malformations is directly proportional to the pre-pregnancy hemoglobin A_{1c} (HbA_{1c}) concentration. Females with diabetes who are planning pregnancy are therefore advised to aim for a HbA_{1c} below 48 mmol/mol, provided it can be achieved without hypoglycaemia. Those with an HbA_{1c} greater than 86 mmol/mol are advised not to conceive because of the high risk of congenital malformation. Because congenital malformations occur as early as 5 to 6 weeks after conception, the importance of pre-pregnancy counselling cannot be understated. Pre-pregnancy counselling affords the opportunity to:

- Discuss the need for optimized glycaemic control and safe and effective contraception until this is achieved.
- Provide high dose folic acid (5 mg/day) to reduce the risk of neural tube defects.
- Assess for microvascular (retinopathy and nephropathy) complications.
- Review medication and make a plan for pregnancy including discontinuation of oral antidiabetes drugs, glucagon-like peptide-1 receptor agonists, ACE inhibitors, ARBs, and statins.
- Offer an opportunity for structured diabetes education.

Other fetal risks include macrosomia with associated shoulder dystocia and birth injury, polyhydramnios, neonatal unit admission, stillbirth, and perinatal death. Impaired postnatal metabolic adaptation leads to higher rates of neonatal hypoglycaemia, jaundice, and respiratory distress. Maternal glucose crosses the placenta, but insulin does not; this leads the fetal islets to hypersecrete insulin to combat maternal hyperglycaemia, so rebound hypoglycaemia may occur when the umbilical cord is cut. In the longer term, the offspring of pregnant persons with diabetes have a higher risk of obesity and diabetes in adulthood.

Maternal risks include increased rates of pre-eclampsia, hypoglycaemia, diabetic ketoacidosis (which occurs at lower glucose concentrations compared with outside of pregnancy and is associated with a 40% risk of fetal death), and worsening of microvascular complications.

Management of diabetes during pregnancy

Maintaining normoglycaemia without causing hypoglycaemia is the main aim of treatment. A multidisciplinary team including a physician, obstetrician, and diabetes specialist nurse or midwife undertakes a clinical review at least every 2 weeks in pregnancy.

Antenatal glucose concentration targets are less than 5.3 mmol/L fasting (95 mg/dL), less than 7.8 mmol/L (140 mg/dL) 1 hour after meals, and less than 6.4 mmol/L (115 mg/dL) 2 hours after meals. Continuous glucose monitoring increases the time in the target glucose range and is associated with improved fetal outcomes. This is recommended for females with type I diabetes in pregnancy. Closed-loop systems with insulin pumps also improve glycaemic control, and they should be offered where available. Females with type II diabetes typically monitor by capillary blood testing. HbA_{1c} is not used to monitor blood glucose control in the second and third trimesters of pregnancy.

Glycaemic control is challenging in pregnancy because insulin requirements change. In the first trimester, insulin requirement decrease as a result of hormonal changes and reduced appetite from pregnancy-associated nausea and vomiting. For reasons that are not entirely clear, early pregnancy is also associated with impaired awareness of hypoglycaemia, which increases the risk and severity of hypoglycaemia. Understanding this risk and the self-management of hypoglycaemia is important. In the second and third trimesters, pregnancy induces a state of insulin resistance, which typically increases insulin requirements by 50% to 100%. Ketoacidosis in pregnancy carries a 40% fetal mortality rate, and it should therefore be diagnosed and managed rapidly. Self-monitoring of blood ketone concentrations is taught at the outset of pregnancy.

Pregnancy can accelerate the progression of retinopathy, so screening is undertaken in the first and third trimester and in the second trimester if retinopathy is present.

Low-dose aspirin (150 mg/day) from the 12th week of pregnancy reduces the risk of preterm pre-eclampsia.

Enhanced fetal surveillance is needed, and a detailed ultrasound scan for fetal anomalies, including a four-chamber view of the fetal heart and outflow tracts, should be offered at 18 to 20 weeks' gestation. Growth scans and other tests of fetal wellbeing should be performed during the third trimester.

Most babies are born before term after an induction of labour or elective caesarean section to reduce the risk of stillbirth. Glycaemic control during labour in those with type 1 diabetes is usually maintained by a variable rate intravenous insulin infusion, but pumps are being increasingly used as an alternative.

After birth, maternal insulin requirements drop rapidly to preconception levels within hours of the birth and may even be lower if a female chooses to breast-feed. Because neonates are at increased risk of hypoglycaemia, early feeding should be encouraged and neonatal glucose checked regularly until it is clear that the pre-feeding glucose levels are being maintained.

Gestational diabetes

Gestational diabetes is defined as diabetes occurring for the first time in pregnancy, not caused by diabetes before pregnancy. It is usually diagnosed in the second and third trimester of pregnancy after a 75-g oral glucose tolerance test. Gestational diabetes occurs when pancreatic β cells are unable to secrete sufficient insulin to counter pregnancy-induced insulin resistance. Risk factors for gestational diabetes are the same as for type 2 diabetes. Gestational diabetes therefore reveal risk for the development of future type 2 diabetes. Not all diabetes presenting in pregnancy is gestational. Screening may identify previously undiagnosed type 2 diabetes. Less commonly, gestational diabetes may unmask pre-clinical type 1 diabetes or maturity-onset diabetes of the young.

Diagnosis of gestational diabetes

There is no international consensus regarding diagnosis with varying diagnostic thresholds (**Box 48.8**). Universal screening for diabetes is recommended in many countries, but in the UK, NICE recommends screening only when there are risk factors (**Box 48.9**) and considering screening when urine dipstick testing shows glycosuria of 2+ or above on 1 occasion or glycosuria of 1+ or above on two or more occasions.

Management of gestational diabetes

Glucose targets are the same as for pre-existing diabetes, but management begins with diet and lifestyle modifications including healthier carbohydrates (low glycaemic index), reduced sugar intake, normal meal portions sizes, and daily exercise. If this does not achieve adequate maternal glycaemia within 1 to 2 weeks, pharmacological therapy is indicated, typically with metformin and insulin if indicated.

HbA_{1c} testing should be undertaken 3 months after delivery to exclude undiagnosed pre-existing type 2 diabetes. Gestational diabetes confers a 50% chance of developing diabetes within 10 years of the pregnancy. This risk and the measures needed to reduce it should be discussed. Screening with an annual HbA_{1c} is recommended. At present, the uptake of postnatal interventions and screening is low, and research into more effective risk management is needed.

Box 48.8 Comparison of the NICE and IADPSG diagnostic criteria for gestational diabetes

Association	Glucose load (g)	No. of high readings needed	Fasting glucose	1-h glucose	2-h glucose
IADPSG/WHO	75	≥1	5.1 mmol/L (92 mg/dL)	10.0 mmol/L (180 mg/dL)	8.5 mmol/L 153 mg/dL)
NICE	75	≥1	5.6 mmol/L (100 mg/dL)		7.8 mmol/L (140 mg/dL)

IADPSG, International Association of Diabetes and Pregnancy Study Groups; *NICE*, National Institute for Health and Care Excellence; *WHO*, World Health Organization.

Box 48.9 Risk factors for gestational diabetes: National Institute for Health and Care Excellence (NICE) recommendations for the 75-g oral glucose tolerance test

NICE recommends undertaking a 75-g oral glucose tolerance test at 24 to 28 weeks' gestation in those with the following risk factors:

- Body mass index >30 kg/m^2
- Previous macrosomic baby (≥4.5 kg)
- Previous gestational diabetes
- First-degree relative with diabetes
- Family origin with a high prevalence of diabetes, e.g., South Asian, Black Caribbean, Middle Eastern

Further reading

International Association of Diabetes in Pregnancy Study Groups Consensus Panel. International association of diabetes in pregnancy study groups recommendations on the diagnosis and classification of hyperglycemia in pregnancy. *Diabetes Care* 2010; 33:676–682.
International Diabetes Federation. *Managing Older People with Type 2 Diabetes*. Global guideline; https://www.idf.org/e-library/guidelines/78-global-guideline-for-managing-older-people-with-type-2-diabetes.html.
Joint British Diabetes Societies. *Management of Adults with Diabetes Undergoing Surgery and Elective Procedures: Improving Standards*; https://www.diabetes.org.uk/resources-s3/2017-09/Use%20of%20variable%20rate%20intravenous%20insulin%20infusion%20in%20medical%20inpatients_0.pdf.
Joint British Diabetes Societies. *The Use of Variable Rate Intravenous Insulin Infusion (VRIII) in Medical Inpatients*; https://www.diabetes.org.uk/resources-s3/2017-09/Use%20of%20variable%20rate%20intravenous%20insulin%20infusion%20in%20medical%20inpatients_0.pdf.
National Institute for Health and Care Excellence. *Diabetes in Pregnancy: Management from Preconception to The Postnatal Period*; https://www.nice.org.uk/guidance/ng3/resources/diabetes-in-pregnancy-management-from-preconception-to-the-postnatal-period-pdf-51038446021.

Thyroid and parathyroid disease

Thyrotoxicosis

Untreated thyrotoxicosis is dangerous for both the pregnant person and fetus, increasing the risk of miscarriage, fetal growth restriction (FGR), and preterm delivery. Graves' disease affects about 0.2% of pregnancies and often improves during pregnancy, but there is a risk of postpartum flare. Both carbimazole and propylthiouracil (PTU) cross the placenta and in high doses may cause fetal hypothyroidism and goitre. The lowest possible maintenance dose of anti-thyroid drug should be used. Guidelines advise avoidance of carbimazole in the first trimester because of the risk of teratogenesis. PTU is usually not given because of the risk of liver failure in the general population but may be used specifically in those planning pregnancy or in the first trimester. Switch from one anti-thyroid drug to another is not indicated before or during pregnancy because of the risk that switch compromises disease control, with no evidence for improved outcomes. For those with disease controlled with dosages of carbimazole of less than 15 mg/day or PTU less than 150 mg/day, pregnancy outcomes are usually good. Breast-feeding is safe on these doses of anti-thyroid drugs. Beta-blockers are also safe if required for thyrotoxic symptoms. Neonatal or fetal thyrotoxicosis caused by transplacental passage of thyroid-stimulating antibodies is rare but occurs most commonly in those with active disease in pregnancy and those with high concentrations of thyroid-stimulating antibodies.

Hypothyroidism

Hypothyroidism affects about 1% of pregnancies. Severe untreated hypothyroidism is associated with infertility, an increased rate of miscarriage, and fetal loss. For those on adequate replacement therapy, maternal and fetal outcomes are usually good and unaffected by hypothyroidism. Very little thyroxine crosses the placenta, and the fetus is not at risk of thyrotoxicosis from maternal thyroxine replacement therapy. Those who are euthyroid at the beginning of pregnancy usually do not require any adjustment to thyroxine dose during pregnancy or in the puerperium. Dose changes should be based on thyroid function tests, interpreted using pregnancy-specific reference intervals (or thyroid-stimulating hormone [TSH] <4.0 mU/L after 7 weeks' gestation). There is no evidence to support older recommendations for a target TSH of less than 2.5 mU/L in pregnancy.

Subclinical hypothyroidism affects 5% of the general population and is more common in females, particularly those who have antithyroid antibodies. Studies fail to demonstrate a consistent association between any adverse pregnancy outcome and subclinical hypothyroidism in pregnancy, and there is insufficient evidence that thyroxine replacement in subclinical hypothyroidism benefits pregnancy outcome. However, most authorities advocate thyroxine replacement if the TSH is greater than 4.0 mU/L.

Postpartum thyroiditis has a prevalence of 5% to 7% and is more common in those with a family history of hypothyroidism, thyroid peroxidase antibodies, and type 1 diabetes. Presentation is usually between 3 and 4 months postpartum. Disease may present with symptoms of hyper- or hypothyroidism, but a high index of suspicion is needed. The condition is caused by a destructive autoimmune lymphocytic thyroiditis. In most cases, treatment is not required, and recovery is complete, but in a small proportion, progression to permanent hypothyroidism will occur. Postpartum thyroiditis often recurs and is a significant predictor of future hypothyroidism.

Hyperparathyroidism

Hypercalcaemia may be improved by pregnancy and the fetal demand for calcium. Risks include acute pancreatitis and hypercalcaemic crisis, especially postpartum, when maternal transfer of calcium to the fetus stops abruptly. There are also increased risks of miscarriage, intrauterine death, and preterm labour. Fetal mortality rates are up to 40% when the maternal hypercalcaemia is severe (>3.5 mmol/L). Up to 25% of pregnancies complicated by hyperparathyroidism develop hypertension or pre-eclampsia. Suppression of fetal parathyroid hormone by high maternal calcium concentrations confers a risk of neonatal tetany and hypocalcaemic convulsions secondary to hypocalcaemia, usually presenting 5 to 14 days after birth but delayed by up to 1 month if the infant is breast-fed. The ideal treatment is surgical removal of the parathyroid adenoma. This may be safely performed in pregnancy, usually in the second trimester, which removes the maternal and neonatal risks associated with hypercalcaemia.

Further reading

Alexander EK, Pearce EN, Brent GA et al. Guidelines of the American Thyroid Association for the diagnosis and management of thyroid disease during pregnancy and the postpartum. *Thyroid* 2017; 27:315–389.

Pituitary disease

Prolactinomas

Prolactinomas are the most common pituitary tumours encountered in pregnancy but rarely cause problems. The prolactin concentration should not be measured during pregnancy or breast-feeding because it is invariably raised. Those with microprolactinomas usually have no problems in pregnancy. The risk of tumour enlargement

during pregnancy is increased with macroprolactinomas. Dopamine receptor agonists are safe for use in pregnancy and are usually continued for macroprolactinomas. Visual fields should be measured in each trimester in those with macroprolactinomas. If tumour enlargement is suspected, magnetic resonance imaging (MRI) of the pituitary is indicated. If those with macroprolactinomas wish to breast-feed, their dopamine receptor agonists may be discontinued in the last few weeks of pregnancy and during breast-feeding, but serial visual field examinations and pituitary MRI are required, and cabergoline or bromocriptine should be re-introduced if there is concern regarding tumour expansion.

Diabetes insipidus

Established or subclinical diabetes insipidus may worsen in pregnancy because of increased vasopressinase production by the placenta or decreased vasopressinase breakdown by the liver. Extended fluid deprivation tests should be avoided in pregnancy. Close observation with paired urine and plasma osmolality measurements may be sufficient to exclude DI. Desmopressin is safe for use in pregnancy for diagnosis or treatment of DI. Transient DI may occur in pregnancy often associated with acute fatty liver of pregnancy and, more rarely, with pre-eclampsia and HELLP syndrome.

Hypopituitarism

Hypopituitarism is rare in pregnancy but may be caused by:

- Pituitary surgery
- Radiotherapy
- Pituitary or hypothalamic tumours
- Pituitary haemorrhage
- Postpartum pituitary infarction (Sheehan's syndrome)
- Lymphocytic hypophysitis

Sheehan's syndrome usually presents after delivery after postpartum haemorrhage and may lead to partial or complete pituitary failure. There may be a delay in diagnosis because the symptoms may be attributed to the postpartum state.

Lymphocytic hypophysitis is an uncommon autoimmune disorder caused by extensive infiltration of the anterior pituitary by chronic inflammatory cells, predominantly lymphocytes, causing pituitary expansion. Most patients present in the last month of pregnancy or the first 2 weeks postpartum. It presents with features suggestive of an expanding pituitary tumour, and 85% of cases have endocrine hypofunction.

Adrenal disease

Conn's syndrome and hyperaldosteronism

This is rare in pregnancy, but hypertension that is difficult to control, the absence of a positive family history, and pre-pregnancy hypokalaemia are indications for screening. Renin and aldosterone concentrations are increased in pregnancy, and diagnosis requires the use of pregnancy-specific reference ranges, which are not commonly available. High concentrations of progesterone in pregnancy may antagonize aldosterone and ameliorate the hypokalaemia. Screening is therefore usually undertaken postpartum. Hypertension is controlled using labetalol, modified-release nifedipine, and/or methyldopa. Patients with hypokalaemia are treated with potassium supplementation or potassium-sparing diuretics. Amiloride is safe to use in pregnancy, and high doses (e.g., 20 mg) may be needed. Spironolactone should be avoided because of the risk of feminization of a male fetus because of antiandrogen effects. Surgery for adrenal adenomas is usually deferred until after delivery.

Phaeochromocytoma and paraganglioma

Phaeochromocytoma and paraganglioma are rare but dangerous causes of hypertension in pregnancy. Those with hypertension associated with unusual features of palpitations, anxiety, sweating, or headache should be screened. Management involves adequate α-blockade with phenoxybenzamine or doxazosin to control hypertension followed by β-blockade, if required, to control tachycardia. α-Blockade is imperative before delivery. Surgical removal is the only cure, and optimal timing of tumour resection depends on the gestational age when the diagnosis is made. There is an increasing vogue to delay tumour resection until the puerperium provided control with medical management is possible.

Further reading

Nana M and Williamson C. *Pituitary and Adrenal Disorders of Pregnancy*. Endotext. October 10 2022.

KIDNEY DISEASE

Urinary tract infection

Asymptomatic bacteriuria is detected in 2% to 6% of pregnancies, with concern that this increases the risks of pyelonephritis and preterm delivery. However, treatment of asymptomatic bacteriuria is based on historical data, with no contemporary randomized control data to guide practice. Symptomatic bacteriuria is always treated, regardless of symptoms.

Chronic kidney disease

Females with CKD are at increased risk of pre-eclampsia, FGR, preterm delivery, neonatal death, and the risk of permanent deterioration in kidney function. The risks of preterm delivery are increased by the presence of hypertension.

In females with ***moderate or severe kidney impairment*** (pre-pregnancy CKD stage 3–5, eGFR <60 mL/min/1.73 m^2), 20% to 90% of infants are born preterm, and there is a risk of accelerated decline in kidney function equivalent to 2 to 5 years of kidney disease in the absence of pregnancy. An increase in the degree of proteinuria is very common in pregnancy and does not necessarily imply pre-eclampsia or worsening kidney disease.

Management should include regular monitoring of BP, kidney function, and fetal wellbeing. In view of the increased risk of pre-eclampsia, treatment is with low-dose aspirin (150 mg/day from week 12–36 weeks).

Kidney transplantation

If graft function is normal, pregnancy outcomes are excellent, and there is no adverse long-term effect on kidney allograft function or survival. The chance of successful pregnancy outcome is reduced and the risk of long-term deterioration in graft function increased with poor baseline graft function. Pregnancy outcome is optimal in those without hypertension, proteinuria, or recent episodes of graft rejection.

Doses of immunosuppressive drugs should maintain pre-pregnancy concentrations. Prednisolone, azathioprine, ciclosporin, and tacrolimus are safe for use in pregnancy without any reported increase in risk of congenital malformations. Mycophenolate mofetil and sirolimus are contraindicated. Prophylactic antibiotics should

be given if there is urinary tract infection in pregnancy. The risks of pre-eclampsia, graft rejection, FGR, preterm delivery, and infection are increased. Caesarean section is required for obstetric indications, but the rate is increased.

Acute kidney injury

Because serum creatinine concentration falls in pregnancy, AKI is defined as an abrupt deterioration in kidney function with a serum creatinine of more than 77 μmol/L. Mild degrees of AKI are not uncommon, particularly in the immediate postpartum period. The most common underlying causes are:

- ***Infection:*** septic abortion, puerperal sepsis, acute pyelonephritis
- ***Blood loss:*** postpartum haemorrhage, placental abruption
- ***Volume contraction:*** pre-eclampsia, hyperemesis gravidarum, diarrhoea
- ***Renal tract obstruction:*** ureteric damage or obstruction
- ***Drugs:*** non-steroidal anti-inflammatory drugs (NSAIDs), antibiotics (e.g., aminoglycosides)

Females with pre-eclampsia are particularly prone to pulmonary oedema; therefore, management protocols include fluid restriction, even in the presence of oliguria. If these patients develop AKI, this is usually rapidly reversible on recovery and less dangerous than pulmonary oedema, which may result from repeated fluid challenges.

SKIN DISORDERS

Common skin changes that occur in pregnancy include striae, spider naevi, melasma (pigmentation on the face), and linea nigra (midline pigmentation on the abdomen). A shift in maternal immune profile from predominantly Th1 to Th2 may underlie the tendency for atopic dermatitis to develop or worsen during pregnancy. Psoriasis usually improves. Up to 20% of pregnant females develop pruritus, which is most common on the lower legs. Intrahepatic cholestasis (see 'Intrahepatic cholestasis of pregnancy' section) of pregnancy should be excluded by measuring bile acid concentrations.

The specific dermatoses of pregnancy are classified as shown in **Box 48.10**.

Further reading

Vaughan Jones S, Ambros-Rudolph C, Nelson-Piercy C. Skin diseases in pregnancy. *Br Med J* 2014; 348:g3489.

RHEUMATIC DISEASE

Systemic lupus erythematosus and rheumatoid arthritis are the commonest rheumatic diseases encountered in pregnancy. Active inflammatory disease reduces fertility, and active disease at the time of conception increases the risk of miscarriage, pre-eclampsia, FGR, and preterm delivery. Therefore, the emphasis should be on achieving good disease control before conception and throughout pregnancy. Up to 50% of females with rheumatoid arthritis go into remission in pregnancy, but 50% flare during the postpartum period. The risk of lupus flare is increased in pregnancy. Many disease-modifying anti-rheumatic drugs are compatible with pregnancy. Females should take advice from their rheumatologists, obstetric physicians, and obstetricians ***before*** conception. **Box 48.11** shows the use of drugs for rheumatoid arthritis in pregnancy.

Further reading

Götestam Skorpen C, Hoeltzenbein M, Tincani A et al. The EULAR points to consider for use of antirheumatic drugs before pregnancy, and during pregnancy and lactation. *Ann Rheum Dis* 2016; 75:795–810.
Russell MD, Dey M, Flint J et al. BSR Standards, Audit and Guidelines Working Group. British Society for Rheumatology guideline on prescribing drugs in pregnancy and breastfeeding: immunomodulatory anti-rheumatic drugs and corticosteroids. *Rheumatology (Oxford)*. 2023;62(4):e48–e88. Erratum in: Rheumatology (Oxford). 2022 Dec 13.

Box 48.10 Classification of specific dermatoses of pregnancy

Dermatosis	Description	Management
Atopic eruption of pregnancy Atopic eczema Prurigo of pregnancy Pruritic folliculitis of pregnancy	Usually starts on the abdomen in the third trimester but may persist for some months after delivery Clustered excoriated papules (prurigo-like lesions) occur on the abdomen and extensor surfaces of the limbs Cause is unknown Can recur in subsequent pregnancies Itchy folliculitis that looks similar to steroid-induced 'acne' Not associated with any increased risk	Topical steroids and oral antihistamines Topical benzoyl peroxide Topical hydrocortisone cream
Polymorphic eruption of pregnancy (pruritic urticated papules of pregnancy)	A relatively common, intensely itchy rash that starts on the abdomen, often within striae, in the third trimester Associated with multiple births Uncomfortable but harmless	Symptomatic topical therapy and antihistamines are helpful until after delivery, when the rash resolves
Pemphigoid gestationis	A rare disease analogous to bullous pemphigoid (see Chapter 45) Associated with prematurity and stillbirth; transplacental passage of pathogenic immunoglobulin G antibodies can lead to transient blistering in the neonate Often starts around the umbilicus with itchy, inflamed papules; then blistering appears and becomes generalized Onset in the first or second trimester; typically flares postpartum Diagnosis confirmed with direct immunofluorescence Recurrences may occur with subsequent pregnancies and oral contraceptive pill, which should be avoided	Systemic steroid therapy is required

Box 48.11 Drug use during pregnancy in treatment of patients with rheumatoid arthritis and systemic lupus erythematosus

- ***Paracetamol:*** the oral analgesic of choice
- ***Oral NSAIDs and selective COX-2 inhibitors:*** can be used short term up until 30 weeks' gestation if symptoms justify their use
- ***Corticosteroids:*** may be used to control disease flares (main risks are hypertension, glucose intolerance, infection, and preterm premature rupture of membranes)
- ***DMARDs***
  - ***Considered safe:*** sulfasalazine, hydroxychloroquine, azathioprine, ciclosporin, tacrolimus
  - ***Unsafe:*** mycophenolate mofetil, methotrexate, leflunomide, and cyclophosphamide (in the first trimester)
  - ***Biological agents:*** good safety data during pregnancy for infliximab, adalimumab, etanercept, and certolizumab pegol
  - Safety for newer agents (e.g., belimumab, tocilizumab, abatacept) not yet established; used if benefit outweighs risk
- ***Contraindicated in breast-feeding:*** methotrexate, leflunomide, cyclophosphamide, mycophenolate mofetil

COX, Cyclo-oxygenase; *DMARD,* disease-modifying anti-rheumatic drug; *NSAID,* non-steroidal anti-inflammatory drug.

PRESCRIBING IN PREGNANCY

Many clinicians are reluctant to prescribe drugs in pregnancy. This relates mainly to concern regarding teratogenic risk. The following general principles should be remembered:

- ***Older generic drugs.*** Consider older generic drugs within each class because there are likely to be more data on use in pregnancy.
- ***Dosages.*** Resist the temptation to prescribe lower doses. Higher doses may be required because of increased kidney and liver clearance.
- ***Disease control.*** Appropriate drug therapy to control underlying diseases, including arthritis, inflammatory bowel disease, lupus, epilepsy, asthma, and thyrotoxicosis, is likely to improve fetal and neonatal outcomes, including reductions in FGR and preterm deliveries.
- ***Treatment for unfamiliar diseases.*** Ask the specialist physician, 'What would you do if this female was not pregnant'? Then assess the risks of this strategy in pregnancy.
- ***Risks to the fetus.*** These must be balanced against potential benefits to the pregnant person (and therefore indirectly to the fetus).

The ***teratogenic potential*** of some drugs classified as ***absolutely contraindicated*** is sufficiently high to justify discussion of termination of a pregnancy after inadvertent exposure; these drugs include mycophenolate mofetil, sodium valproate, and thalidomide. For others, there are theoretical reasons to avoid their use in pregnancy, but they carry a low risk of teratogenesis, so there is no justification for termination (e.g., rubella vaccine, simvastatin, ACE inhibitors).

For drugs that are ***relatively contraindicated***, there are situations in which use is appropriate when no safer alternative exists (e.g., warfarin in prosthetic heart valves or AEDs in epilepsy). ***Diuretics*** are avoided in the treatment of patients with hypertension but are appropriate in the treatment of those with pulmonary oedema.

Further reading

National Institute for Health and Care Excellence. *NICE Clinical Guideline 192: Antenatal and postnatal mental health: Clinical management and service guideline*. NICE 2018; https://www.nice.org.uk/guidance/cg192/chapter/1-Recommendations.

Questions for this chapter can be found on Elsevier eBooks+.

49 Medicine of the older adult

Adam Harper, Iain Stewart Ayrton Wilkinson, and Joanna Preston

Background 1694
Ageing 1695
Why do we age? 1695
Theories related to limited cell replication 1695
Theories related to evolution 1695
Calorie restriction 1696
Cell senescence 1696
Applied anatomy and physiology of ageing 1696
Clinical approach to the older adult 1696
Frailty, multimorbidity, and disability 1696
Frailty 1697
Multimorbidity 1697
Sarcopenia 1698
Investigations in older adults 1699
Comprehensive geriatric assessment 1699
Multidisciplinary team working 1703
Advance care planning 1704
Common issues facing older people 1704
Polypharmacy 1704
Falls and instability 1705
Postural hypotension 1710
Delirium 1710
Urinary tract infections 1712
Urinary incontinence 1714
Complex discharge planning 1715
Important considerations 1715
Social and functional support 1717
Ageing well and advantages of ageing 1719
Societal and economic advantages of an active older population 1719
Health and wellbeing 1719

CORE SKILLS AND KNOWLEDGE

Geriatric medicine focuses on the medical care of older people. Increasingly, the criteria for specialist geriatric care are defined by the presence of multiple problems or frailty rather than chronological age. Medical care is provided within a multidisciplinary team (MDT).

Geriatricians cover the range of general medicine in older adults from acute hospital settings (emergency departments and specialist wards) to the community (rehabilitation wards and patients' homes). Geriatricians support other specialists with specific issues such as the management of multimorbidity, delirium, and complex discharge planning. Subspecialty work includes falls, orthogeriatrics, stroke medicine, and Parkinson's disease, along with evolving areas such as oncogeriatrics and surgical liaison.

Key skills in geriatric medicine include:

- Managing the complexity inherent in ageing physiological systems with interacting diseases and treatments
- Making pragmatic and holistic decisions for individual patients in the face of uncertainty, including understanding and individualizing the evidence base
- Working with allied health professionals to create holistic multidisciplinary plans to optimize patients' function and independence
- Caring and advocating for the most vulnerable and frail adult patients.

BACKGROUND

There have been incredible changes in human life expectancy over the past 200 years, with increasing numbers of people enjoying fulfilling lives up until their 90s and beyond. These changes reflect numerous factors from improved sanitation and nutrition to the development of antibiotics. Preventative measures, such as reductions in cardiovascular disease through blood pressure (BP) control, can postpone the onset of disease. Many diseases that previously killed in middle age are now treatable or even curable, including many cancers and chronic infections such as human immunodeficiency virus (HIV). In turn, many people are living into older age with one or more chronic conditions, which may require ongoing treatment and intervention to reduce negative impacts on the individual's quality of life. These interventions may have their own side-effects, and ageing itself has an impact on the body and its function, requiring a specific approach.

In the 1960s, Bernard Isaacs described four ***'geriatric giants'***, a fifth being added later as medical interventions progressed:

1. Incontinence
2. Immobility
3. Intellectual impairment
4. Instability
5. Iatrogenesis (when patients may experience harm as a result of medical interventions)

More recently, ***'5 Ms'*** have been proposed by Mary Tinnetti (Box 49.1).

Presentations of illnesses often differ in older age. For example, the physical symptoms and signs associated with depression are too common and non-specific to be easily differentiated from organic causes. A diagnosis of depression may be signalled by reduced social interaction and by the fact that the patient does not leave their home when this is unexplained by any co-morbidities or physical disability.

Box 49.1 The five Ms of geriatric medicine

- ***Mind:*** mentation, dementia, delirium, depression
- ***Mobilization:*** impaired gait and balance, fall injury prevention
- ***Medications:*** polypharmacy, optimal prescribing and deprescribing, adverse drug reactions
- ***Multicomplexity:*** multimorbidity, complex bio- or psychosocial situations
- ***Matters most:*** each individual's own meaningful health outcome goals and care preferences

Adapted from Tinetti M, Huang A, Molnar F. The geriatrics 5M's: A new way of communicating what we do. J Am Geriatr Soc *65(9): 2017.*

Investigation and interpretation of results must also be balanced against pragmatic considerations. For example, finding a pulmonary nodule in a frail person should prompt the questions, 'If this were malignant, what would the patient want? Are they fit for surgery or chemo- or radiotherapy? Will their present co-morbidities and disabilities allow them to tolerate further investigations? Will it increase their life expectancy'? Another common scenario is a person who is experiencing postural hypotension because of antihypertensives. It may be better to accept a higher BP than guidelines would suggest to stop the patient's falling or living in fear. These decisions must be discussed with the patient as part of shared decision-making, but finding a balance can prove difficult for patient and clinicians alike. At its heart, geriatric medicine is the interplay between the patient, the medical teams, and the declining physiology of the ageing process.

AGEING

Ageing is a progressive physiological decline associated with living beyond roughly the age of 30 years. Bernhard Strehler defined ageing as:

- ***Universal:*** The process must happen to all members of a species.
- ***Intrinsic:*** It takes place via endogenous processes and does not depend on extrinsic factors.
- ***Progressive:*** Changes leading to ageing must occur progressively throughout a person's lifespan.
- ***Deleterious:*** It is bad for the individual.

Virtually all physiological mechanisms lose efficiency with ageing and lose the capacity to maintain homeostasis when faced by external challenges. Whereas a younger person stumbling over a kerb would have the strength and reflexes to adjust and maintain their balance, an older person is more likely to fall.

Ageing happens at differing rates between individuals. Some aspects are easily apparent, such as the hair turning grey. Other aspects are less obvious, such as accumulation of atheromatous plaques within blood vessels. Ageing eventually occurs in all of us but can be influenced by a number of factors, from genetic make-up to physical activity levels and the external environment. Some factors are associated with accelerated ageing, such as smoking and diabetes. Others are protective and slow the ageing process, such as a balanced diet and regular exercise.

Why do we age?

Theories of ageing can broadly be split into two: those concerned with the limited replication of cells and those related to evolution.

Theories related to limited cell replication

Telomere shortening

With each cell replication, a small piece of the telomere is removed until it eventually becomes too short to continue dividing without risk of damage to the chromosome. At this stage, it is no longer able to prevent errors in DNA replication, which may lead to the development of cancer.

Damage accumulation theory

Ageing is a result of fault accumulation at cellular and molecular levels caused by a limitation of maintenance and repair mechanisms. The primary problems do not lie with the cells themselves but with the regulation of mechanisms keeping the cell division process efficient and safe.

Free radical theory

Free radicals, derived from chemical reactions involving oxygen molecules (oxidation), occur more frequently in older cells. Species with high levels of free radicals have shorter life expectancies. Free radicals are balanced by antioxidants, which protect against oxidative damage to proteins, DNA, and lipids, all of which are vital for normal cell function.

The mitochondrial theory of free radicals in ageing suggests that damage by reactive oxygen species causes oxidative stress that overwhelms the antioxidant cellular defences, leading to ***cellular senescence*** (see later). Chronic oxidative stress throughout the lifespan plays a critical role in ageing overall. There is some evidence that antioxidants can reduce this damage, but no clinically relevant outcomes have been observed from their use.

Theories related to evolution

Evolutionary theories are based on the idea that the biological processes associated with ageing provide an advantage earlier in life in terms of reproduction and survival of the species. Because humans now regularly survive beyond reproductive age, the negative biological changes for the individual that occur later in life may be seen as payback for this group advantage.

Disposable soma theory

This suggests that an organism must budget for the limited energy available to it and preferentially focus on reproduction rather than maintaining the mature body (or soma, i.e., the rest of the body that is not directly involved with reproduction), leading to inefficiencies in cell replication later in life, when the soma is no longer needed because the organism is beyond reproductive age.

Mutation accumulation theory

This theory supports the disposable soma idea by asserting that DNA mutations taking place after reproductive age cannot affect future generations, meaning that there is no evolutionary imperative to correct them or develop protection against consequent defects.

Antagonistic pleiotropy theory

Pleiotropy is when a single gene can have two different effects. The antagonistic pleiotropy theory takes the idea one step further, suggesting that genes that are protective or beneficial in early life become detrimental (i.e., antagonistic) later in life. An example is sickle cell disease, which is protective against malaria in earlier life.

Calorie restriction

Calorie restriction is thought to prolong life expectancy, mediated through reduction of oxidative damage and anti-inflammatory benefits. This is supported by animal studies and by the example of the population of the Japanese island of Okinawa, where there are cultural calorie restriction and the highest prevalence of centenarians worldwide.

Cell senescence

Cell senescence, when cells stop dividing but remain metabolically active, leads to a pro-inflammatory state. Cells with senescent properties are seen in the tissues of those with age-related diseases, such as osteoarthritis, pulmonary fibrosis, atherosclerosis, and Alzheimer's disease. Exactly what the mechanism is that leads cells to become senescent is unclear.

Senescence is different from apoptosis, which is programmed cellular death followed by removal of the cell from the system in an organized way. Senescent cells are resistant to apoptosis and may instead continue to remain metabolically active, secreting pro-inflammatory cytokines, or may undergo necrosis instead, which is in itself pro-inflammatory.

Senescence is not just seen in the context of ageing. Another example is wound healing. After a wound is healed, the myofibroblast cells go into acute senescence to prevent excessive fibrosis at the site. Senescence also occurs when the blood supply for an embryo is being built. In these circumstances, it is a programmed process, triggered by specific stimuli and targeting specific types of cells. In ageing, the process appears unscheduled and more random. Clearance of aged cells that have undergone senescence is much less efficient. This may be attributable to the reduced immune function associated with ageing.

APPLIED ANATOMY AND PHYSIOLOGY OF AGEING

One of the key aspects of providing medical care for older people is distinguishing between normal ageing and disease. Ageing affects each organ differently. Presentations that probably arise as single-organ pathology in younger adults are likely to have more than one cause in older adults. For example, anaemia in a young woman is likely to be related to menstrual bleeding. In an older woman, anaemia might represent chronic disease, such as myelofibrotic bone marrow that is unable to respond quickly, exacerbated by a poor nutritional state with low iron and folate levels. Older adults need to have their physical conditions assessed and managed alongside their functional and social circumstances.

What commonly differentiates people with the same apparent condition(s) is the ability of their whole system to balance these conditions in a way that keeps the person functioning. This difference is usually explained by the presence of frailty as an underlying physiological process.

As we age, our organs function less well, with the percentage function of each individual system reducing by the order of 0.5% to 1% per organ system per year from the age of 30 years, as demonstrated by changes in respiratory function in a large international cohorts (Fig. 49.1). Similar graphs can be drawn for maximum usage of oxygen (VO_2 max), cardiac output, maximum heart rate, or renal function.

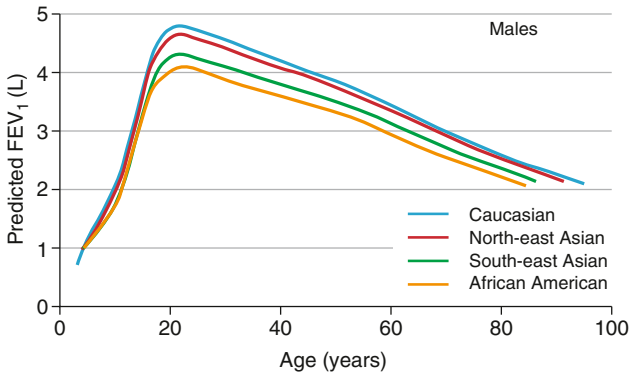


Fig. 49.1 Changes in forced expiratory volume in 1 second (FEV_1) with age. *(From Stanojevic S, Quanjer P, Miller MR et al. The Global Lung Function Initiative: dispelling some myths of lung function test interpretation.* Breathe *2013; 9:462–474.)*

Further reading

Pellegrino ED, Thomasma DC. *The Virtues in Medical Practice*. Oxford, New York: Oxford University Press; 1994.
Strehler BL. *Time, Cells and Aging*, 2nd edn. New York: Academic Press; 1977.
Wilkinson I, Preston J. MDTea podcasts—*Episode 4.01 Theories of ageing.* https://thehearingaidpodcasts.org.uk/episode-4-01-theories-of-ageing/.

CLINICAL APPROACH TO THE OLDER ADULT

When symptoms can be non-specific and many conditions present in an atypical manner, it can be difficult to know where to start. Asking open questions such as, 'What is your biggest concern today'? can help. In those with confusion, timelines of presentation may not be reliable, but you can usually get a feel for immediate concerns, such as pain or breathlessness. Do not be afraid to ask about difficult issues (e.g., bladder and bowel dysfunction, death and dying); patients often want to explore these areas but are hesitant to bring them up themselves.

A key aspect of geriatric medicine is the ability to think in detail about the current issue while maintaining a longer-term, bigger picture of the person and their life. Geriatricians often manage patients with interacting multimorbidity, polypharmacy, and physiological decline. They need the ability to focus on aspects that might be troubling the patient (e.g., constipation or a dry mouth) while being aware of overall progressive decline, reservations about interventions, and the potential need for advanced care planning.

FRAILTY, MULTIMORBIDITY, AND DISABILITY

These are distinct entities that often coexist. The difference is important in the management of patients.

- ***Frailty*** describes a risk of a deterioration when faced with a physiological stressor.
- ***Multimorbidity*** involves two or more medical conditions that may or may not interact.
- ***Disability*** is physical or mental impairment that has a substantial and long-term negative effect on a person's ability to carry out normal daily activities.

Frailty

Frailty is 'a state of increased vulnerability to stressors due to age-related declines in physiologic reserve across neuromuscular, metabolic, and immune systems'. It is the effect of systemic ageing (the underlying biological ageing process) rather than chronological age. It is possible to be chronologically old and not be frail and likewise to be young and very frail. As an individual organ's ability to react to a physiological stressor reduces with age, the impact of a given illness on that organ system has a greater effect. A minor infection (e.g., an uncomplicated urinary tract infection [UTI]) has little impact on a young and fit patient but may cause serious delirium and morbidity in those with long-term cognitive impairment, especially when combined with frailty or serious multimorbidity.

Identifying frailty

A number of models have been developed to describe frailty and the manner in which it evolves. Frailty is relatively simple to recognize when it is severe but more challenging when mild. Its recognition ensures that patients are managed according to their functional level rather than their chronological age.

Frailty recognition is an effective way of identifying people who may be at greater risk of future hospital admission, care home placement, or death; those with severe frailty have more than a four times greater annual risk for these outcomes than individuals of the same biological age who are not frail. Interventions targeted at actively managing frailty around a stressor, such as elective surgery, show better outcomes.

The key benefits to patients of a definition of frailty and an active search for it are twofold:

- It identifies those who might benefit from an objective detailed and evidence-based assessment (Comprehensive Geriatric Assessment [CGA]; see later)
- It acts as a trigger for the point at which medicine stops following standardized protocols and becomes a balance between the two questions that Edmund Pellegrino termed 'prudential': 'What *can* be done for this patient'? and 'What *should* be done for this patient'?

Identification and stratification of population-level frailty could assist in planning for future demands on health and social care while allowing targeted interventions to help people age well. Upskilling all those working with older adults is a priority. This runs in parallel with geriatric teams supporting specialist services, from perioperative care to specific patient populations in areas such as oncogeriatrics and HIV medicine.

Two main theories can also be used to identify frailty—frailty as a phenotype and frailty as an accumulation of deficits.

Frailty phenotype

In 2001, Linda Fried published her model of frailty and demonstrated that becoming frail was associated with negative outcomes in terms of length of life and institutionalization (**Box 49.2**).

Frailty index

The Frailty Index (FI) defines frailty from a list of medical conditions, physiological parameters, and observations on factors that become more common with increasing age and are associated with negative health outcomes. This method has been used largely in research, but now with the development (in the UK) of the e-Frailty Index, it can be run on routinely collected data from general practitioners' (GPs') records.

Box 49.2 Fried's phenotypic frailty assessment[a]

Domain	Suggested assessment
Weight loss	Loss of ≥4.5 kg or ≥5% of body mass in the previous year
Gait speed	6 sec or more to walk 4 m (0.66 m/sec) (slowest 20% by gender, height)
Energy expenditure	Low levels of energy expenditure or physical activity in lowest 20% of population (males, <383 kcal/week; females, <270 kcal/week)
Subjective feeling of lethargy	Subjective feeling of exhaustion: Felt that everything they did was 'an effort' in the past week Could not 'get going' in the past week
Muscle strength	Grip strength as a score within the lowest 20%, stratified by gender and body mass index

[a]Three or more define frailty.

From Fried LP, Tangen CM, Walston J et al., Cardiovascular Health Study Collaborative Research Group. Frailty in older adults: evidence for a phenotype. Gerontol A Biol Sci Med Sci *2001; 56(3):M146–156.*

The Clinical Frailty Scale

The Clinical Frailty Scale (CFS) correlates with the more in-depth FI. It uses clinical descriptors and images to stratify older adults according to level of vulnerability. It has been validated in a number of clinical settings but only after a CGA has been carried out. It is increasingly being used for identification purposes, however, because of its ease of use (**Fig. 49.2**).

Multimorbidity

Older people are more likely to have multiple medical issues. By 2035, around two-thirds of people older than 65 years of age will have multimorbidity. Within the UK, for example, such patients currently account for 50% of general practitioner consultations, 55% of hospital admissions, and 80% of prescriptions. Each condition often requires a number of medications, with treatments potentially interacting adversely. This can lead to ***prescribing cascades***, whereby medications are given for one condition and then further medications are required to counter the side-effects (e.g., the co-prescription of a proton pump inhibitor with aspirin).

Meanwhile, traditional healthcare models were usually designed around single-organ pathology. For example, a patient is admitted under cardiology for their acute coronary syndrome and treated well for this, with a selection of medications to reduce the risk of further cardiac events. However, angiotensin-converting enzyme (ACE) inhibitors may cause a cough, which might aggravate their stress incontinence, which could then increase their frequency of falls, and the beta-blockers may aggravate their longstanding depression. A more holistic approach could be to be more selective in medication choices rather than automatically starting all the usual medications given after a cardiac event and focusing on patient priorities and goals for optimizing their quality of life. This would include a discussion with the person about the benefits and harms from each while also involving the rest of the MDT for further ways optimize their independence, for example, to reduce their risk of falls and explore ways to counter depression.

Rockwood Frailty Scale

Clinical Frailty Scale

1 Very Fit – People who are robust, active, energetic and motivated. These people commonly exercise regularly. They are among the fittest for their age.

2 Well – People who have no active disease symptoms but are less fit than category 1. Often, they exercise or are very active occasionally, e.g., seasonally.

3 Managing Well – People whose medical problems are well controlled, but are not regularly active beyond routine walking.

4 Vulnerable – While not dependent on others for daily help, often symptoms limit activities. A common complaint is being "slowed up", and/or being tired during the day.

5 Mildly Frail – These people often have more evident slowing, and need help in high order IADLs (finances, transportation, heavy housework, medications). Typically, mild frailty progressively impairs shopping and walking outside alone, meal preparation and housework.

6 Moderately Frail – People need help with all outside activities and with keeping house. Inside, they often have problems with stairs and need help with bathing and might need minimal assistance (cuing, standby) with dressing.

7 Severely Frail – Completely dependent for personal care, from whatever cause (physical or cognitive). Even so, they seem stable and not at high risk of dying (within ~ 6 months).

8 Very Severely Frail – Completely dependent, approaching the end of life. Typically, they could not recover even from a minor illness.

9 Terminally Ill – Approaching the end of life. This category applies to people with a life expectancy <6 months, who are not otherwise evidently frail.

Scoring frailty in people with dementia

The degree of frailty corresponds to the degree of dementia. Common **symptoms in mild dementia** include forgetting the details of a recent event, though still remembering the event itself, repeating the same question/story and social withdrawal.

In **moderate dementia,** recent memory is very impaired, even though they seemingly can remember their past life events well. They can do personal care with prompting.

In **severe dementia,** they cannot do personal care without help.

Fig 49.2 The Clinical Frailty Scale. *IADL,* Instrumental activity of daily living. *(© 2007–2009 Version 1.2 All rights reserved. Geriatric Medicine Research, Dalhousie University, Halifax, Canada.)*

Sarcopenia

Sarcopenia is closely related to frailty (**Box 49.3**). It is a progressive and generalized muscle disorder associated with an increased risk of adverse outcomes, including falls and fractures. Muscle fibres reduce in number, fast-twitch type II muscle fibres in particular. The result is loss of explosive power, which is needed for functions such as sprinting, standing from a chair, or even coughing.

In sarcopenia, VO_2 max within muscle reduces, as does gait speed. Reduction in physical activity leads to a reduction in energy usage; when coupled with the relative anorexia of ageing, this leads to an energy deficit. With this energy deficit, cells that are predetermined to undergo apoptosis do not have sufficient free adenosine triphosphate (ATP) to allow this process to occur. Instead, the cells undergo necrosis, which uses more ATP and worsens the energy deficit. When this occurs in muscle tissue, it leads to worsening sarcopenia and development of the cycle of frailty (**Fig. 49.3**).

INVESTIGATIONS IN OLDER ADULTS

Physiological parameters have different implications and require different interpretation in older adults compared with younger people. For example, there is a 'U-shaped' association between BP and death in older people, and aggressive lowering of BP into 'normal' ranges is inappropriate in those at risk for falls. Similarly, oxygen saturations can fall with age and co-morbidities, so levels of 92% to 95% may be more physiologically normal.

Normal ranges stated by laboratories account for the values expected for the majority of the population, but older adults (like children) are outliers compared with the whole population. The altered anatomy and physiology in older adults mean that routine investigations need to be interpreted with this in mind (**Box 49.4**).

An unenhanced computed tomography (CT) scan of the head is often used to investigate confusion or cognitive impairment and may reveal non-specific findings of cerebral atrophy or findings suggestive of specific causes of dementia (e.g., small vessel disease or atrophy of the temporal horns, suggesting Alzheimer's disease).

Box 49.3 European Working Group on Sarcopenia in Older People (2018)

Criterion	Definition
1. Low muscle strength 2. Low muscle quantity or quality 3. Low physical performance	***Probable sarcopenia*** is identified by criterion 1. ***Confirmed sarcopenia*** requires criteria 1 and 2. ***Severe sarcopenia*** requires criteria 1, 2 and 3.

Small vessel disease is a common finding on CT scans in older adults and is often dismissed, though it is associated with cognitive impairment, a higher risk of delirium, gait dyspraxia and instability, and incontinence. There may be evidence of specific and potentially reversible causes of acute confusion (e.g., a subdural haematoma or space-occupying lesion).

Further reading

National Institute for Health and Care Excellence. *NICE Guideline 56: Multimorbidity: Clinical assessment and management*. NICE 2016; https://www.nice.org.uk/guidance/ng56.
Wilkinson I, Preston J. MDTea podcasts—*Episode 2.01 frailty*. https://thehearingaidpodcasts.org.uk/episode-2-1-frailty/.

COMPREHENSIVE GERIATRIC ASSESSMENT

After a patient is identified as being frail, what should be done? Although some evidence suggests that people can transition from more frail to less frail, the vast majority progress through advancing stages of frailty. There is usually no single treatment to cure frailty, but multicomponent interventions may slow progression and even reverse the decline. Inaction is likely to result in worsening frailty, functional impairment, and disability. Therefore, a proactive assessment of each of the potential clinical and subclinical deficits is needed.

CGA is not something that is just delivered by doctors; it is a multidimensional, interdisciplinary diagnostic *process* focused on determining a person's medical, psychological, and functional capability to develop a coordinated and integrated plan for treatment and long-term follow-up. As such, it is firmly rooted in the biological,

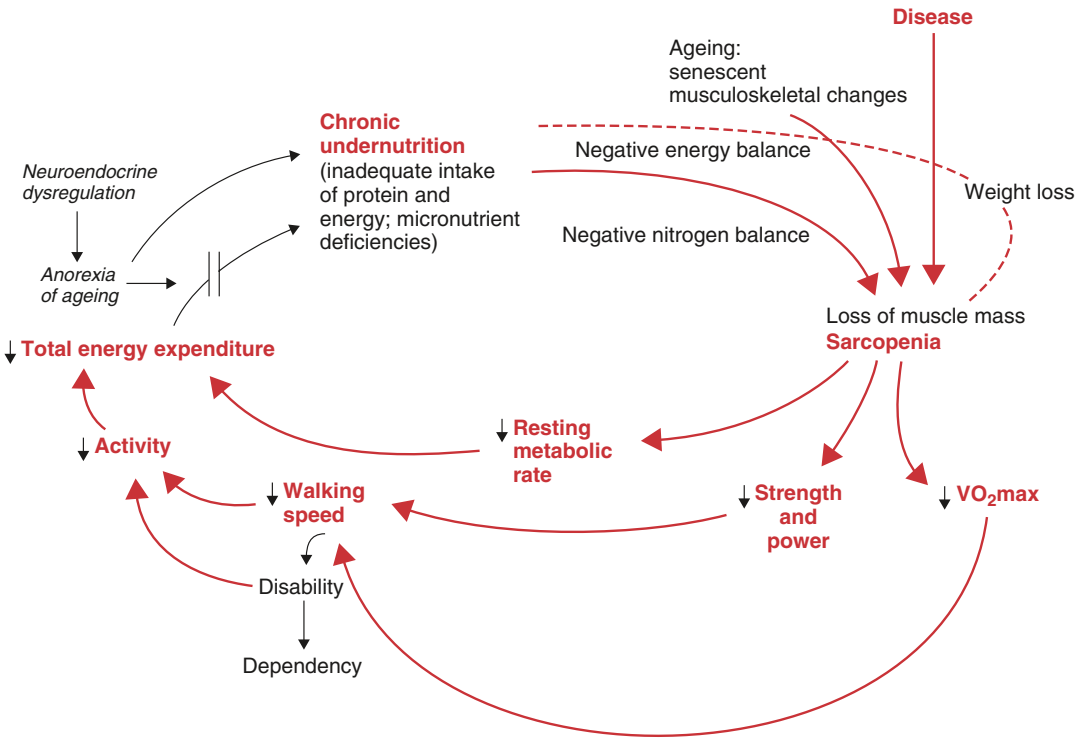


Fig. 49.3 The frailty cycle. *(Adapted from Clegg A, Young J. The frailty syndrome.* Clin Med *2011; 11:72–75.)*

Box 49.4 Blood tests in older adults

Blood test	Normal ranges in older adults	High values	Low values
Full blood count			
Haemoglobin	Often lower in advanced ages	**Higher concentration per volume** May suggest dehydration **Increased production** Polycythaemia (rubra) vera Secondary polycythaemia caused by chronic hypoxia	**Blood loss** Review medication for antiplatelets and anticoagulation that increase bleeding risk and for gastric irritants, e.g., corticosteroids or NSAIDs **Inadequate nutrients or substrates** Haematinic deficiencies, e.g., pernicious anaemia **Reduced production** Bone marrow disorders Anaemia of chronic disease
White blood cell count	Remains normal	**Increased need** Infection **Increased production** Haematological malignancies; commonest in older adults is CLL, causing chronically high lymphocyte counts Steroid use or stress-induced release	**Reduced production** *Acute* Overwhelming sepsis *Subacute or chronic* Bone marrow failure, e.g., myelodysplasia, myelofibrosis, infiltration of bone marrow by malignancy **Bone marrow suppression** Medication related **Ethnicity** Black people have lower total WBC counts; this presents over the course of their lives, so check past results
Platelets	Remain normal	**Increased need (acute)** Acute-phase reactant to inflammation **Increased production (subacute and chronic)** Essential thrombocythaemia	**Reduced production** Medications Bone marrow failures Liver failure **Increased consumption** Sepsis **Decreased survival** Hypersplenism Autoimmune
Serum creatinine and electrolytes			
Sodium	Remains normal	Dehydration associated with reduced homeostasis of thirst axis Hyperosmolar, hyperglycaemic states	Addisonism Fluid overload Medications, e.g., diuretics, SSRIs
Potassium	Remains normal	Renal failure Artefact (haemolysed sample) Medications e.g., ACE inhibitors, aldosterone antagonists, heparin Haemolysis High platelet counts (pseudohyperkalaemia)	Medications, e.g., thiazide diuretics, salbutamol GI losses: diarrhoea and vomiting **Low values may impair muscle function:** Skeletal muscle: reduced mobility, falls Smooth muscle: constipation Cardiac muscle: arrhythmias
Creatinine	'Normal' values may hide kidney injury in those with low muscle mass	AKI: often dehydration, sepsis, or urinary retention CKD Medication related, e.g., trimethoprim	Low muscle mass from sarcopenia
Estimated glomerular filtration rate	Reduction may indicate normal ageing, as well as renal disease Consider calculating creatinine clearance, especially in very frail and very old patients		Less useful in the very underweight and very overweight

Box 49.4 Blood tests in older adults—cont'd

Blood test	Normal ranges in older adults	High values	Low values
Liver function tests			
Total protein		**Myeloma**	**Malnutrition**
Albumin	Remains normal		*Acute:* catabolic states, e.g., infection, cancer *Chronic:* poor nutritional state
ALP	Remain normal	*Subacute or chronic:* Bony metastases (prostate, breast), vitamin D deficiency, Paget's disease *Acute:* acute-phase reactant, bone fracture	
AST, ALT, GGT	Remain normal	Chronic alcohol excess Hepatic congestion caused by right heart failure Hepatotoxic medications MAFLD Ischaemic hepatitis	
Bilirubin	Remains normal	Obstructive bile duct disease, e.g., gallstones, pancreatic cancer	
Bone profile			
Calcium	Normal range, though often at the lower end	Dehydration Malignancy Primary hyperparathyroidism Calcium or vitamin D supplements	Poor dietary intake Recent parenteral treatment for osteoporosis (e.g., denosumab, zoledronic acid) Magnesium deficiency
Vitamin D	Normal range, but low levels common	Supplements, not usually of clinical significance	Nutritional deficiency Reduced sunlight exposure Fat malabsorption
Other			
Erythrocyte sedimentation rate	Normal values increase with age	>100 mm/hour should prompt investigation for malignancy or vasculitis, e.g., polymyalgia rheumatica or giant cell arteritis	
Creatine kinase	Normal range	Rhabdomyolysis after fall and long lie Medications, e.g., statins, neuroleptics	
C-reactive protein	Normal range	Sepsis Inflammatory disorders Physiological stress, e.g., surgery Malignancy	
Thyroid function (TSH, T_4)	Normal range	Various causes of hyperthyroidism May cause confusion, delusions, tremor	Various causes of hypothyroidism May cause confusion, constipation, myopathy
Iron studies (ferritin or TI)	Normal range	Ferritin is an acute-phase reactant; consider checking TI	Ferritin low in iron-deficient anaemia (may be disproportionately normal in acute infection and iron-deficiency anaemia)

ACE, Angiotensin-converting enzyme; *AKI,* acute kidney injury; *ALP,* alkaline phosphatase; *ALT,* alanine transaminase; *AST,* aspartate aminotransferase; *CKD,* chronic kidney disease; *CLL,* chronic lymphocytic leukaemia; *GGT,* gamma-glutamyl transferase; *GI,* gastrointestinal; *MAFLD,* metabolic dysfunction–associated fatty liver disease; *NSAID,* non-steroidal anti-inflammatory drug; *SSRI,* selective serotonin reuptake inhibitor; T_4, thyroxine; *TI,* transferrin index; *TSH,* thyroid-stimulating hormone.

psychological, and sociological approach to thinking about patients and their conditions. CGA involves exploring a number of domains (**Box 49.5**). The further roles of the MDT are explored later.

In older adults with frailty, there is often a large overlap between symptoms caused by disease, by treatment of disease, by normal ageing, and by accelerated ageing. More practically, CGA can be thought of as a process of good holistic care delivered by a focused MDT, which goes above and beyond simply managing the acute problem with which the person has presented. If an older person is not managed in this way, care becomes reactive to a series of acute problems rather than a proactive problem-solving process.

Compared with those given 'usual' care, those who undergo CGA during an acute admission have favourable outcomes. At 6

Box 49.5 Domains of the Comprehensive Geriatric Assessment

Domain	Assessment area	Assessment	Examples of assessment tools
Physical symptoms	Pain	Location, severity, impact on quality of life	Numerical scales Visual scales, e.g., Abbey Pain Scale (for patients with dementia)
	Continence	Exploration of both storage and voiding; lower urinary tract symptoms	Bladder diary *or* Postvoid residual bladder volume
	Sensory impairment	Hearing assessment Check for ear wax	
		Visual assessment	Snellen chart Optician or ophthalmology review (including glaucoma, cataracts, macular degeneration)
	Musculoskeletal problems	Mobility, range of movement	Berg Balance Scale Timed Up and Go test 6-m walk test
	Multimorbidity management	Multimorbidity management approach Identify disease vs treatment burden Discuss with patient and agree a shared management plan Review on a regular basis	Polypharmacy assessment tools, e.g., STOPP-START, Beers criteria Ascertain patient's priorities of care: symptom relief (AF rate control) versus future prevention (AF anticoagulation)
Mental health symptoms	Mood assessment	Mental health assessment, including full mental state examination	Geriatric Depression Scale (GDS) Hospital Anxiety and Depression Scale (HADS) Cornell Depression Scale (for patients with cognitive impairment)
	Cognitive assessment	Detailed cognitive assessment, including collateral history from a relative or carer	Abbreviated Mental Test score Mini-Mental State Examination (MMSE) Six-item Cognitive Impairment Test (6-CIT) Montreal Cognitive Assessment (MOCA) Clock drawing task (e.g., CLOX 1) Addenbrooke's Cognitive Assessment-Revised (ACE-R) (more detail; 15 min) Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE)
Functional abilities	Activities of daily living (ADL) assessment	Personal ADLs: Washing and dressing Bed transfers Toileting needs Domestic ADLs: Food preparation Housework	Barthel Index (personal ADLs) Nottingham Extended ADL Scale (includes domestic ADLs and higher-level functional assessments)
Living environment	Home visit or access visit Usually led by occupational therapists	Assessment for hazards around home (e.g., lighting, flooring)	
		Assessment for safety (e.g., heating, safe use of cooker, stair rails)	
Social support network	Financial assessment	Often performed by social services	
	Carers' assessment	Care Act (UK)	
Future wishes	Advanced care planning	Nomination of power of attorney, CPR decisions and advance decision to refuse treatment (see Chapter 3, 'Advance decisions' section)	DNA CPR; ceilings of treatment (e.g., ward-based ceiling of care) ReSPECT document Locally agreed advanced care plan

AF, Atrial fibrillation; *CPR,* cardiopulmonary resuscitation; *ReSPECT,* Recommended Summary Plan for Emergency Care and Treatment; *STOPP-START,* Screening Tool Of Older People's Prescriptions (STOPP)-Screening Tool to Alert to Right Treatment (START).

Box 49.6 Common members of the multidisciplinary team and their roles

Job	Role	Clinical examples
Dietician	Assessment of nutritional status and support of healthy weight during illness and health	Falls: ensuring adequate calorie and protein intake for muscle maintenance Cognition: adaptations to meal routine to optimize nutrition, e.g., opportunistic mealtimes
Discharge coordinator	In-hospital coordination of complex discharges	Coordination of multiple agencies, e.g., social services, specialist equipment, funding streams, safeguarding
Doctor	Medical diagnoses and optimization, including reversible conditions and prescribing Prognostication	Falls: causes and interventions, bone health assessment and prescription Polypharmacy: medication review Inputs into most Comprehensive Geriatric Assessment domains
Nurse	Delivery of hands-on care for older patients; assessment of holistic and physical needs, including medication compliance, skin care, continence	Continence: assessment and management Skin: wound care and pressure damage Medication: delivery and compliance
Occupational therapist	Functional assessments and management (including cognition) aiming for a greater degree of independence	Falls: assessments and adaptations for maximizing independence; optimization of environment, e.g., ensuring furniture of a reasonable height to stand from easily; movement-sensitive lighting, pendant alarms, 'key safe' (small box with combination lock outside property containing house keys for carers) Cognition: cognitive assessment, activities of daily living assessments, kitchen function and safety assessments
Pharmacist	Optimization of medicine concordance Addressing polypharmacy and interactions	Falls: review of contributing medications
Physiotherapist	Optimization of physical functioning in relation to musculoskeletal function	Falls: gait analysis, exercise programmes to improve balance and strength, walking aids Continence: pelvic floor exercises
Podiatrist	Assessment and management of foot health	Falls: reduction of foot pain, suitable footwear for deformities
Psychologist	Instigation and direction of talking therapies, e.g., cognitive-behavioural therapy	Falls: fear of falling counseling Mood: talking therapies, including in early dementia
Social worker	Assessment of social functioning and safety	Falls: home adaptations, provision of care Cognition: capacity assessments in relation to welfare and finances Carer strain: assessments as part of Care Act (UK) Welfare: assistance with finances
Speech and language therapist	Swallowing and communication	Communication strategies, support for capacity decisions when communication is impaired Dysphagia: recommendations for minimizing risks and advance care planning
Volunteers	Supporting social interaction and function	Falls: exercise classes Cognition: dementia cafés, dementia friends Nutrition: assistance with shopping

and 12 months, respectively, the numbers needed to treat (NNTs) are 20 and 25 to prevent institutionalization and 17 for unnecessary death. The NNT for many medications is in the hundreds to thousands range.

One area that remains to be fully explored is how best to identify patients who would benefit from a CGA who are not in an acute hospital. The proactive assessment of patient frailty is one means of doing this. After patients are identified as frail (or becoming frailer), then a proactive CGA, in the community, day hospital, or inpatient setting, seems to be a pragmatic approach.

Multidisciplinary team working

To deliver true CGA, geriatricians need to work as part of an expert MDT. Communication between the whole of the MDT is essential to ensure that patient care plans are discussed and developed together. Each team member's specialist skills are better used, making for a more egalitarian approach to care rather than the traditional hierarchical one with the consultant as the leader (Box 49.6). Learning about geriatric medicine therefore requires an understanding not only of the role of the doctor in delivering CGA but also of the roles of each member of the MDT.

ADVANCE CARE PLANNING

Advance care planning (ACP) involves discussing the approach to a patient's future care, often before their becoming unwell or more unwell (see Chapter 50, 'Advance care planning: planning for the future' section). ACP is a continuous process of communication between patients, their loved ones, and healthcare professionals. It is not always possible or appropriate to cover everything in one conversation, so this may take some time and be an iterative process. Patients may want to think through often complex issues and discuss their plans with their loved ones. Plans should also be subject to regular review as a person's priorities and goals change.

ACPs ideally should be offered to all older people if they wish to explore their future wishes for care. ACP discussions can be started at any time if the patient is willing and ready. They should be specifically considered in patients diagnosed with progressive conditions, such as neurodegenerative diseases (e.g., dementia, Parkinson's disease), when there is the potential for capacity to be lost in the latter stages. This is particularly different to younger patients in whom terminal diagnoses (e.g., cancer) may preserve cognitive function for much longer.

This can also be a good time to highlight the importance of ensuring that wills and power of attorney are considered. These legal frameworks vary between countries but broadly speaking allow a person to either state what sort of interventions they would not like in the future or allocate someone to decide on their behalf in the event they become unable in the future.

ACPs often focus on the dying phase but should be expanded to include what constitutes living well and matters most for that person. Statements of wishes and preferences can range from where a person lives (e.g., in their own home or in a care home) to their favourite foods.

One of the most common decisions concerns resuscitation status. Many older people recognize that they are coming toward the end of their lives (around one in five people older than 70 years old die within 1 year of an emergency admission). Even though they are in good health, many would like to be allowed a natural death in the event of a cardiorespiratory arrest. Others may want to live for as long as possible and ask for all that medicine can do to prolong their lives. Events that may precede resuscitation should be planned for, too, such as hospitalization for treatment of infections.

Knowing when to start this process can be tricky: too early, and it may not feel relevant to the patient; too late, and they may have missed out on the chance to make their wishes clear. Patients, though, are very often happy to have a discussion about the process even if they do not wish to record their views at that point in time. Starting such a discussion with people gives them the ability to be able to be involved with decisions around their future medical care while they are still able to communicate their wishes. Taking this proactive approach allows a discussion of care toward the end of life and complex choices such as resuscitation to be thought about in detail.

In the UK, ACPs are now commonly recorded in a ReSPECT (Recommended Summary Plan for Emergency Care and Treatment) form, which is kept by the patient in their home. This allows the patient to choose their level of care on a scale from an approach in which prolongation of life should be done at all costs, including surgery and intensive care if needed, to being kept at home and palliated if they were to deteriorate. The balance is sometimes between length of life and quality of life, though it is important to realize that this is not a binary choice.

Any outcomes agreed upon should be documented and available to all who need access to them, including the patient and their family, ambulance staff, GPs, community teams, and hospital teams.

COMMON ISSUES FACING OLDER PEOPLE

Polypharmacy

Prescribing in older people

Although prescribing for older people is largely the same as for younger people, there are some key differences. The approach should be based around the patient's physiological rather than chronological age. There are numerous mechanisms behind the altered pharmacokinetics in older people (**Box 49.7**), but the net result is that older patients tend to have increased side-effects from their medications.

Drug interactions

One-third of patients older than 65 years of age living in the UK take 6 or more medications, with 5% taking more than 10. Polypharmacy leads to increased drug interactions and increased potential for ***drug–drug*** (e.g., cholinesterase inhibitors and anticholinergics) and ***drug–disease*** (e.g., non-steroidal anti-inflammatory drugs worsening hypertension) interactions. Polypharmacy accounts in part for up to 6.5% of all iatrogenic hospital admissions.

Box 49.7 Reasons for altered pharmacokinetics in older people

Drug absorption
- Gastric pH gradually increases with age (exacerbated by medications such as proton-pump inhibitors).
- Gastric emptying slows.

Drug distribution

The main physiological changes that occur with ageing and lead to adverse drug reactions include:
- Increased body fat and decreased body water (as a proportion of total body weight)
- Increased volume of distribution for fat-soluble drugs, so medications may accumulate because of an increased elimination half-life; this means that their effects may continue even though the drug has been stopped
- Decreased volume of distribution for water-soluble drugs, so drugs such as digoxin need a reduced loading dose
- Altered permeability of the blood–brain barrier, so the brain of an older adult can be exposed to higher levels of medication

Drug metabolism
- Reduced hepatic blood flow (and hence clearance) of medications that have high hepatic excretion ratio, e.g., amitriptyline

Elimination by the kidneys
- With age, there is a reduction in renal mass and renal blood flow.
- eGFR decreases by 0.5% per year after the age of 20 years, so 'normal' creatinine may be associated with significant renal impairment.
- This affects clearance of water-soluble drugs such as diuretics, NSAIDs, and digoxin.

eGFR, Estimated glomerular filtration rate; *NSAID*, non-steroidal anti-inflammatory drug.

Box 49.8 Optimizing compliance and concordance

Understand what medications are and are not being taken
- The patient may not be taking all prescribed medications.
- Why might they not want particular medications?
- Exercise caution when increasing medications without understanding what is currently being taken to avoid causing iatrogenic complications.
- Review side-effects (e.g., diuretics causing incontinence).
- Review the aims of treatment (see later).

Understand problems around taking the medications
- Dysphagia: Is there a liquid or patch alternative?
- Dexterity: Can the patient take the medication from the packaging? Aids are available for inhalers and eye drops.
- Cognition: Is the patient forgetting to take the medication or taking a second dose of one already taken? Are they managing complicated instructions, e.g., for oral bisphosphonates, or variable-dosing medications such as insulin or warfarin?
- Visual impairment: Consider large-font labelling.

Simplify regimens
- Is a particular time of day better to take the medications?
- Consider once-a-day, modified-release preparations.
- Reduce inappropriate polypharmacy.
- Agree which medications the patient wishes to take.
- Use compliance aids (e.g., 'dosette boxes') to reduce the need to coordinate multiple medications and aid monitoring of concordance.

Review care needs
- Is supervision or prompting required?

Patient choice

Polypharmacy may leave patients unhappy with their medication burden, which can lead to poor concordance overall (**Box 49.8**). Many prefer to take fewer medicines, accepting an increased risk of further events and potentially shortened life expectancy (e.g., foregoing a statin after acute coronary syndrome). A patient's priorities of care (what is most important to them) often rate immediate symptom control over preventative interventions. When prescribing, clarifying the goals and educating patients about the benefits of treatment are key to supporting them in making treatment choices.

Evidence base

Few clinical trials involve patients older than 80 years old, and those who are included tend to be fitter than those encountered in clinical practice, especially those in secondary care. Reported levels of benefit and risk of harm from medications studied in younger, non-frail, non–co-morbid groups cannot therefore be assumed to be the same as in older, frailer cohorts with higher levels of co-morbidity. Research that includes frailer populations with multimorbidity is inherently very difficult because of the confounding issues seen in such cohorts and the practicalities of obtaining consent or attending follow-up visits. However, it would be wrong to deny patients the benefits of treatment on the grounds that there is no evidence base; the absence of an evidence base is not the same as the absence of benefit.

Benefits of therapy

The levels of benefit from interventions differ between age groups. For example, there is increasing evidence in older patients with type 2 diabetes treated with insulin that tighter target hemoglobin (A_{1c}) HbA_{1c} levels below 50 mmol/mol (<6.7%) are associated with worse outcomes (increased mortality, increased weight gain) than in those patients with a more relaxed target HbA_{1c}. In older people with diabetes, the avoidance of immediate complications (e.g., hypoglycaemic events) should be prioritized over prevention of longer-term consequences, although this must be balanced with minimizing hyperosmolar symptoms (e.g., polydipsia, polyuria, recurrent infections, fatigue), which are seen when glycaemic control is very poor.

Similarly, although targets for BP tend to be around 130/80 mmHg in younger patients, outcomes for older and frailer patients are better with a higher target; there is a U-shaped curve with an optimum of 160/90 mmHg (i.e., there is an increased risk of adverse events and mortality when BP is significantly higher or lower than this; **Fig. 49.4**). Some of these adverse events relate to the side-effects of antihypertensives, such as orthostatic hypotension, falls, and associated morbidity and mortality.

Clinical decision-making

Many tools exist to aid clinical decision-making with respect to polypharmacy. They aim to summarize the evidence for a given medication in older adults and allow a negotiated approach between patient and clinician regarding the stopping and starting of medications. Examples include the Beers criteria, Screening Tool of Older People's potentially inappropriate Prescriptions-Screening Tool to Alert doctors to Right Treatments (STOPP-START), anticholinergic burden index, and Scottish polypharmacy app (see **Box 49.5**). **Fig. 49.5** shows the potential hierarchy of utility for medications that might be considered when undertaking a polypharmacy review.

Falls and instability

Most of us take for granted the amazing complexity of the process that maintains an upright posture. High levels of functioning are needed across multiple systems, including musculoskeletal, neurological, and cardiovascular. Although anyone of any age can fall, it is said that 'young people trip; old people fall'. For many older or frail adults, there are numerous, cumulative risk factors for falling that, individually, would be unlikely to produce major instability (**Boxes 49.9 and 49.10**; **Fig. 49.6**). Assessment of falls is one of the most complex areas for geriatricians and can take significant time. Asking about falls should nevertheless be part of routine practice for all professionals working with older people.

Approximately 5% of fallers older than 65 years of age sustain an injury that is serious enough to limit normal activities. Most falls occur during the daytime, when people are most active, and only around 20% take place overnight. Risk factors include cognitive impairment, medications, alcohol excess, and a change of environment (e.g., being admitted to hospital or moving to new accommodation).

Breaking down the contributing risk factors helps provide some structure for assessing and developing a management plan for a person who has fallen (**Box 49.11** and **Fig. 49.7**) and for targeting modifiable risk factors. An approach to the assessment of falls is summarized in **Box 49.12**. After clarifying the risk factors, it is then important to recognize which are reversible (e.g., cataracts), which are modifiable (e.g., glaucoma), and which are untreatable (e.g., dry macular degeneration).

Transient loss of consciousness (TLOC) is a worrying symptom, but of patients who do actually lose consciousness, more than 30% are unaware that it occurred. Often patients try to rationalize what has happened, stating, 'I must have tripped.' Normally, people put their arms out when falling to protect themselves. Be suspicious that a transient loss of consciousness may have caused falls when this has not happened and there are facial injuries. Items to check include lying and standing BP,

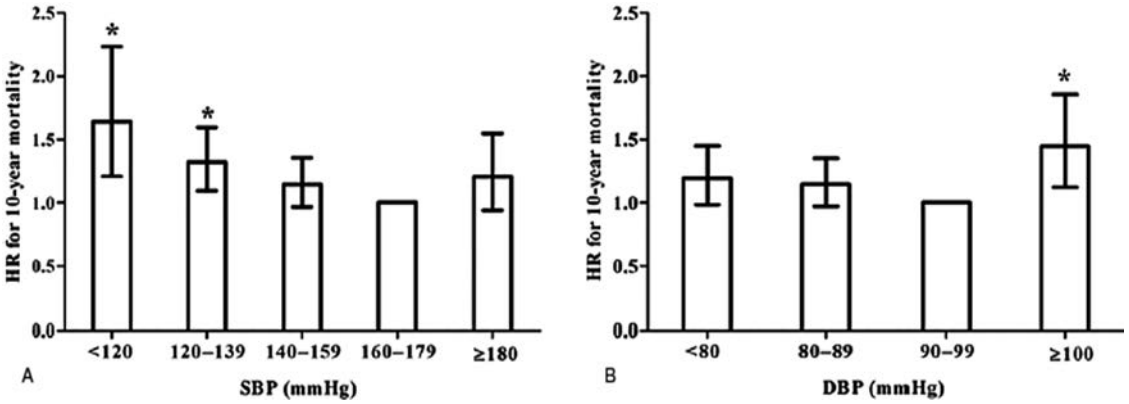


Fig. 49.4 Risk of 10-year mortality in systolic blood pressure (SBP) and diastolic blood pressure (DBP) categories. Bars represent hazard ratios (95% confidence interval). The category of SBP 160 to 179 mmHg and the category of DBP 90 to 99 mmHg were set as references in A and B, respectively. (*Taken from: Giulia Ogliari et al 2015, Blood pressure and 10-year mortality risk in the Milan Geriatrics 75+ Cohort Study: role of functional and cognitive status.* Age and Ageing. *44 [6], 932–937.*)

Utility	Medications that:	Examples
More useful	Provide immediate relief for distressing symptoms	Analgesics, antiemetics
	Modify an acute condition that is life-threatening or will soon result in distressing symptoms if not treated	Antibiotics for severe sepsis Diuretics for acute heart failure
	Modify a chronic condition that may progress to become life-threatening or progress to cause significant symptoms if not treated	Methotrexate for rheumatoid arthritis
Less useful	Have the potential to prevent serious disease without a symptomatic benefit	Antihypertensives, antiplatelet agents
	Are unlikely to be useful in either the short or the long term	Fish oils, glucosamine
	Are used for indications where non-pharmacological therapy is equally or more effective	Physiotherapy for back pain Sleep hygiene vs long-term benzodiazepines

Fig. 49.5 Hierarchy of utility of medications. *(Adapted from Scott IA, et al. Deciding when to stop: towards evidence-based deprescribing of drugs in older populations.* Evid Based Med *2013 18[4]: 121–24.)*

i Box 49.9 Falls in the UK

- One-third of people older than 65 years fall at least once per year.
- Half of people aged older than 80 years fall at least once per year.
- More than 250 000 emergency hospital admissions each year relate to falls.
- Falls are estimated to cost the National Health Service £2.3 billion per year.
- Falls from 2 m or less are the most common mechanisms of injury for major trauma in older people in the UK.

i Box 49.10 Categories of falls

Category	Optimizable	Static
Intrinsic	Concurrent illness, cardiac syncope, medications	Weakness from previous stroke
Extrinsic	Trip hazards, grab rails	Stairs into the house

electrocardiogram, and medications such as antihypertensives or diuretics. Management of TLOC is largely the same at all ages (see Chapter 11, 'Dizziness and blackouts' section).

Consequences of falls

Most falls result in minor or no injuries. However, consequences can include fractures and trauma.

Fear of falling

Even after minor falls, patients can develop a fear of falling (reduced 'falls efficacy'). This is characterized by anxiety around mobilizing and is associated with an increasing loss of confidence and a hesitant gait (see **Box 49.10**). As a result, people often reduce the amount of walking they do and stop going out of the house. In turn, they develop reduced muscle strength and an increased risk of falling. There is also a danger of social isolation and depression. Arguably, fear of falling is not all bad; it can just reflect an increased need for caution.

Fragility fractures

Fragility fractures occur after an injury from a standing height or lower and range from vertebral fractures (when there may even be no fall) to wrist, proximal humeral, pubic rami, and neck of femur fractures. Most of these are associated with osteoporosis, and enquiry should be made about:

- Previous fractures, including site, history, and age at the time
- Age of menopause for female patients
- History of previous osteoporosis treatment

After this is known, an assessment of osteoporosis risk can be calculated by using the FRAX (fracture risk assessment) score (see Box 33.8). This is widely available, and risk is adjusted for a number of countries around the world.

Sometimes there is no history of trauma. These atraumatic fractures can be caused by severe osteoporosis, but other causes should be considered, particularly malignancy (e.g., breast, prostate, myeloma, lung), and investigated accordingly. Long-term antiresorptive bone therapy (i.e., bisphosphonates or denosumab) can cause bones to become unnaturally brittle, especially if it is given as a parenteral treatment or at high doses for malignancy. This can result in unusual fractures, often atraumatic ones, such as fractures of the femoral shaft.

Neck of femur fractures

In England, there are approximately 60 000 neck of femur fractures per year, accounting for 15% of all fragility fractures (**Fig. 49.8**). They represent a significant injury for older persons:

- They account for 1.8 million inpatient bed days per year.
- Some 20% of patients sustaining a neck of femur fracture require transfer to long-term care within 1 year.
- The mortality rates are 5% at 30 days and 30% at 1 year.

In a number of countries, there has been a drive to develop geriatric liaison services to deliver CGA for patients with hip fractures. The geriatrician's role in perioperative care covers the issues outlined in **Box 49.12** and includes:

- Pain management
- Management of venous thrombotic risk

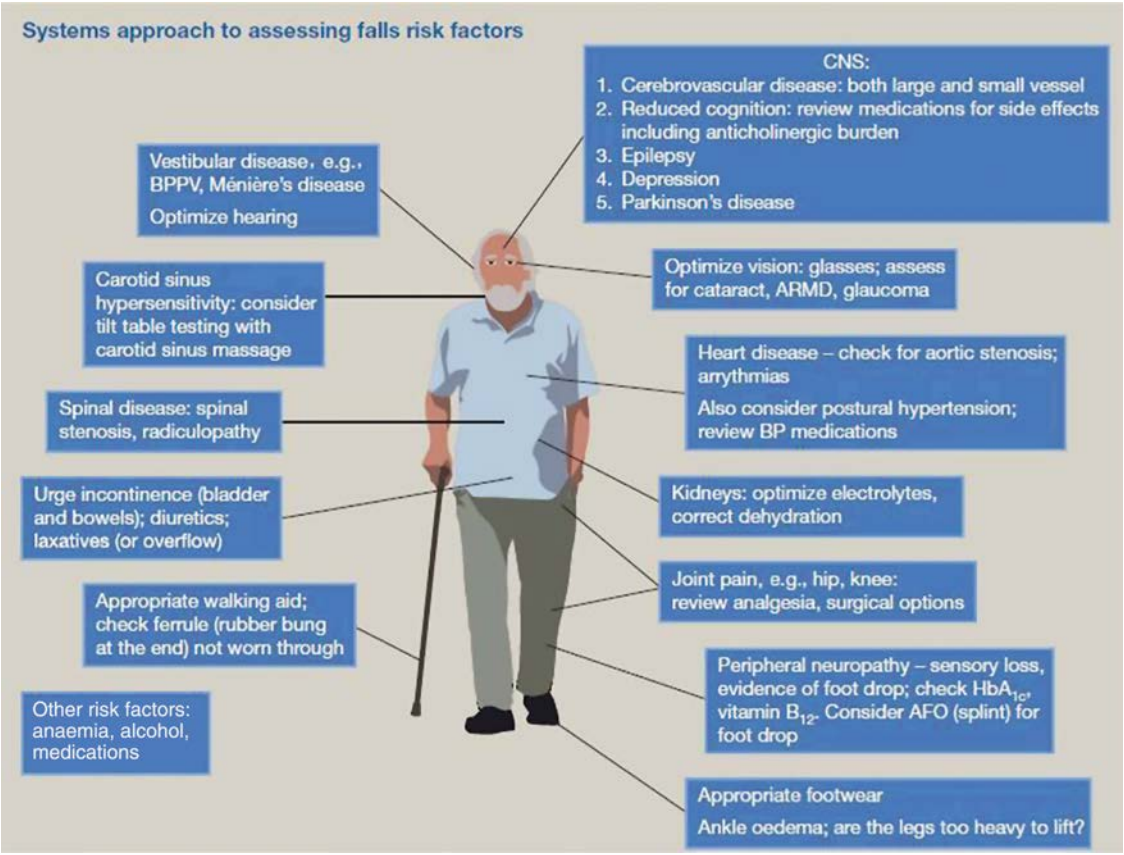


Fig. 49.6 Systems approach to assessing falls risk factors. *AFO,* ankle-foot orthosis; *ARMD,* age-related macular degeneration; *BP,* blood pressure; *BPPV,* benign paroxysmal positional vertigo; *CNS,* central nervous system.

Box 49.11 Some risk factors for falls with possible management strategies

Risk factor	Causes	Management
Unsteady gait or poor balance	Varied: Gait assessment for speed, sway, hesitancy, degree of heel strike, time with both feet on the ground Balance assessment, including cerebellar examination	Physiotherapy assessment: Balance and strength exercises Review of walking aids Encouragement to keep active Appropriate and well-fitting footwear
Leg weakness	Multiple causes Spinal disease Sarcopenia Disuse or deconditioning Motor neurone disease Myopathy	Neurological assessment Clinical assessment Imaging, e.g., MRI Physiological, e.g., EMG or nerve conduction studies Targeted physiotherapy
Peripheral sensory impairment	Peripheral neuropathy, e.g., diabetes, vitamin B_{12} deficiency Reduced proprioception	Optimization of risk factors Physiotherapy and occupational therapy for adaptive measures
Joint instability	Greater likelihood of tripping and falling because of pain or reduced range of movement of hips and knees in osteoarthritis	Analgesia: topical, oral, intra-articular; consider joint replacement Walking aids to reduce stress though joint Physiotherapy to improve strength around joint
Pain	Antalgic gait Side-effects of pain medication, e.g., drowsiness from opioids	Optimization and individualization of pain control
Cerebrovascular disease	Large vessel: causing focal visual, sensory, and motor deficits Small vessel: frequently dismissed Both associated with: Gait dyspraxia Marche à petits pas Cognitive impairment Urinary incontinence	Optimization of vascular risk factors
Cognitive impairment	Inability to pre-empt hazards, resulting in reduced safety awareness Impaired executive function in planning complex tasks Increased activity: 'wandering' Inability to use or remember walking aids	Screening for reversible causes of confusion Review of medication, e.g., opiates, sedatives, anticholinergic burden Occupational therapy for home safety
Depression	Association with reduced activity and consequent weakness, poor nutritional intake and poor sleep Impact on cognitive or executive function Interplay between depression and fear of falling	Activity can improve mood and increase interaction with social world Antidepressants may be required but can contribute further to falls Avoidance of medications most associated with postural hypotension (e.g., tricyclics) Monitoring for hyponatraemia and QTc prolongation
Visual impairment	Visual field defects: stroke, macular degeneration, glaucoma Visual neglect: stroke Reduced visual acuity: lens changes, retinopathy *Reduced hazard awareness*	Optician review, ideally annually Improvement of lighting around house; consider movement-sensitive lights for nocturia
Hearing impairment	Conductive: wax Sensorineural loss: commonly age-related plus other causes Social isolation *Reduced hazard awareness*	Audiological assessment for hearing aids Treatment for ear wax
Medications and polypharmacy	Exacerbation of postural hypotension Dehydration Electrolyte imbalances Confusion and sedating effects	Medication review (see earlier) Over-the-counter medications Alcohol history
Postural hypotension	Frequent association with antihypertensive treatment; may be intermittent Symptoms can be vague	Ambulatory BP monitoring if associated with hypertension Review of target BP—loosening control is often a pragmatic solution Minimization of vasodilating and dehydrating medications, e.g., ACE inhibitors, diuretics, nitrates

Box 49.11 Some risk factors for falls with possible management strategies—cont'd

Risk factor	Causes	Management
Electrolyte imbalances	Mild–moderate derangements often dismissed but associated with non-specific muscle weakness, malaise, constipation Severe derangements can lead to arrhythmias, seizures, myopathies presenting as a fall Often medication or hydration related	Blood tests Correction of abnormalities Review of medications Review of diet
Anaemia	Causes fatigue, weakness, and breathlessness Increased risk of postural hypotension if volume loss-associated Pernicious anaemia causes vitamin B_{12} deficiency, which can lead to peripheral sensory neuropathy and SACD Frequent association with chronic disease and iron deficiency	Establishment of underlying cause Consideration of: Blood transfusion Iron replacement Vitamin B_{12} and folate supplementation
Urinary incontinence	Overactive bladder Results in rushing to toilet—especially problematic if there are also gait problems	Treatment of continence; consider specialist referral Minimization of diuretics
Diabetes mellitus	Multiple falls risk, including: Hypoglycaemic episodes Visual impairment: retinopathy or cataracts Peripheral sensory neuropathy and autonomic dysfunction	Check of blood sugar levels: diary HbA_{1c} higher target for frail older adults: higher mortality rate associated with HbA_{1c} <50 Retinal screening Check for neuropathy and postural BP
Brady- and tachyarrhythmia (e.g., AF or CHB)	Reduction of cerebral perfusion and increase in falls risk	If ECG is abnormal, cardiac monitoring may be required (e.g., 24-h/7-day tape, implantable loop recorder)
Fear of falling	Very common, especially after major falls, but can be independent of falling Association with changes in gait pattern to be less steady, further potentiating fear Physical activity may reduce, exacerbating muscle weakness Complex interplay between depression and fear of falling	Cognitive-behavioural therapy Exercise (particularly tai chi) Treatment of underlying anxiety or depression Review of walking aids to aid confidence and independence
Vertigo, e.g., BPPV	Vestibular disorders in older people common and frequently unrecognized	Head impulse test: peripheral vs central cause Dix–Hallpike: BPPV Epley manœuvres or vestibular exercises (e.g., Cooksey–Cawthorne) ENT or specialist physiotherapy input
Foot health	Foot deformities and pain contribute to falls risk and poorly fitting footwear Leg length discrepancies	Podiatry referral Trimming of toenails Well-fitting footwear; may need specialist input

ACE, Angiotensin-converting enzyme; *AF,* atrial fibrillation; *BP,* blood pressure; *BPPV,* benign paroxysmal positional vertigo; *CHB,* complete heart block; *ECG,* electrocardiogram; *EMG,* electromyography; *ENT,* ear, nose, and throat; *HbA_{1c},* hemoglobin A_{1c}; *MRI,* magnetic resonance imaging; *SACD,* subacute combined degeneration of the spinal cord.

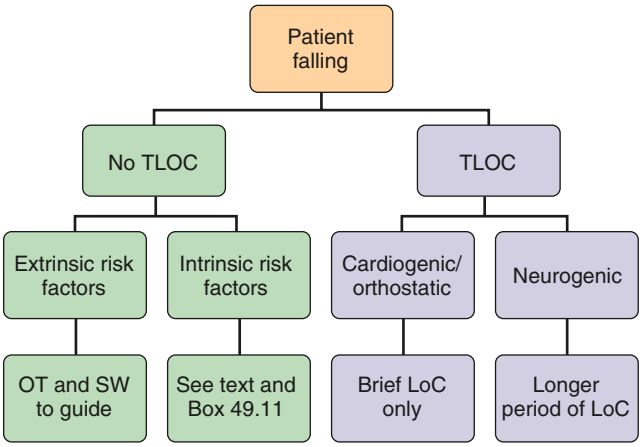


Fig. 49.7 Falls assessment. *LoC,* Loss of consciousness; *OT,* occupational therapist; *SW,* social worker; *TLOC,* transient loss of consciousness.

- Recognition and management of delirium
- Management of bowels and bladder
- Assessment and management of bone health and osteoporosis.

In the UK, the widespread establishment of multidisciplinary orthogeriatric services has contributed to a halving of 1-month mortality for hip fractures (from ~10% to ~5%), and similar models of care are increasingly being developed for other cohorts of surgical patients.

Institutionalization

Falls and their consequences contribute to approximately 40% of admissions to care homes. The fact of being in a care home does not in itself improve gait or remove the risk of falling; indeed, the risk of falling may initially increase on moving into a care home because this is a new and unfamiliar environment. However, it does mean that there are people on hand to help reduce the risk and to assist if someone does fall.

Box 49.12 Assessment of falls

Immediate management
- Treatment of acute illness (ABC approach)
- Assessment for injuries
- Pain control: consider regional anaesthesia (fascia iliaca nerve block for hip fractures) to minimize opiate use
- Preparation for safer surgery if required, e.g., optimizing fluid balance, correcting clotting (if on anticoagulants)

History and examination (ideally including collateral information)
- Events before the fall; previous falls or near misses; baseline mobility (e.g., any aids, asking why they are needed—muscle weakness, balance, other)
- Any loss of consciousness
- Medication review; alcohol history
- Full examination; consider risk factors, e.g., aortic stenosis, evidence of Parkinsonism, cerebellar issues, visual screen
- Postural blood pressure
- Inspect gait: even without formal training, think about how unsteady this looks, including turning, and how well a walking aid is used and whether the patient looks comfortable with it
- List risk factors for falls; address when possible
- Evidence of harm resulting from fall: pain, bruising, deformity, evidence of head injury

Investigations
- Bloods tests:
  - Full blood count (e.g., anaemia, evidence of sepsis; fractures can cause a neutrophilia)
  - Serum creatinine and electrolytes (evidence of acute kidney injury, electrolyte imbalance)
  - Glucose and hemoglobin A_{1c}
  - Bone profile and vitamin D
  - Thyroid function, micronutrient levels (vitamin B_{12}, folic acid, iron studies)
- Electrocardiogram (ECG), e.g., conduction abnormalities
- Imaging
  - Plain X-rays and cross-sectional imaging for possible fractures
  - Computed tomography of the head (evidence of large- or small-vessel cerebrovascular disease; exclusion of trauma, e.g., subdural haemorrhage)

Consider need for acute hospital admission
- Major trauma, including 'silver trauma' (older people with multiple injuries, often from low-impact trauma, such as falling from standing height), head injuries, fractures (e.g., hip)
- Possible or likely acute illness requiring inpatient care (e.g., intravenous antibiotics, fluids)
- Inability to mobilize safely
- Care needs that cannot be met in the community
- Not all fractures require admission, depending on mobility and social circumstances (e.g., wrist, neck of humerus, pubic rami can be managed in the community).

Factors that should prompt further assessment
- Frequent falls without recent input from a specialist team when the risk of further falls is believed to be high
- Possible syncope
- Polypharmacy
- Referral to relevant specialists, depending on findings from assessments (e.g., Parkinson's disease specialist)

Further management
- Provide oral and written information about falls
- Assess for and manage osteoporosis (e.g., DEXA scan, optimization of calcium/vitamin D levels, anti-resorptive bone medications or teriparatide)
- Advise regular optician review (e.g., screen for cataract, glaucoma)

DEXA, Dual-energy X-ray absorptiometry.

Postural hypotension

Cerebral autoregulation is responsible for maintaining stable blood flow through the brain. As we age, this becomes less well controlled, causing the brain to become more vulnerable to changes in systemic BP and in turn to postural instability and falls. The prevalence of postural (orthostatic) hypotension in the community is one in five older adults.

On standing, physiological responses should ensure that BP remains the same. This requires:
- Normal plasma volume, i.e., the person must not be dehydrated (younger people compensate well for this)
- Intact baroreceptors and reflexes
- Venomotor and arterial tone, i.e., the ability of blood vessels to respond to changes in blood flow.

All of these factors may be impaired in older people. Symptoms of postural hypotension (**Box 49.13**) typically occur on or soon after standing; they are rare when a person is sitting or lying down unless they have been lying for a long period and then attempt to sit upright or stand. Some definitions of 'significant drop' are provided in **Box 49.14**.

Aetiology

Acute postural hypotension is seen with a reduced intravascular volume, such as decreased oral intake, excessive diuresis, or excessive fluid losses as in diarrhoea, vomiting, or haemorrhage.

Chronic causes are principally:
- ***Medications:*** classically, diuretics, beta-blockers and other anti-arrhythmics, antihypertensives, and vasodilators.
- ***Autonomic dysfunction:*** associated with diabetes mellitus, Parkinson's disease, and multisystem atrophy. (Medications for Parkinson's disease may exacerbate this further.) Some autoimmune conditions can also cause autonomic dysfunction.
- ***Poor vascular compliance:*** seen in ageing blood vessels caused by a reduction in elasticity and increased calcification.

Management

The evidence base behind management of postural hypotension is poor, and only a couple of medications are licensed for use. The goals are to minimize symptoms and improve standing time to enable activities of daily living through improvement of standing BP without excessive supine hypertension (**Box 49.15**).

Delirium

The prevalence of cognitive impairment increases with age, which is reflected in older populations both at home and in hospital (up to 75% of inpatients older than 90 years old in some cohorts). This adds an additional layer of complexity to their healthcare management, with cognitive impairment being associated with an increase in mortality rate and the risk of dependency.

The presence of cognitive impairment makes ongoing management more challenging because of the potential lack of clarity in the patient's recollection of events and their history, their ability to comply with complex treatment pathways, and the increased levels of care that may be needed. For this reason, it is important when assessing an older person to have an objective assessment of their cognition. A number of cognitive screening and assessment tools can be used for this (see **Box 49.5** for some examples).

A collateral history around a baseline functional level is also vital. This helps distinguish dementia, a chronic progressive and typically

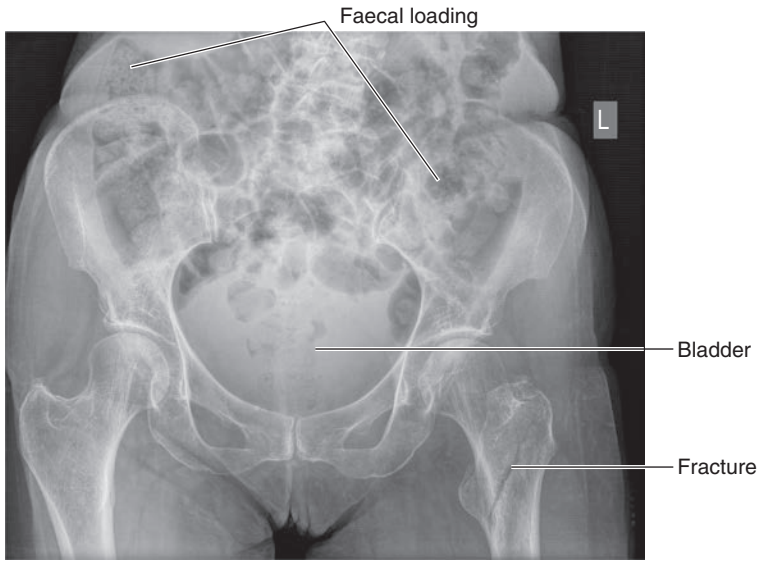


Fig. 49.8 A left-sided intertrochanteric hip fracture. Also of note, there is a large bladder (smooth shadowing in the lower pelvis), likely associated with urinary retention, for which the patient would need a catheter. Faecal loading is also visible, so laxatives should be started.

Box 49.13 Measurement of postural blood pressure

Preparation
- Use a manual sphygmomanometer.
- Lie the patient flat for 5 minutes before starting measurements.

Lying measurement
- With the patient's arm at the level of their heart, record blood pressure.

Standing measurement
- Ask the patient to stand up slowly and steadily.
- Be ready to support the patient if they feel dizzy or faint.
- Measure the blood pressure 1 minute and 3 minutes after standing.
- If the blood pressure continues to drop, continue to record at 2-minute intervals until it stabilizes.
- Record any symptoms the patient experiences.

Interpretation
- A drop of >20 mmHg systolic or >10 mmHg between lying and standing is generally considered to be significant drop.

Box 49.14 Definitions of a 'significant' drop in blood pressure

- ***Classic:*** reduction of 20/10 mmHg within 3 min of standing
- ***Initial:*** reduction of 40/20 mmHg in first 15 sec—a beat-to-beat recording is required
- ***Delayed:*** no agreed definition but occurring >3 min
- ***In hypertensive patients:*** reduction of 30 mmHg systolic; variability is higher in this group

untreatable neurodegenerative disease, from delirium, a short-term change in cognitive function often associated with reversible pathology. The collateral history also helps guide longer-term outcomes and realistic goals for the future. One tool to aid collateral history is MAPLE-V (Fig. 49.9).

People with cognitive impairment are at higher risk for the development of delirium when there is additional physiological load (e.g., in the presence of infection or constipation). In this context, delirium can be particularly long-lasting and florid.

Treatment for the cause of the delirium should be considered for older people as for younger people, depending on the balance of benefits and harms from the treatments and the patient's wishes when they can be ascertained. It is important to have a clear, honest, and supportive explanation of dementia (if applicable) and its effects in older people with the patient and their advocates. An acute hospital setting is not the ideal time for this because delirium may significantly change the patient's presentation, so any cognitive and functional assessments at this time are unlikely to give an accurate representation of their true baseline and may lead to inaccurate or inappropriate diagnoses.

When people are living with dementia, their healthcare can be complex, and optimal management requires multiprofessional input (skilled nursing, therapy, and perhaps psychiatry inputs in particular). Similar to the management of falls, the management of a patient living with dementia in the acute setting is an example of CGA in action. For example, if a person with dementia stops eating and drinking during a hospital admission, the assessment and management of this is complex and needs to take into account numerous factors, such as their general physical health, their oral health, the presence or absence of delirium, medication-related factors, and being in an unusual environment. Skilled specialist healthcare teams learn to work through this complexity while keeping family, carers, and patients up to date.

For patients in hospital with pre-existing cognitive impairment, there are a number of ways to minimize their risks for delirium. These are aimed at optimizing a patient's orientation while addressing any health needs, including nutrition, analgesia, and bowel and bladder care. These relatively simple strategies can reduce the risk of delirium by 35% (Box 49.16).

Managing challenging behaviours

When a person living with dementia starts to experience different or exaggerated behaviours that would not have been usual for them, it can have a significant impact not just for the person with dementia but also their family or carers.

Many of the behaviours, which are called 'behavioural and psychological' symptoms of dementia (BPSD), are not actually

Box 49.15 Management options in postural hypotension

Recommendation	Detail	Explanation
***A*bdominal and leg binders**	'Corsets' beneficial in ~50%; usefulness comes down to practicalities Compression stockings (TEDs): full-length, grade 2 and above; logistically difficult to put on or to tolerate; below-knee TEDS unhelpful because amount of blood in the calves is relatively small	Large volume of blood and baroreceptors in splanchnic and mesenteric circulation
***B*olus of water**	Drinking water: conflicting evidence: needs to be fluid bolus, 500 mL of water within 5 min with benefit lasting around 90 min; useful if timing of hypotension is predictable, e.g., first thing in the morning	Fluid bolus causes a change in the osmolality of the portal blood supply and induces an autonomic response in blood pressure
***B*ed elevation**	Elevating head of bed to 30 degrees from horizontal	Lower renal perfusion pressure leading to decreased diuresis and therefore fluid retention to maintain peripheral capillary pressure on standing up Another benefit is avoidance of supine hypertension in patient with impaired baroreceptor responsiveness lying flat
***C*ounter-manœuvres**	Standing on the toes; pelvic and abdominal clenches Crossing the legs and squeezing the buttocks, thighs, and calves while standing Standing slowly: education programmes with physiotherapy	Physical manœuvres to increase blood return to the heart by improving peripheral resistance Engaging venous return of mesenteric and splanchnic blood volume Unclear mechanism
***D*rugs (limited evidence)**	Fludrocortisone Midodrine	Increases fluid retention; can lower potassium Alpha-adrenergic agonist increasing peripheral vascular resistance: can cause dangerous supine hypertension, aggravate angina, or precipitate urinary retention
***E*ducation for self-management**	STOP what you are doing when you recognize symptoms coming on SIT down: allow blood pressure to recover DRINK some water, the most modifiable factor in the immediate term THINK about triggers: recognize patterns to manage	Patient understanding that their blood pressure responses are not normal and that they need to find ways to adapt to live with this in longer term
***E*xercise**	Moderate exercise Practicing standing slowly to avoid symptoms and getting up from floor	Improves orthostatic tolerance
***F*luid intake**	Keeping well hydrated	Maintains plasma volume

symptoms of dementia itself but are other symptoms or problems that cannot be communicated because of the impact of dementia on language and communication. BPSD can affect up to 90% of people with dementia at some course in their illness, and symptoms can include:

- Agitation: up to 75%
- Wandering: up to 60%
- Depression: up to 50%
- Psychosis: up to 30%
- Screaming or calling out: up to 25%
- Aggression: up to 20%
- Sexual disinhibition: up to 10%
- Others: inappropriate urination or defecation, disrobing, perseveration, repetitive questioning, hoarding or hiding, eating inedible objects, resisting personal care

It can be difficult sometimes to determine if symptoms may be related to delirium or BPSD. The latter is more likely if the answers to all of these questions are yes:

- Does the person have known dementia?
- Is it in the severe stages?
- Are the behaviors present for a longer than a short period?
- Has a physical cause been ruled out?
- Is there significant persistent agitation, psychosis, or apathy?

Understanding potential causes guides management and the PIECES framework help provide a structure for this (see **Box 49.17**).

Using the information gained in **Box 49.16** will help to piece together patterns and a better understanding of the situation and therefore how it can be managed in a person-centred way. It is important to understand exactly what the 'problem' is, in a similar way that you would approach understanding someone's pain: What are the behaviours that challenge, and for how long have they been present?

- Are there triggers, alleviators, or exacerbators?
- What has been the natural course of this?
- What, if anything, is tipping the balance now?
- Where does this sit in terms of a person's premorbid personality, any known diagnoses of dementia, its severity, and existing support informally and formally (including medications)? Where are the opportunities for optimizing these?

Urinary tract infections

UTIs are common in older adults. However, they are also commonly misdiagnosed or overdiagnosed without clear evidence. This can result in potentially dangerous overuse of antibiotics and overlooking (or missing) the actual cause of the patient's presentation.

Fig. 49.9 **Taking a collateral history for cognition.**

'Recurrent' UTIs in particular should be questioned. Often evidence is lacking, and alternative diagnoses, such as vulvovaginal atrophy, should be investigated.

A diagnosis of UTI is likely with a short history of new-onset dysuria alone or two or more of:

- Frequency or urgency
- Incontinence
- Delirium
- Worsening mobility
- Suprapubic discomfort
- Fever
- Visible haematuria

Investigations

Urinalysis in the form of a urine dipstick, looking for nitrites, white blood cells (WBCs), or protein, has traditionally been used to 'diagnose' UTIs, but a positive result for presence of bacteria may indicate colonization rather than infection. A negative result has a better predictive value for the absence of an infection.

A midstream specimen of urine (MSU) should guide antibiotic use before their initiation if there is diagnostic uncertainty. This may help reduce microbial resistance in the longer term. Urine dipstick analysis is of limited value (**Box 49.18**).

Asymptomatic bacteriuria is bacterial colonization without active infection in the urinary tract. It can cause false-positive results on urine dipsticks and raised WBC counts on MSU samples.

Management

- If possible, wait for urinary culture result and tailor initial antibiotics to this.
- If treatment is needed earlier (e.g., in sepsis) or if clinical suspicion is high, treat with empirical antibiotics until the MSU is available; then review with MSU sensitivities.
- Consider changing a catheter as soon as possible if it has been *in situ* for more than 7 days before the onset of infection.

Box 49.16 Management of cognitive impairment in the hospital setting

Interpersonal skills	Introduce yourself each time you see them and talk through what you are doing. Use visual clues as you go through the history (e.g., point to the abdomen when asking about bowels or bladder function). Think about validation therapy (e.g., if the patient is asking for their parents who are known to be deceased, ask the patient to talk about them; this fosters a similar feeling to their being present). Use non-verbal communication techniques. Use distraction and behavioural approaches when agitation builds.
Orientation	Ensure hearing aids are *in situ* and have a functional battery. Examine for ear wax and address it if necessary. Clean glasses and put them on. Make calendars and clocks available, accurate, and clearly visible. Provide orientation boards (with name of location, date, names of staff, and so on). Avoid moves either within the ward or between wards.
Bowels and bladder	Ensure the bowels are open regularly. Remove urinary catheters whenever possible, Encourage regular toileting.
Medication use	Avoid sedation and long-term antipsychotics. Avoid opioids unless needed for analgesia. Review changed medications (including alcohol withdrawal). Minimize anticholinergic burden.
Pain	Regular paracetamol has been shown to reduce agitation in those who may have pain they are unable to articulate. Non-verbal pain scales (e.g., Bolton or Abbey) may be helpful in assessment.
Promote a regular daily cycle, including sleep–wake pattern	Keep to regular routines on the ward or care setting. Keep disturbances overnight to a minimum (e.g., lights off, healthcare staff not chatting loudly, routine observations not done while the patient sleeps). Ensure mealtimes are regular. Try to avoid long or late siestas. Melatonin can be used in the short term to correct day–night reversal.

Recurrent suspected urinary tract infections

- Hold off antibiotics if possible and send two or three MSU samples.
- Consider alternative diagnoses if MSUs are repeatedly negative, including vulvovaginal atrophy in peri- and postmenopausal females, sexually transmitted infections, urethritis, and prostatitis.
- Prostatitis should also be considered in males with recurrent UTIs and positive MSUs, and a longer course of antibiotics may be required.
- Ensure that non-specific symptoms are not being inappropriately attributed to UTIs.
- In truly recurrent UTIs, specific management is often required (**Box 49.19**).

Box 49.17 PIECES Framework for assessment in behavioural and psychological' symptoms of dementia

Physical problem or discomfort	Acute medical problem: delirium Drugs (polypharmacy, side-effects), alcohol, other substances Chronic disease symptoms Pain (under-recognized and undertreated) Basic human needs (e.g., tiredness, hunger, thirst
Intellectual or cognitive changes	Type of dementia The As: anosognosia, amnesia, aphasia, apraxia, agnosia, altered perceptions, apathy
Emotional	Depression Anxiety Psychosis Psychological trauma from earlier experiences Adjustment difficulties Difficult interactions with carers (important to identify burnout in carers; acute presentation to services may not actually be because of the behaviour getting worse but because of carers being worn out or having their own mental health problems) Aggression
Capacities	Boredom Unused energy or previous habits (e.g., 'wandering' may actually be quite purposeful even if mistimed or out of sync with other care home routines) Changes in interest or difficulty keeping up with previous interests Demands exceed capabilities (frustration and catastrophic reactions)
Environment	Relocation, feeling lost, needing to get 'home' Ambience: excessive or distressing noise, unfriendly, confusing environment, poor lighting
Spiritual and cultural	Spiritual and cultural needs Language (e.g., carers not speaking the right language) Returning to previous language Family and friends

Hung L, Lee PA, Au-Yeung AT, Kucherova I, Harrigan M. Adopting a clinical assessment framework in older adult mental health. J Psychosoc Nurs Ment Health Serv *2016; 54(7):26–31.*

Urinary incontinence

Up to 50% of females will be incontinent at some time in their lives. For males, urinary tract symptoms and difficulty passing urine are also common but less well studied. Incontinence is under-recognized and undertreated. Many people do not volunteer their symptoms; even if asked directly, not everyone is comfortable talking about it.

Incontinence has a huge impact on a person's quality of life and often on their relationships with partners and family. Poorly controlled urinary and faecal incontinence are common reasons for admissions to residential care because of carer strain. It is independently associated with falls and is a modifiable risk factor for them. For those who are bedbound, incontinence can cause local irritation in areas prone to pressure damage, increasing the risk of breakdown of skin integrity.

Box 49.18 Urine dipsticks in the diagnosis of urinary tract infection

Public Health England guidance (2018) suggests not using urine dipsticks in the over-65 population.

- In conjunction with symptoms, dipsticks have a sensitivity of around 77%.
- In the absence of symptoms (e.g., as a screen in a patient with confusion), sensitivity falls to 44%.
- The frequency of positive urine dips in institutional care is very high: up to 50% of females in nursing homes have asymptomatic bacteriuria. In non–care-home residents, the figure is lower, ranging from 6% to 16% in females aged 65–90 years.
- The negative predictive value of dipsticks is 100% for nursing home residents.
- In the presence of long-term catheters, dipsticks will almost always be positive (colonization).

There are four main types of urinary incontinence, but there is often overlap between them.

Stress incontinence

Urine leaks from the bladder due to higher pressures than the pelvic muscles are able to contain. This is commonly caused by weakness or damage to the pelvic floor muscles, and previous pregnancy is a risk factor. Typically, stress incontinence occurs on coughing, laughing, or straining.

Management may be conservative or surgical. Conservative measures can cure stress incontinence and so are first-line treatment (**Box 49.20**). If they fail, surgical referral should be considered. Surgery aims to support the pelvic floor either through suspension of the bladder higher up from the external sphincter or by bulking the pelvic floor itself.

Urge incontinence (overactive bladder)

The bladder becomes irritable, causing the person to feel that they urgently need to empty it when the bladder is not actually full. Symptoms include frequency, urgency, and nocturia. If these are of new onset, the trigger may be a UTI. Chronic infection should be excluded as a cause.

When a history is taken, the questions need to be broader than in stress incontinence. Many people will have adapted their lifestyles to minimize the risk of being incontinent, so questions might include:

- Do you ever not make it to the toilet in time?
- When you feel the urge to go to the toilet, are you able to hold it, or do you need to go quickly? What happens if you do not?
- If you are going out, do you plan your day around the availability of toilets?
- Do you use continence pads?

Urodynamic studies can be used to monitor the pressure within the bladder, and electrical activity of the bladder muscles. Intravesicular pressure monitoring shows inappropriate pressure rises associated with electromyography (EMG) evidence of bladder muscle contraction despite small bladder volumes.

Management begins with conservative measures, as for stress incontinence (see **Box 49.20**) Additionally, it includes avoiding bladder irritants, including dehydration (concentrated urine) and caffeine. Bladder training aims for a slow increase of the amount of time from feeling the urge to pass urine to micturition. If these measures fail, some patients benefit from medications such as antimuscarinics (e.g., oxybutynin, tolterodine). Around half of patients will have stopped taking the medication by 3 months, and only one-third will

Box 49.19 Recurrent urinary tract infection

- ***Structural abnormalities*** require imaging ± urological referral for cystoscopy.
- ***Vulvovaginal atrophy*** mimics symptoms, as well as increasing the risk of urinary tract infections; topical oestrogen reduces frequency.
- ***Constipation*** is a cause of obstruction.
- ***Hygiene practices*** require a review of care needs if not self-caring.
- ***Post-treatment midstream specimen of urine*** permits a check for bacteriuria and resistances.
- ***Prophylactic antibiotics*** may reduce events; however, there are risks from long-term antibiotic use (e.g., pulmonary fibrosis from nitrofurantoin).
- ***Increased infection susceptibility*** is associated with poorly controlled diabetes and immunosuppression.

Box 49.20 Conservative measures for managing stress incontinence

- ***Pelvic floor strengthening*** has a higher success rate when supervised but can be performed independently.
- A ***bladder diary*** determines whether any other types of overactive bladder are present that require treatment.
- ***Review of fluid intake*** will note caffeine, high fluid intake, and diuretics, which increase urine output and therefore enhance the potential for stress incontinence.
- ***Weight loss*** (body mass index >30 kg/m^2) puts less strain on muscles.
- ***Optimization of respiratory disease management*** reduces coughing, if present; stop angiotensin-converting enzyme inhibitors.

still be taking it at 1 year because of anticholinergic side-effects, which include dry eyes or mouth, constipation, and confusion. The latter means that these drugs are relatively contraindicated in those with cognitive impairment. Beta-3-agonists, such as mirabegron, should not be used in people with significant hypertension.

Overflow incontinence

This is caused by bladder outflow tract obstruction. The most common cause is prostate enlargement, and thus this is a disorder commonly seen in older males (see Chapter 47, 'Benign enlargement of the prostate gland' section). Other precipitants include medications (e.g., anticholinergics) and constipation. Symptoms include difficulty initiating micturition, poor stream, and terminal dribbling (i.e., further urinary leakage after completion of micturition). Overflow incontinence can be associated with renal failure caused by obstructive uropathy when the system is under high pressure. It can be overlooked because patients still pass some urine but without complete bladder emptying; often patients pass small volumes of urine frequently. Catheterization may be required, either in the short or long term, to alleviate the obstruction.

Functional incontinence

This involves the patient's being physically unable to reach the toilet rather than there being a primary urogenital problem. For example, someone with poor mobility may be unable to get to the toilet in time but would otherwise be continent if their mobility improved. In states of impaired cognition, such as delirium or advanced dementia, the signals for micturition are no longer interpreted normally, resulting in the passing of urine without conscious processing and thus incontinence. Management requires CGA to address any modifiable components, such as pain control, to allow improved mobilization. This can include environmental adaptations, which may make toileting easier. Regular toileting can be helpful for those with advanced dementia and good mobility; this involves prompting the person to go to the toilet at regular intervals, before they need to go urgently, to minimize the risk of incontinence.

Box 49.21 Non–home-based living options in the UK

- ***Care home:*** This can mean a residential or nursing home.
- ***Residential home:*** This is suitable for people who need minimal assistance with mobility, transfers, and some personal care.
- ***Nursing home:*** This is suitable for people requiring significant support, supervision, and care, with trained nurses present to support delivery.
- ***Warden-assisted living or sheltered accommodation:*** Residents have their own flats. A warden is accessible at certain times and is able to provide some support, but not with personal care. Often there are communal areas.
- ***Extra-care sheltered accommodation:*** This is sheltered accommodation that also has on-site carers who can provide additional care more flexibly.
- ***EMI home:*** These homes are for the 'elderly mentally infirm' and specialize in caring for patients with challenging behaviours, typically caused by dementia.

COMPLEX DISCHARGE PLANNING

Complex discharge planning does not just refer to a process that will need a lot of paperwork. It is also a complex balance of managing risk, expectations, and resources. It requires close multidisciplinary working across primary and secondary care to fit best with a patient's priorities for the way they wish to live their lives to the fullest.

In its simplest form, discharge destination from an acute hospital can be directly to home or to an intermediate care facility for further rehabilitation or convalescence. (The most common discharge destinations in the UK are outlined in **Box 49.21**.) A number of factors affect a patient's discharge and destination, which interact to make it potentially complex (**Box 49.22**). There is often very little for patients to do in hospital or a rehabilitation unit, so they do not do very much. Meanwhile, at home, they are more likely to do more, with day-to-day 'pottering around' in their own homes contributing to enhanced recovery, provided they have the necessary support in place to do so.

For discharge planning to be truly effective, it needs to be started at the beginning of a patient's admission and regularly reviewed with the MDT caring for the patient, with an understanding that the initial plans may be adapted and changed. Regular discussion with both patients and their relatives and carers is important. Traditionally, when patients are preparing for discharge, assessments are carried out in hospital; however, this may involve exploring the patient's mobility in a place where distances may be significantly greater than at home, and kitchen assessments take place in an unfamiliar environment. Occupational therapy home visits evaluate a patient's ability to function within their own home environments, and people often do significantly better there. Some schemes, such as the original Discharge to Assess (D2A) programme, complete all assessments in the patient's home on the presumption that they will do better in their own environment, with essential equipment and care being put in place on the same day if they are going to stay there. This can reduce time spent in hospital and often results in lower care needs than were predicted from hospital-based assessments.

Important considerations

Rehabilitation goals

Rehabilitation is a multidisciplinary process through which an individual's functional level is optimized, including physical, psychological,

Box 49.22 Factors affecting discharge planning

Patient factors

- ***Patient wishes:*** For some people, their priority is to remain in their own homes for as long as possible, but others may feel very trapped and isolated there, preferring somewhere that has other people around, such as a care home.
- ***Medical issues and prognosis:*** Has the issue leading to hospital admission resolved (e.g., hip fracture, pneumonia), or is it likely to be progressive (e.g., lung cancer)? What is the likely prognosis? For those with a short prognosis (e.g., a few weeks), it may be better to take more risks and maximize time at home as opposed to aiming for rehabilitation when there is a longer prognosis.
- ***Functional status:*** How much are patients able to do for themselves? If a patient can walk to the toilet and back, support can be put in place for most other issues. Were they struggling at home before admission? What support do they need overnight, such as with toileting or wandering behaviours? If there are issues with continence, this can make things particularly difficult at home, with faecal incontinence in particular. Mobility is a major issue, with falls being some of the leading reasons for admission to a care home. Falls can occur anywhere, including at home, in hospital, or in a care home, though the latter two have staff to reduce the risk or help after the event.
- ***Cognition:*** This affects a patient's ability to care for themselves and their safety awareness, as well as their ability to call for help appropriately, either through pendant alarms or telephones. Some may need increasing support with meals, either ready-made for the oven or microwave or delivered to the home. Some patients may forget or have poor insight into their cognitive problems, in which case options such as turning off the gas supply to the cooker can make discharge home safer. Reversal of a patient's sleep pattern (day–night reversal) can make it much harder for a spouse to offer support because their own sleep pattern will also be affected. Patients with cognitive impairment tend to function far better in a familiar environment than elsewhere, and home visits from the occupational therapists should be considered before any permanent moves to a care home.

Home factors

- ***Suitability for frailer people or those with cognitive impairment:*** Some homes are more suitable than others. Cramped spaces can be difficult, although it is easier for people with reduced mobility to hold on to the walls. Some beautiful rickety cottages with little steps between each room can be a challenge! Modern purpose-built flats for older people are usually fine.
- ***Stairs:*** These can be a challenge, though having a toilet on every floor makes things easier.
- ***Potential for adaptations:*** Is there space for a hoist? Can the bathroom be adapted to house a walk-in shower? Can a stairlift be installed?

Social factors

- ***Co-habitation:*** This can provide physical support, such as assistance with activities of daily living (from washing and dressing to housework), as well as someone to call for help if there are any problems. Be aware of the physical and cognitive level of the spouse, who may also be frail with multiple co-morbidities. Live-in carers are another option, though they are expensive.
- ***Financial status:*** Is the patient able to pay for support and equipment?
- ***Family wishes:*** Although patient wishes are vitally important, the family members are the ones who pick up the pieces when things do not work out. Balancing a patient's wishes with what is realistic can be tricky, as can balancing them with the family's wishes, which may not be the same as the patient's. When a patient has capacity for decisions regarding specific issues, these should be respected wherever possible. Beware of abuse (financial abuse and neglect) when the family agree only to an insufficient care package in their attempts to preserve their inheritance.

Resources

- Availability of and funds for carers
- Availability of and funds for equipment
- Access to rehabilitation units
- Local schemes for domiciliary care: can rehabilitation be done from the patient's own home?

and social aspects of function. It can take place anywhere, from the patient's home to acute inpatient settings and even in an intensive care unit, where there are interventions to help preserve longer-term function. Rehabilitation is dependent on working closely with the patient and the wider MDT, including (where relevant) community- and hospital-based services.

Although the emphasis is on returning patients to their premorbid function, a limitation of functional regain, such as after a major stroke, must be recognized. In such cases, helping the patient adjust to the new functional level is a fundamental aspect of the process.

Most commonly, we think of rehabilitation occurring after an acute illness, such as a stroke or pneumonia. However, it also plays a key part in chronic conditions, such as Parkinson's disease and reduced mobility. Increasingly, its role in preparing people for interventions such as surgery is being recognized; this is known as 'prehabilitation' and aims to expedite recovery and improve longer-term outcomes.

A number of factors affect the likely outcome from rehabilitation:

- ***Premorbid status.*** A frailer patient is less likely to improve as much as those who are stronger before their acute illness.
- ***Reasons for deterioration.*** An acute injury or illness, from which full or nearly full recovery might be expected, should be compared with progressive underlying illness, such as cancer or motor neurone disease, in which it is possible to achieve short-term gains, but patients deteriorate over the longer term.
- ***Co-morbidities.*** When possible, these should be addressed through CGA. Areas that bring particular benefit include pain management (e.g., in arthritis of the knee), BP (too low can increase falls risk), and Parkinson's disease (timing sessions to periods of the day when control is better).
- ***Cognitive impairment.*** This may reduce the patient's ability to follow instructions in the intervals between sessions. However, significant improvements may occur, and cognition alone should not preclude rehabilitation. The clinician should also be aware that recognizing and treating pain in patients with dementia can be difficult but that poor control has an impact on progress.
- ***Depression.*** Be suspicious when patients do not progress as expected. Trying to discriminate between depression and an adjustment reaction after an acute illness (e.g., loss of function after a stroke) can be difficult, and monitoring over a period of weeks is often necessary.
- ***Motivation.*** Rehabilitation takes effort on the part of the patient, and there is great variability in how much individuals are willing to engage.

A distinction should be made between patients who have a realistic chance of making a meaningful improvement and those who have no realistic chance of seeing a significant change. One of the key principles of rehabilitation is goal setting, with clear and achievable targets over appropriate timescales. For example, in a patient needing a Zimmer frame and assistance from two people to transfer between a bed and a chair, rather than the goal being 'To improve transferring', a better target would be 'To be able to transfer with assistance of one person and a Zimmer frame within 3 weeks'.

Managing risk

Managing risk is a crucial part of discharge planning. Just as life itself inherently contains risk, risk cannot be removed completely from any discharge. The aim is to recognize harmful, predictable risks and offer ways to minimize them. It is important to remember that everyone has different priorities for what they consider a true risk to be and have differing levels of risk tolerance. Respect should

be shown for people's autonomy to make unwise decisions. The key here is to ensure that the person is informed as to the options so they can make their own decision.

Cognitive impairment

Cognitive impairment can impact decisions about discharge in a few different ways. Regarding decision-making and risk evaluation, the right to autonomy is fundamental but is closely linked with the mental capacity to understand and weigh up these risks. If a risk is not perceived to be real, then an informed decision around it and implications for optimizing safety cannot be made. In this circumstance, involving an advocate is vital. This can be anyone who knows the person well enough to share what the person would be likely to express for themselves if they were able to. This is usually a family member but does not need to be; it can be any person who knows the patient well enough to act in their best interests. Under UK law, 'next of kin' has no legal standing, but their opinion should be given significant weighting in decision-making; legally, only an allocated power of attorney or court of protection–appointed guardian can make a decision on behalf of a person, as though they are that person.

When considering discharges for those with cognitive impairment, it is important to consider whether the current presentation is caused by delirium or established dementia. In the latter, care and support needs are likely to be relatively stable and predictable. If the person has delirium, then their care and support needs are likely to change as they recover, and a higher intensity support option in the shorter term is likely to be necessary. Remaining in hospital may be the physically safest option but may hinder cognitive recovery. Hospitals can be inherently 'deliriogenic' and disrupt a person's usual orientation and coping mechanisms, so the two need to be balanced carefully.

Incontinence can become a particular challenge posed to being able to be at home, especially in advanced dementia. If a person is no longer able to recognize when they need to go to the toilet or are unable to get to the toilet unaided or obtain help when needed, incontinence of both urine and faeces may occur. This is challenging for those caring for, providing support to, or living with this person from many perspectives by affecting hygiene, physical health, mental wellbeing, and relationships. When someone needs assistance to go to the toilet in a home environment, a key discharge decision will need to include what will happen if they are alone and unable to toilet themselves (e.g., between care calls, when a caregiver is out, overnight). This is often a deciding factor in opting to move into residential support.

Social and functional support

If someone is returning home or a decision is being made between returning home or to another environment, the availability of social and functional support needs to be clear so an informed decision can be made. This might include the use of equipment and adaptations around the home to make activities of daily living easier and provide formal health or social care support (e.g., community nursing for complex medications or skin care, carer visits for personal care). Ensuring the right level of support from the right people is particularly important when an individual is physically, psychologically, or cognitively vulnerable and may not be able to advocate for themselves fully to ensure the maintenance of their wellbeing and safety.

Carer strain

In the UK, there are around 5 million unpaid carers who provide informal support to family or friends and around 1 million paid carers working in the social care workforce. Many unpaid carers may struggle with the demands of their role. They may lose their sense of self and their own wellbeing or experience an enhanced level of anxiety and stress, particularly if they are subject to physical or verbal abuse or being woken during the night.

Other factors that have a major impact include depression in the patient and incontinence, with faecal incontinence being particularly difficult for carers to manage. Carers may, in addition, have medical issues of their own. Assumptions should not be made regarding a carer's capacity or willingness either to start taking responsibility or to continue to care for a person. Carers should be actively involved in decisions surrounding a patient's care.

Ways to help reduce carer strain include:

- Recognition and appreciation of what the carer is doing
- **Financial support**, e.g., a carer or attendance allowance
- **Maximizing performance status**, e.g., making transfers easier
- **Addressing co-morbidities**, e.g., treating depression
- **Periods of respite**, e.g., day centres, where the carer can take some time off; a sitting service, where someone besides the main carer stays with the patient for a few hours to give the main carer a break; rolling respite, where the patient spends 1–2 weeks every 2 months (e.g., in a care home); and asking friends and family to help for periods, either regularly or on a more ad hoc basis
- Holidays, provided by companies who specialize in holidays for people with health conditions, with carers on site; often, the patient and carer will travel together.

Older adult abuse and safeguarding

Abuse of older adults is defined as a single or repeated act that causes harm or distress to an older person or as a lack of appropriate action. It can take place within any relationship in which there is an expectation of trust. There are a number of different types of abuse (**Box 49.23**), and it is quite common for a number of these to occur simultaneously.

Older people are particularly vulnerable to abuse. Worldwide, the World Health Organization (WHO) has found that approximately 1 in 10 older people experiences abuse each month. The WHO found that two-thirds of older people with dementia have been abused. Obtaining clear data is very difficult, and there is concern that these figures are underestimates. Some risk factors for abuse are shown in **Box 49.24**.

Older people are often afraid to talk about cases of abuse to family, friends, or the authorities, with fewer than 1 in 20 cases being reported. Even when abuse is recognized, many victims are unwilling to discuss or pursue complaints, especially if family members are involved.

Cognitive impairment is a major risk factor for abuse, in which recall of specific issues around the abuse is patchy and unreliable, making clarification of concerns difficult. It is essential to be aware of the issues and to have a low threshold for escalating possible concerns.

Although older males have the same risk of abuse as females, older females are at higher risk of neglect and financial abuse in some cultures where females have unequal social status, particularly when they are widowed. Females may also be at higher risk of more persistent and severe forms of abuse and injury.

Safeguarding is the protection of people's health and wellbeing, enabling them to live safely and free from neglect or abuse in conjunction with their own wishes and beliefs. Preventative strategies include:

Box 49.23 Some types of abuse

Type	Features	What to look for
Physical	Hitting, scratching Misuse of medicine, both undermedicating (e.g., inadequate analgesia) and overmedicating (e.g., excessive sedation) Forcible feeding or withholding of food Restraint	Unusual pattern of injuries Absence of clear explanation Frequent injuries for which medical treatment has not been sought
Sexual	Any sexual activity the patient cannot or does not consent to	Bruising, particularly to the thighs, buttocks, and upper arms Bleeding, infections, pain, sexually transmitted disease, or itching in the genital region
Psychological or emotional	Social isolation Removal of mobility or communication aids Preventing the patient from meeting their religious needs	Change in psychological state in the presence of another person Low self-esteem or withdrawn appearance
Financial or material	Preventing the patient from accessing their own money or assets Arranging less care than is needed to save money (e.g., to maximize an inheritance) Moving into a person's home rent-free or under duress	Recent changes, e.g., to property ownership Obtaining lasting power of attorney after patient is assessed as lacking mental capacity to manage their finances
Discriminatory	Unequal treatment based on age, cognition, and so on	
Organizational or institutional	Settings such as nursing and residential homes or hospitals Lack of respect for dignity and privacy, e.g., during bathing or toileting	Inappropriately prioritizing care of younger patients Precluding all patients with dementia from rehabilitation regardless of their potential to improve
Neglect or acts of omission	Denial of access to basic needs, including food, shelter, clothing, heating, and personal and medical care Failure to administer medications appropriately	Malnourished appearance Untreated medical conditions or injuries Pressure damage
Self-neglect	Failure to carry out personal hygiene, when behaviour impacts on safety Squalid or unsanitary conditions Unkempt appearance Non-compliance with care	Health risk to neighbours

Box 49.24 Risk factors for abuse

Patient factors
- Frailty or any serious health condition
- Physical or sensory disability, e.g., reduced mobility, poor vision
- Memory problems or dementia, other mental health problems, learning disability
- Previous episodes of abuse
- Low income or socioeconomic status

Social factors
- Substance misuse by abuser
- Mental health problems in abuser
- Shared living situation
- Abuser's dependency on the older person (often financial)
- Difficult family relationships
- Social isolation

Care provider factors
- Poor staff training or support, e.g., in institutions
- Policies operating in the interests of the institution rather than the residents
- Poor resources, e.g., to pay for carers or provide suitable environment

- Clear support, education, and guidance for patients and caregivers
- Education at an institutional level and clear guidelines on the management of specific conditions, e.g., BPSD
- Involvement of patients and advocates in the development of policies and guidelines
- External assessments or care providers, e.g., inspections by the Care Quality Commission within the UK.

Recognizing abuse can be difficult. There should be a low threshold for escalating concerns to relevant specialists, who can then investigate and act when necessary. Often, this is led by social services.

Loneliness

As people age and experience associated disability and frailty, their personal worlds may progressively shrink, in some cases to a single room. This is commonly associated with an increase in social isolation and loneliness. Approximately 50% of the population older than 75 years of age in the UK live alone, with around 10% of older people reporting loneliness. There are two main facets to loneliness:

- ***Emotional:*** stemming from the absence of an intimate relationship or a close emotional attachment (e.g., a partner or a best friend)
- ***Social:*** stemming from the absence of a broader group of contacts or an engaging social network (e.g., friends, colleagues, and people in the neighbourhood)

Loneliness may be regarded as a new 'geriatric giant' and is linked with numerous negative health outcomes. The negative impact of loneliness on health is substantial: it is equivalent to smoking 15 cigarettes a day, and people who have a high degree of loneliness are twice as likely to develop Alzheimer's disease. Conversely, people taking part in health- and independence-maintaining behaviours are less likely to feel isolated and more likely to feel that their community is a good one to grow old in. Loneliness has complex causes, and the most effective schemes to deal with it take a broad view, targeting higher-risk groups, including those from lower socioeconomic backgrounds, widowed people, the physically

isolated, people who have given up driving, and people with sensory impairments.

Further reading

Ahmed H, Farewell D, Jones HM et al. Antibiotic prophylaxis and clinical outcomes among older adults with recurrent urinary tract infection: cohort study. *Age Ageing* 2019; 48:228–234.
Anyanwagu U, Mamza J, Donnelly R et al. Relationship between HbA_{1c} and all-cause mortality in older patients with insulin-treated type 2 diabetes: results of a large UK Cohort Study. *Age Ageing* 2019; 48:235–240.
Schaeffer AJ, Nicolle LE. Urinary tract infections in older men. *New Engl J Med* 2016; 374:562–571.

AGEING WELL AND ADVANTAGES OF AGEING

Amid the common negative rhetoric around the challenges that face healthcare and society at large, we should not forget that older adults are exactly that: adults who have grown to an older age, bringing their personalities, experiences, and individuality with them. The National Health Service in the UK was originally set up to deliver emergency and life-saving treatments for predominantly single-organ conditions. Emergency department attendances are increasingly characterized by relapses in chronic conditions, the consequences of a fall, a breakdown in social care, or functional difficulties such as a decrease in mobility. The nature of inpatient work has shifted because of the incredible successes that healthcare has delivered over the past century, reflecting its fundamental goals: to help all members of society live their lives to their greatest participation within society and to optimize their wellbeing.

Societal and economic advantages of an active older population

In high-income countries, there is an increasing and active population of older people who are making major contributions to society. An Age UK study found that in 2013, older people contributed an estimated £61 billion to the economy in England (>5% of gross domestic product), which was six times the amount spent by local authorities on social care. The main financial contributions came through continued working, informal caring, providing childcare for grandchildren to enable their children to work, and volunteering. Clearly, the benefits extend far beyond the financial aspects, with older people's skills, knowledge, and experience supporting many aspects of society.

Health and wellbeing

The Institute for Public Policy Research (2009) has identified five essential elements of wellbeing:

- Resilience
- Independence
- Health
- Income and wealth
- Having a role and time

Building resilience within an individual and their support network is increasingly seen as a buffer against the effects of frailty. This may be achieved by optimizing cognition, mood, social networks, mobility, and physical reserves.

The presence of disease should not be equated with a lack of wellbeing or meaningful life. A study by Strawbridge et al. in 2002 found that 50.3% of older adults considered themselves to be ageing successfully compared with 18.8% classified as ageing successfully using 'objective' criteria. Although absence of chronic conditions and maintenance of functioning were positively associated with successful ageing on both definitions, many participants with chronic conditions and with functional difficulties still rated themselves as ageing successfully.

There is good evidence to support mid-life interventions as being beneficial in improving health and reducing frailty in later life. These include:

- Smoking cessation
- Promotion of physical activity
- Reduction of alcohol consumption
- Improvement of diet
- Maintenance of a healthy weight

Healthcare systems need to adapt to the changing demographics of the ageing population, shifting towards greater prevention of disease and proactive care with better integration of health and social care services. This would mean that older people age well and stay well and will have the support they need for the management of complex co-morbidities. Key to this is information sharing because older people's healthcare is highly complex and integrated.

Further reading

McCormick J with Clifton J, Sachrajda A et al. *Getting On: Wellbeing in Later Life*. Institute for Public Policy Research 2009; https://www.ippr.org/files/images/media/files/publication/2011/05/getting_on_1744.pdf.
National Institute for Health and Care Excellence. *NICE Guideline 16: Dementia, Disability and Frailty in Later Life—Mid-life Approaches to Delay or Prevent Onset.* NICE 2015; https://www.nice.org.uk/guidance/ng16/chapter/1-Recommendations.
Oliver D, Foot, C, Humphries R. Making our health and care systems fit for an ageing population Kings fund. 2014.
Strawbridge W, Wallhagen M, Cohen R. Successful aging and wellbeing: self-rated compared with Rowe and Kahn. *Gerontologist* 2002; 42:727–733.

Questions for this chapter can be found on Elsevier eBooks+.

50 Palliative care and symptom control

Chloe Chin, Catherine Moffat, and Richella Ryan

Background and general aspects 1720
Who provides palliative care? 1721
When should palliative care needs be assessed: problems rather than prognosis? 1721
What are the patient's needs, and what is the patient's understanding? 1722
Advance care planning: planning for the future 1722
Difficult conversations 1722
Do not attempt cardiopulmonary resuscitation orders and treatment escalation plans 1723
Symptom control 1723
Pain 1723
Gastrointestinal symptoms 1726
Anorexia, weight loss, and malaise 1727
Respiratory symptoms 1727
Other physical symptoms 1728
Psychosocial issues 1729
Palliative care in non-malignant disease 1729
Heart failure 1729
Chronic respiratory disease 1731
Renal disease 1731
Liver disease 1731
Neurological disease 1733
Palliative care for young adults 1733
Recognizing dying 1733
Care of the dying 1733

CORE SKILLS AND KNOWLEDGE

Palliative care is the active total care of patients who have advanced, progressive, life-shortening disease. It is based on needs, not diagnosis, and is required in non-malignant conditions as well as cancer. Knowledge of the principles of palliative medicine is essential for all healthcare professionals and can be applied throughout medical practice so that all patients, across all care settings, have access to and receive good care. All specialties treat people who are sick enough to die, and the recent COVID-19 pandemic has highlighted the importance of clinicians' ability to provide good end-of-life care.

Palliative care is delivered by multidisciplinary teams (MDTs) that include doctors, nurses, therapists, chaplains, psychologists, and social workers. It may be delivered in patients' own homes; during hospital admissions or clinic visits; and through hospices, which provide both inpatient and outpatient services. It aims to address all aspects of a patient's care needs, promote independence, and support patients and their families through the dying process.

Key skills to acquire in palliative care include:

- Learning to talk openly and sensitively with patients and families about disability, uncertainty, deterioration, and death
- Learning to manage common symptoms such as pain and nausea
- Working within an MDT to provide holistic care for patients.

BACKGROUND AND GENERAL ASPECTS

A palliative care approach is relevant to both malignant and non-malignant conditions and may be applicable well beyond the last weeks of life. It is increasingly common for palliative care to be delivered alongside active treatment in recognition that palliative and active approaches may have complementary effects. This dual approach is particularly relevant for patients with chronic non-malignant disease, such as organ failures (heart, lung, and kidney) and neurodegenerative diseases, who may live many years with a symptom burden similar to or greater than people with cancer, as well as the increasing numbers of cancer patients who are surviving longer but requiring support with long-term sequelae of their disease and its treatment. An early palliative care approach with access to specialist palliative care (SPC) teams for complex problems may therefore be beneficial (**Box 50.1**).

The ***goal*** of palliative care is to achieve the best possible quality of life for patients and their carers by:

- Managing physical, psychological, social, and spiritual problems to provide excellent symptom control
- Enabling patients to be cared for and to have the best chance of dying in their preferred place
- Enabling the acceptance of death as a normal process when life-prolonging treatments no longer improve or maintain quality of life
- Providing the opportunity to say goodbye and reduce complicated grief.

Box 50.1 Key points in palliative care and key components of a modern palliative care service

Key points in palliative care

- Patients should always be involved in decisions about their care.
- Quality of life is increased when treatment goals are clearly understood by everyone, including the patient and carer.
- The multidisciplinary team provide a high standard of care, but there must be realism and honesty about what can be achieved.
- End-of-life care is delivered in all settings: the patient's usual place of residence, hospitals, and hospices.
- Care at home should be encouraged for as long as possible, even if the patient's preferred place of death is elsewhere.
- Discussions about end-of-life care planning are best held outside times of crisis, with clinicians with whom the patient has a good relationship. An unplanned admission can be a trigger for advance care planning discussions.
- Discussions must be recorded and made known to everyone involved in the patient's care.

Key components of a modern palliative care service

- Management based on *needs*, not diagnosis: the symptom burden of non-malignant disease often equals that of cancer.
- Care that is independent of the patient's location and that helps patients to remain at home if possible, avoiding unwanted admissions to hospital.
- Rehabilitation for people with advanced disease.
- Support for carers.
- Bereavement care for people with complicated grief.
- Advice for other clinicians; dissemination of palliative care knowledge.
- Education (surrounding palliative and end-of-life care) for clinicians of all disciplines and all levels.

Who provides palliative care?

All healthcare providers should have basic palliative care skills and know how to access SPC services (see **Box 50.1**) for support with more complex problems. A hallmark of a SPC service is the MDT because single professionals cannot provide the breadth of necessary expertise. The emotional demands of working in this area require team support to enable balanced, compassionate, and professional care. Members of the team can include clinical nurse specialists, doctors, psychologists, therapists (occupational, physical, music, and creative), chaplains, and social workers. In some centres, the term ***supportive care*** is used instead of 'palliative care'. Teams may be based in hospitals, hospices, or in the community because palliative care crosses all healthcare settings. Clinicians should be aware of the services that their local SPC teams offer and recognize when referral is appropriate.

Patients may have close relationships with their usual medical team(s), and an integrated approach, across a number of settings, may be needed to allow optimization of disease-directed interventions as well as palliative care. Patients may have access to supportive and palliative care clinics, sometimes called 'enhanced supportive care', alongside hospital medical clinics early in their healthcare journey. They may use inpatient SPC services in hospices for a limited period (days to weeks) for complex problems to be intensively addressed, or they may be supported by community palliative care teams in their homes. Excellent communication between all members of the healthcare team and between the patient and carer underpins the successful management of advanced disease and end-of-life care. Effective liaison between the hospital and community teams is also essential.

Box 50.2 Problems arising when specialist palliative care is delayed until the end of life

- There is insufficient time to achieve good symptom control by combining non-pharmacological and pharmacological components.
- Specialist palliative care services are deemed less acceptable by patient and family, being associated with 'dying', 'giving up', or 'giving in' to illness.
- Psychological distress and physical symptoms become intractable and contribute to complex grief.
- It becomes too late to adopt a rehabilitative approach or teach or use non-pharmacological interventions that need a degree of training and patient motivation (e.g., cognitive-behavioural approaches, mindfulness meditation, attendance at day therapy).

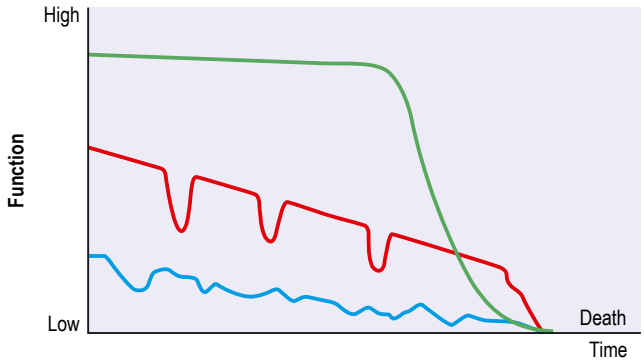


Fig. 50.1 Trajectories of different disease types. *Green* indicates cancer, *red* indicates chronic non-malignant disease, and *blue* indicates multimorbidity and increasing frailty. *(From Murray SA, Kendall M, Boyd K et al. Illness trajectories and palliative care.* BMJ *2005; 330:1007.)*

When should palliative care needs be assessed: problems rather than prognosis?

Palliative care can be beneficial from the time of diagnosis of potentially life-threatening or life-limiting illness even when cure is considered possible ('hoping for the best, preparing for the worst'). A problem-based approach to disease management ensures that patients and carers obtain access to appropriate support services, including SPC, and avoid an either/or approach ('either curative treatment or palliative care'). Palliative care is most effective when considered as part of routine care alongside disease-specific therapy. Early referral links palliative care with quality-of-life improvements; positive associations increase the likelihood that patients and families continue to use palliative care services when they need them over the course of the illness. In cancer, there is good evidence that integrating palliative care and anti-tumour treatment soon after diagnosis reduces long-term distress and even increases survival in selected cases.

If palliative care is seen as relevant only to the end-of-life phase, patients may be denied expert help for complex symptoms. Timely management of physical and psychosocial issues earlier in the course of disease prevents intractable problems later (see **Box 50.2**). It can be difficult to know when to consider palliative care in non-malignant conditions because of their variable disease trajectories (**Fig. 50.1**), but using the 'Surprise Question' (**Box 50.3**) may help clinicians decide. This forms the basis of the UK's ***Gold Standards***

Framework tool that is commonly used to help tailor care to the right time and place.

What are the patient's needs, and what is the patient's understanding?

The causes of a patient's symptoms are often multifactorial, and a holistic assessment (**Box 50.4**) is central to optimum management. Assessing the patient's grasp of the disease, understanding their wishes, and acknowledging their concerns help the team plan and implement effective support. Validated questionnaires, such as the integrated Palliative care Outcome Scale (iPOS), have been developed and are routinely used to aid this.

Further reading

Hui D, Bruera E. Models of palliative care delivery for patients with cancer. *J Clin Oncol*. 2020; 852–865.
Kelley AS, Morrison RS. Palliative care for the seriously ill. *N Engl J Med*. 2015; 373:747–755.

ADVANCE CARE PLANNING: PLANNING FOR THE FUTURE

Throughout the course of illness, it is essential to give patients and families a chance to discuss their possible future options and wishes. Having the skills to support these conversations is everybody's responsibility. These conversations can be had at any point, but they should not be left to the last weeks and days of a patient's life: early discussion of difficult choices and management of uncertainty surrounding their condition(s) is helpful. These discussions are ideally held when the patient is relatively well and outside an acute episode of care. Discussions delayed until the crisis of acute admission may lead to acceptance of invasive and potentially non-beneficial treatment that is later regretted by the patient or their loved ones (see **Box 50.2**).

One in three adults admitted to hospital and 80% of care home residents in the UK are in their last year of life. Talking about dying is not easy, but it allows for future care planning, which can improve patients' quality of life, reduce burdensome treatments at the end of life, and reduce emergency hospital admissions. If a patient is anticipated to lose capacity or the ability to communicate, healthcare professionals must be proactive in initiating discussions.

The content of the conversation may cover ways that patients can plan their care and support in advance, perhaps by recording advance statements, decisions to refuse specific treatments, or assigning lasting power of attorney to individuals important to them who could make decisions on their behalf (see Chapter 3 for more details).

Box 50.3 The 'Surprise Question'

Would you be surprised if this patient died within the next 12 months?

Further resources

Downar J, Goldman R, Adhikari N. The 'surprise question for predicting death' in seriously ill patients: a systematic review and meta-analysis. *CMAJ* 2017; 189:E484–493.

The Gold Standards Framework. https://www.goldstandardsframework.org.uk.

Difficult conversations

Finding a quiet place and allocating plenty of time for the conversation are necessary first steps. It is important to note that conversations may need to take place over several consultations and may not reach a firm conclusion. This is especially the case if there is uncertainty in the trajectory of the medical condition(s) and outcome of current treatments. Honesty and sensitivity are key, and patients generally appreciate it if there is no firm or 'correct' answer to a question as long as it is accompanied by a careful explanation. Before sharing information, it is important to check the patient's understanding of their disease(s), to clarify how much information they would like to receive, and to check if they would like to have a family member or friend present. Patients have differing needs for information and deal with 'bad news' in different ways. A sensitive approach, respecting individual requirements,

Box 50.4 Domains of a holistic assessment[a]

Mnemonic	Domain	Things to consider
P	Physical	Symptom assessment: may involve specific symptom assessment tools or assessment by other professionals (e.g., physiotherapists, specialist clinicians)
E	Emotional	Psychological assessment, including understanding patient expectations: may involve psychological assessment tools (e.g., distress thermometer) and onward referral for appropriate support
P	Personal	Cultural, ethnic, spiritual, and religious needs
S	Social support	Work and finance, family and close relationships, recreational activities: may benefit from social services referral or occupational therapy advice
I	Information and communication	Patient understanding of the plan, multidisciplinary team outcomes communicated, identification of key workers
C	Control and autonomy	Mental capacity to make decisions, patient choice regarding treatment options and plans, preferred place of care and death
O	Out of hours	Advance care planning, awareness of where to access services out of hours
L	Living with your illness	Rehabilitation, management of activities of daily living, future expectations, and end-of-life care planning
A	Aftercare	Funeral arrangements, bereavement needs, future support for the family

[a]In each domain, the tool identifies potential anticipated patient issues and concerns, cue questions to ask patients and carers, and resources for professionals to signpost to.
Adapted from the Gold Standards Framework. https://www.goldstandardsframework.org.uk.

is crucial. Language must be clear and euphemisms avoided, both in discussion and subsequent documentation of the conversation. The importance of silence should not be underestimated; it can be hugely comforting, and pauses can invite a response from the patient. Involving other relevant health professionals in the conversation may also be useful and appropriate.

Do not attempt cardiopulmonary resuscitation orders and treatment escalation plans

Part of talking about the future may include discussion about resuscitation and appropriate escalation plans. The resuscitation status of every patient should be discussed by senior doctors at the time of admission to hospital or when deterioration is noted that is not thought to be reversible. This decision must be clearly documented in the notes. Treatment escalation plans and 'ceilings of care' determining what treatments are likely to be beneficial or futile should be discussed at the same time in line with patient wishes.

Deciding a person's resuscitation status is a careful balance of risk versus benefit. Co-morbidities, pre-morbid quality of life, and the patient's wishes should be taken into account. Although Do Not Attempt Cardiopulmonary Resuscitation (DNACPR) orders and treatment escalation plans are medical treatment decisions, the patient and family must be involved in discussions unless they have specifically requested not to be, and the reasoning behind decisions should be explained. In the UK in 2014, the Court of Appeal ruled in the case of *Tracey v Cambridge University Hospitals NHS Foundation Trust* that a failure to discuss the making of a DNACPR order with a patient who had expressed a wish to be involved in decision-making around their care constitutes a breach of their human rights. Concern regarding possible patient distress is no justification for not discussing resuscitation, though it will guide the way the subject is considered with the patient. Only if it is thought that there will be significant psychological or physical harm resulting from the discussion can it be postponed. If the patient requests that cardiopulmonary resuscitation is not performed in the event of cardiopulmonary arrest, this should be respected.

Remember that a decision not to resuscitate a patient is *not* the same as a decision to withhold other treatment. Patients may still benefit from active treatments such as antibiotics, fluids, endoscopy, and even surgery. Management should remain positive and proactive, allowing the patient to die free of distress and with dignity.

ReSPECT (Recommended Summary Plan for Emergency Care and Treatment) is an initiative implemented in many hospitals and communities. It is a process that creates and records personalized recommendations for a person's clinical care in a future emergency when they are unable to make or express choices. It is useful in those with complex health needs, integrating discussion surrounding DNACPR and treatment escalation.

Further reading

Etheridge Z, Gatland E. When and how to discuss 'do not resuscitate' decisions with patients. *BMJ* 2015; 350:h2640.
Etkind S, Bone A, Gomes B et al. How many people will need palliative care in 2040? Past trends, future projections and implications for services. *BMC Med* 2017; 15:102–112.
Fritz Z, Slowther A, Perkins G D. Resuscitation policy should focus on the patient, not the decision *BMJ* 2017; 356:j813.
General Medical Council. *Treatment and Care Towards the End of Life: Good Practice in Decision Making*. London: GMC; 2010. https://www.resus.org.uk/respect/respect-healthcare-professionals.

Box 50.5 Examples of reversible problems that can be treated to improve symptoms

Patient 1
A patient with metastatic prostate cancer has known spinal metastases at multiple levels. He complains of worsening back pain and leg weakness. This warrants urgent magnetic resonance imaging to rule out metastatic spinal cord compression, which could be treated with steroids and radiotherapy, leading to an improvement in his symptoms and functioning.

Patient 2
A woman with metastatic breast cancer describes a history of general deterioration, nausea, abdominal pain, constipation, and low mood. This warrants a set of blood tests that may reveal hypercalcaemia, which can improve with intravenous fluids and bisphosphonates; there may even be resolution of symptoms.

Patient 3
A woman with metastatic ovarian cancer has a similar history to patient 2. A thorough clinical examination may reveal tense ascites as a cause of abdominal discomfort, nausea, and constipation, which could be alleviated by paracentesis.

SYMPTOM CONTROL

Good palliative care integrates non-pharmacological approaches, such as anxiety management and rehabilitation, with pharmacological control of symptoms. Medications should be rationalized to reduce polypharmacy. A key, and usually initial, component of palliative care is actively identifying and managing reversible problems that could be contributing to inadequate symptom control (**Box 50.5**). Symptom management runs in parallel with these active treatments.

Pain

The International Association for the Study of Pain defines pain as 'an unpleasant sensory and emotional experience associated with or resembling that associated with actual or potential tissue damage'. It may be classified as 'acute pain' (see Chapter 28) and 'chronic pain'. Chronic pain is 'pain that persists or recurs for longer than 3 months' (ICD-11). It may be **further** subclassified as 'chronic secondary pain', when it is caused by an underlying disease, or 'chronic primary pain', when there is no identified underlying disease. It is extremely important to make this distinction because the management of 'chronic secondary pain' caused by cancer or advanced disease differs significantly from the management of 'chronic primary pain', which requires a distinct approach, focusing on a range of multidisciplinary non-pharmacological therapies. Importantly, opioids and other commonly used adjuvant analgesic agents are not recommended for use in patients with 'chronic primary pain'.

Chronic secondary pain is the subcategory of pain usually seen in the palliative care setting, where it is most commonly caused by cancer, though it can also have non-cancer causes. At least two-thirds of people with cancer have significant pain, and it is the most feared symptom at the end of life. It may be caused by the cancer directly (e.g., pressure on surrounding structures) or indirectly (e.g., caused by weight loss or pressure ulcers) and can result from co-morbidity such as arthritis. Within each of these aetiologies, pain mechanisms may be nociceptive (visceral or somatic) or neuropathic (see Chapter 28), generating different pain characteristics, which can require different treatment approaches. Various scales can be used to assess a patient's pain (**Fig. 50.2**). Emotional and spiritual distress may be expressed as and can exacerbate physical pain.

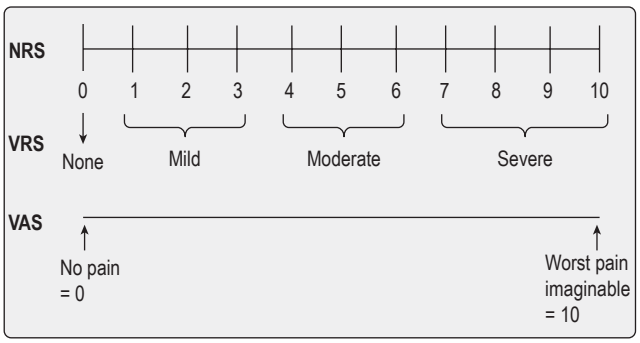


Fig. 50.2 Commonly used one-dimensional scales for pain assessment: an 11-point numerical rating scale (NRS), visual analogue scale (VAS, which can be 0–10 or 0–100), and verbal rating scale (VRS). *(Adapted from Breivik H, Borchgrevink PC, Allen SM et al. Assessment of pain.* Br J Anaesthesia *2008; 101:17–24.)*

The term 'total pain' encompasses a variety of influences that contribute to pain:

- ***Biological:*** the disease itself, disease-modifying therapy (drugs, surgery, radiotherapy)
- ***Social:*** family distress, loss of independence, financial problems
- ***Psychological:*** fear of dying, of pain, or of being in hospital; anger at dying or at the process of diagnosis and perceived delays, with anxiety or depression stemming from all of these
- ***Spiritual:*** fear of death, questions about life's meaning, guilt, asking 'why me'?

Pain and the use of opioids

Opioids are effective for the control of acute pain (see Chapter 28) and in the management of severe chronic cancer pain. They may also have a role in the short-term management of chronic non-cancer pain in the last weeks of life. In general, they should not be used outside of these contexts and are likely to cause more harm than benefit if used long-term in chronic non-cancer pain. This is because they have limited efficacy in chronic non-cancer pain, and there is a high risk of developing tolerance, addiction, and long-term side-effects. This cannot be overemphasized in the wake of the recent opioid epidemic in the USA, where there were 6000 deaths from opioid overdose in 2018.

Cancer pain and the World Health Organization analgesic ladder

Most cancer pain can be managed with oral or commonly used transdermal preparations. The World Health Organization (WHO) cancer pain relief ladder guides the choice of analgesic according to pain *severity* (Fig. 50.3).

If *regular* use of optimum dosing (e.g., paracetamol 1 g four times daily) for step 1 does not control pain, then an analgesic from the next step of the ladder is prescribed. Because pain has different physical aetiologies, an adjuvant analgesic may be needed, for example, a gabapentinoid for neuropathic pain (see Box 50.9).

Strong opioid drugs

Dose titration and route

Morphine is the drug of choice and in most circumstances should be given regularly by mouth. The dose should be tailored to the individual's needs by allowing 'as required' doses up to every 1 to 2 hours. If a patient has needed further doses in addition to the regular daily dose *and* they have been effective, then the amount in the additional doses can be added to the following day's regular dose. This up-titration can be repeated until the daily requirement becomes stable (Box 50.6). When a stable daily dose requirement has been established, the morphine can be changed to a modified-release preparation and the patient given 'as required' doses of immediate-release morphine, which are usually one-sixth of their total daily morphine dose.

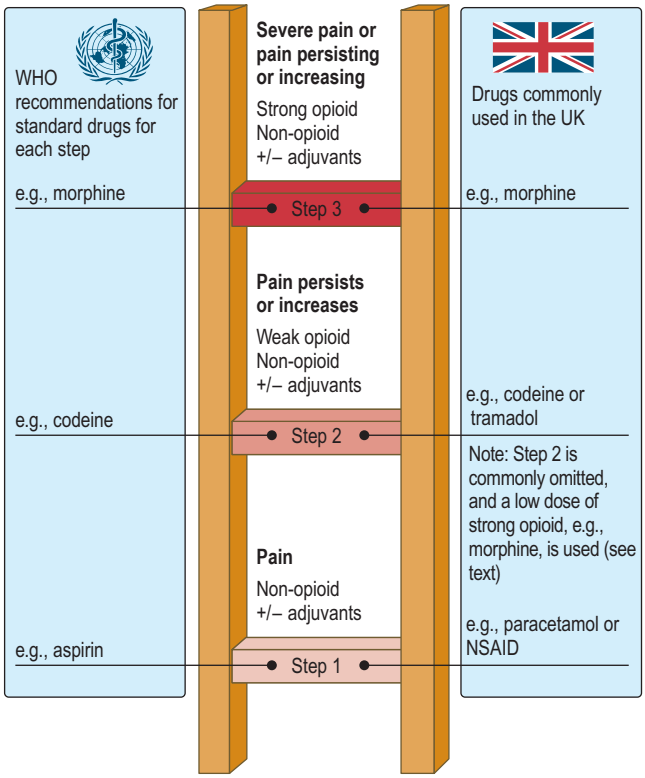


Fig. 50.3 The World Health Organization (WHO) analgesic ladder for cancer and other chronic pain. Adjuvant drugs are listed in Box 50.9. *NSAID,* Non-steroidal anti-inflammatory drug.

The starting dose of morphine is usually 2.5 to 5 mg every 4 hours, depending on patient size, renal function, and whether a weak opioid is already being given. If an individual is already on step 2 of the WHO ladder, explaining conversions between medications can be helpful to gain patient confidence. For example, maximum-dose codeine (60 mg four times daily) is equivalent to a total of 24 mg morphine per day (Box 50.7). If there is renal dysfunction, morphine should be used in low doses or oxycodone considered in mild to moderate renal dysfunction (hepatic metabolism). However, if renal dysfunction is significant (estimated glomerular filtration rate <30 mL/min), an alternative opioid, such as alfentanil, should be considered because of the risk of metabolite accumulation. If there is concurrent significant hepatic dysfunction, oxycodone should be avoided. Dosing is not an exact science in palliative care; advice can be sought from the SPC team.

Strong opioids come in transdermal preparations, such as buprenorphine or fentanyl *patches* (Box 50.8). These are useful in patients with stable pain. Serum levels do not change quickly with transdermal patches; they can therefore be hazardous if the pain stimulus is removed and are cumbersome to titrate in patients with escalating or unstable pain. The patient should be reassessed and reasons for the escalation in pain sought. Be aware that transdermal

Box 50.6 Opioid titration examples

Scenario 1

A patient regularly takes 20 mg of morphine sulphate immediate release every 4 hours. What long-acting preparation could they be started on?

20 mg every 4 hours =120 mg morphine per day

= 60 mg twice daily of a 12-h morphine sulphate modified-release (MR) preparation

Scenario 2

The same patient has started morphine sulphate MR 60 mg twice a day with 'as required' immediate-release morphine 20 mg every 1–2 hours. A few months later, their pain escalates because of disease progression, and they require an additional four doses per day of immediate-release morphine. What should their long-acting morphine be increased to?

20 mg × 4 = additional 80 mg morphine per day

Total daily morphine dose including morphine MR = 200 mg

= 100 mg twice daily of a 12-h morphine sulphate MR

Top tip: Always thoroughly check dose conversions. Dosing is not an exact science; discuss with specialist palliative care if unsure.

Box 50.7 Opioid equivalences

Drug and route	Dose (mg)	Potency to oral morphine equivalent
Oral codeine	100	1 : 10
Subcutaneous diamorphine	3	3 : 1
Oral morphine	10	1 : 1
Subcutaneous morphine	5	2 : 1
Oral oxycodone	6.6	1.5–2 : 1
Subcutaneous oxycodone	3.3	3 : 1
Oral tramadol	100	1 : 10
Subcutaneous alfentanil	0.3	30 : 1

Adapted from National Institute for Health and Care Excellence. https://bnf.nice.org.uk/medicines-guidance/prescribing-in-palliative-care/.

patches may be administered every 72 hours or weekly, depending on the dose and brand; always check this when prescribing.

If a patient is unable to take oral medication because of weakness, swallowing difficulties, or nausea and vomiting, opioids can be given parenterally. This may be via continuous subcutaneous infusion, whereby the opioid dose is administered continuously via a syringe pump over the course of 24 hours.

Doctors and patients may have worries, such as the fear of addiction, which mean that adequate doses of opioids are not prescribed or taken. Iatrogenic addiction in the context of cancer pain is very rare (<0.01% risk), and the adverse effects and morbidity from uncontrolled pain are much higher. Patients and carers should be reassured about their medications and the rationales for their use, and it is good practice to provide written information regarding opioids. Regular review is necessary to achieve optimal pain control, including regular assessment to distinguish pain severity from distress caused by pain.

Side-effects

The most common side-effects of strong opioids are:

- ***Nausea and vomiting*.** These can be managed with use of antiemetics.

Box 50.8 Buprenorphine and fentanyl patches

Oral morphine equivalent/day	Patch
12 mg	Buprenorphine 5 μg/h
24 mg	Buprenorphine 10 μg/h
36 mg	Buprenorphine 15 μg/h
30 mg	Fentanyl 12 μg/h
60 mg	Fentanyl 25 μg/h
120 mg	Fentanyl 50 μg/h
180 mg	Fentanyl 75 μg/h

Adapted from National Institute for Health and Care Excellence. https://bnf.nice.org.uk/medicines-guidance/prescribing-in-palliative-care/.

- ***Constipation*.** This is common and should be anticipated; best practice guidelines advocate that *all* opioids should be co-prescribed with laxatives. Laxatives for opioid-induced constipation, peripheral mu-receptor opioid antagonists, can be used if response to other laxatives is inadequate.

  If side-effects are intractable, a change of opioid is often helpful. Some patients find transdermal preparations of opioids less constipating.

Toxicity

Confusion, persistent and undue drowsiness, myoclonus, nightmares, and hallucinations indicate opioid toxicity. This may follow rapid dose escalation and respond to dose reduction and slower retitration. It may indicate pain that is poorly responsive to opioids and the need for adjuvant analgesics. Patients should be assessed for reversible factors contributing to reduced tolerance of opioids, such as sepsis. Importantly, respiratory depression (respiratory rate <8 breaths/min) is a late sign and indicates significant toxicity that needs management as an emergency. Consider stopping any long-acting opioid preparation and administering naloxone if respiratory depression is severe. Dependency on opioids used in the longer term does occur and means that, unless absolutely necessary, sudden reversal (with naloxone) or withdrawal of opioids is not advisable.
Antipsychotics such as haloperidol may help settle the patient's distress while waiting for toxicity to resolve. Some individuals tolerate an alternative opioid such as oxycodone better or an alternative route such as subcutaneous injection.

Adjuvants and alternatives

Box 50.9 outlines the most commonly used adjuvant analgesics. Neuropathic pain often responds well to adjuvants in the antiepileptic or tricyclic antidepressant class. Other treatments, such as radio- or chemotherapy, interventional pain strategies (e.g., nerve blocks, epidurals), and transcutaneous electrical nerve stimulation, may be useful in selected patients. MDT working is essential to manage patients' pain optimally, and joint working between palliative care and pain teams, sometimes via specialist pain clinics, helps to achieve this.

Additionally, to improve their symptoms and wellbeing, patients facing incurable illness are increasingly turning to complementary therapies to serve as alternatives to prescribed medical treatments. These include acupuncture and herbal and massage therapies. Although there may not be a strong evidence base to support some of these, this does not mean that they do not have a beneficial contribution to make to the holistic care of patients. For example, research into cannabinoids suggests they may be beneficial in a

Box 50.9 Commonly used adjuvant analgesics

Drugs	Indication
NSAIDs, e.g., naproxen (250–500 mg twice daily)	Bone pain, inflammatory pain
Antiepileptics, e.g., gabapentin (300–3600 mg/day in divided doses) or pregabalin (50–600 mg/day in divided doses)	Neuropathic pain
TCAs, e.g., amitriptyline (10–75 mg/day)	Neuropathic pain
Bisphosphonates, e.g., zoledronic acid, denosumab (a RANKL inhibitor)	Metastatic bone disease
Dexamethasone	Neuropathic pain, inflammatory pain (e.g., liver capsule pain), headache from cerebral oedema caused by brain metastases

NSAID, Non-steroidal anti-inflammatory drug; *RANKL,* receptor activator of nuclear factor kappa-B ligand; *TCA,* tricyclic antidepressant.

variety of conditions; can modulate symptoms, including pain, nausea and vomiting, and sleep disorders; and have a good risk-to-benefit profile. Certain countries have legalized cannabis to allow its prescription. As ever, more research is needed into these compounds and therapies to demonstrate their effectiveness clearly, and the picture will evolve over time.

Further reading

Caraceni A, Hanks G, Kaasa S et al. Use of opioid analgesics in the treatment of cancer pain: evidence-based recommendations from the EAPC. *Lancet Oncol* 2012; 13:e58–e68.

National Institute for Health and Care Excellence. *NICE Clinical Guideline CG140: Palliative Care for Adults: Strong Opioids for Pain Relief*. NICE 2016; https://www.nice.org.uk/Guidance/CG140.

National Institute for Health and Care Excellence. *NICE Clinical Guideline CG193; Chronic pain (primary and secondary) in over 16s: assessment of all chronic pain and management of chronic primary pain. NICE 2021:* https://www.nice.org.uk/guidance/ng193.

Treede RD, Winfried R, Barke A, et al. Chronic pain as a symptom or a disease: the IASP Classification of Chronic Pain for the International Classification of Diseases (ICD-11). *Pain* 2019; 160:19–27.

Gastrointestinal symptoms

Nausea and vomiting

These symptoms are common and may have numerous causes: both physical and psychological factors can contribute. Careful choice of antiemetic targeted toward the underlying cause of the nausea or vomiting is essential.

Nausea and vomiting are associated with:

- ***Chemotherapy*.** Certain agents confer a high risk of nausea and vomiting. A 5-hydroxytryptamine (5-HT_3) antagonist, such as ondansetron 4–8 mg up to three times daily, may help during the acute emetic stimulus of chemotherapy administration. Aprepitant, a neurokinin-1 antagonist, can also be given.
- ***Chemical causes***, e.g., hypercalcaemia, hepatic failure, uraemia or drugs. Haloperidol 0.5–1.5 mg as required every 4 hours, up to a maximum of 5 mg/day, is the first choice.
- ***Gastric stasis and distension*.** Metoclopramide 10 mg three times daily is a prokinetic and promotes gastric emptying. The prokinetic effects can also ameliorate constipation.
- ***Raised intracranial pressure***, e.g., brain tumour. Cyclizine 50 mg three times daily is most useful.
- ***Mechanical causes of vomiting*.** It may be necessary to start antiemetic therapy parenterally by continuous subcutaneous infusion to gain control. If the patient has gastrointestinal (GI) obstruction or poor oral or GI absorption, this route may need to be continued. Haloperidol is the antiemetic of choice in patients with complete malignant bowel obstruction. Levomepromazine is a second-line antiemetic if initial therapy is not successful; 6.25 mg as required, up to a maximum of 25 mg/day, is recommended. Other medications to aid nausea and vomiting include corticosteroids, particularly in the case of bowel obstruction, in which reducing tumour oedema may alleviate the obstruction and improve symptoms.

Constipation

This is common, occurring in up to 45% of hospice patients and 90% of patients taking opioids. It is best managed with a combination of pharmacological and non-pharmacological strategies. Non-pharmacological measures can include maintaining hydration, mobility, dietary advice, optimal positioning, privacy, and comfort to aid relaxation when trying to open the bowels.

Pharmacological strategies including laxatives (**Box 50.10**) are targeted to the cause of the constipation, and a careful history is required. Often, patients with opioid-induced constipation respond to a combination of stool softener and stimulant (e.g., docusate and senna). Doses require individual titration. Some patients may need suppositories or enemas. Those with reduced sensation and mobility (e.g., caused by spinal cord compression) may require a bowel regimen of regular suppositories and maintenance of a relatively firm stool to facilitate best care. Many are likely to need to continue regular laxatives because they will develop troublesome constipation again on cessation of these medications.

Bowel obstruction

This is common in advanced cancer (3%–15%), especially gynaecological (<42%) and colorectal (<24%) cancers. It is caused by mechanical obstruction of the bowel lumen or peristaltic failure caused by infiltration of the nerves. Presenting symptoms include pain, vomiting, colic, and constipation. It can be partial or complete, at single or multiple sites, and carries a poor prognosis. Management of bowel obstruction should be individualized, focussing on improving symptoms (which may involve medications) and interventions to decompress the upper GI tract (e.g., a large-bore nasogastric tube). Consideration should be paid to hydration and nutrition, including using artificial methods (e.g., intravenous fluid) when appropriate. The patient's prognosis, performance status, and further treatment options, among other things, should be taken into account.

Metoclopramide's prokinetic effect can be used in partial obstruction to encourage the bowels to function again. However, it should be avoided in complete obstruction when the bowel is physically blocked because it will worsen colic. In this situation, an antispasmodic, such as hyoscine butylbromide, is preferred, along with regular analgesia and an antiemetic such as haloperidol. A dose of 60–120 mg/24 hours subcutaneously is usually recommended, but much higher doses (<300 mg) may be needed because parenteral hyoscine butylbromide can be rapidly inactivated.

Octreotide (a somatostatin analogue) is used as second-line treatment in bowel obstruction to improve symptoms, which it does by reducing gut secretions and the volume of vomitus.

Physical measures, such as a defunctioning colostomy or a venting gastrostomy, may be helpful. Occasionally, a lower bowel obstruction is resolved with insertion of a stent or transrectal resection of tumour in selected individuals.

Box 50.10 Types of laxatives

Type	Name	Suggested dosage
Stimulant	Senna Bisacodyl	15–22.5 mg at night 5–10 mg
Softener	Docusate sodium	100 mg twice daily
Osmotic	Macrogol	1–3 sachets daily
Suppository: soft loading	Bisacodyl Phosphate enema	10 mg 1
Suppository: hard loading	Glycerol	1

In ***advanced disease***, when the obstruction is irreversible, patients should be encouraged to drink and take small amounts of a soft diet for comfort as they wish. With good mouth care (see later), the sensation of thirst can be managed, and parenteral fluids may not be needed. If oral food and fluids are taken, they can be introduced as a trial to improve symptoms and patients counselled about the benefits and risks, including their possible contribution to increased gut secretions and volume of vomitus; in this case, the fluids may be ceased.

Anorexia, weight loss, and malaise

These result from the cachexia syndrome of advanced disease and are poor prognostic factors. Although attention must be paid to nutrition, the syndrome is mediated through chronic stimulation of the acute phase response and tumour-secreted substances. Thus, calorie–protein support alone gives limited benefit; parenteral feeding has been shown to make no difference to patient survival and can lead to increased complications such as infections.

Management is generally supportive, and advice from dieticians may be helpful. Patients and their carers should be reassured that decreased appetite is to be expected, and advising 'little and often' and 'what you want, when you want' can be useful.

Corticosteroids are occasionally used as appetite stimulants. However, the weight gained is usually fluid retention, and muscle catabolism is accelerated, causing proximal myopathy. Any benefit in appetite stimulation tends to be short-lived, which must be explained clearly to patients and carers. Thus, use of corticosteroids should be limited to the short term and reviewed regularly.

Other interventions, such as a rehabilitative palliative care programme encompassing nutritional supplementation and exercise in advanced cancer, are being trialled.

Further reading

Obita GP, Boland EG, Currow DC et al. Somatostatin analogues compared with placebo and other pharmacologic agents in the management of symptoms of inoperable malignant bowel obstruction: a systematic review. *J Pain Symp Manage* 2016; 52:901–919.

Roila, F, Molassiotis A, Herrstedt et al on behalf of the participants of the MASCC/ESMO Consensus Conference Copenhagen 2015. 2016 MASCC and ESMO guideline update for the prevention of chemotherapy- and radiotherapy-induced nausea and vomiting and of nausea and vomiting in advanced cancer patients. *Annals Onc* 2016; 27(Suppl 5):v119–v133.

Respiratory symptoms

Breathlessness

Chronic breathlessness (dyspnoea) is a common but frightening symptom of many advanced malignant and non-malignant diseases but often goes unrecognized and untreated. Its features have been described as 'breathlessness that persists despite optimal treatment of the underlying pathophysiology and that results in disability'. It is distressing for carers, who often feel helpless. Carer anxiety may exacerbate a patient's breathlessness, and it is therefore essential to involve them in treatment strategies and support them in their own right. Recent advances have helped to establish effective treatments to reduce the impact of breathlessness, but palliative care or specialist breathlessness services are often needed to provide optimal care.

The sensation of being breathless is generated in the central nervous system through the integration of peripheral and central signals by a complex network involving structures in the midbrain, higher cortical regions, and the limbic system (**Fig. 50.4**). Individual assessment is key, as is reversal of any contributory factors, such as optimizing treatment of underlying medical conditions, although in advanced disease, many of these are irreversible.

The first step in assessing breathlessness (**Box 50.11**) is to ask directly about it because many patients with severe breathlessness are not troubled at rest. Best practice in managing breathlessness rests on a complex intervention composed of a number of non-pharmacological and pharmacological approaches (**Boxes 50.11 and 50.12**). Careful assessment of the patient—for example, using the 'Breathing, Thinking, Functioning' clinical model (**Fig. 50.5**)—guides the choice of initial non-pharmacological approach (see **Box 50.12**). Pharmacological measures should be reserved for those with severe breathlessness and end-stage progressive disease. Morphine is the drug that has been most studied; other treatments, such as mirtazapine, are being investigated. Benzodiazepines should not be used routinely to manage breathlessness because there is no evidence of their benefit; they may, however, be prescribed in those at the end of life or when there is a limited prognosis. Oxygen is not helpful in relieving the sensation of breathlessness but may be essential in managing the underlying medical condition and improving exercise capacity when hypoxia is present. There are specific assessment tools, such as the Dyspnoea-12 and numerical rating scales, which can help initial management and subsequent monitoring of treatment strategies.

Cough

Cough is an unpleasant symptom that can affect the lives of both the patient and carer. Drug approaches (e.g., opioids or gabapentin) have limited impact but may be needed. Non-pharmacological interventions can include breathing exercises, cough suppression techniques, education, and counselling to help patients gain greater control of their cough.

Secretions

Excessive respiratory secretions at the end of life can be treated with ***glycopyrronium*** 200 to 400 μg every 2 to 4 hours (maximum, 1.2–2.4 mg per 24 hours) or ***hyoscine butylbromide*** 20 mg every 1 to 2 hours (maximum, 120 mg per 24 hours). These can be given by continuous subcutaneous infusion. Both give a dry mouth, and care must be taken not to dry secretions significantly in a patient who is still able to cough well because they may become too sticky to expectorate, leading to significant discomfort.

Further reading

Maddocks M., Brighton L.J., Farquhar M. et al. *Holistic Services for People with Advanced Disease and Chronic or Refractory Breathlessness: A Mixed-Methods Evidence Synthesis.* Southampton (UK): NIHR Journals Library; 2019. PMID: 31241880.

Parshall MB, Schwartzstein RM, Adams L et al. An official American Thoracic Society statement: update on the mechanisms, assessment, and management of dyspnea. *Am J Respir Crit Care Med* 2012; 185:435–452.

Spathis A, Booth S, Moffat C et al. The Breathing, Thinking, Functioning clinical model: a proposal to facilitate evidence-based breathlessness management in chronic respiratory disease. *NPJ Prim Care Respir Med* 2017; 27:27.

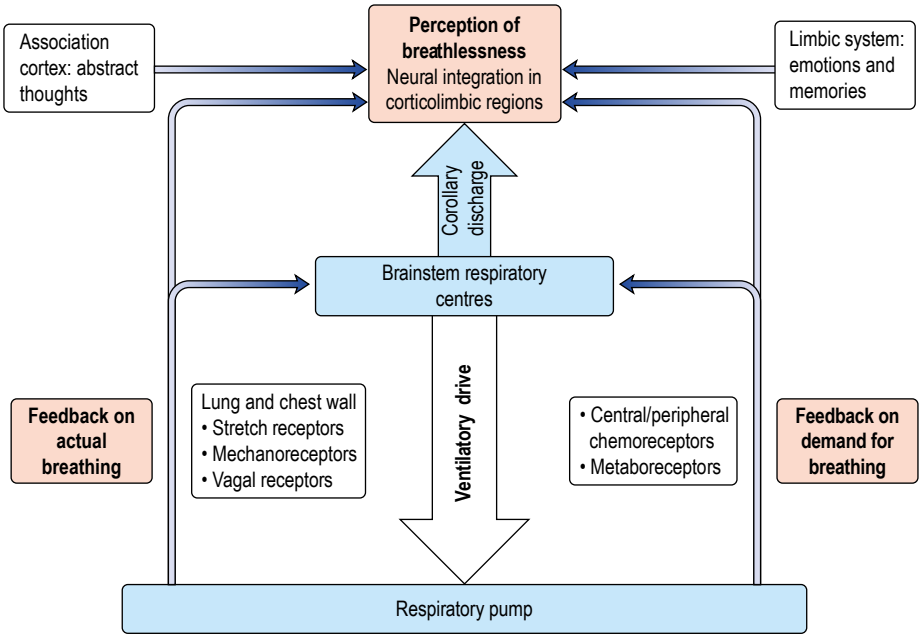


Fig. 50.4 Genesis of breathlessness. *(Adapted from Moosavi S, Booth S. The genesis of breathlessness. In:* Cambridge Breathlessness Intervention Service Treatment Manual. *Cambridge University Hospitals NHS Foundation Trust; 2011.)*

Other physical symptoms

People with cancer may develop other physical symptoms caused by the tumour directly (e.g., hemiplegia caused by brain metastases) or indirectly (e.g., bleeding or venous thromboembolism caused by coagulopathy). Symptoms may also result from treatment, such as lymphoedema in breast cancer or heart failure secondary to chemotherapy. The principles of holistic assessment, treatment of reversible factors, and appropriate involvement of the multiprofessional team should be applied.

Fatigue

Fatigue is a significant and debilitating problem for palliative patients. It has physical, cognitive, and affective components; unlike normal tiredness, it is not relieved by usual sleep or rest. An assessment for reversible contributory factors, such as anaemia, hypokalaemia, or over-sedation caused by poorly optimized medication, should be undertaken. Management strategies are mainly non-pharmacological: relaxation, sleep hygiene, resting 'proactively' rather than collapsing when exhausted, planning, pacing and prioritizing daily activities, and use of aids and adaptations.

Mouth symptoms

Dry mouth is common in advanced disease and may be a result of medications used to treat other symptoms, such as pain or nausea. It can be managed with artificial saliva substitutes and good mouth care to keep mucous membranes moist. Careful inspection of the mouth identifies any reversible causes contributing to oral discomfort (e.g., candidiasis, which is common in immunocompromised patients but treatable with antifungals). As patients lose weight, dentures may become ill-fitting, causing ulceration to the gums. This can be managed with analgesia, denture cushions, and topical treatments to numb the sore parts of the mouth.

Patients receiving courses of radiotherapy for haematological disorders or head and neck cancers can develop painful mucositis (pain and inflammation of the mucous membrane, which may present as painful mouth ulceration affecting any or all intraoral surfaces). This is a common reason for referral to hospital palliative care teams even when the treatment aims are curative. Management is with meticulous mouth care, including mouthwashes to maintain oral hygiene and to provide pain relief. Some patients with severe mucositis require enteral feeding and management of their pain with continuous subcutaneous infusion until it settles.

Loss of function, disability, and rehabilitation

Some of the most pressing concerns include increasing physical frailty, loss of independence, and a perception of being a burden on others. Evidence suggests that functional problems are not routinely assessed and not as well managed as other symptoms.

Palliative rehabilitation in long-term incurable illness aims to enable patients to achieve their functional and social priorities in the context of physical decline, as well as to adjust to changes in their physical ability, role, and sense of self. Patients and carers often appreciate a focus on what is still possible rather than what is lost.

Rehabilitation may include exercise to address reversible physical decline and fatigue, which may be especially evident after an exacerbation or relapse. Understandably, sustaining motivation to maintain physical activity can be difficult. Setting small and achievable goals may help maintain motivation and promote self-efficacy. Involving carers in rehabilitation sessions may reduce over-protectiveness.

Rehabilitation includes supporting adaptation by encouraging patients to do things differently, including the use of mobility aids or equipment. Aids and adaptations may be seen as symbols of 'giving up' by some patients. They need to be presented in a positive light as 'enablers' that assist people to achieve their goals and that may also reduce carer burden. Psychological support to accept change may be required, such as acceptance and commitment therapy.

Hospices take a 'rehabilitative approach', with most having physical and occupational therapists and exercise equipment.

Box 50.11 Assessment and management of breathlessness

Step 1

- Ask directly about breathlessness: 'Are you troubled by breathlessness'?
- Ask about other respiratory symptoms: 'Are you troubled by a cough'?
- Ask about fatigue: 'Are you troubled by fatigue or extreme tiredness that does not go away with rest or sleep'?
- Take a full history of each symptom; aggravating and easing factors; and 24-hour pattern, including nocturnal symptoms.
- Ask about the carer's views, thoughts, and feelings relating to the patient's breathlessness.
- Find out what happens when the patient has an 'attack' of breathlessness, including related emotions such as panic, anxiety, or frustration and what helps and what doesn't help.

Step 2

- Assess which component of breathlessness is the most troubling or disabling for the patient: the breathlessness itself ('Breathing'), deconditioning, or difficulty managing activities of daily living ('Functioning'), or the patient's thoughts and feelings about being breathless ('Thinking').
- Choose an initial treatment based on that.

Initial treatment

All patients

- Handheld fan
- Education about breathlessness and its causes

Breathing cycle

- Breathing techniques and breathing retraining from physiotherapist
- Positioning

Functioning cycle

- Pulmonary rehabilitation
- Individual exercise programme from palliative care or community therapy or groups
- Fatigue management advice from occupational therapist
- Aids and adaptations for mobility and activities of daily living

Thinking cycle (both patient and carer)

- Address misconceptions
- Relaxation exercises and mindfulness
- Specialist psychological help

Pharmacological treatment for severe breathlessness or at end of life (seek specialist palliative care advice)

- Use oral or parenteral opioids.
- Consider mirtazapine.
- Consider benzodiazepines.
- Dying patients with breathlessness need subcutaneous opioids and midazolam.

Referral to hospice day therapy (sometimes known as a Living Well service), pulmonary rehabilitation, community exercise groups, or community physical or occupational therapy may be appropriate. Some group programmes may have the option of being delivered virtually for those who find it difficult to leave the home. However, patients must have an active desire to engage in rehabilitation. Some may see therapy as an additional unwanted drain on their time, and their wishes should be ascertained, as well as discussing the merits and disadvantages of engaging.

Functional problems and fatigue should not be seen as inevitable, unavoidable, and insoluble. All healthcare professionals have a role in promoting the benefits of patients keeping as active as they are able to help maintain their strength and ability in long-term incurable illness. Barriers to this, including misconceptions about exercise and activity being harmful, should be discussed and addressed, as well as considering psychological barriers such as anxiety and depression. Throughout illness, there is a need to take into account changing performance status, as well as changes in goals and priorities.

Poor sleep

This is common, and patients may have longstanding difficulties with sleep that worsen during their illness. Discomfort in bed, leading to worsening symptoms, and psychological concerns surrounding their condition and the future, can compound the problem. Eliciting and listening to patients' concerns, explaining the concept of sleep hygiene, and discouraging lengthy day-time naps can be helpful, as well as therapy team review of the bed and provision of equipment that will optimize symptom management. Pharmacological measures (e.g., hypnotics such as zopiclone or melatonin) should be used only temporarily and are not long-term solutions to sleep problems. Complementary therapies are available in hospices and some hospitals and may aid sleep.

Psychosocial issues

Depression is a common feature of life-limiting and disabling illness and is often missed or dismissed as 'understandable'. However, it may respond to antidepressant drugs or to non-pharmacological measures such as cognitive-behavioural therapy, increased social support, and support for family relationships. Antidepressants can take several weeks to take effect, and a patient's prognosis must therefore be considered if they are initiated. The 'side-effects' of certain antidepressant medications are sometimes used to benefit patients. For example, mirtazapine can improve sleep, anxiety, appetite, and low mood but may also be helpful to treat pain and nausea; evidence is also emerging for its benefits in gastroparesis. Such interventions can make a big difference to the patient's quality of life and ability to manage their situation.

Further reading

Hudson PL, Remedios C, Thomas K. A systematic review of psychosocial interventions for family carers of palliative care patients. *BMC Palliat Care* 2010; 9:17.

Montagnini M, Javier NM, and Mitchinson A. The role of rehabilitation in patients receiving hospice and palliative care. *Rehabil Oncol* 2020; 38(1)9–21.

PALLIATIVE CARE IN NON-MALIGNANT DISEASE

Heart failure

There are special considerations with respect to cardiac medication in advanced disease:

- ***Drugs.*** Many commonly used palliative care medications can prolong the QT interval and should be used with caution in patients with heart failure; they include amitriptyline and many antiemetics. Others may precipitate worsening symptoms; for example, non-steroidal anti-inflammatory drugs (NSAIDs) may exacerbate fluid retention. However, use of certain medications that can be harmful in the long term may become appropriate when the patient is dying, and the risks and benefits are weighed up and discussed with the patient and family.
- ***Sudden death.*** This is more common than in patients with malignancy. There are increasing numbers of patients with implanted defibrillators because the indications for them has expanded. If present, the device should be reprogrammed to pacemaker mode in people with advanced disease. Implanted

Box 50.12 Key non-pharmacological interventions for breathlessness

Intervention	Putative mechanism of action	Most useful
Facial cooling with a handheld fan	Cools area served by second and third branches of trigeminal nerve	For reducing length of episodes of SOB on exertion or at rest
	Reduces temperature of air flowing over nasal receptors, altering signal to brainstem respiratory complex	For giving patient and carer confidence by having an intervention they can use Use and evidence base need to be explained
Breathing techniques (e.g., pursed-lip breathing, diaphragmatic breathing)	Improves mechanical effectiveness of respiratory system	Any breathlessness caused by chronic pathology
Breathing pattern retraining	Improves mechanical efficiency of breathing pattern	When breathlessness is disproportionate to disease severity because of breathing pattern disorder
Walking aids	Improves the mechanical effectiveness of respiratory system through forward lean and upper limb positioning	For breathlessness during walking; four-wheeled walkers may also provide a portable seat to rest
Exercise	Stops spiral of disability developing	In patients who are still quite mobile to maintain independence and to maintain function and comfort in those who are less mobile
	Changes muscle structure: less lactic acid produced	Lessening or deferring symptoms by reducing deconditioning
Pacing (finding a balance between activity and rest, prioritizing daily activities)	Avoids over-exertion, which can lead to exhaustion, inactivity, and subsequent deconditioning	In patients who are able and willing to modify daily routines
Anxiety reduction, e.g., CBT (needs a skilled clinician to administer) or simple relaxation therapy	Works on central perception of breathlessness, reducing the impact	In people with higher levels of anxiety at baseline (i.e., when first seen)
	Interrupts the panic–anxiety cycle	In patients willing to persevere with learning a new skill
Carer support	Reduces carer anxiety and distress, which are part of 'total' anxiety–panic cycle	When the carer is isolated or under extra pressures (e.g., looking after an older adult parent, going through divorce)

CBT, Cognitive-behavioural therapy; *SOB*, shortness of breath (breathlessness).

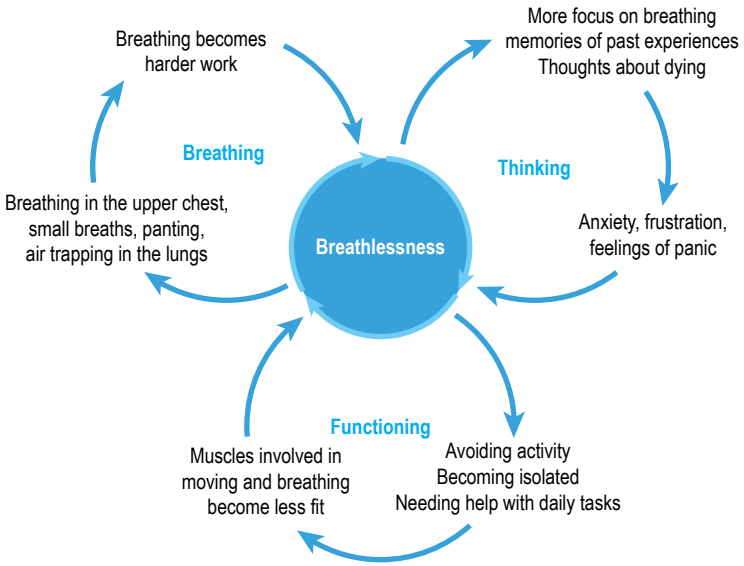


Fig. 50.5 The Breathing, Thinking, Functioning clinical model. *(From Spathis A, Booth S, Moffat C et al. The Breathing, Thinking, Functioning clinical model: a proposal to facilitate evidence-based breathlessness management in chronic respiratory disease.* NPJ Prim Care Respir Med *2017; 27:27. Reproduced with kind permission of Dr Anna Spathis, University of Cambridge and Cambridge University Hospitals NHS Foundation Trust.)*

defibrillators have not been shown to improve survival in patients with severe heart failure, and it could be distressing for patient, family, and staff if the defibrillator discharges as the patient is dying.

- ***Peripheral oedema.*** This can become a major problem and may be more resistant to diuretic therapy; careful balancing of medication regimens is required. Ultimately, symptom relief is prioritized over renal function.
- ***Cardiac medications.*** Drugs should be regularly rationalized to reduce polypharmacy. Those prescribed to reduce long-term secondary risk (e.g., statins) can be stopped, but those that help with symptom control (including angiotensin-converting enzyme

inhibitors, which also benefit survival) can be continued. Antihypertensives may have to be stopped if patients experience symptomatic hypotension.

Chronic respiratory disease

Chronic obstructive pulmonary disease (COPD) is the most common chronic respiratory disease. Patients with COPD may live increasingly restricted lives for years, rather than months or weeks, when they become breathless. This can lead to depression, isolation, frustration, and sometimes guilt associated with previous smoking habits. As with many chronic conditions, patients can reach older age before becoming very disabled, and an older spouse often has to carry significant physical carer burdens.

Palliative care breathlessness support services can be very helpful for patients unable to manage pulmonary rehabilitation or those with complex psychosocial needs that require individual assessment. Emergency admissions to hospital for non-medical reasons are often caused by severe anxiety (in the patient or carer), and the support offered by community palliative care services or breathlessness services working with respiratory teams can help reduce these.

In advanced disease, patients may use non-invasive ventilation (NIV) regularly at home. This may provide considerable symptom relief, but early discussions should take place regarding management of NIV during exacerbations of COPD and at the end of life.

Patients with interstitial lung diseases are often troubled by dry cough as well as breathlessness and may rapidly deteriorate at the end of life. Early palliative care should be considered to help with symptom control and anxiety. Ambulatory or long-term oxygen therapy may be necessary to manage severe hypoxaemia and maintain function in these patients.

Opioid titration in non-malignant respiratory disease

In non-malignant respiratory disease, opioids may be considered after non-pharmacological management of breathlessness has been maximized or if a patient is in the terminal phase. Individual assessment is essential. When starting opioids, a cautious approach with regular monitoring is recommended. Lower doses tend to be used (e.g., 1 mg of oral morphine twice a day and as required <10 mg/24 hours). Titration may need to follow a different pattern from that used in patients with malignant disease because many patients with cancer are already taking opioids for pain control before they develop breathlessness. Significant respiratory depression has not been noted in various trials of opioids in advanced respiratory disease, but other side-effects such as constipation should be anticipated.

COVID-19

The COVID-19 pandemic placed significant demands on health systems and led to a huge increase in deaths globally. There was a need to adapt quickly in response to service delivery challenges, and it became clear that skills for managing patients at the end of life were essential to all healthcare professionals. Palliative care services were a vital part of the pandemic response, not just in caring for patients but also in providing support to staff not used to dealing with high numbers of deaths. Medications used to manage symptoms, such as breathlessness, anxiety, and agitation, were largely similar to those used before COVID-19. However, communication with patients and families in the context of layers of personal protective equipment, lockdowns, social isolation, fear of contracting the virus, and restricted visiting was challenging. When difficult conversations would have previously taken place face to face with relatives present to support patients, telephone conversations and alternative methods, such as video calls, became the norm, and guidance for having compassionate phone conversations has been developed (Fig. 50.6). Finally, palliative care services continue to support patients in managing troubling symptoms such as chronic breathlessness caused by long COVID.

Further reading

Janssen DJA. Palliative care in COVID-19. *Curr Opin Support Palliat Care* 2021; 15(4):199–204.
Johnson MJ, Booth S. Palliative and end-of-life care for patients with chronic heart failure and chronic lung disease. *Clin Med (London)* 2010; 10:286–289.

Renal disease

There is limited evidence that dialysis improves either quality or length of life in older patients or those with significant co-morbidity, so frailer patients with advanced chronic kidney disease (CKD) should be supported to receive maximum supportive care rather than renal replacement therapy if it is their choice. This may preserve independence and time spent at home and limit unnecessary hospital visits and procedures.

Palliative care in this context consists of managing the complications of advanced renal insufficiency (including nausea, itch, anaemia, oedema, and restless legs syndrome) while preserving residual renal function for as long as possible.

Prescribing is complicated, particularly in those receiving dialysis. Drugs are variably removed by dialysis and this must be taken into account; the renal drug database (see Further reading) is helpful. Specific considerations may be relevant for analgesia: NSAIDs may dramatically reduce residual renal function and should be avoided. Morphine is renally excreted and can accumulate in those with advanced CKD, so alternatives such as oxycodone or alfentanil may be preferred.

Withdrawal of dialysis

Withdrawal of dialysis is necessary when the effort of attendance becomes burdensome, there is little improvement in quality of life, and the impact of other co-morbidities becomes intrusive.

If there is no residual renal function (i.e., the patient is anuric), survival after withdrawal of dialysis may be only a few days, but if there is some residual function, survival may be several weeks. Patients and carers need to understand these differences to make informed choices.

Further reading

Kane PM, Vinen K, Murtagh FE. Palliative care for advanced renal disease: a summary of the evidence and future direction. *Palliat Med* 2013; 27:817–821.
https://renaldrugdatabase.com *Prescribing guide for renal practitioners*.

Liver disease

Symptom burden in patients with advanced liver disease is high, and mortality rates from chronic liver disease are rising. Decompensated cirrhosis requiring admission to hospital is associated with a 50% mortality rate at 12 months. Early discussion with patients and their families is important to understand their priorities, allow for advanced care planning, and optimise symptom management. Palliative care is useful in conjunction with active management of chronic liver disease to improve quality of life and can provide support when patients deteriorate. Additionally, patients with advanced liver disease and their carers often have significant psychological,

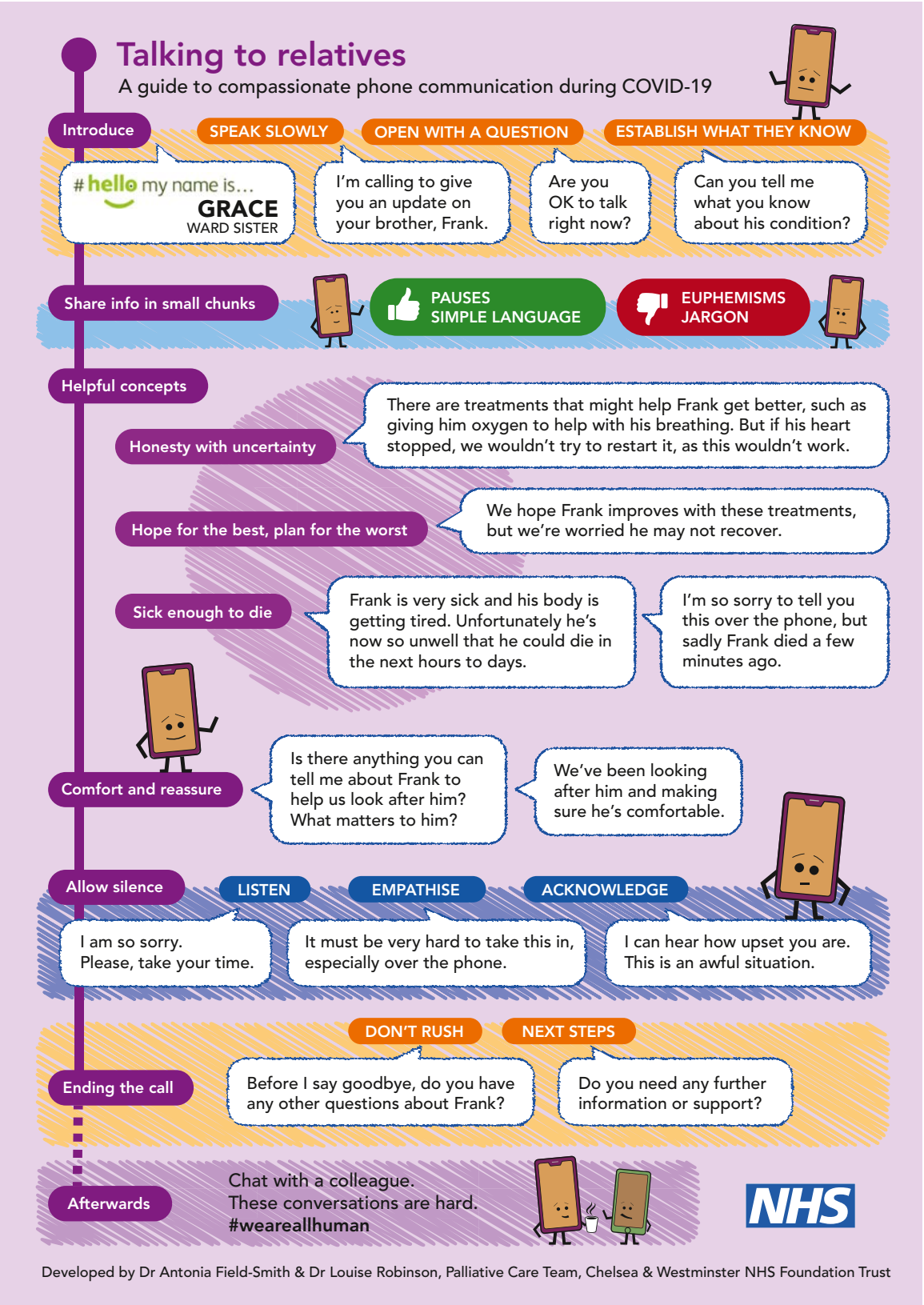


Fig. 50.6 Talking to relatives. A guide to compassionate phone communication during COVID-19.

social, and financial needs that need to be identified and supported as part of their holistic management.

Many commonly used medications are hepatotoxic, and there may be concurrent renal impairment. Therefore, drug choice and dosing must be considered carefully, with dose reductions often required. Regular monitoring is important, especially when titrating medications, and SPC advice should be sought.

Further reading

Woodland H, Hudson B, Forbes K, McCune A, Wright M; British Association for the Study of the Liver (BASL) End of Life Special Interest Group. Palliative care in liver disease: what does good look like? *Frontline Gastroenterol*. 2019; 11(3):218–227.

Neurological disease

People with chronic degenerative neurological diseases have a considerable burden of palliative care needs, which may include:

- Difficulties in swallowing (e.g., in motor neurone disease and multiple sclerosis)
- Excessive difficulty clearing secretions and weak cough, leading to a fear of choking
- Progressive loss of physical function with associated pain and vulnerable pressure areas
- Loss of mental capacity, anxiety, and depression

The MDT is vital to the care of these patients and may provide physiotherapy and occupational therapy to manage loss of function, as well as speech and language therapy to optimize communication. Discussions regarding the patient's care preferences and practical wishes (e.g., writing a will) should take place early so that these can be supported if the patient is able to participate and when communication is easier.

Motor neurone disease

Motor neurone disease is usually rapidly progressive, often requiring hospice support. Percutaneous endoscopic gastrostomy feeding may be needed. Secretions, such as excess saliva, can be problematic, and treatment may entail careful titration of an anti-secretory medication such as hyoscine hydrobromide, which is available as a transdermal preparation. In addition, if ventilatory failure develops, NIV may be offered. Patients and their families need to understand:

- Why this treatment has been offered (to prevent hypercapnia and associated morning headache and confusion; it can confer a survival benefit)
- When this treatment will be withdrawn (when it is no longer helping to maintain or improve quality of life in the face of advancing disease; guidelines are widely available to help with these specialist discussions)

Patients need to have a clear understanding of what alternative symptom control will be offered at withdrawal.

Multiple sclerosis

Pain is often prominent because of muscle spasm. This is managed with regular physiotherapy; medications; and in severe cases, botulinum toxin, nerve blocks, or surgery. Patients may become too disabled to attend outpatient clinics, requiring a proactive palliative care approach to ensure they do not miss out on surveillance or support. Hospice day therapy services, rehabilitation, and support for the family can have a huge impact on quality of life. A person can live for many years disabled and symptomatic, and regular surveillance and support are essential.

Dementia

Dementia-related palliative care needs arise in the context of neurological conditions that:

- Tend to occur in older people (Alzheimer's disease and vascular dementia)
- Also affect younger people (e.g., Parkinson's disease, multiple sclerosis, Huntington's disease).

It is often difficult to ascertain whether these patients are in pain. A tool that assesses behavioural response to pain can be useful, such as the Abbey or Doloplus pain scale; these are based on vocalization (e.g., groaning), facial expression (e.g., frowning), and body language (fidgeting, rocking), behaviour (e.g., confusion, refusal to eat), and physical changes.

In England and Wales, Deprivation of Liberty Safeguards (DoLS) may need to be considered for people lacking mental capacity to consent to care arrangements in a care home or hospital to ensure any deprivation of liberty is necessary to keep them safe from harm and in the person's best interest.

Further reading

Hussain J, Allgar V, Oliver D. Palliative care triggers in progressive neurodegenerative conditions: an evaluation using a multi-centre retrospective case record review and principle component analysis. *Palliat Med* 2018; 32:716–725.

Palliative care for young adults

Increasing numbers of children reach adulthood with life-threatening and life-limiting illness, such as cystic fibrosis and severe cerebral palsy. Palliative care supports these young adults and their families to achieve the best possible quality of life and care. The model of palliative care in young people is different to that in services built for older adults because many young people are supported for years by hospices; generally, they are cared for in their own homes, and the number of inpatient deaths is small compared with that in older adults. Many conditions are extremely rare and may be familial. As well as symptom management, hospices can provide short respite breaks for families caring for seriously ill young people. Ensuring a smooth transition from paediatric to adult palliative care services is an evolving area of development.

Further reading

https://www.togetherforshortlives.org.uk *Help to support seriously ill children who are expected to have short lives.*

RECOGNIZING DYING

After reversible causes of deterioration have been excluded, the dying phase can be recognized from a combination of the features in **Box 50.13**. It is important to recognize this phase to enable the right support to be given to patients and families.

Care of the dying

Most people express a wish to die in their own homes, provided their symptoms are controlled and their families are supported. However, patients may die in any setting, and all healthcare professionals should therefore be proficient in end-of-life care.

Hospitals and communities have developed individualized or personalized care plans for the last days of life that help to prompt

Box 50.13 Recognizing dying

	Features
Weeks ↓ Days	Ongoing deterioration despite active treatments may indicate irreversibility
	Extreme tiredness or weakness, meaning help with all personal care is required
	Becoming bedbound
	Little interest in food or fluids, weight loss
	Sleeping for most of the day or being comatose
	Difficulty swallowing medications
	Skin changes caused by peripheral vasoconstriction
	Noisy breathing from secretions collecting in the throat or chest
	Changes in breathing pattern, e.g., irregular rate, apnoeic pauses

discussions surrounding the dying process and how physical, psychological, social, spiritual, and family concerns will be addressed (**Box 50.14**). The decision to provide support through a personalized care plan for the last days of life is reached by a multiprofessional team through careful assessment of the patient, exclusion of reversible causes of deterioration, and excellent communication with family.

Further reading

General Medical Council. *Treatment and Care Towards the End of Life: Good Practice in Decision Making.* London: GMC; 2022.
National Institute for Health and Care Excellence. *NICE Guideline NG31: Care of Dying Adults in the Last Days of Life.* NICE 2015; https://www.nice.org.uk/guidance/ng31.

Box 50.14 Components of personalized care plans for the last days of life

- Recognition of the dying phase and communication with patient and carers
- Initial assessment (which includes the patient's and carers' understanding, wishes, and psychological state)
- Resuscitation status and treatment escalation plans
- Discussion surrounding how symptoms will be managed
- Discussion about feeding and fluids
- Ongoing assessment and monitoring
- Support for the carers after the patient's death

Bibliography

British National Formulary. *Guidance on Prescribing: Prescribing in Palliative Care. British National Formulary*. London: BMJ Group and RPS Publishing; http://www.bnf.org.
Twycross R, Wilcock A, Howard P. *Palliative Care Formulary*, 6th edn; 2018; www.pharmpress.com.
Watson M, Armstrong P, Back I et al. *Palliative Adult Network Guidelines*, 4th edn; 2016; https://book.pallcare.info.

Significant websites

http://uk.sagepub.com/en-gb/eur/journal/palliative-medicine *Palliative medicine.*
http://www.cancerresearchuk.org *UK charity.*
https://www.cuh.nhs.uk/our-services/breathlessness-intervention-service/ *Cambridge University Hospitals breathlessness information.*
http://www.macmillan.org.uk *UK patient organization.*
http://www.palliativedrugs.com *Palliative drugs information.*

Questions for this chapter can be found on Elsevier eBooks+.

A Chapter questions

Chapter 2: Clinical communication

1. Professional–patient communication is a form of institutional communication. What is the best way to prevent difficulties arising from institutional communication in healthcare settings?
   a. Confirm appointment times with the patient before attendance.
   b. Have a policy that all healthcare staff wear name badges.
   c. Make the healthcare setting look as welcoming as possible.
   d. Provide information to the patient about what is going to happen.
   e. Put up signs explaining the different uniforms that health professionals wear.
2. Medical empathy is an important feature of healthcare communication. What is the single best way to determine whether the health professional has been empathic in a consultation?
   a. Ask the patient if they felt understood.
   b. Confirm the professional's tone of voice became softer when discussing emotional content.
   c. Determine that there has been more eye contact time than non–eye contact time between the professional and patient during the consultation.
   d. Get the health professional to complete a reliable and valid empathy questionnaire.
   e. Note the presence of 'empathic touch'.
3. Patients need good-quality information to make decisions about their healthcare. What is the single best way for a health professional to share information?
   a. Ensure there are plenty of pauses between sentences.
   b. Give medical terms first and then the lay explanation.
   c. Go sufficiently slowly so that there is no need to repeat information.
   d. Provide information first and then ask for questions at the end.
   e. Start with what the patient already knows and build from there.
4. A patient attends their GP to discuss a new medical problem. What is the single best way for the health professional to gather information?
   a. Ask a series of specific questions using a clerking checklist.
   b. Have the patient explain their problem to the receptionist when booking the appointment.
   c. Review the patient's previous medical problems with the patient.
   d. Start with open questions.
   e. Use plenty of silence and eye contact.
5. Medical treatments may have adverse side-effects. Which is the single best way to communicate the likelihood that a treatment will have side-effects?
   a. Use commonly understood words (e.g., 'sometimes', 'usually', 'occasionally').
   b. Use natural frequencies (e.g., 4 of 10 people).
   c. Use percentages (e.g., 40% of people).
   d. Use reassuring language (e.g., 'hardly any side-effects').
   e. Use relative risks (e.g., half the side-effects of other treatments).

Chapter 3: Medical ethics

1. The GMC states that doctors have a 'duty of candour'. Which is the best definition of 'candour'?
   a. Speaking openly without trying to conceal information
   b. Voluntarily telling the truth, especially about difficult or embarrassing subjects
   c. Answering questions honestly, without hiding important information
   d. Raising concerns about patient safety even if it means disclosing information about the performance of colleagues
   e. Being honest and telling the truth, especially about difficult or embarrassing subjects, and speaking frankly without trying to conceal information or put a positive spin on it for our own advantage
2. The GMC blog called 'Making decisions together—the implications of the Montgomery judgement' gives advice on how to ensure clinical practice is compliant with the judgement: 'Discussing **risk** with a patient shouldn't be limited to a numbers game; it should be part of a dialogue. The key is to understand what matters—or is likely to matter—to the individual patient'. What term is used to discuss **risk** that matters or is likely to matter to a patient?
   a. Qualifying risk
   b. Material risk
   c. Legal risk
   d. Evidential risk
   e. Specific risk
3. If a case of clinical negligence occurs, which body administers the indemnity scheme for NHS Trusts in England?
   a. British Medical Association
   b. Department of Health
   c. NHS Resolution
   d. Clinical Negligence Indemnity Scheme
   e. Clinical Negligence Scheme for Trusts

Chapter 4: Evidence-based practice

1. Which of the following is a primary source of evidence?
   a. Medline
   b. Cochrane Database of Systematic Reviews
   c. Research papers
2. Which of the following is not a component in the PICO approach to developing structured clinical questions?
   a. Patient
   b. Comparison
   c. Observations
3. Which of these three study types is regarded as the strongest form of evidence in the 'hierarchy of evidence'?
   a. Randomized controlled trials
   b. Case series
   c. Observational studies

4. What is the goal of randomization in randomized controlled trials?
   a. To ensure that study groups are as similar as possible
   b. To ensure that allocation to the intervention or control arm is fair
   c. To ensure that similar numbers of participants are allocated to each study arm
5. Which of the following is not true of pragmatic randomized controlled trials?
   a. Outcome measures chosen represent the full range of health gains relevant to patients and clinicians.
   b. Pragmatic trials should represent the spectrum of patients to whom the treatment will be applied.
   c. Patients and clinicians must be blinded to the intervention that is being delivered.
6. Which of the following statements is true of cohort studies?
   a. No adjustments are made to participants' treatments or lifestyle.
   b. Cohort studies are always retrospective (i.e., looking back in time).
   c. Cohort studies cannot be used to study rare outcomes.
7. Which of the following statements is true of case-control studies?
   a. There is a high risk of recall bias affecting the results.
   b. Careful matching of cases and controls eliminates the risk of unknown confounding factors affecting the results.
   c. Information is usually gathered from cases and controls via focus groups.
8. Qualitative research methods (e.g., interviews and focus groups) can be useful to explore which of the following type of question?
   a. The number of patients living with a condition
   b. Whether a new intervention is acceptable to patients
   c. The likely length of time between diagnosis of a condition and death
9. In a study with high internal validity, which of the following is correct?
   a. The study findings are relevant to patients in other settings.
   b. The study is sufficiently free from bias.
   c. The study team is reputable.
10. In a study with high external validity, which of the following is correct?
    a. The study findings are relevant to patients in other settings.
    b. The study is sufficiently free from bias.
    c. The study team is reputable.
11. Which of the following is an example of potential information bias?
    a. Participants in a cohort study are all patients at a major teaching hospital.
    b. Participants in a case-control study are more likely to remember relevant past history if they have the disease of interest.
    c. The average age of participants in the intervention arm of a randomized controlled trial is 10 years younger than that in the control group.
12. What does an odds ratio describe?
    a. The odds of an event occurring in the intervention group divided by the odds in the control group
    b. The frequency with which an event occurs in the intervention group versus the control group
    c. The probability of seeing a result, such as the one obtained if the intervention had no real effect
13. What does a risk ratio describe?
    a. The number of people to whom the treatment would have to be given to see an effect
    b. The proportion of people in the treatment group who experience the outcome event divided by the same proportion in the control group
    c. The likelihood that an intervention will have no effect versus the control
14. In a study of all-cause mortality in patients who have received a novel cancer drug, compared with conventional medication, the odds ratio is 1.14. What does this suggest?
    a. There is no difference between the intervention and the control.
    b. The control performs better than the intervention.
    c. The intervention performs better than the control.
15. What does a 95% confidence interval tell us?
    a. There is a 95% probability that the 95% confidence interval contains the true value.
    b. The results are 95% correct, which is sufficient for us to be confident in their validity.
    c. The study has involved 95% of the possible participants in the population, so we can confidently report that the results are true for this group.
16. The *P* value reported for a randomized controlled trial is .01. What does this tell us?
    a. There is strong statistical evidence against the null hypothesis.
    b. The intervention is not effective.
    c. The intervention should be used in routine clinical practice.

Chapter 5: Human genetics

1. A patient presents with palmoplantar keratoderma and seeks a genetic diagnosis. What would be an appropriate first test to pursue?
   a. Whole-genome sequencing.
   b. Targeted panel of genes.
   c. SNP array.
   d. Single-cell RNA sequencing.
   e. Methylation array.
2. A genetic test report is returned on a neonate and reveals a new pathogenic variant in keratin 1, compatible with a diagnosis of epidermolytic ichthyosis. Neither parent is known to be affected. Which one is not a possible explanation?
   a. Gonadal mosaicism.
   b. Cutaneous and gonadal mosaicism.
   c. Nonpaternity.
   d. Both parents are unaffected carriers.
   e. *De novo* postzygotic mutation.
3. A mother and son attend the dermatogenetic clinic with a history of familial melanoma. The son, aged 38 years, has just had his first melanoma, which was a stage 1B tumour. His mother had a stage 3 tumour with lymph node involvement at age 58 years, has received immunotherapy, and is now in remission. Her father died of pancreatic cancer at the age of 42 years. What would you do next?
   a. Family history questionnaire.
   b. Test the mother's germline DNA.
   c. Test the son's germline DNA.
   d. Do not offer a test.
   e. Test the mother's archival skin tumour.
4. A couple are referred to the genetic counselling clinic because they are planning to have a family in the near future. The wife, who is healthy, has a family history of haemophilia A with both her brothers having the diagnosis. After a genetic test, the wife

but not the husband is a carrier for the common haemophilia A pathogenic variant. Which statement is true?
   a. The factor VIII gene is located on chromosome 2.
   b. Fifty percent of girls will be carriers.
   c. All boys will be affected.
   d. It is an X-linked dominant condition.
   e. There is no risk of passing on haemophilia A to the boys.

5. A saliva sample is taken from a boy with congenital hearing loss. Both parents have normal hearing. The sample is sent to the laboratory for a *GJB2* gene test involving polymerase chain reaction and sequencing. What is not needed to perform this genetic test?
   a. DNA.
   b. Oligonucleotide primers.
   c. Bisulphite modification of DNA.
   d. Capillary electrophoresis.
   e. Heat-stable DNA polymerase.

Chapter 6: Immunity

1. Which of the following is a key cellular component of the innate immune system responsible for phagocytosis and the initial defense against microbial infections?
   a. T cells
   b. B cells
   c. Natural killer cells
   d. Macrophages
   e. Helper T cells

2. Which of the following cells is primarily responsible for presenting antigens to T cells and initiating the adaptive immune response?
   a. Neutrophils
   b. Macrophages
   c. Natural killer cells
   d. B cells
   e. Dendritic cells

3. What is the primary function of antibodies in the immune system?
   a. Phagocytosis
   b. Cytokine production
   c. Antigen presentation
   d. Neutralization of pathogens
   e. Oxygen transport

4. Checkpoint inhibitors in the immune system primarily function by
   a. enhancing phagocytosis.
   b. suppressing cytokine production.
   c. inhibiting T-cell activation.
   d. blocking the activity of antibodies.
   e. releasing the brakes on T-cell activity.

5. Anaphylaxis is a severe and potentially life-threatening allergic reaction that involves the release of large amounts of histamine and other mediators from
   a. neutrophils.
   b. B cells.
   c. T cells.
   d. mast cells.
   e. macrophages.

Chapter 7: Therapeutics and prescribing

1. Which route of administration is associated with the highest bioavailability?
   a. Oral
   b. Intravenous
   c. Buccal
   d. Intramuscular
   e. Topical

2. Which of these is not a recognized phase in pharmacokinetics?
   a. Absorption
   b. Distribution
   c. Metabolism
   d. Inhibition
   e. Excretion

3. Which of the following features is most closely associated to a type B adverse reaction?
   a. Dose independent
   b. Cause teratogenicity
   c. Occurs on drug cessation
   d. Occurs only after prolonged exposure only
   e. Invariably benign in clinical outcome

4. For a drug X at steady state that has an elimination half-life of 4 hours, how long would it take for effective complete drug elimination after stopping the medication?
   a. 2 hours
   b. 4 hours
   c. 12 hours
   d. 20 hours
   e. 48 hours

5. Postmarketing studies to evaluate drug safety are most commonly referred to as what phase drug study?
   a. I
   b. II
   c. III
   d. IV
   e. V

Chapter 8: Radiology

1. What is the investigation of choice for the initial identification of gallstones in a patient?
   a. CT of the abdomen
   b. HIDA scan
   c. MRI of the abdomen or MR cholangiopancreatogram (MRCP)
   d. Ultrasonography of the abdomen
   e. X-ray of the abdomen

2. What is the first investigation if the clinical concern is bacterial meningitis?
   a. CT of the brain
   b. X-ray of the brain
   c. CT angiography of the brain
   d. MRI of the brain
   e. None of the above

3. Which of the following examinations is contraindicated if the patient has renal impairment (estimated glomerular filtration rate of <30 mls/min/1.73 m^2)?
   a. Non-contrast CT
   b. Non-contrast MRI
   c. Ultrasonography
   d. Contrast-enhanced CT
   e. Contrast-enhanced MRI (gadolinium)

4. Which of the following statements is false on a normal chest X-ray (CXR)?
   a. You can estimate the heart size only on a CXR taken with posteroanterior (PA) projection.
   b. A well-inspired CXR has 8 to 10 posterior or 5 to 6 anterior ribs visible.
   c. The mediastinum measures 10 cm.
   d. Lucency (or 'blackness') between the lungs could be different because of patient rotation on CXR.
   e. The trachea should be equidistant between the medial ends of the clavicles in a non-rotated CXR.

5. Which of the following regarding MRI is not true?
   a. The MRI scanner is never switched off.
   b. Pacemakers are contraindicated for MRI.
   c. Cochlear implants are contraindicated for MRI.
   d. Pregnancy is an example of relative contraindication for MRI.
   e. Implanted metallic objects (e.g., bullets, shrapnel, occupational iron embedded in the eyes) are contraindicated for MRI.

Chapter 9: Public health

1. The Bradford Hill criteria on the strength of an association between a disease and its proposed cause state that causality is more likely if
   a. the effect size is small.
   b. the effect occurs before the cause.
   c. the association is demonstrable in only a specific timeframe.
   d. greater exposure leads to an equivalent effect.
   e. the effects could be proposed to be similar to other situations.
2. The Wilson and Jungner screening criteria include
   a. there should be an inexpensive treatment.
   b. the condition should be rare.
   c. there should be no latent stage.
   d. the natural history of the condition should be well understood.
   e. case finding should occur annually.
3. Which of the following is a property of screening tests?
   a. False-positive results occur when individuals test negative but do have disease.
   b. False-negative results occur when individuals with no disease test positive.
   c. Sensitivity is the percentage of true negatives.
   d. Positive predictive value varies with prevalence of the disease.
   e. Specificity is the ability of the test to detect all cases of disease.
4. What is currently included within the *routine* UK vaccination schedule for *school-aged* children?
   a. Rotavirus
   b. Influenza
   c. Pertussis
   d. Measles, mumps, and rubella
   e. Hepatitis B
5. The Index of Multiple Deprivation can be used to
   a. compare small areas across England.
   b. quantify how deprived a small area is.
   c. explore the causes of deprivation.
   d. identify individuals who are deprived.
   e. look at change in absolute deprivation over time.

Chapter 10: Global and planetary health

1. Over the period 2000 to 2020, average life expectancy
   a. has increased by 10 years for women and men.
   b. is between 60 and 65 years in lower- and middle-income countries.
   c. has decreased because of economic austerity.
   d. mirrors the increase in healthy life expectancy.
   e. is directly related to the per capita spend on healthcare provision.
2. Which of the following is a Bellwether procedure in global surgery?
   a. Treatment of 'closed' fractures
   b. Laparotomy
   c. Placement of chest drains
   d. Amputation for diabetic gangrene
   e. Haemostasis for major arterial bleeding
3. Which of the following statement is true for global food security?
   a. Deficiencies in transport and storage of food account for 10% of food wastage in lower- and middle-income countries.
   b. Land exploitation is beneficial for food production.
   c. Child malnutrition is responsible for 20% of deaths in children up to 5 years of age.
   d. One-third of all food produced is wasted each year.
   e. Food security is promoted by industrial development.
4. With regard to the security of water and sanitation 2023, which of the following statements is false?
   a. Seventy percent of the global population now has access to a 'safely managed drinking water source'.
   b. Access to water supplies is a hazardous process for women.
   c. Access to water and sanitation is not a UN-recognized human right.
   d. Collection of water is usually undertaken by child labourers.
   e. One billion people collect water from unprotected sources.
5. The following international agreements are not legally binding:
   a. The International Covenant on Civil and Political Rights.
   b. The Committee on the Elimination of Discrimination against Women.
   c. The UN Convention on the Rights of the Child.
   d. The Universal Declaration of Human Rights.
   e. The International Covenant on Economic, Social and Cultural Rights.

Chapter 11: Acute medicine

1. A 64-year-old male with a body weight of 125 kg is referred to Same Day Emergency Care (SDEC) with a 4-day history of pleuritic chest pain and breathlessness. He has a history of deep vein thrombosis (DVT) previously, which had been treated at that time with warfarin therapy for 6 months.

On examination, he has a raised JVP and oxygen saturation of 90% (94%–98%) breathing room air.

Investigations:

ECG	Sinus rhythm, 105 beats/min with right bundle branch block
CXR	Normal
CTPA	Dilated right ventricle and pulmonary arteries; bilateral filling defects in segmental branches of upper and lower pulmonary artery

What is the single most appropriate measurement to help decide if he can be considered for management in an ambulatory fashion?
   a. Body mass index (BMI)
   b. D-dimer value
   c. QRS duration on ECG
   d. Simplified PE severity index (SPESI)
   e. Wells score
2. An 83-year-old male with a history of dementia and reduced mobility presents with shortness of breath and right-sided chest pain. Clinical examination reveals a dull right base, with reduced breath sounds and bronchial breathing above the area. It is 20:00 on a Friday evening. Which of the following features

is most likely to make the treating clinician consider inserting a chest drain out of hours?

a. CXR showing right-sided pleural effusion occupying the lower half of the thoracic cavity
b. Clinical evidence of a deep vein thrombosis (DVT)
c. Needle aspiration showing pleural fluid with a pH of 7.3
d. Requiring 24% oxygen to keep oxygen saturations above 94%
e. Septations and debris on the right side on point-of-care ultrasound (POCUS)

3. A 41-year-old female with no known medical history is admitted with jaundice and pruritus (itching). She recently had dark coloured urine and had been treated for pneumonia with amoxicillin and clavulanic acid 2 weeks previously. She had also been taking paracetamol up to a maximum dose of 1000 mg four times a day for the past 5 days. She regularly drinks approximately 12 units of alcohol per week.

On examination, she has a BMI of 29.0 kg/m^2 (18.0–25.0). She is icteric with excoriations. There is no organomegaly noted. The respiratory examination is unremarkable.

Investigations:

Haemoglobin	127 g/L (130–180)
White blood cells	8.0 × 10^9/L (4.0–11.0)
Neutrophils	4.9 × 10^9/L (1.5–7.0)
Lymphocytes	1.7 × 10^9/L (1.5–4.0)
Platelets	360 × 10^9/L (150–400)
Serum total bilirubin	152 μmol/L (1–22)
Serum alanine aminotransferase	85 U/L (5–35)
Serum aspartate aminotransferase	65 U/L (1–31)
Serum alkaline phosphatase	99 U/L (45–105)
Serum CRP	5 mg/L (<10)
Ultrasound scan of abdomen	Thin-walled gallbladder, multiple gallstones; common bile duct measuring 4 mm

What is the most likely cause of her jaundice?

a. Chronic liver disease caused by steatosis
b. Cytomegalovirus (CMV) primary infection
c. Idiosyncratic reaction to amoxicillin and clavulanic acid
d. Recently passed impacted gallstones
e. Staggered paracetamol overdose

4. A 22-year-old female presents with shortness of breath associated with nausea and vomiting. She has had diabetes mellitus since the age of 12 years. Her diabetes has been treated by basal bolus insulin regimen, but she has not taken it today because of not being able to eat. She weighs 60 kg.

Examination shows a soft, non-tender abdomen but clinical dehydration. Chest examination is unremarkable apart from a respiratory rate of 28 breaths/min.

Investigations:

Haemoglobin	161 g/L (130–180)
White blood cells	7.0 × 10^9/L (4.0–11.0)
Neutrophils	4.9 × 10^9/L (1.5–7.0)
Lymphocytes	1.2 × 10^9/L (1.5–4.0)
Platelets	210 × 10^9/L (150–400)
Serum sodium	124 mmol/L (137–144)
Serum potassium	2.8 mmol/L (3.5–4.9)
Serum urea	13.8 mmol/L (2.5–7.0)
Serum creatinine	140 μmol/L (60–110)
Serum CRP	4 mg/L (<10)
Fasting plasma glucose	28.0 mmol/L (3.0–6.0)
Haemoglobin A_{1c}	66 mmol/mol (20–42)
Blood ketones	6.0 mmol/L (<0.6)

Arterial blood gases, breathing air:

PaO_2	14.1 kPa (11.3–12.6)
$PaCO_2$	2.8 kPa (4.7–6.0)
pH	7.02 (7.35–7.45)
Bicarbonate	14 mmol/L (21–29)
Base excess	–8.0 mmol/L (±2.0)
Lactate	3.0 mmol/L (0.5–1.6)
Oxygen saturation	99% (94–98)
Serum cortisol (22.00 h)	92 nmol/L (50–180)

With regular monitoring, what is the best management strategy?

a. 0.9% saline only
b. 0.9% saline with potassium and 3 units per hour of fixed-rate insulin and usual long-acting insulin
c. 0.9% saline with 3 units per hour fixed-rate insulin and i.v. hydrocortisone
d. 0.9% saline with potassium and 6 units per hour of fixed-rate insulin
e. 0.9% saline with potassium and 6 units per hour of fixed-rate insulin and usual long-acting insulin

5. The acute medicine team receives an alert from the on-call biochemist that a 62-year-old female with baseline eGFR of 28 mL/min/1.73 m^2 (>60) had an acute kidney injury on blood tests from the outpatient clinic. She has stage 4 chronic kidney disease and has blood taken in the outpatient clinic, where she is being seen regularly for long-standing osteomyelitis of the foot. The observations from the clinic were within normal limits, and it is noted that she has a long-term catheter in situ.

Investigations 2 weeks previously:

Serum sodium	142 mmol/L (137–144)
Serum potassium	4.2 mmol/L (3.5–4.9)
Serum urea	7.9 mmol/L (2.5–7.0)
Serum creatinine	280 μmol/L (60–110)

Investigations from today:

Serum sodium	145 mmol/L (137–144)
Serum potassium	6.1 mmol/L (3.5–4.9)
Serum urea	8.1 mmol/L (2.5–7.0)
Serum creatinine	370 μmol/L (60–110)

What is the most likely reason for the worsening of her biochemistry results?

a. Co-trimoxazole therapy
b. Dehydration
c. Potassium permanganate soaks
d. Red flag sepsis
e. Urinary obstruction

Chapter 12: Sepsis and the treatment of bacterial infection

1. A ventilated patient in the intensive care unit develops progressive respiratory failure. They have already received 5 days of treatment with piperacillin/tazobactam, and *Acinetobacter baumannii* has now been isolated from a sputum sample. Which of the following antibiotics could be used as empirical treatment of ventilator-associated pneumonia?

a. Azithromycin
b. Ceftazidime
c. Meropenem
d. Ampicillin/sulbactam
e. Ceftolozone/tazobactam

2. Which of the following antibiotics are likely to be active against *Enterococcus faecium* and could be used as monotherapy for a complicated intra-abdominal infection.
   a. Teicoplanin
   b. Amoxicillin/clavulanate
   c. Cefotaxime
   d. Metronidazole
   e. Tigecycline
3. A 35-year-old male presents with a 3-day history of lower limb cellulitis. He has no known penicillin allergy. On examination, he has cellulitis from his ankle to his midthigh. He is referred to the OPAT service. Which intravenous antibiotic should he be given?
   a. Benzylpenicillin
   b. Ceftriaxone
   c. Daptomycin
   d. Flucloxacillin
   e. Teicoplanin
4. A 45-year-old female presents with a 5-day history of fever, rigors, and left flank pain. On examination, she is alert and orientated, and her blood pressure is 135/75 mmHg. Blood cultures grow *Escherichia coli* (ESBL positive), susceptible to meropenem, ertapenem, and amikacin. The patient is referred to the OPAT service. Which intravenous antibiotic should she be given?
   a. Amikacin
   b. Ceftriaxone
   c. Ertapenem
   d. Meropenem
   e. Tazocin
5. Septic shock is best defined as
   a. septic cardiomyopathy leading to distributive shock.
   b. positive blood cultures and low blood pressure.
   c. sepsis with an unsupported mean arterial blood pressure below 65 mmHg at presentation.
   d. sepsis with an unsupported mean arterial blood pressure below 65 mmHg and an arterial blood lactate concentration above 2 mmol/L after fluid resuscitation.
   e. sepsis with an arterial blood lactate concentration above 2 mmol/L.

Chapter 13: Critical care medicine

1. A previously fit and well 25-year-old female has been admitted to the intensive care unit with sepsis secondary to a urinary tract infection. Despite appropriate fluid therapy, she remains hypotensive. She has both invasive arterial and central venous pressure monitoring. She has a pulse of 120 beats/min, blood pressure of 80/40 mmHg, and central venous pressure of 10 mmHg. A cardiac output monitor demonstrates a cardiac index of 4.9 L/min/m^2 with a stroke volume variation of less than 5%. Her urine output has been only 20 mL for the past 4 hours. Arterial blood gas analysis demonstrates a lactate of 4.6 mmol/L. What is the most appropriate next intervention?
   a. Administration of a further fluid bolus of 250 mL of compound sodium lactate.
   b. Start an infusion of dobutamine.
   c. Administer a bolus of furosemide 40 mg i.v.
   d. Start an infusion of norepinephrine.
   e. Start an infusion of vasopressin.
2. A 74-year-old male is admitted in cardiac arrest after collapsing at home shortly after complaining of some chest pain. CPR has been ongoing for 20 minutes, and he has been intubated by the pre-hospital team. Mechanical ventilation is ongoing on 100% oxygen. He is in ventricular fibrillation on his arrival and successfully cardioverted with a single 360J DC shock. An arterial blood gas is obtained shortly after return of circulation:

pH	7.12
$PaCO_2$	4.75 kPa
PaO_2	61.20 kPa
$CHCO_3$	8 mmol/L
Base excess	–15 mmol/L

Which statement best describes this patient's metabolic disturbance?
   a. Respiratory acidosis
   b. Respiratory alkalosis
   c. Metabolic acidosis with associated hypoxia
   d. Metabolic acidosis without associated hypoxia
   e. Metabolic alkalosis
3. Concerning the acute respiratory distress syndrome:
   a. the syndrome can be defined by clinical parameters alone when the radiology findings are clear.
   b. steroids have been shown to be effective in the initial treatment of patients with the condition.
   c. a PaO_2/FiO_2 ratio of less than or equal to 300 mmHg without PEEP is diagnostic of ARDS.
   d. the condition does not occur in patients who are immunosuppressed because it is primarily mediated by an adverse immune reaction.
   e. improved survival has been demonstrated with lower ventilatory tidal volumes.
4. A 22-year-old male is admitted acutely unwell to hospital. His blood results show the following:

pH	7.10
$PaCO_2$	4.23 kPa
PaO_2	12.08 kPa
$CHCO_3$	12 mmol/L
Base excess	–10 mmol/L
Na	143 mmol/L
K	4.9 mmol/L
Cl	95 mmol/L
Urea	4.8 mmol/L
Creatinine	143 micromol/L
Glucose	4.9 mmol/L
Urine ketones	negative

What is the likely diagnosis?
   a. Lactic acidosis
   b. Diabetic ketoacidosis
   c. Salicylate poisoning
   d. Diarrhoea
   e. Ethylene glycol toxicity
5. A 45-year-old male is an inpatient on intensive care after an inferior myocardial infarction. On examination, his pulse is 100 beats/min, and blood pressure is 80/40 mmHg. Clinical examination reveals an elevated JVP.

   A Swan-Ganz catheter is inserted, and the following values are obtained:

Right atrial pressure	10 mmHg	(8–12)
Pulmonary artery pressure	24/0 mmHg	(20/0–35/0)
Pulmonary capillary wedge pressure	5 mmHg	(5–10)
Mixed venous oxygen saturation	62%	(62%–65%)
Cardiac index	3.4 L/min/m^2	(>2.5)

These findings are consistent with what diagnosis?

a. Cardiac tamponade
b. Hypovolaemia
c. Ventricular septal defect
d. Mitral regurgitation
e. Cardiogenic shock

Chapter 14: Anaesthesia and perioperative medicine

1. A 75-year-old male patient with type 2 diabetes mellitus and hypertension underwent elective open right hemicolectomy for bowel cancer 3 days ago. It was surgically complex and lasted 5 hours. He has not yet mobilized from bed because of pain and is struggling to cough effectively. His oxygen saturations have fallen over the past 12 hours, such that he now needs 10 L/min of oxygen to maintain SpO_2 above 94%. What is the most likely cause of his hypoxia?

a. Pneumothorax
b. Myocardial infarction
c. Pneumonia
d. Pleural effusion
e. Anaphylaxis

2. An 82-year-old female is scheduled for elective lumpectomy of a breast lesion under general anaesthetic. She takes amlodipine for hypertension, with a preoperative BP of 131/78 mmHg, senna, and vitamin D supplements. Which is the most accurate ASA score for this patient?

a. 1
b. 2
c. 3
d. 4
e. 5

3. A 63-year-old male is scheduled for elective radical cystectomy for bladder cancer. He has a body mass index of 34.1 kg/m^2, smokes 10 cigarettes per day, and has had a previous myocardial infarction with percutaneous stent insertion. He is referred to the local prehabilitation service. Select the best description of the core components of prehabilitation.

a. Psychology, exercise, and nutrition
b. Psychology, chemotherapy, and occupational therapy
c. Exercise, blood testing, and echocardiography
d. Echocardiography, psychology, and smoking cessation
e. Drinking, eating, and mobilizing

4. A 72-year-old female undergoes elective right nephrectomy for a renal cell carcinoma, with planned enhanced recovery. She receives total intravenous anaesthetic (TIVA) and is intubated easily. The surgeon requests placement of a nasogastric tube. She is given 2 L of balanced crystalloid, during the surgery. Thromboprophylaxis is with intermittent compression devices, then low-molecular-weight heparin postoperatively. Which element of her care should have been avoided based on the plan for enhanced recovery?

a. TIVA
b. Intubation
c. Nasogastric tube
d. Balanced crystalloid
e. Intermittent compression devices

5. A 53-year-old female is being assessed by the perioperative team in advance of a major head and neck cancer surgery. Which of the following would be the best way for the team to provide risk stratification for this patient?

a. Acute Physiology and Chronic Health Evaluation (APACHE)
b. European System for Cardiac Operative Risk Evaluation (EuroSCORE)
c. C-reactive protein levels
d. Surgical Outcome Risk Tool (SORT)
e. Montreal Cognitive Assessment (MOCA)

Chapter 15: Toxicology

1. What is the preferred antidotal treatment for a patient presenting within 2 hours of ingestion of 100 mL of antifreeze (100% ethylene glycol)?

a. Ethanol
b. Acetylcysteine
c. Fomepizole
d. Pralidoxime
e. Sodium bicarbonate

2. Which toxic gas is produced when the rodenticide aluminium phosphide reacts with moisture in the air?

a. Carbon monoxide
b. Chlorine
c. Hydrogen cyanide
d. Phosphine
e. Hydrogen sulfide

3. Which of the following is NOT a clinical sign of organophosphate pesticide poisoning?

a. Bradycardia
b. Miosis
c. Dry mouth
d. Bronchorrhea
e. Muscle fasciculations

4. Which electrocardiographic changes are consistent with severe tricyclic antidepressant toxicity?

a. Prolonged PR interval
b. Prolonged QT interval
c. Prolonged QRS duration
d. Peaked T waves
e. Down-sloping ST segment depression ('reverse tick')

5. Which of the following antidotes does not work by forming an inert complex with the poison?

a. Desferrioxamine
b. Digoxin-specific antibody fragments
c. Hydroxocobalamin
d. Succimer (dimercaptosuccinic acid)
e. Methylthioninium chloride ('methylene blue')

Chapter 16: Environmental medicine

1. Which of the following can be a medical problem in relation to aviation and space medicine?

a. Acceleration
b. Radiation exposure
c. Circadian dysrhythmia
d. Thermal stress
e. All of the above

2. In relation to acute high-altitude cerebral oedema, which of the following is the best treatment?

a. Dexamethasone and immediate descent to a lower altitude
b. Dexamethasone and atenolol
c. Nifedipine and oxygen
d. Sildenafil and a non-steroidal anti-inflammatory drug
e. None of the above

3. Hypothermia is defined as a core temperature of less than 35°C. Which of the following occurs in hypothermia?
   a. Oxygen dissociation curve shifts to the right
   b. Pulse rate rises
   c. Respiratory rate rises
   d. Coordination is lost
   e. QT interval shortens
4. In the diagnosis of heat stroke,
   a. the core temperature must be above 42°C.
   b. sweating will always be present.
   c. hypersalivation is common.
   d. intravascular coagulation and multiorgan failure can occur.
   e. neurological symptoms and signs are not present.
5. A person is rescued from immersion in water. Which of the following apply?
   a. The most common cause of death is from hypothermia.
   b. Immediate treatment is to drain water from the victim's lungs.
   c. The victim must be lifted vertically from the water.
   d. Immediate treatment is to give rescue breaths and oxygen if available.
6. In relation to human lightning strikes, which of the following is true?
   a. Fifty percent of humans struck by lightning die.
   b. Fern-shaped patterns on the skin should be treated as burns.
   c. Bystanders are at risk of electric shock from touching the patient after a lightning strike.
   d. Prolonged CPR may be required because of respiratory muscle paralysis.
   e. A defibrillator should not be used because of the thermal damage to the heart.

Chapter 17: Cardiology

1. A 40-year-old smoker presents to the emergency department with a 1-hour history of crushing central chest pain that radiates to his jaw. He is not on any regular medication. His heart rate is 90 beats/min, blood pressure is 145/95 mmHg and oxygen saturations are 98% on room air. What is the most appropriate next investigation?
   a. Cardiac enzymes
   b. Computed tomography pulmonary angiography
   c. D-dimer
   d. Echocardiogram
   e. Electrocardiography
2. A 76-year-old female is referred to cardiology outpatients with palpitations, which have been intermittent for the past week. She is on treatment for hypertension (perindopril 4 mg) and type 2 diabetes (metformin 1 g). Her electrocardiogram shows atrial fibrillation with a ventricular rate of 110 beats/min. Her blood pressure is 140/85 mmHg. Her total cholesterol is 6.4 mmol/L. Her HbA_{1c} is 7% or 53 mmol/mol. What is the best next management step to reduce her risk of future stroke?
   a. Atrial fibrillation ablation
   b. Clopidogrel
   c. Direct oral anticoagulant
   d. Left atrial appendage occlusion device
   e. Statin therapy
3. A 65-year-old male with ischaemic cardiomyopathy is being seen in the outpatient department. He develops shortness of breath after walking 50 m on the flat. His medications include ramipril, bisoprolol, eplerenone, aspirin, and atorvastatin. He has severely impaired left ventricular systolic function. He has moderate mitral regurgitation. His blood pressure is 110/75 mmHg, and his heart rate is 65 beats/min. His electrocardiogram confirms sinus rhythm and a narrow QRS duration. What is the next most appropriate step to reduce his risk of cardiovascular death and heart failure admission?
   a. Internal cardioverter-defibrillator
   b. Ivabradine
   c. Sodium-glucose transport protein 2 inhibitor
   d. Transcatheter mitral valve intervention
   e. Vericiguat
4. A 25-year-old healthy male is referred for a new diagnosis of hypertension. A 24-hour ambulatory blood pressure monitor has confirmed a daytime average of 150/95 mmHg. Clinical examination reveals a regular right radial pulse and a systolic murmur. It is difficult to palpate his femoral pulse. What is the likely relevant diagnosis?
   a. Aortic stenosis
   b. Coarctation of the aorta
   c. Conn's disease
   d. Mitral regurgitation
   e. Peripheral vascular disease
5. A 19-year-old university student presents to the emergency department with 3 hours of severe chest pain. It is worse with inspiration and better sitting up. There is ST segment elevation on the electrocardiograph. The troponin I assay is negative at 1 hour after arrival in the department. The remaining blood tests, including a full blood count, renal function, C-reactive protein, and D-dimer, are grossly unremarkable. What is the best management?
   a. Aspirin and prednisolone
   b. Aspirin and clopidogrel
   c. Direct oral anticoagulant
   d. Ibuprofen and colchicine
   e. Penicillin G injection

Chapter 18: Hypertension

1. Which of these medications is most likely to interfere with blood pressure control?
   a. Paracetamol
   b. Ibuprofen
   c. Chlorphenamine
   d. Codeine
   e. Omeprazole
2. Which of these dietary patterns is associated with hypertension?
   a. Low salt intake
   b. Low vegetable and fruit intake
   c. Low saturated fat intake
   d. Low simple carbohydrate intake
   e. Low animal protein intake
3. Which of these parameters is measured in millimetres of mercury (mmHg)?
   a. Jugular venous pressure
   b. Central venous pressure
   c. Diastolic blood pressure
   d. Heart rate
   e. Respiratory rate
4. Which of the following medications is least appropriate to prescribe to uncomplicated hypertension?
   a. Amlodipine
   b. Clonidine
   c. Lisinopril
   d. Spironolactone
   e. Indapamide

5. A patient with a daytime ambulatory blood pressure of 137/80 mmHg is best classified as having what type of blood pressure or hypertension?
   a. Normal blood pressure
   b. Elevated blood pressure
   c. Grade 1 hypertension
   d. Grade 2 hypertension
   e. Severe hypertension

Chapter 19: Lipid and metabolic disorders

1. Which of the proteins below is found only in chylomicron particles?
   a. Apolipoprotein B48
   b. Apolipoprotein B100
   c. Apolipoprotein A1
   d. Apolipoprotein E
   e. Apolipoprotein C3
2. Which lipid sub-fraction shows the strongest association with risk of cardiovascular events?
   a. High-density lipoprotein cholesterol (HDL-C)
   b. Low-density lipoprotein cholesterol (LDL-C)
   c. Non-HDL cholesterol
   d. Remnant cholesterol
   e. Triglycerides
3. Which of the following conditions is NOT associated with a secondary elevation in cholesterol?
   a. Excess alcohol
   b. Newly diagnosed diabetes
   c. Hypothyroidism
   d. Leukaemia
   e. Extreme weight loss
4. Which of the following conditions commonly shows autosomal dominant inheritance?
   a. Familial chylomicronaemia syndrome
   b. Familial hypercholesterolaemia
   c. Lecithin-cholesterol acyltransferase deficiency
   d. Abetalipoproteinemia
   e. Familial combined hyperlipidaemia
5. Which of the following medications does not lower low-density lipoprotein cholesterol (LDL-C) levels?
   a. A statin (HMG-CoA reductase inhibitor)
   b. Ezetimibe
   c. Proprotein-convertase subtilisin kexin-9 (PCSK9) inhibitor
   d. Icosapent ethyl (omega-3 fatty acid)
   e. Bempedoic acid

Chapter 20: Respiratory medicine

1. A 18-year-old male with a history of childhood asthma repeatedly attends the emergency department (ED) with symptoms of wheeze, cough, and shortness of breath. He is well enough to be discharged from the ED after salbutamol nebulizer because his peak expiratory flow is 80% of predicted. What is the best plan on hospital discharge?
   a. Prescribe oral prednisolone. He should also use a salbutamol inhaler hourly, see his general practitioner in 1 month, and seek medical advice if he deteriorates in the interim.
   b. Prescribe oral prednisolone, ensure a preventer and reliever therapy prescription, and advise to see general practitioner within 1 week and seek medical advice if his condition fails to improve.
   c. Advise follow-up with his general practitioner as soon as possible and make an urgent referral to the respiratory team for an outpatient review.
   d. Advise to re-attend the ED if he has further symptoms.
   e. Arrange an ambulatory care review the next day and make a referral to the respiratory clinic.
2. A 74-year-old male presents with pneumonia. Which of the following statements is correct?
   a. The most common causative agent is *Streptococcus pneumoniae.*
   b. Severity of pneumonia may be assessed with CURB 65.
   c. A follow-up chest radiograph should be requested in 6 weeks.
   d. Cigarette smoking is an independent risk factor for community-acquired pneumonia.
   e. All of the above.
3. A 38-year-old female who has recently moved to the UK is seen in the emergency department with fever, cough, night sweats, haemoptysis, and weight loss over a 6-week period. The chest radiograph shows upper lobe consolidation. Which of the following is TRUE?
   a. The patient should have care in a side room.
   b. One sputum sample that is negative for acid-fast bacilli excludes tuberculosis.
   c. The likely diagnosis is non-tuberculous mycobacterial disease.
   d. Bronchoscopy is mandatory to make the diagnosis.
   e. A positive pregnancy test result means that treatment for tuberculosis should be delayed.
4. A 54-year-old male attends the general practitioner surgery. He has a 40-pack-year smoking history. He has had a cough for 4 weeks. He may have lost some weight. What are your next steps?
   a. Assess for postnasal drip and prescribe a nasal steroid spray.
   b. Trial a proton pump inhibitor and review again in a few weeks.
   c. Order a chest radiograph and blood tests and send a 2-week wait referral for suspected lung cancer.
   d. Prescribe a course of antibiotics.
   e. Reassure and ask to make another appointment if the symptoms do not resolve.
5. A 68-year-old male who is a 40 pack-year smoker has a new diagnosis of chronic obstructive pulmonary disease with no previous history of asthma or atopy. Lung function tests shows an FEV_1/FVC ratio of 55%. What feature of his clinical history helps guide initial inhaled treatment?
   a. Symptomatic improvement when treated with oral prednisolone
   b. A breathlessness score (e.g., MRC or CAT) and frequency of exacerbations
   c. Whether he has received influenza and SARS-CoV-2 vaccinations within the past 12 months
   d. Whether he has completed pulmonary rehabilitation
   e. How often he uses a salbutamol inhaler

Chapter 21: Ear, nose, and throat

1. A 54-year-old male develops an upper respiratory tract infection (URTI) with nasal obstruction and rhinorrhea. A few days later, he develops hearing loss and tinnitus in his right ear, although this is not painful. The general practitioner performs tuning fork tests. Weber's test lateralizes to the right ear. Rinne's test is positive on the left and negative on the right. What should you tell the patient?
   a. This is sudden sensorineural hearing loss on the left, and urgent steroids are required.
   b. This is sudden sensorineural hearing loss on the right, and urgent steroids are required.
   c. This is 'glue ear' on the right and should resolve in the next few weeks.
   d. This is 'glue ear' on the left and should resolve in the next few weeks.
   e. The best treatment is antibiotics and steroids for 1 week.

2. A 67-year-old female falls and hits her head. A few weeks later, she wakes up with dizzy spells lasting several seconds, which continue to occur multiple times a day, particularly when she reaches up for items in the kitchen. She denies vomiting and hearing loss. She is otherwise fit and well. Which is the correct diagnosis?
   a. Posterior fossa mass
   b. Ménière's disease
   c. Vestibular migraine
   d. Benign paroxysmal positional vertigo
   e. Multiple sclerosis
3. An 87-year-old male with a history of chronic hypertension develops sudden-onset disabling dizziness and vomiting, which lasts all day. His hearing is also reduced. He is unable to get up to walk to the toilet. His head thrust test shows no catch-up saccade, but he does have direction changing nystagmus. A test for skew is positive. What is the most likely diagnosis?
   a. Vestibular neuritis
   b. Ménière's disease
   c. Persistent postural perceptual dizziness
   d. Posterior circulation stroke
   e. Benign paroxysmal positional vertigo
4. A 45-year-old male presents with a 6-week history of persistent right unilateral throat pain and difficulty swallowing. He gets occasional pain radiating to the right ear. On examination, there are a unilaterally enlarged tonsil and palpable lymph nodes in the neck. Which test is most likely to be diagnostic?
   a. Throat swab for microscopy and culture
   b. Computed tomography imaging of the neck
   c. Fine-needle aspiration of the lymph node
   d. Serum sample for inflammatory markers
   e. Skin-prick testing
5. A 30-year-old builder has a 1-year history of left-sided hearing loss and tinnitus. An audiogram shows a 40-DB sensorineural hearing loss on the left. What would you do next?
   a. Repeat a hearing test in 3 months.
   b. Give the patient a course of nasal steroids and antibiotics.
   c. Give the patient oral steroids for 1 week.
   d. Request an MRI scan of the internal auditory meatus.
   e. Refer for hearing aids.

Chapter 22: Gastroenterology

1. Which of the following pathogens is not a common cause of acute diarrhoea?
   a. Norovirus
   b. *Escherichia coli*
   c. *Giardia* spp.
   d. *Streptococcus* spp.
   e. *Clostridium difficile*
2. Which of the following is not an 'alarm' sign or symptom that should prompt urgent upper gastrointestinal endoscopy?
   a. Weight loss
   b. Belching
   c. Persistent vomiting
   d. Iron-deficiency anaemia
   e. Epigastric mass
3. Which of the following genes is most commonly associated with coeliac disease?
   a. *HLA-DQ2*
   b. *HLA-DQ8*
   c. *HLA-DR2*
   d. *HLA-DR3*
   e. *HLA-DP3*
4. Infliximab and adalimumab target which pro-inflammatory molecule?
   a. IL-6
   b. CD20
   c. TNF-α
   d. IL-2
   e. IL-23
5. Which of the following best describes the action of the gut regulatory peptide ghrelin?
   a. Stimulates gastric acid secretion
   b. Inhibits glucagon secretion and gastric emptying and stimulates insulin synthesis
   c. Inhibits secretion and action of most gut hormones
   d. Stimulates appetite and increases gastric emptying
   e. Causes gallbladder contraction and sphincter of Oddi relaxation

Chapter 23: Hepatology

1. If a patient is hepatitis C antibody positive and hepatitis C RNA positive, which of the following is true?
   a. The patient has been exposed to hepatitis C and requires monitoring.
   b. The patient is at risk of cirrhosis within the next few months.
   c. The patient's sexual contacts should be immunized against hepatitis C.
   d. The patient should be informed that curative tablet-based therapy is available.
   e. A liver biopsy must be performed.
2. In a patient with chronic hepatitis B (HBsAg positive, HBeAg negative, HBV DNA = 800 000 IU/mL), which of the following is true?
   a. The patient's family should be offered prophylactic medication.
   b. A liver biopsy is mandatory.
   c. The risk of live cancer is minimal, and no screening is indicated.
   d. An HIV test should be performed before initiation of therapy.
   e. Treatment with interferon should be introduced.
3. In a patient with abnormal liver function test results, which of the following is NOT immediately required?
   a. A liver ultrasound scan
   b. An assessment of body mass index and diabetes
   c. A careful drug history
   d. An endoscopic retrograde cholangiopancreatography examination
   e. A family history, including details of any neurological disorders
4. The ALT:AST ratio is a useful clinical indicator, but which of the following is NOT true?
   a. In patients with viral hepatitis, ALT is greater than AST unless cirrhosis is present, in which case AST is greater than ALT.
   b. In patients with alcoholic liver disease and steatohepatitis, the ALT is often greater than the AST.
   c. In patients with viral hepatitis, an AST:ALT ratio of more than 1 indicates cirrhosis.
   d. In patients with liver disease without cirrhosis in whom AST is greater than ALT, alcohol or obesity is the most likely aetiological agent.

5. Which of the following is NOT true?
   a. Ascites is the most common complication of cirrhosis, and this clinical presentation represents a watershed moment with patients' median survival falling from greater than 12 years for compensated cirrhosis to approximately 2 years.
   b. The Model for End-stage Liver Disease (MELD) focuses on serum bilirubin, creatinine, and the International Normalized Ratio and is widely used as a predictor of death in patients awaiting liver transplantation.
   c. After liver transplantation for autoimmune hepatitis, the disease never recurs.
   d. Elective liver transplantation in low-risk patients has a 90% 1-year survival rate and a 70% to 85% 5-year survival rate.
   e. Approximately 90% of patients with cirrhosis develop gastro-oesophageal varices over 10 years, but only one-third of these will bleed.

Chapter 24: Biliary tract and pancreatic disease

1. A 42-year-old female presents with colicky right hypochondrium pain, associated with nausea and vomiting. The pain was precipitated when she was eating her dinner and is controlled with analgesics. She denies any fever or rigors, and Murphy's sign is negative. What is the diagnosis?
   a. Typical peptic ulcer disease
   b. Renal colic
   c. Cholecystitis
   d. Ascending cholangitis
   e. Biliary colic
2. A 40-year-old patient with a history of alcohol dependence presents with upper abdominal pain. The pain started in the epigastric region and is associated with nausea and vomiting with radiation through to the back. The pain began after an alcoholic binge. What is the diagnosis?
   a. Peptic ulcer
   b. Acute pancreatitis
   c. Biliary colic
   d. Acute cholecystitis
   e. Mirizzi syndrome
3. A 40-year-old female presents with 3-month history of dull abdominal pain that radiates to the back and is associated with weight loss and progressive anorexia. She describes how she obtains partial relief by sitting forward. What is her diagnosis?
   a. Chronic pancreatitis
   b. Acute pancreatitis
   c. Carcinoma of the head of the pancreas
   d. Carcinoma of the ampulla of Vater
   e. Carcinoma localized to the body or tail of the pancreas
4. A 40-year-old male presents with 3-month history of worsening abdominal pain radiating to the back and associated weight loss and progressive anorexia. He describes how his wife has told him that he is looking more yellow. What is the diagnosis?
   a. Chronic pancreatitis
   b. Acute pancreatitis
   c. Carcinoma of the head of the pancreas
   d. Carcinoma localized to the body or tail of the pancreas
   e. Gastric tumour
5. The pancreas can be associated with cysts and some are predisposed to becoming malignant. Which one of the following is NOT known to become malignant?
   a. Pseudocyst
   b. Serous cyst
   c. Mucinous cyst
   d. Intraductal papillary mucinous cyst
   e. None of the above

Chapter 25: Nutrition

1. Which of the following statements regarding re-feeding syndrome is correct?
   a. Re-introduction of carbohydrate after a period of starvation leads to suppression of insulin release.
   b. A high body mass index is a risk factor for re-feeding syndrome.
   c. Intravenous vitamin K should be administered to patients at high risk of re-feeding syndrome.
   d. Re-feeding results in hypokalaemia, hypomagnesemia, hypophosphatemia, and thiamine deficiency.
   e. Patients at risk of re-feeding syndrome should be fed with excessive calories to account for their period of starvation.
2. The clinical syndrome of pellagra is associated with deficiency of which vitamin?
   a. B_1
   b. B_3
   c. Vitamin C
   d. Folate
   e. B_{12}
3. Which of the following is NOT a cause of intestinal failure?
   a. Short bowel syndrome
   b. Extensive small bowel Crohn's disease
   c. Acute severe ulcerative colitis
   d. Postoperative ileus
   e. Radiation enteropathy
4. What is the support a patient with 75 cm of short bowel to a jejunostomy will likely require?
   a. Nasogastric feeding
   b. Intravenous fluids
   c. Parenteral nutrition with or without intravenous fluids
   d. Parenteral nutrition
   e. Oral glucose solution
5. Which of the following is NOT considered an indication for intestinal transplant?
   a. Intestinal failure–associated liver disease
   b. Metastatic colorectal cancer
   c. Ultra-short bowel syndrome
   d. Frequent catheter-related bloodstream infections
   e. Central venous thrombosis of two or more central veins

Chapter 26: Kidney and urinary tract disease

1. A 29-year-old male presents with joint pain and a rash over his legs, having had a sore throat in the preceding week. He has a blood pressure of 155/93 mmHg and urine dipstick testing shows haematuria and proteinuria. What is the most likely diagnosis?
   a. Urinary tract infection
   b. IgA nephropathy
   c. Hypertensive crisis
   d. Meningococcal septicaemia
   e. Systemic Lupus Erythematosus

2. A 21-year-old female presents with fever, rigors, and left-sided flank pain. On examination, she is febrile (38.5°C) and has both suprapubic and left renal angle tenderness. What is the most likely diagnosis?
   a. Renal cell carcinoma
   b. Pyelonephritis
   c. Pancreatitis
   d. Acute renal colic
   e. Pelvic inflammatory disease
3. A 60-year-old male presents with visible, painless haematuria. Blood tests show a haemoglobin of 191 g/L. What is the most likely diagnosis?
   a. Renal calculi
   b. Renal cell carcinoma
   c. Prostate carcinoma
   d. Bladder carcinoma
   e. Haemolysis
4. A 71-year-old female presents with a severe headache. She is known to have chronic hypertension, diabetes, and coronary artery disease. Her blood pressure is 170/95 mmHg. Test results show:

Haemoglobin	99 g/L
Creatinine	212 μmol/L
Urine protein:creatinine ratio	410 mg/mmol
Ultrasound	normal-size kidneys

What is the most likely diagnosis?
   a. Hypertensive kidney disease
   b. Renal artery stenosis
   c. Reflux nephropathy
   d. Diabetic nephropathy
   e. Polycystic kidney disease
5. A previously fit 83-year-old male with stage 5 chronic kidney disease on peritoneal dialysis presents with confusion and constipation. On examination, his abdomen is tender, and there is a yellow discharge at his peritoneal catheter exit site. What is the most likely diagnosis?
   a. Ischaemic bowel disease
   b. Lower urinary tract infection
   c. Peritoneal dialysis-associated peritonitis
   d. Bowel obstruction
   e. Abdominal wall cellulitis

Chapter 27: Water balance, fluids, and electrolytes

1. Which of the following diuretics blocks the epithelial sodium channels in the distal collecting duct?
   a. Furosemide
   b. Acetazolamide
   c. Metolazone
   d. Amiloride
   e. Spironolactone
2. Arginine vasopressin (AVP) leads to urinary concentration through what mechanism?
   a. Inhibition of the sodium glucose co-transporter in the proximal tubule delivering sodium to the distal nephron
   b. Increasing the water permeability of the cortical and medullary collecting ducts
   c. Activation of the renin–angiotensin–aldosterone system
   d. Inhibition of the sodium–potassium-2-chloride channel in the thick ascending limb of Henle
   e. Inhibition of the sodium potassium ATP pump to allow more filtrate to flow to the collecting duct
3. Which is not a biological effect of SGLT2 inhibitors?
   a. Weight loss
   b. Hyperglycaemia
   c. Caloric deficit
   d. Hypotension
   e. Natriuresis
4. Hypophosphataemia is not associated with which of these conditions?
   a. Post-parathyroidectomy
   b. Dent's disease
   c. Re-feeding syndrome
   d. Vitamin D deficiency
   e. Respiratory acidosis

Chapter 28: Neurology

1. A 20-year-old male is brought into hospital by his friends. His friends state that he has not got out of bed for 3 days after celebrating his birthday. He admits to drinking alcohol and taking drugs at the time. He is now complaining of headaches and myalgia and states that sometimes he can hear voices. On examination, he becomes aggressive and makes a detailed neurological examination impossible, but he is noticed to be febrile. There is no rash on examination. What is the likely diagnosis?
   a. Cocaine-induced cerebrovascular event
   b. Drug-induced psychosis
   c. Herpes simplex encephalitis
   d. Subdural haemorrhage
   e. Variant Creutzfeldt–Jakob disease
2. A 34-year-old male presents to the emergency department with a severe left-sided headache that woke him from his sleep. He is now complaining of severe pain on the back of his left eye. He states that he has had this before, and it typically lasts for some minutes. He has had these attacks for the past few nights. On this occasion, he denies vomiting but states that he has vomited with previous episodes. He is also complaining that his left nostril feels congested. On examination, you find nothing abnormal. What is the likely diagnosis?
   a. Cluster headache
   b. Migraine
   c. Tension headache
   d. Subarachnoid haemorrhage
   e. Sinusitis
3. A 55-year-old female goes to see her primary care physician complaining of right-sided headaches and a transient episode of right-sided blindness 1 day ago, which has completely resolved. She notices that the pain in her head is exacerbated when she is combing her hair. On neurological examination, you find nothing abnormal. What is the likely diagnosis?
   a. Horner's syndrome
   b. Giant cell arteritis
   c. Malignant hypertension
   d. Amaurosis fugax
   e. Migraine
4. A 35-year-old female is referred to the neurologist with a 3-year history of progressive, perceptive deafness, with worsening symptoms of rotation dizziness and new onset of left facial weakness. Further examination reveals left facial numbness and a wide gait on walking. What is the likely diagnosis?
   a. Left parietal lobe glioma
   b. Cerebrovascular event
   c. Schwannoma (acoustic neuroma)
   d. Left cerebellar haemangioblastoma
   e. Glomus tumour of the jugular bulb

5. A 30-year-old obese female presents to her doctor with symptoms of headaches and complaining of some visual blurring. On fundoscopy, you notice papilloedema. Computed tomography is arranged for her, and the results do not show any mass or any increase in ventricular size. What condition is most likely?
   a. Idiopathic intracranial hypertension
   b. Sagittal sinus thrombosis
   c. Glioma
   d. Meningioma
   e. Subdural haematoma
6. A 25-year-old female presents to her primary care physician with pain in both her upper limbs, which is worse when coughing. You notice that she has burn marks in different areas of her upper arm, which she hasn't really noticed. On examination, her light touch is still intact, but she has loss of sensation to temperature. She has reduced upper limb reflexes bilaterally, and you notice some muscle wasting in the thenar muscle bilaterally. What is the likely diagnosis?
   a. Glioma of the spinal cord
   b. Vitamin B_{12} deficiency
   c. Syringomyelia
   d. Syringobulbia
   e. Spinal infarct
7. A 46-year-old male with a history of excess alcohol consumption presents to the emergency department with confusion and loss of sensation in his lower limbs and having an ataxic gait and nystagmus. He also complains of weakness in his lower limbs. What is the likely diagnosis?
   a. Lathyrism
   b. Konzo
   c. Vitamin B_{12} deficiency
   d. Vitamin B_1 deficiency
   e. Diabetic neuropathy
8. A 34-year-old female has been referred to the neurologist with symptoms of nasal regurgitation of fluids and choking. She is otherwise fit and well and has no other past medical history. On examination, you find she has a fibrillating tongue and a weak palate. What condition is most likely?
   a. Acute cerebrovascular accident
   b. Myasthenia gravis
   c. Progressive bulbar palsy
   d. Progressive muscular atrophy
   e. Cerebellar ataxia
9. A 20-year-old male presents to the doctor because he has noticed some lumps on his skin. He states that he has always had a few, but they have increased in number the older he gets. He is otherwise fit and well. He states that this seems to run in the family, and his father and grandfather both had a similar condition. On examination, there is nothing abnormal except that you notice some patches of skin that seem darker than the rest of his skin. What is the likely diagnosis?
   a. Neurofibromatosis type II
   b. Neurofibromatosis type I
   c. Tuberous sclerosis
   d. Sturge–Weber syndrome
   e. Vitiligo

Chapter 29: Stroke medicine

1. A 75-year-old male presents with a basal ganglia haemorrhage. His past medical history includes hypertension, and diabetes. He has a family history of aneurysm. Which of the following is the most likely cause of the intracranial haemorrhage?
   a. Cerebral amyloid angiopathy
   b. Arteriovenous malformation
   c. Cerebral aneurysm rupture
   d. Hypertension
   e. Venous sinus thrombosis
2. A 70-year-old smoker presents with mild right-sided facial weakness and mild word-finding difficulties (NIH Stroke Scale score, 2). His history includes diabetes and high cholesterol. Magnetic resonance imaging of the brain confirms a small infarct in the left frontal lobe without associated haemorrhage. MR angiography shows no evidence of intracranial stenosis but identifies a 20% to 49% left carotid stenosis. Twenty-four hours of cardiac rhythm monitoring shows normal sinus rhythm. Renal function is normal. What is the optimal antithrombotic strategy for secondary stroke prevention in this case?
   a. Prescribe warfarin with an international normalized ratio (INR) target of 2 to 3.
   b. Start apixaban immediately.
   c. Start aspirin 300 mg/day continued for 2 weeks then low-dose aspirin life-long
   d. Start aspirin and clopidogrel mg for 3 weeks and then single-antiplatelet treatment.
   e. Start aspirin and clopidogrel for 3 months and then single-antiplatelet treatment.
3. A 65-year-old female smoker presents with acute onset of an unsteady gait and vertigo. Blood pressure is normal. Electrocardiogram shows normal sinus rhythm. Computed tomography (CT) of the brain shows hypoattenuation in the left cerebellum without haemorrhage. CT angiography on admission confirmed left carotid stenosis (50%–69%) and a small-calibre left vertebral artery. What is the optimal intervention strategy?
   a. Left vertebral artery stenting
   b. Left carotid artery endarterectomy within 2 weeks
   c. Left carotid artery stenting within 2 weeks
   d. Left carotid endarterectomy within 24 hours
   e. Aggressive vascular risk factor control
4. A 24-year-old pregnant female (24 weeks' gestation) presents with right-hand weakness, headache, and drowsiness with a subsequent episode of seizure. Computed tomography of the brain identifies a left juxtacortical 1-cm haemorrhage. Taking into consideration the most likely diagnosis, what is the most appropriate treatment plan?
   a. Basilar thrombectomy
   b. Immediate neurosurgical decompression
   c. Commence immediate heparin
   d. Commence immediate aspirin
   e. Treat seizures and observe

Chapter 30: General hospital psychiatry

1. You are called to see a confused elderly male on the acute admissions medical ward. He had a high temperature (38.1°C) recorded last night on presentation to the emergency department. He refused physical examination then and now (4 a.m.). You are under pressure to examine his vital signs and his chest, as well as ruling out any injury. How will you proceed?
   a. Let him be and leave it until the morning shift comes in, just 4 hours from now.
   b. Ask psychiatry to place him under Mental Health Act legislation so you can examine him and give intravenous antibiotics.
   c. Phone the hospital legal department and ask for a case conference.
   d. Establish if he has the mental capacity to refuse vital signs and examination; document that he does not have capacity to refuse a basic examination and make a *best interests decision* to examine him.
   e. Call security to restrain him; then examine him.

2. The same patient (71-year-old male, known diabetes with angioplasty last year) is confused on presentation. Other than a fast heart rate, his temperature and vital signs are normal, and your physical examination did not find a source of infection. What features establish a delirium diagnosis?
   a. High blood glucose
   b. Auditory hallucinations recorded in his pre-angioplasty assessment last year, and he looks like he might be hearing voices right now
   c. Fluctuating consciousness, disorientated in time, and fleeting visual hallucinations
   d. Taking six different medications (polypharmacy), and his prescription differs from hospital records
   e. He is angry that you woke him up and is getting aggressive with nursing staff and you
3. In relation to prescribing a new psychiatric medication for your patient, which of the following is a source of potential harm to him or her?
   a. Asking for allergies, drug reactions, and intolerances to previous medications
   b. Changing to haloperidol from newly prescribed olanzapine because the hospital has run out
   c. Checking hospital and community records, including pharmacy if available, for current and past medications
   d. Reducing their selective serotonin reuptake inhibitor slowly over weeks before introducing a second-line antidepressant
   e. With new lithium prescription, requesting weekly blood levels with a renal profile
4. Identify the option that features only the cognitive symptoms of anxiety.
   a. Chest tightness, palpitations, and 'latchkey' urinary urgency
   b. Difficulties with concentration, episodes of the mind going blank, and indecision
   c. Fear of dying, fearing the worst, and a conviction that something bad is going to happen
   d. Hypervigilance, performance anxiety, and avoiding situations that provoke(d) anxiety
   e. Low mood, anger or irritability, and flashbacks

Chapter 31: Ophthalmology

1. Which of the following will not give a relative afferent pupillary defect?
   a. Retinal detachment
   b. Diabetic macular oedema
   c. Optic neuritis
   d. Non-arteritic anterior ischaemic optic neuropathy
   e. Ischaemic central retinal vein occlusion
2. Binocular diplopia is not caused by
   a. third nerve palsy.
   b. orbital floor fracture.
   c. sixth nerve palsy.
   d. thyroid eye disease.
   e. cataract.
3. Misuse of contact lens wear may lead to
   a. endophthalmitis.
   b. orbital cellulitis.
   c. acute glaucoma.
   d. keratitis.
   e. uveitis.
4. Sudden painless central loss of vision is likely caused by
   a. central retinal artery occlusion.
   b. retinal detachment.
   c. central retinal vein occlusion.
   d. wet macular degeneration.
   e. papilloedema.
5. Which of the following will not lead to optic disc swelling?
   a. Anterior ischaemic optic neuropathy
   b. Diabetic papillopathy
   c. Papilloedema
   d. Central retinal artery occlusion
   e. Central retinal vein occlusion

Chapter 32: Rheumatology

1. A 49-year-old female presents to her general practitioner with pain in her hands. The pain began gradually 5 weeks ago, is worse in the morning, and is associated with 1 hour of stiffness when she wakes from sleeping. The pain improves after activity. What is the most likely diagnosis in this patient?
   a. Psoriatic arthritis
   b. Axial spondylarthritis
   c. Osteoarthritis
   d. Gout
   e. Rheumatoid arthritis
2. A 73-year-old female presents to her general practitioner with 12 months of increasing pain in her left knee. There are no other joint symptoms. The knee pain began gradually and is worse after activity. It has begun to restrict her ability to climb stairs. What is the most likely finding on knee X-ray in this patient?
   a. Marginal erosions
   b. Marginal osteophytes
   c. Syndesmophytes
   d. Periarticular bone resorption
   e. Preserved joint space
3. A 23-year-old male under rheumatology care with axial spondylarthritis has ongoing severe inflammatory low back pain, fatigue, and prolonged morning stiffness. He has tried two different non-selective non-steroidal anti-inflammatory drugs at full doses for 3 months each with no improvement in symptoms. He does not have any joint pain in his arms or legs. What is the most appropriate treatment to consider next?
   a. Adalimumab
   b. Etoricoxib
   c. Hydrotherapy
   d. Methotrexate
   e. Prednisolone
4. A 35-year-old female presents to her general practitioner with pain all over her body for the past 2 years. Her sleep is disturbed, and she feels unrefreshed on waking. There are multiple tender spots on her body but no swollen joints. All blood tests and X-ray findings are normal. What is the most likely diagnosis in this patient?
   a. Polymyalgia rheumatica
   b. Fibromyalgia
   c. Osteoarthritis
   d. Polymyositis
   e. Rheumatoid arthritis
5. A 25-year-old female presents to an emergency department with bilateral ankle oedema, raised blood pressure, and a rash over her face. She has a previous history of pleural effusion

and recurrent mouth ulcers. Investigations show raised anti–double-stranded DNA levels, low albumin, and a high urine protein-creatinine ratio. What is the most likely diagnosis in this patient?

a. Rheumatoid arthritis
b. Sjögren's syndrome
c. Systemic sclerosis
d. Granulomatosis with polyangiitis
e. Systemic lupus erythematosus

Chapter 33: Bone disease

1. Which of the following is a risk factor for osteoporosis?
a. Age
b. Early menopause
c. Coeliac disease
d. Glucocorticoids
e. All of the above

2. Symptoms of osteoporosis include
a. bone pain.
b. fracture.
c. fatigue.
d. weakness.
e. all of the above.

3. Diagnosis of osteoporosis can be made by
a. DEXA scan.
b. fracture history.
c. bone biopsy.
d. a or c.
e. a, b, or c.

4. Which of the following is not a recommended initial treatment for an individual with T-score of –2.6 and no fractures?
a. Alendronate
b. Denosumab
c. Teriparatide
d. Zoledronate
e. c and d

5. Which of the following is the correct course of action after 5 years of denosumab treatment?
a. Drug holiday; monitor dual-energy x-ray absorptiometry (DXA) scan
b. Change to teriparatide
c. Change to alendronate
d. Continue denosumab
e. Change to zoledronate

Chapter 34: Endocrinology

1. A young female is referred by the primary care physician with oligomenorrhoea and galactorrhoea. On examination, she has bitemporal hemianopia.

Which one of the following is the best radiological investigation for pituitary tumours?
a. Computed tomography (CT) of the head
b. Lateral skull X-ray
c. Positron emission tomography (PET)
d. Magnetic resonance imaging (MRI) of the pituitary
e. Nuclear uptake scan

2. A 23-year-old Filipino female presents with irritability, weight loss of 6.5 kg in the past month, dry eyes, and increased sweating and thirst. On examination, you note that she is tremulous and warm. She has a heart rate of 100 beats/min, respiratory rate of 16 breaths/min, and blood pressure of 140/80 mmHg. There is a palpable diffuse lump in her neck with mild exophthalmos.

Results of thyroid function tests for this female are as follows:

Investigations:

Thyroid-stimulating hormone	<0.05 mU/L (0.3–3.5 mU/L)
Free T_4	49 pmol/L (10–25 pmol/L)
T_3	6.2 nmol/L (1.2–3.1 nmol/L)

What do these results imply?
a. Primary hypothyroidism
b. Thyroid-stimulating hormone deficiency
c. TSH-oma
d. Thyrotoxicosis
e. Compensated euthyroidism

3. Which one of the following statements about primary hyperparathyroidism is true?
a. Parathyroid hormone (PTH) is usually low
b. A high phosphate level is common
c. Conservative management is never appropriate
d. Histology usually reveals parathyroid hyperplasia
e. It is commonly seen in patients with multiple endocrine neoplasia type 1 (MEN 1)

4. A 20-year-old male presents to the emergency department with severe palpitations, headache, sweatiness, nausea, and panic attacks. On examination, his blood pressure is 200/120 mmHg with a heart rate of 125 beats/min. Computed tomography shows a calcified 6-cm adrenal mass. Which one of the following is the most likely diagnosis?
a. Primary hyperaldosteronism
b. Essential hypertension
c. Congenital adrenal hyperplasia
d. Renal artery stenosis
e. Phaeochromocytoma

5. A 40-year-old female complains of depression and insomnia, easy bruising, and weakness at the tops of her muscles. On examination, her skin is thin, and there are multiple abdominal striae. What is the most likely diagnosis?
a. Conn's syndrome
b. Cushing's syndrome
c. Addison's disease
d. Hypopituitarism
e. Prolactinoma

Chapter 35: Diabetes mellitus

1. A 21-year-old student presents with a 4-week history of thirst, polyuria, polydipsia, and weight loss. The laboratory glucose level is 24.3 mmol/L. The clinical diagnosis is type 1 diabetes. What is the investigation of choice to confirm he has type 1 diabetes?
a. HbA_{1c}
b. Blood ketones
c. Islet autoantibodies
d. C-peptide
e. Genetic testing

2. A 56-year-old female has a 10-year history of type 2 diabetes and has chronic kidney disease stage 3. Which class of oral antidiabetes agent would be the treatment of choice?
a. Metformin
b. Thiazolidinedione
c. Sulfonylurea
d. SGLT-2 inhibitor
e. GLP-1 receptor agonist

3. Cancer has recently overtaken cardiovascular disease as the leading cause of death in people with diabetes in several high-income countries. Which cancer occurs more commonly in people with diabetes?
   a. Lung
   b. Pancreatic
   c. Prostate
   d. Melanoma
   e. Colon
4. Insulin is the primary treatment of type 1 diabetes. The most commonly used regimen is a basal-bolus regimen, which can be administered by multiple daily injections or an insulin pump. Which of the following is true about pump therapy?
   a. Pump therapy reduces HbA_{1c} in those with elevated HbA_{1c}.
   b. Pump therapy is cheaper than multiple daily injections.
   c. Pump therapy is associated with a higher rate of hypoglycaemia.
   d. Hybrid closed loop systems are no more effective than standard pump therapy.
   e. Pump therapy can only be used for people with type 1 diabetes.
5. The pregnancy outcomes in females with pre-existing diabetes are worse than females without diabetes.

What interventions have been shown to improve pregnancy outcomes in females with pre-existing diabetes?
   a. Interventions to improve glucose levels starting before conception
   b. Prepregnancy folic acid 5 mg/day
   c. Use of continuous glucose monitoring
   d. Regular contact with a specialist multidisciplinary team jointly led by a physician and obstetrician
   e. All of the above

Chapter 36: Metabolic disease

1. Which of the following statements is not true?
   a. Pregnancy in a person with phenylketonuria should be treated as a medical emergency.
   b. Ammonia should be measured as a first-line test in all patients presenting with encephalopathy.
   c. Acute metabolic decompensation in several inherited metabolic diseases can be triggered by fasting or decreased calorie intake.
   d. A diagnosis of inherited metabolic disease is very unlikely in patients presenting with symptoms for the first time in adulthood.
   e. Acute decompensation in several inherited metabolic diseases is treated by giving adequate calorie intake in the form of oral carbohydrates and intravenous 10% dextrose.
2. Immediate management of a patient presenting with high ammonia in suspected or confirmed inherited metabolic disease includes
   a. stopping all oral protein intake.
   b. ensuring adequate calorie intake through oral or intravenous glucose.
   c. ammonia scavenger drugs.
   d. hemofiltration.
   e. all of the above.
3. Which of the following does not present with symptoms of acute porphyria?
   a. Acute intermittent porphyria
   b. Variegate porphyria
   c. Hereditary coproporphyria
   d. ALA dehydratase deficiency porphyria
   e. Porphyria cutanea tarda
4. Which of the following statements about Fabry disease is FALSE?
   a. It is caused by a deficiency of alpha-galactosidase A enzyme.
   b. It has an X-linked inheritance.
   c. Females are not clinically affected.
   d. Genetic confirmation in females is needed because biochemistry is unreliable.
   e. Treatment options include enzyme replacement therapy and symptomatic management.
5. Which of the following is not a feature of carbohydrate metabolism disorders?
   a. Hypoglycaemia
   b. Cardiomyopathy
   c. Myalgia
   d. Liver disease
   e. Hyperammonemia

Chapter 37: Obesity

1. A 34-year-old female presents to the general practitioner with weight gain, tiredness, and occasional constipation for the past few weeks. She does not get her periods because she has a progesterone implant. Her blood pressure is 150/90 mmHg, pulse is 65 beats/min, temperature is 36.5°C, and oxygen saturation is 96%. She eats a healthy proportion of food and is physically active. Which of the following conditions will you consider excluding first in this female?
   a. Hypothyroidism
   b. Cushing's syndrome
   c. Type 2 diabetes mellitus
   d. Dyslipidemia
   e. Vitamin D deficiency
2. Which of the following presents with early-onset obesity and hyperphagia (insatiable hunger)?
   a. Monogenic obesity
   b. Polygenic obesity
   c. Hormonal obesity
   d. Drug-associated obesity
   e. Endocrine-disrupting chemical–associated obesity
3. Which of the following complications arise from obesity?
   a. Hypertension
   b. Obstructive sleep apnoea
   c. Osteoarthritis
   d. Type 2 diabetes
   e. All of the above
4. A 35-year-old female with anxiety, type 1 diabetes, asthma, and schizophrenia presents to the obesity clinic with an inability to lose weight. Her blood pressure is 130/75 mmHg, pulse is 100 beats/minute, temperature is 37°C, and oxygen saturation is 97%. Her body mass index is 55 kg/m^2. She takes insulin, olanzapine, a regular steroid inhaler, and fluoxetine. Her full blood count, serum biochemistry, and HbA_{1c} are normal. Endocrine tests are completed, and her thyroid-stimulating hormone, and 9 a.m. serum cortisol are normal; however, her prolactin is 700 miU/L. What do you think could possibly be causing her weight gain?
   a. Prolactin
   b. Olanzapine
   c. Insulin
   d. Fluoxetine
   e. Options b, c, and d

5. A 50-year-old male presents to the general practitioner with tiredness and generalized pain. He is a vegetarian, is physically active, doesn't smoke, drinks two to four glasses of alcohol daily, and works full time. His body mass index is 45 kg/m^2, blood pressure is 155/100 mmHg, pulse is 69 beats/min, temperature is 36.5°C, and oxygen saturation is 97%. On examination, he has a plethoric, round face; a hump at the back of his neck; and increased abdominal girth with purple stretch marks. His investigations reveal a thyroid-stimulating hormone level of 2.05 mU/L, 9 a.m. serum cortisol of 700 nmol/L, low vitamin D level, and a low testosterone. After a dexamethasone suppression test, the cortisol is reduced to 35 nmol/L. What is the likely diagnosis?
   a. Cushing's syndrome
   b. Hypothyroidism
   c. Obesity
   d. Hypogonadism
   e. Pseudo-Cushing's syndrome

Chapter 38: Haematology

1. In trying to determine the cause of a patient's anaemia:
   a. A low MCV (mean cell volume) is diagnostic of iron deficiency anaemia.
   b. A raised reticulocyte count indicates that the bone marrow is working.
   c. A bone marrow biopsy is a last-resort test.
   d. A blood film is only helpful if the white blood cell or platelet counts are also abnormal.
   e. A low serum iron concentration is sufficient to indicate the presence of iron deficiency.
2. In haemolytic anaemias:
   a. G6PD deficiency cannot manifest in female patients as the gene is on the X-chromosome.
   b. Pyruvate kinase deficiency requires splenectomy to avoid blood transfusions.
   c. Drug-induced haemolytic anaemias are usually diagnosed by detecting the presence of drug-induced red cell antibodies.
   d. A low haptoglobin concentration rules out extravascular haemolysis.
   e. High LDH concentrations suggest increased red cell destruction.
3. For patients with haemoglobinopathies:
   a. The most severe form of beta thalassaemia major presents *in utero*.
   b. Patients with sickle cell disease and thalassaemia are at increased risk of strokes in childhood.
   c. Iron overload in transfusion-dependent thalassaemia can only be avoided by bone marrow transplantation at an early age.
   d. In sickle cell disease, patients are at risk of repeated parvovirus B19-induced aplastic crises.
   e. HbF concentrations can be naturally elevated in patients with sickle cell disease and thalassaemia.
4. Concerning red cell antibodies:
   a. Destruction of red cells by IgG autoantibodies requires complement fixation.
   b. A positive DAT is diagnostic of autoimmune haemolytic anaemia.
   c. A blood film showing spherocytes in conjunction with a negative DAT is sufficient to diagnose hereditary spherocytosis.
   d. A delayed transfusion reaction is unlikely to happen after a patient's first ever transfusion.
   e. Haemolytic disease of the newborn is best diagnosed with a cord sample at birth.
5. Concerning clotting disorders:
   a. Some types of malabsorption can mimic warfarin treatment on clotting tests.
   b. An elevated fibrinogen concentration is a sign of disseminated intravascular coagulation.
   c. Platelet transfusions are indicated to prevent bleeding in thrombotic thrombocytopenic purpura if the platelet count falls to less than 10 × 10^9/L.
   d. Lupus anticoagulants, which prolong the APTT, are associated with intra-articular bleeds.
   e. Patients with Type 2 Von Willebrand disease are at increased risk of postoperative thrombosis and require high doses of low molecular weight heparin to prevent venous thromboembolism.

Chapter 39: Haematological oncology

1. A 48-year-old male presents to the emergency department with a wrist fracture after being hit by a car while cycling home after work on a Friday night. A routine blood test shows:
   - Haemoglobin: 130 g/L
   - White blood cell count: 96 × 10^9/L
   - Platelet count: 250 × 10^9/L

   Results of biochemistry and clotting tests are normal. Apart from his wrist fracture, he feels well. He has never had a blood test before. Which of the following tests will give the most helpful information?
   a. Bone marrow investigation
   b. Blood film
   c. PET/CT scan
   d. Immunophenotyping on peripheral blood sample
   e. Blood tests for infection: CRP, ESR, and blood culture
2. A 23-year-old female goes to her general practitioner for the third time with persistent fevers and feeling run-down. She thinks she has lost weight. She has completed two courses of antibiotics for tonsillitis. She has noted a lump in her neck, which is not painful. She had COVID-19 5 months ago. She is not worried and thinks that being a single parent while trying to finish her degree are contributing to her symptoms. On examination, she has prominent, non-tender lymph nodes in her neck, the largest of which is 1.5 cm. What is the best course of action for the general practitioner?
   a. Reassure her and give her antibiotic treatment that covers atypical bacteria.
   b. Reassure her that her symptoms are likely caused by her previous COVID-19 infection and that no treatment is required.
   c. Request imaging of her neck and arrange to see her when the results are available.
   d. Refer her to the ear, nose, and throat team for their assessment and opinion.
   e. Request imaging of her neck and refer via the suspected cancer pathway.
3. A 78-year-old female requests an emergency general practitioner appointment on a Monday morning after developing increasingly severe back pain over the weekend. She had been advised to do this by the out-of-hours service when she contacted them for advice on Sunday. She feels the pain might be improving, but she is following up on the advice she was given. She last attended the general practitioner for a routine health check 1 year ago, but she has had two telephone appointments about back pain in the past 4 months. Blood tests were requested, but she had not been able to attend to have them done because she is the sole career for her husband, who has recently been unwell. She feels that the pain is higher up than the pain she experienced in her

hip and lower back before a successful hip replacement 5 years ago. She is reluctant to take analgesia because she has been constipated for the past few weeks. She thinks that her mind has not been as sharp as usual for the past few weeks. What is the best course of action?

a. Organize blood tests and arrange to ring her with the results in a couple of weeks.
b. Refer her for assessment by the community physiotherapy team and advise her that if her pain becomes more severe, she should go to the local emergency department.
c. Reassure her and organize assessment by the district nursing team to see if she could benefit from additional support at home. You will ask the district nurses to do blood tests at the same time.
d. You are worried about her back pain and request an urgent magnetic resonance imaging scan to take place within the next few weeks. You will then see her face to face with the results.
e. You give her an urgent appointment for face-to-face review and blood tests to take place the next day. You ask her to come in the morning so that her results are available the same day.

4. At the end of your busy emergency department (ED) shift, one of the nurses gives you a print-out of test results on a patient who came in overnight but left before being seen. The patient had presented with leg pain after flying back from holiday. Tests showed:

- Haemoglobin: 130 g/L
- White blood cell count: 9.8 × 10^9/L
- Platelet count: 1173 × 10^9/L
- Biochemistry: not processed because the sample was haemolyzed

What is the best course of action?

a. Nothing. She self-discharged, so her symptoms can't have been that bad. She will likely present to the ED again if her symptoms get worse.
b. You ask your colleague to contact the patient by telephone to inform her that she needs to make an appointment with her general practitioner. You document the test results and this plan.
c. You telephone the patient and ask her to come back. You explain that you are worried about a deep vein thrombosis (DVT), and you need to examine her and repeat her blood tests.
d. You write a letter to her general practitioner informing them that self-discharged from the ED and ask that they see her urgently.
e. You requests a Doppler scan to take place the same day and for the report to be sent urgently to her general practitioner so they can organize treatment if needed.

5. While working as a doctor in training, you receive a call on Thursday morning from the hospital laboratory about a blood test result on a 56-year-old female, sent non-urgently by her general practitioner on Tuesday. The request form states '?hypothyroidism'. The full blood count shows:

- Haemoglobin: 98 g/L
- White blood cell count: 38 × 10^9/L
- Platelet count: 21 × 10^9/L
- White blood cell count differential: not available; 'flag' raised by analyzer and blood film made
- Blood film: report awaiting authorization

The full blood count result was communicated by telephone to the out-of-hours general practitioner service late on Tuesday evening. Her home address is between two different hospitals (with different computer systems). No repeat blood test or clinical assessment has taken place at the hospital where you work, though she could have attended the other nearby hospital. What is the best course of action?

a. You do not need to do anything further. The out-of-hours general practitioner service was contacted appropriately, and a message would have been communicated to her usual general practitioner.
b. You contact the patient's general practitioner to ensure that the results have been communicated to them to action.
c. You requested that the blood film is examined and that that result is communicated to the general practitioner the same day for them to action.
d. You telephone your colleague who was on call on Tuesday evening and ask them look at the blood film and follow-up the result.
e. You go to the laboratory to look at the blood film yourself. You contact the patient with the result.

Chapter 40: Venous thromboembolic disease

1. What is the best combination of anticoagulant and reversal agent for a patient with life-threatening gastrointestinal bleeding?

a. Rivaroxaban and intravenous vitamin K
b. Low-molecular-weight heparin and intravenous vitamin K
c. Dabigatran and andexanet alfa
d. Rivaroxaban and andexanet alfa
e. Low-molecular-weight heparin and prothrombin complex concentrate

2. Which of the following is true?

a. Heterozygous factor V Leiden is rare (less than 1%) in individuals of White European ethnicity.
b. Antithrombin deficiency is the most common heritable thrombophilic tendency.
c. It is important to test for heritable thrombophilia in individuals who have had a deep vein thrombosis after major surgery.
d. Detection of a heritable thrombophilia in a patient with venous thromboembolism indicates the need for life-long anticoagulation.
e. Thrombosis in people with antiphospholipid syndrome can be venous, arterial, or microvascular.

3. Which of the following statements is true in patients presenting with a suspected deep vein thrombosis (DVT) or pulmonary embolism (PE)?

a. A normal D-dimer concentration excludes a diagnosis of DVT or PE.
b. Lower limb DVT can be excluded by a single negative ultrasound scan of the proximal venous system in a patient whose two-level Wells score is 'DVT likely'.
c. A normal resting oxygen saturation on air excludes the diagnosis of PE.
d. Superficial femoral vein thrombosis requires treatment as a DVT.
e. Most patients with a PE have an associated symptomatic DVT.

4. Which of the following is true regarding venous thromboembolism (VTE)?

a. The annual incidence of VTE in the population is about 1 per 1000 people.
b. The annual incidence of VTE peaks in middle age.
c. About 10% of episodes of VTE are unprovoked.

d. All patients with an unprovoked episode of VTE should be screened for underlying cancer with a computed tomography scan of the chest, abdomen, and pelvis.
e. Long-haul travel is a stronger risk factor for VTE than major surgery.

Chapter 41: Malignant disease

1. An 80-year-old male presents to the emergency department with a fall. On questioning, he complains of progressive worsening back pain and sudden weakness in his legs in the past week. He tells you he has a history of hypertension and prostate cancer. What is the most important diagnosis to exclude?
   a. Disc prolapse
   b. Discitis
   c. Metastatic spinal cord compression
   d. Acute spinal fracture
   e. Guillan-Barré syndrome
2. A 53-year-old female presents to her general practitioner with a change in bowel habit and weight loss. She is suspected to have bowel cancer, and a computed tomography scan and colonoscopy confirm this on biopsy. What tumour marker is associated with bowel cancer?
   a. CA 19-9
   b. PSA
   c. CA 125
   d. LDH
   e. CEA
3. You are the ward doctor assessing an male with a history of a cough with blood-stained sputum. He is a smoker and has become so breathless he cannot leave the house. Which of these conditions are most associated with small cell lung cancer?
   a. Lambert-Eaton syndrome
   b. Amyloidosis
   c. Dermatomyositis
   d. Endocarditis
   e. Hypoglycaemia
4. A 34-year-old female presents to her general practitioner with a breast lump. She is referred via her general practitioner to a one-stop clinic. Ultrasonography, magnetic resonance imaging, and biopsy confirm a triple-negative breast cancer with lymph node involvement. She is started on both chemotherapy and immunotherapy while awaiting genetic testing. Which genetic disorder is most associated with breast cancer?
   a. Gorlin's syndrome
   b. Lynch syndrome
   c. *BRCA* mutation
   d. von Hippel-Lindau
   e. Hereditary non-polyposis colorectal cancer
5. A 47-year-old male with base of tongue cancer sees you in the accident and emergency department with a fever and productive cough. He is currently receiving chemotherapy and had his last chemotherapy 7 days ago. His heart rate is 124 beats/min, and his blood pressure is 97/64 mmHg with a temperature of 38.6°C. What is the most appropriate initial step in management?
   a. Order chest computed tomography
   b. Await his blood test and lactate results before starting intravenous fluids and antibiotics
   c. Start intravenous antibiotics and fluids without waiting for blood results
   d. Start intravenous fluids and paracetamol
   e. Obtain a sputum culture

Chapter 42: Infectious disease

1. A 42-year-old female presents to primary care in the UK with a 2-day history of fever with uncontrollable chills, dry cough, loose stools, and vomiting. She had returned 2 weeks previously from a 1 month stay in Lagos, Nigeria, visiting her friend and relatives in the city and in a village 3 hours out of town. Apart from a temperature of 39°C and a heart rate of 106 beats/min, the examination is unremarkable. Which of the following is the recommended management?
   a. Send the patient to the nearest emergency department for urgent malaria rapid diagnostic test and blood film.
   b. Obtain routine blood tests, including FBC, U&E, LFTs, and reassess in 48 hours.
   c. Obtain chest X-ray to exclude bacterial pneumonia and TB. Obtain stool analysis for *Salmonella* spp. and other bacterial pathogens.
   d. Start a presumptive antibiotic course to treat travellers' diarrhoea.
2. A 54-year-old patient with insulin-dependent diabetes presents to the emergency department with a 2-day history of fever and confusion. On examination, he is febrile, with a signs of meningism and a reduced conscious level but no focal neurology. There is no rash nor any murmurs. Investigations reveal a raised neutrophil count of 16×10^9/ml and a C-reactive protein (CRP) of 175 mg/L. A computed tomography scan of the head is normal, and a lumbar puncture is planned. What is the appropriate combination of urgent intravenous antimicrobial therapy?
   a. Ceftriaxone and aciclovir
   b. Ceftriaxone, amoxicillin, and aciclovir
   c. Amoxicillin and aciclovir
   d. Ceftriaxone and amoxicillin
   e. Benzylpenicillin and acyclovir

Chapter 43: Sexually transmitted infections

1. A 25-year-old female presents with an unpleasant-smelling vaginal discharge. She has no vulval symptoms and no abnormal bleeding. Her last unprotected vaginal sex was 4 weeks ago with a casual male. She did online chlamydia and gonorrhoea testing 10 days ago, and the results were negative. What is the most likely cause of her vaginal discharge?
   a. Bacterial vaginosis
   b. Candidiasis
   c. Cervical cancer
   d. Retained tampon
   e. Trichomonas vaginalis
2. A 19-year-old male from Leeds, UK, has never been tested for sexually transmitted infections (STIs). He has no symptoms but has had unprotected vaginal sex with several female partners also from Leeds. What STIs would you recommend he is tested for?
   a. Chlamydia and gonorrhoea
   b. Chlamydia, gonorrhoea, syphilis, and HIV
   c. Chlamydia, gonorrhoea, syphilis, HIV, and hepatitis B
   d. Chlamydia, gonorrhoea, syphilis, HIV, and hepatitis C
   e. Chlamydia, gonorrhoea, syphilis, HIV, and hepatitis B and C
3. In 2020, how many people did the World Health Organization (WHO) estimate acquired a sexually transmitted infection each day?
   a. About 50 000
   b. About 100 000
   c. About 500 000
   d. About 1 000 000
   e. About 5 000 000

4. A 24-year-old male presents with burning when passing urine and a whitish urethral discharge for 5 days. Microscopy of Gram-stained urethral secretion shows greater than 5 polymorphonuclear leucocytes per high-power field, but no bacteria are seen. What is the most likely cause of his urethral discharge?
   a. *Chlamydia trachomatis*
   b. Herpes simplex virus
   c. *Neisseria gonorrhoeae*
   d. *Trichomonas vaginalis*
   e. Urinary tract infection
5. *Neisseria gonorrhoeae* antimicrobial resistance is increasing worldwide, so gonococcal culture for antibiotic susceptibility testing should be taken before treatment. When antimicrobial susceptibility is not known before treatment, which antibiotic should be used?
   a. Azithromycin
   b. Ceftriaxone
   c. Ciprofloxacin
   d. Doxycycline
   e. Gentamycin

Chapter 44: Human immunodeficiency virus

1. An HIV antibody test is indicated in all adults diagnosed with what condition?
   a. Viral pneumonitis
   b. Pancreatitis
   c. Type I diabetes
   d. Non-Hodgkin's lymphoma
   e. Crohn's disease
2. A 20-year-old male has fever, sore throat, and lymphadenopathy. He has had condomless sex with several males in the preceding 3 weeks. Investigations show:
   Total white blood cell count: 6.6×10^9/L
   Neutrophils: 5.0×10^9/L
   Lymphocyte 0.5×10^9/L
   What is the most likely diagnosis?
   a. Acute Epstein-Barr virus infection
   b. Gonorrhoea
   c. Syphilis
   d. Acute HIV infection
   e. Acute cytomegalovirus infection
3. Excluding those with evidence of recent infection, the threshold definition of late HIV diagnosis currently used in the UK is a CD4 count within 91 days of diagnosis in what range?
   a. Below 500 cells/mm^3 of blood
   b. Below 450 cells/mm^3 of blood
   c. Below 400 cells/mm^3 of blood
   d. Below 350 cells/mm^3 of blood
   e. Below 200 cells/mm^3 of blood
4. A 60-year-old male with HIV is diagnosed with deep vein thrombosis. He is taking darunavir/cobicistat/emtricitabine/tenofovir alafenamide (Symtuza). Which drug should not be prescribed?
   a. Heparin
   b. Streptokinase
   c. Aspirin
   d. Warfarin
   e. Tinzaparin
5. A 48-year-old female with HIV has been virally suppressed for several years on stable anti-retroviral therapy. She describes gradual onset of poor sleep, night sweats, difficulty in concentrating at work, and low mood. What is the most likely diagnosis?
   a. HIV-associated dementia
   b. Progressive multifocal leucoencephalopathy
   c. Clinical depression
   d. Perimenopause
   e. Anti-retroviral toxicity

Chapter 45: Dermatology

1. When considering dermatological disease, what is TRUE?
   a. Skin diseases are never fatal.
   b. Skin cancer is the second most common cancer for patients of all ages.
   c. Patients with psoriasis are 2.0 to 2.5 times more likely to consider suicide than patients who do not have psoriasis.
   d. Prenatal diagnosis of skin disease is not possible because access to the fetal skin in only available after delivery of the child.
   e. Photodermatoses (skin diseases related to visible and ultraviolet light exposure) do not affect patients with skin of colour.
2. A 26-year-old chef presents to the outpatient clinic with a 2-month history of a scaly, itchy dermatitis on the hands. He has been applying a moisturizer after washing his hands. What is the best therapeutic option to offer next?
   a. Narrowband UVB phototherapy 3 times a week for 20 treatments
   b. 1 week of oral anti-staphylococcal antibiotics
   c. Topical permethrin 5% cream applied as two applications 1 week apart
   d. Moderately potent topical steroid applied daily for 2 to 4 weeks
   e. Weekly oral methotrexate 0.4 mg/kg
3. In bullous pemphigoid, which of the following statement is FALSE?
   a. Bullous pemphigoid predominantly affects older adult patients.
   b. Blisters form within the epidermis and present as flaccid erosions with mucosal involvement.
   c. Autoantibodies against 230 kDA and ColXVII can be found.
   d. Localized disease can be treated with high-potency topical corticosteroids.
   e. Intact blisters are best drained.
4. Which of the following statements related to skin cancer is TRUE?
   a. Basal cell carcinoma is the most common skin cancer in people with lighter skin tones, is more common in immunocompromised people, and metastasizes more frequently than cutaneous squamous cell carcinoma.
   b. The incidence of cutaneous squamous cell carcinoma is reducing in most countries because of more widespread use of sunscreen and other photoprotective measures.
   c. Atypical mole syndrome is only associated with an increased risk of melanoma if there is also a family history of melanoma.
   d. Mycosis fungoides is the most common form of primary cutaneous lymphoma.
   e. Cutaneous squamous cell carcinoma in situ (Bowen's disease) progresses to invasive cutaneous squamous cell carcinoma in at least 50% to 60% of cases.

Chapter 46: Young adult medicine

1. A 14-year-old female asks to be prescribed an oral contraceptive without her parent's knowledge. Which of the following is the best approach?
   a. Sexual activity with a 14-year-old is a criminal offense. You refuse to prescribe an oral contraceptive and make a child protection referral.
   b. It is not in the best interests of a 14-year-old to take an oral contraceptive, so you do not prescribe it. You inform whoever has parental responsibility for the young person about their request.
   c. You establish a rapport with the young person. You judge her to be competent and prescribe the oral contraceptive. You try to persuade her to inform her parents.
   d. You establish a rapport with the young person and judge her to be competent. You advise that she can collect a prescription for the oral contraceptive when she has informed her parents.
   e. You do not prescribe the oral contraceptive. You advise using a barrier contraceptive method instead.
2. Which of the following best applies in the UK in relation to legal rights, duties, powers, responsibilities, and authorities a parent has in relation to their child?
   a. The local council has full parental responsibility for looked after children.
   b. No more than two people can have parental responsibility for the same child.
   c. Same-sex partners automatically have joint parental responsibility of their children.
   d. A father only has parental responsibility if he is married to the mother or named as the father on the child's birth certificate.
   e. A father no longer has legal parental responsibility if he divorces his child's mother.
3. Which of the following best describes the transfer of care from paediatric to adult services for long-term medical conditions?
   a. Most young adults do not need to continue to attend healthcare services regularly after moving to adult care.
   b. A period of preparation, ideally starting in the early teenage years, is an important determinant of improved long-term outcomes.
   c. Parents need to relinquish responsibility so that young people can take on the management of their health needs.
   d. Transition is a single event in which young people and their carers meet with the adult team to hand over their ongoing care.
   e. Young people should have their care transferred to adult services before their 17th birthday.
4. Which of the following statements is TRUE about smoking and vaping in the UK?
   a. Increasing numbers of young people smoke cigarettes.
   b. It is illegal to sell any vaping product, whether it contains nicotine or not, to those younger than 18 years of age.
   c. Most young people who vape do so to help them give up smoking.
   d. Vaping nicotine-containing products is not as addictive as smoking nicotine-containing cigarettes.
   e. Most smokers need to use a nicotine replacement product to enable them to give up smoking.
5. Which of the following is most accurate in relation to risk-taking behaviour among young people in the UK?
   a. Alcohol use is associated with unplanned pregnancy.
   b. Death from intentional self-harm is more common than accidental death.
   c. Cannabis use is unlikely to progress to wider substance misuse behaviours.
   d. Alcohol misuse is an increasing healthcare problem.
   e. Delayed sleep phase syndrome is a rare cause of difficulty getting to sleep at night and waking in the morning.

Chapter 47: Sex and gender in medicine

1. A 23-year-old female presents 100 hours after unprotected sexual intercourse. She has epilepsy and takes levetiracetam for this. She is allergic to penicillin. She has had two previous surgical terminations. What is the best emergency contraception for her?
   a. Levonelle 1.5 mg (levonorgestrel)
   b. EllaOne (ulipristal acetate 30 mg)
   c. Combined oral contraceptive pill tricycle
   d. Intrauterine copper device
   e. Levonelle 3 mg (levonorgestrel)
2. A 35-year-old nulliparous female presents with a history of heavy, regular menstrual bleeding with associated dysmenorrhoea. She does not experience intermenstrual or postcoital bleeding. She does not have any non-cyclical pelvic pain. Her human papillomavirus (HPV) screening was normal 2 years ago. She has one regular male sexual partner who uses condoms for contraception. On examination, her abdomen is soft, and there is a pelvic mass palpable to the level of the umbilicus.

Investigations show:

Plasma haemoglobin: 85 g/dL

Urine bHCG test: negative

What is the most appropriate next investigation for this patient?
   a. HPV (smear) test
   b. Computed tomography scan of the abdomen and pelvis
   c. Pelvic ultrasound scan
   d. Hysteroscopy
   e. Pelvic magnetic resonance imaging
3. In male patients, chronic urinary retention is suggested by
   a. postvoid bladder volume of 150 mL.
   b. pelvic pain with visible haematuria.
   c. painless inability to void.
   d. terminal dribbling.
   e. prostate volume greater than 80 cc.
4. Which of the following statements about testicular torsion is true?
   a. It requires urgent treatment with antibiotics.
   b. It is a time-critical surgical emergency.
   c. It is not seen in adult males.
   d. It requires confirmation with ultrasound examination of the testes.
   e. It is typically caused by uropathogens in males older than 35 years of age.
5. Hormonal management of gender incongruence
   a. is required for all people who change the social expression of their gender.
   b. requires ongoing assessment of cardiovascular risk.
   c. means additional hormonal contraception is no longer required to prevent unplanned pregnancy.
   d. leads to maximal breast development within 3 to 6 months of treatment in transgender females.
   e. leads to an increase in the risk of breast cancer in transgender females compared with cisgender females.

Chapter 48: Obstetric medicine

1. Which of the following conditions is an indication for the use of low-dose aspirin from 12 weeks' gestation to prevent preterm pre-eclampsia?
   a. Grave's disease
   b. Hypothyroidism
   c. Previous venous thromboembolism provoked by the contraceptive pill
   d. Chronic hypertension
   e. Epilepsy
2. A female with type 2 diabetes mellitus presents for pre-pregnancy counselling. Which of the following medications should she be advised to continue in pregnancy?
   a. Metformin
   b. Empagliflozin
   c. Semaglutide
   d. Ramipril
   e. Simvastatin
3. Which of the following physiological parameters are unaffected by pregnancy?
   a. Heart rate
   b. Blood pressure
   c. Respiratory rate
   d. Glomerular filtration rate
   e. Cardiac output
4. Which of the following antiepileptics drugs carries the highest risk of congenital malformations?
   a. Lamotrigine
   b. Levetiracetam
   c. Carbamazepine
   d. Sodium valproate
   e. Phenytoin
5. Which of the following test results do not influence pregnancy outcome in females with systemic lupus?
   a. Protein:creatinine ratio
   b. Anti-Ro/SSA
   c. Double-stranded DNA
   d. Lupus anticoagulant antibody
   e. Alkaline phosphatase

Chapter 49: Medicine of the older adult

1. An 82-year-old female tripped on a curb and fractured her hip. She denies losing consciousness. She has a medical history including diabetes, hypertension, osteoarthritis of her knee, and atrial fibrillation. She normally walks with a stick. Her medications include metformin, losartan, tramadol, bisoprolol, and apixaban. On examination, she is in pain. Her abbreviated mental test score is 8 of 10. Her left leg is shortened and externally rotated. She has a graze on her left palm where she had landed. What is the most likely cause of her fall?
   a. Arrythmia
   b. Hypoglycaemia
   c. Mechanical fall
   d. Multifactorial fall
   e. Simple trip
2. A 95-year-old female is admitted with increasing confusion. Her medical history includes hypertension, arthritis of her spine, and gallstones. Her medications include aspirin and amlodipine. She was recently started on a low–dose buprenorphine patch. On examination, she looks well. Her abbreviated mental test score is 4 of 10. Cardiovascular, respiratory, and abdominal examinations are unremarkable, apart from some mild abdominal distension. Her weight is 40 kg.
   Her blood results include:
   Haemoglobin: 105 g/L (115–165 g/L)
   White blood cell count: 8.6 × 10^9/L (4–10 × 10^9/L)
   Urea: 5.3 mmol/L (2.8–8.1 mmol/L)
   Creatinine: 85 μmol/L (44–80 μmol/)
   Alkaline phosphatase: 130 IU/L (35–104 IU/L)
   Bilirubin: 22 μmol/L (0–21 μmol/L)
   Thyroid-stimulating hormone: 6.23 μ/L (0.27–4.2 μ/L)
   Free T_4: 16 pmol/L (11.1–22)
   What is the most likely cause of her confusion?
   a. Hypothyroidism
   b. Inappropriate dosing
   c. Liver dysfunction
   d. Reduced body fat
   e. Renal impairment
3. A 80-year-old female complains of urinary incontinence. She gets very little warning of when she needs to pass urine, and she finds it difficult to control. She is independent with her activities of daily living, though she gets out less than she used to because of the incontinence. She walks with a stick. There is no evidence of current urinary infection. A postvoid bladder scan shows a residual volume of 130 mL. What is the cause of her incontinence?
   a. Functional incontinence
   b. Mixed incontinence
   c. Overflow incontinence
   d. Stress incontinence
   e. Urge incontinence
4. A 76-year-old male has been struggling at home after the death of his wife 6 months previously. His daughter helps where she can, but she has three young children and struggles to help with his needs. He walks with a stick, though he is occasionally a bit wobbly on his feet and has nearly fallen on a couple of occasions. He occasionally gets breathless when walking to the shops. His daughter is worried about him doing his own cooking because his memory is not as sharp as it used to be, and he left the oven on once. She also says there have been occasions when he has not changed his clothes or washed as much as he used to. He has also stopped getting out as much as usual, and his daughter is worried that he is getting depressed after his wife's death. What is the most important next step with his management?
   a. Comprehensive geriatric assessment
   b. Geriatrician assessment
   c. Occupational therapist assessment
   d. Physiotherapy assessment
   e. Social worker assessment
5. An 87-year-old male is struggling at home because of early dementia. He previously did well with a twice-daily care package to help with his meals and washing and dressing. This was stopped 1 month ago. When asked, he said that he did like them coming round and had felt quite lonely since they stopped calling in. However, his daughter had been worried about how expensive it was and that 'it would use up all my inheritance'. She had stopped this and started claiming benefits for providing his care herself. Social services called a few days after the care package was cancelled and had no concerns. However, he mentioned that she had not been able to call in very often because she was often too busy and had just returned from a holiday. After a fall at

home, the ambulance crew that attended raised a safeguarding alert because the patient appeared unkempt, and the house was very dirty. What is the main type of abuse seen here?

a. Emotional abuse
b. Financial abuse
c. Neglect
d. No concerns
e. Organizational abuse

Chapter 50: Palliative care and symptom control

1. Which of these patients is appropriate for referral to a palliative care team?

a. An actively dying patient in the last days of life
b. A patient with metastatic lung cancer with no further oncological treatment options
c. A patient undergoing curative chemotherapy for acute myeloid leukemia with extensive and painful oral mucositis
d. A patient with severe heart failure and troublesome breathlessness who would like to talk about his wishes for the future
e. All of the above

2. A 75-year-old female has severe chronic obstructive pulmonary disease on maximal treatment and repeated hospital admissions with exacerbations. She has discontinued pulmonary rehabilitation; she finds fresh air gives some relief but is increasingly housebound. She is distressed by breathlessness, and has made repeated calls to the general practitioner. What would you consider first?

a. Prescribe morphine immediate release orally 2.5 mg four times daily and as required.
b. Prescribe lorazepam 0.5 mg sublingually as required up to 2 mg/24 hours.
c. Prescribe oxygen therapy at 2 L/min.
d. Refer to community palliative care team to consider non-pharmacological interventions (e.g., a handheld fan and breathing exercises).
e. Prescribe morphine sulphate tablets 5 mg twice daily.

3. Advanced care planning discussions

a. are ideally held during a patient's last weeks of life.
b. are best had during an acute admission.
c. should be considered proactively at every healthcare interaction.
d. must involve completion of an Advanced Decision to Refuse Treatment.
e. can only be completed in one conversation.

4. An 85-year-old nursing home resident with a background of advanced dementia was admitted to hospital with drowsiness, fever, and shortness of breath. He has completed a course of intravenous antibiotics for an aspiration pneumonia, but his condition has continued to deteriorate, and he is bedbound, drowsy, and drinking minimal amounts of fluid. The treating team meets with the family and communicates that they think he is dying and may be in the last days of life. A best interests decision is made to focus on comfort measures, and his ReSPECT form is completed indicating that he will not receive cardiopulmonary resuscitation. What is the most appropriate next step?

a. Start a syringe driver with morphine and midazolam.
b. Cross off all his regular medications.
c. Prescribe i.v. fluids to maintain hydration.
d. Refer to chaplaincy.
e. Develop a personalized care plan, in discussion with his family, to support individualized decisions about hydration and nutrition, spiritual care, and symptom control.

5. A 64-year-old female with metastatic breast cancer and bone metastases attends the oncology clinic with progressive right hip pain. A computed tomography scan confirms a bone metastasis corresponding to the pain site and rules out a fracture. She is treated with palliative radiotherapy and continues on her chemotherapy. She is reviewed 2 months later and continues to report significant pain. She is frightened that her disease is progressing and cannot sleep at night. She is now on morphine MR 30 mg twice daily, which has been titrated up with little effect. She is not experiencing any side-effects and is mobilizing well. What would be the most appropriate next step?

a. Increase the morphine MR dose further because there is no ceiling effect.
b. Switch to a fentanyl patch because it may be stronger.
c. Referral to the hospital palliative care team for advice on starting a syringe driver.
d. Start amitriptyline at night for low mood and nerve pain.
e. Refer to the hospital palliative care team for assessment of other pain dimensions and consideration of adjuvant analgesics and psychological support.

B Answers to chapter questions

Chapter 2: Clinical communication

1. d. Provide information to the patient about what is going to happen.
2. a. Ask the patient if they felt understood.
3. e. Start with what the patient already knows and build from there.
4. d. Start with open questions.
5. b. Use natural frequencies (e.g., 4 of 10 people).

Chapter 3: Medical ethics

1. e. Being honest and telling the truth, especially about difficult or embarrassing subjects, and speaking frankly without trying to conceal information or put a positive spin on it for our own advantage. Summarized from GMC Duty of Candour https://www.gmc-uk.org/ethical-guidance/ethical-guidance-for-doctors/candour---openness-and-honesty-when-things-go-wrong/being-open-and-honest-with-patients-in-your-care-and-those-close-to-them-when-things-go-wrong. Each option listed is accurate in part, but only E covers the full meaning of the GMC guidance.
2. b. Material risk
3. c. NHS Resolution

Chapter 4: Evidence-based practice

1. c. Research papers
2. c. Observations
3. a. Randomized controlled trials
4. a. To ensure that study groups are as similar as possible
5. c. Patients and clinicians must be blinded to the intervention that is being delivered
6. a. No adjustments are made to participants' treatments or lifestyle
7. a. There is a high risk of recall bias affecting the results.
8. b. Whether a new intervention is acceptable to patients
9. b. The study is sufficiently free from bias.
10. a. The study findings are relevant to patients in other settings.
11. b. Participants in a case-control study are more likely to remember relevant past history if they have the disease of interest.
12. a. The odds of an event occurring in the intervention group divided by the odds in the control group
13. b. The proportion of people in the treatment group who experience the outcome event divided by the same proportion in the control group
14. b. The control performs better than the intervention.
15. a. There is a 95% probability that the 95% confidence interval contains the true value.
16. a. There is strong statistical evidence against the null hypothesis.

Chapter 5: Human genetics

1. b. Targeted panel of genes. These panels are updated to reflect the latest informative genes to study for given phenotypes and would be the first choice. Whole-genome sequencing would yield the answer but typically is selected after a targeted panel, especially if the targeted panel did not reveal a result. SNP array would not target rare single nucleotide variants. Single-cell RNA sequencing would not give genetic sequence data required. Methylation array would provide only epigenetic data.
2. d. Both parents are unaffected carriers. Epidermolytic ichthyosis is a dominantly inherited ichthyosis, so this is an explanation for recessive ichthyoses. Gonadal mosaicism is possible and may be revealed by testing of sperm in the father and may help future family planning. Cutaneous and gonadal mosaicism is also possible and both parents should have a complete skin examination; the presence of a linear birthmark would be a clinical clue of cutaneous mosaicism and can be confirmed with a biopsy and testing of the skin for the same pathogenic variant. Nonpaternity is a possible explanation that needs to be addressed with sensitivity. A *de novo* postzygotic mutation is an early genetic event in the forming embryo and could account for the genetic result.
3. a. Family history questionnaire. This is the first step and would identify the youngest person with the most melanomas, who would be the preferred first person to test in the pedigree if this was feasible. Testing the mother's germline DNA may not be conclusive as in familial melanoma some affected parents may not carry the familial pathogenic variant. Testing the son's germline DNA would be the next step if a family history did not reveal anyone else that was affected. To not offer a test may be the appropriate outcome if the patient had an informed discussion and chose this; although typically this would follow a full family history and skin examination of the affected members. Testing the mother's archival skin tumour would give information on somatic changes in the mother's skin tumour but these would need to be demonstrated in the germline to be useful for genetic counselling of the family.
4. b. Fifty percent of girls will be carriers. Haemophilia A is inherited as an X-linked recessive trait and so fifty percent of girls will be carriers if they inherit either affected X chromosome from the mother or 50% non-carriers if they inherit the normal maternal allele. Fifty percent of boys will have haemophilia because they will inherit the pathogenic allele from their mother (and the Y chromosome from their father); the other 50% will be normal because they inherit the normal allele from their mother (and the Y chromosome from their father).
5. c. Bisulphite modification of DNA. This step is used when testing for DNA methylation, which is not needed here. DNA is used to amplify the single protein coding exon of the *GJB2* gene. Oligonucleotide primers that are complementary to the sequences on different strands at each end of the *GJB2* gene are used. Capillary electrophoresis separates the different-sized sequence fragments and captures the fluorescent signals to allow the DNA sequence to be analyzed. The heat-stable DNA polymerase enzyme allows the repeated amplification of the *GJB2* gene, which includes 30 to 35 denaturing cycling steps at a high temperature.

Chapter 6: Immunity

1. d. Macrophages
2. e. Dendritic cells
3. d. Neutralization of pathogens
4. e. releasing the brakes on T-cell activity.
5. d. mast cells.

Chapter 7: Therapeutics and prescribing

1. b. Intravenous. By definition, intravenous administration gives 100% bioavailability.
2. d. Inhibition. Pharmacokinetics is what the body does to the drug. Inhibition is one mechanism by which drugs exert pharmacodynamic effect.
3. a. Dose independent. Type B reactions are idiosyncractic, dose independent, unpredictable, and often serious.
4. d. 20 hours. In four to five half-lives, more than 97% of drug will be eliminated.
5. d. IV. Phase IV refers to post-marketing studies that usually evaluate drug safety but can be used to evaluate real world effectiveness as well.

Chapter 8: Radiology

1. d. Ultrasonography of the abdomen. Ultrasonography is an excellent (>90% sensitivity and specificity) modality at detecting gallstones, which has the added benefits of being readily available, economical, and comfortable for patients. MRCP is excellent (>90% sensitivity and specificity) at detecting choledocholithiasis (bile duct calculus). HIDA uses the intravenous injection of a radioactive tracer to investigate the physiological function of the biliary system. Some indications include biliary leak (e.g., after transplant or gallbladder surgery), chronic abdominal pain (sphincter of Oddi dysfunction or chronic cholecystitis), and biliary atresia (paediatric patient group). Only about 15% of gallstones are visible on X-ray. This is why CT is also poor at identifying gallstones. CT is, however, able to detect complications of gallstones (e.g., biliary obstruction, severe cholecystitis, gallbladder perforation).
2. e. None of the above. Bacterial meningitis is a rapidly evolving condition (see Chapter 11). Confirm a diagnosis of bacterial meningitis based on: clinical features, blood test results and lumbar puncture results. Neuroimaging is indicated only if the person has risk factors for or symptoms or signs of raised intracranial pressure. These include new focal neurological features (including seizures or posturing), abnormal pupillary reactions, a Glasgow Coma Scale score of 9 or less, or a progressive and sustained or rapid fall in level of consciousness. Do not perform a lumbar puncture until these factors have been resolved because of the risk of tonsillar herniation ('coning'). Take bloods, give antibiotics, and stabilize the person before imaging.
3. e. Contrast-enhanced MRI (gadolinium). Gadolinium-based contrast media use in patients with severe chronic kidney disease (CKD) was demonstrated to very rarely cause nephrogenic systemic fibrosis (NSF). Restrictive policies of gadolinium use in CKD and selective use of contrast agents that bind free gadolinium more strongly have resulted in the virtual elimination of cases of NSF. Concerns of gadolinium use in renal impairment are magnified by recent observations demonstrating gadolinium deposition in the brain and a possible systemic syndrome attributed to gadolinium contrast agents. Discuss with radiologists if you think gadolinium-enhanced imaging is required for your patient. Ultrasonography is often indicated if a patient has renal impairment to exclude a structural cause for the impairment (e.g., renal obstruction). Risk of CT iodinated contrast-induced acute injury is estimated to be near 0% at an eGFR greater than 45, 0% to 2% at an eGFR of 30 to 44, and 0% to 17% at an eGFR less than 30 mL/min. Prophylaxis with intravenous fluid therapy may be indicated (Davenport MS et al. Consensus statement American College of Radiology and National Kidney Foundation 2020).
4. c. The mediastinum measures 10 cm. A mediastinum larger than 9 cm on a supine (where the image is magnified when compared to PA) CXR is considered abnormal, and CT should be considered.
5. b. Pacemakers are contraindicated for MRI. Older implanted pacemakers are contraindicated, but newer pacemakers may be conditionally safe. Please discuss which category your patient's pacemaker falls into with MRI technicians.

Chapter 9: Public health

1. e. The effects could be proposed to be similar to other situations.
2. d. The natural history of the condition should be well understood.
3. d. Positive predictive value varies with prevalence of the disease.
4. b. Influenza
5. a. compare small areas across England.

Chapter 10: Global and planetary health

1. c. has decreased because of economic austerity.
2. b. Laparotomy
3. d. One-third of all food produced is wasted each year.
4. c. Access to water and sanitation is not a UN-recognized human right.
5. d. The Universal Declaration of Human Rights.

Chapter 11: Acute medicine

1. d. Simplified PE severity index (SPESI). The SPESI, or PE severity index (PESI), can be used to decide which patients need admission. Additional factors may include immobility and other barriers to treating in the community such as homelessness. The weight and therefore the BMI may affect which anticoagulant can be used (DOAC vs warfarin) or dosage (e.g., split-dose LMWH if weight markedly increased). The D-dimer level is non-specific but may help with ruling out the diagnosis of thromboembolic disease if negative, although in this case, the patient has already undergone a diagnostic test (CTPA). A prolonged QRS duration, most commonly right bundle branch block, raises concern for right ventricular strain but may be present as a normal variant. The Wells score for pulmonary embolism should be used to calculate the risk for possible pulmonary embolism and the need for further investigation. It is not used after diagnosis.
2. e. Septations and debris on the right side on point-of-care ultrasound (POCUS). Debris and septations on POCUS are strongly suggestive of an empyema and as such would support the decision to insert an appropriately sized chest drain in an urgent manner; there may be a need to perform CT to elucidate if it is

loculated. The size of the effusion is not useful when deciding whether to intervene out of hours because the patient may be stable enough to monitor overnight, especially if the condition is chronic. Thromboembolic disease does not usually present as a large pleural effusion unless there is other underlying pathology. Use of anticoagulants may need careful consideration when contemplating a procedure. A pleural pH of 7.2 or below requires urgent management because this is suggestive of empyema. The requirement for relatively low-level oxygen therapy overnight, if monitored closely, would not make an urgent procedure more likely.

3. c. Idiosyncratic reaction to amoxicillin and clavulanic acid. Idiosyncratic reactions to amoxicillin and clavulanic acid classically occur 2 weeks after treatment and can present with a cholestatic picture. Although the BMI is raised, it is not over 30 kg/m^2, and the alcohol consumption is within accepted advisory limits (14 units/week). Primary CMV infection is often asymptomatic and frequently earlier in life than the fifth decade, and it is less common to have organ involvement if immunocompetent. Initial blood tests in acute CMV or EBV infection might show a leucocytosis and thrombocytopenia. Serological tests can be sent to help with the diagnosis. A recently passed gallstone may present with biliary colic and may lead to a dilated bile duct (1 mm per decade is often deemed within normal limits). A staggered paracetamol overdose (or therapeutic misadventure if caused by apparent clinical need) is unlikely here unless the body weight is less than 50 kg or malnutrition, when dose reduction may be needed (to e.g., 500 mg four times a day).

4. e. 0.9% saline with potassium and 6 units per hour fixed-rate insulin and usual long-acting insulin. The patient has diabetic ketoacidosis (DKA) as evidenced by their diabetes history, likely Kussmaul breathing, low bicarbonate and pH, and high plasma glucose and blood ketone levels. Initial rehydration should include i.v. fluids (of note, the sodium may appear low because of pseudohyponatraemia from high blood glucose) to correct dehydration. The fixed rate of insulin therapy for DKA is 0.1 units/kg/hour. Long-acting insulin should be considered at a normal dose to maintain a background level of insulin. The patient's blood potassium level will drop further with insulin therapy driving it intracellularly, so replacement is required here. The cortisol level is in the normal range and is taken at 10 p.m. (rather than in the morning); therefore, it is unlikely that this is an Addisonian crisis.

5. a. Co-trimoxazole therapy. As an acute physician, it is common to be called with abnormal blood results. Co-trimoxazole is a combination of sulfamethoxazole and trimethoprim, which are often combined because they have a synergistic effect on bacteria. A relatively common side-effect of this medication is to increase potassium levels (due to an amiloride-like effect of trimethoprim in the distal renal tubule) as well as cause a rise in serum creatinine because of competitive inhibition of creatinine's tubular excretion. In dehydration, it might be expected that the urea would rise more relatively than the creatinine. Potassium permanganate soaks would not be absorbed in any meaningful fashion and should not affect potassium or creatinine measurements. If the patient had sepsis, they should have had abnormal observations and would have been reviewed for this in the clinic. Urinary obstruction is less likely in the female population, especially because in this case she is catheterized.

Chapter 12: Sepsis and the treatment of bacterial infection

1. d. Ampicillin/sulbactam

2. e. Tigecycline

3. b. Ceftriaxone

4. c. Ertapenem

5. d. sepsis with an unsupported mean arterial blood pressure below 65 mmHg and an arterial blood lactate concentration above 2 mmol/L after fluid resuscitation.

Chapter 13: Critical care medicine

1. d. Start an infusion of norepinephrine. Norepinephrine is a predominantly α1-active agent and will reverse the sepsis-related loss of arterial tone. This patient's CVP of 10 mmHg and stroke volume variation of less than 10% suggest adequate venous filling pressures, and further fluid at this point is unlikely to help. Dobutamine is a predominantly β1-active compound, therefore having effects on heart rate and contractility. The major issue here is of loss of α1-mediated arterial tone. Dobutamine will worsen her tachycardia and may cause some further loss of arterial tone. This patient's poor urine output relates to a lack of renal perfusion that should be corrected. Furosemide will worsen this and may cause further hypotension. Vasopressin may be used to increase arterial tone, but current recommendations suggest it should be used as an adjunct to norepinephrine rather than in place of it.

2. d. Metabolic acidosis without associated hypoxia. The blood gas shows an acidosis because the pH is less than 7.35, which is metabolic in nature. This is evidenced by the normal $PaCO_2$ and low $cHCO_3$. This is most likely caused by a build-up of lactic acid in tissues not perfused during his cardiac arrest. CPR will never be as effective as normal cardiac output in perfusing body tissues. If the patient had respiratory acidosis, we would expect the $PaCO_2$ to be high. The patient is acidotic, not alkalotic, and we would expect the $PaCO_2$ to be low. Although the blood gas shows a metabolic acidosis, the patient is not hypoxic; he is hyperoxic, and we should turn down the oxygen. The patient is acidotic, not alkalotic, and we would also expect the $cHCO_3$ to be high.

3. e. Improved survival has been demonstrated with lower ventilatory tidal volumes. The ARDSnet trial showed that survival was greater with tidal volumes of 6 to 10 mL/kg. This suggests that some of what was initially thought to be ARDS may in fact have been ventilatory-associated lung injury. The diagnosis of ARDS requires the triad of hypoxia, low lung compliance, and widespread infiltrates on chest imaging. Steroids may be useful in later fibrosing disease but have no place in the acute management of the condition, which is largely supportive. A PaO_2/FiO_2 ratio of less than 300 mmHg with a minimum PEEP of 5 cmH_2O is required to diagnose ARDS. Immunosuppression offers no protection against ARDS.

4. d. Diarrhoea. This causes a normal anion gap with acidosis caused by gastrointestinal bicarbonate loss. Lactic acidosis causes a raised anion gap. Diabetic ketoacidosis causes a raised anion gap and is ruled out by the negative ketones. Salicylate poisoning would cause a raised anion gap. Ethylene glycol toxicity causes a raised anion gap.

5. b. Hypovolaemia. The best guide of this is the wedge pressure representative of left atrial filling pressure. Right atrial or central venous pressure is not always representative of this, particularly

in right ventricular dysfunction such as after myocardial infarction. Cardiac tamponade would produce an elevated right atrial pressure, normal or high wedge pressure, and low cardiac index. Ventricular septal defect would produce a significant fall in venous oxygen saturation caused by right-to-left shunt. A low saturation can also be seen in low cardiac output states or anaemia. Mitral regurgitation produces left atrial hypertension, so the wedge pressure and pulmonary artery pressures should be high. Cardiogenic shock produces a fall in cardiac index, usually with elevated wedge pressure.

Chapter 14: Anaesthesia and perioperative medicine

1. c. Pneumonia. This patient has a number of risk factors for postoperative pulmonary complications, including major abdominal surgery, age, and undergoing a prolonged procedure. He has not mobilized and is not coughing well to clear his secretions. He is likely to have postoperative pneumonia caused by atelectasis and sputum retention.

2. b. 2. This patient has mild systemic disease, so has an ASA score of 2. She is only using a single medication for her hypertension and appears to have good blood pressure control, so the score of 3 (severe systemic disease) does not apply here.

3. a. Psychology, exercise, and nutrition. Prehabilitation is built around psychological preparation of the patient for major surgery, structured exercise to improve cardiorespiratory fitness, and nutritional support to optimize metabolic status. Smoking cessation may additionally be relevant, as could be occupational therapy assessment and medical optimization (including investigations such as echocardiography). However, these are not the core parts of a true prehabilitation programme.

4. c. Nasogastric tube. Enhanced recovery or fast-track protocols vary by centre and by type of surgery. However, most advise minimizing lines, drains, and tubes where possible, including nasogastric tubes. Patients are better able to achieve early drinking, eating, and mobilizing ('DrEaMing') in this way.

5. d. Surgical Outcome Risk Tool (SORT). The SORT models 30-day mortality risk for a wide variety of surgeries, including head and neck procedures. APACHE is predominantly used to assess outcome risk in critical care patients, EuroSCORE is for cardiac surgery, and MOCA is for assessing a patient's cognitive condition (e.g., in dementia or delirium). C-reactive protein levels are not useful in isolation for risk prediction in surgery.

Chapter 15: Toxicology

1. c. Fomepizole
2. d. Phosphine
3. c. Dry mouth
4. c. Prolonged QRS duration
5. e. Methylthioninium chloride ('methylene blue')

Chapter 16: Environmental medicine

1. e. All of the above
2. a. Dexamethasone and immediate descent to a lower altitude
3. d. Coordination is lost
4. d. intravascular coagulation and multiorgan failure can occur.
5. d. Immediate treatment is to give rescue breaths and oxygen if available.
6. d. Prolonged CPR may be required because of respiratory muscle paralysis.

Chapter 17: Cardiology

1. e. Electrocardiography
2. c. Direct oral anticoagulant
3. c. Sodium-glucose transport protein 2 inhibitor
4. b. Coarctation of the aorta
5. d. Ibuprofen and colchicine

Chapter 18: Hypertension

1. b. Ibuprofen. Non-steroidal anti-inflammatory drugs interfere with salt and water handling in the kidney and elevate blood pressure.

2. b. Low vegetable and fruit intake. The other patterns are associated with improved blood pressure control.

3. c. Diastolic blood pressure. The unit measure of arterial pressure is millimetres of mercury (mmHg). Venous pressure is measured in centimetres (of water).

4. b. Clonidine. Central agents should be reserved for expert use and are not one of the first four to six classes of medications recommended for routine use in patients with hypertension.

5. c. Grade 1 hypertension. Although the diastolic blood pressure is normal, systolic blood pressure is greater than 135 mmHg, defining grade 1 hypertension for out-of-office methods.

Chapter 19: Lipid and metabolic disorders

1. a. Apolipoprotein B48. Apolipoprotein B48 is made by gut enterocytes and is only found in chylomicron particles.

2. c. Non-HDL cholesterol. This is the best epidemiological predictor of cardiovascular disease risk and is superior to LDL-C.

3. d. Leukaemia. Leukaemia, like many cancers, is not associated with a cholesterol elevation but usually with reduced cholesterol.

4. b. Familial hypercholesterolaemia. The vast majority of cases of familial hypercholesterolaemia are autosomally dominantly inherited. One very rare condition affecting the low-density lipoprotein receptor adaptor protein-1 (LDLRAP1) in lysosomes shows recessive inheritance and is associated with high cholesterol levels.

5. d. Icosapent ethyl (omega-3 fatty acid). Omega-3 fatty acid preparations have no effects on LDL-C levels or may slightly raise them.

Chapter 20: Respiratory medicine

1. b. Prescribe oral prednisolone, ensure a preventer and reliever therapy prescription, and advise to see general practitioner within 1 week and seek medical advice if his condition fails to improve.

2. e. All of the above.

3. a. The patient should have care in a side room.

4. c. Order a chest radiograph and blood tests and send a 2-week wait referral for suspected lung cancer.

5. b. A breathlessness score (e.g., MRC or CAT) and frequency of exacerbations

Chapter 21: Ear, nose, and throat

1. c. This is 'glue ear' on the right and should resolve in the next few weeks. The negative Rinne's test result on the right, with the Weber's lateralizing to the right is suggestive of a conductive hearing loss on the right. In the context of a recent URTI, this is suggestive of glue ear and should resolve. Antibiotics and steroids are not routinely recommended.

2. d. Benign paroxysmal positional vertigo. A vertigo history where episodes last for seconds, triggered by head movement, are typical for benign paroxysmal positional vertigo. This

condition is known to be triggered by head injuries, among other insults (e.g., vestibular neuritis).

3. d. Posterior circulation stroke. This question references parts of the HINTS exam. In an older adult patient, disabling acute vertigo that lasts all day is concerning for a posterior circulation stroke. Other clues are direction changing nystagmus, a positive test of skew, and negative head thrust test result.

4. c. Fine-needle aspiration of the lymph node. Unilateral throat pain and dysphagia with otalgia are head and neck red flag symptoms for cancer. The enlarged tonsil and palpable lymph nodes are concerning for oropharyngeal cancer. Fine-needle aspiration cytology is the only option that provides a cytological diagnosis of cancer.

5. d. Request an MRI scan of the internal auditory meatus. This question refers to a history of asymmetrical hearing loss. This is the most common way in which vestibular schwannomas present and should be investigated with magnetic resonance imaging of the internal auditory meatus. Referring for hearing aids may be appropriate if the patient wants to try these, but this should be done alongside or after the diagnostic investigations, which should be prioritized. There is no role for steroids in hearing loss that has been present for 1 year.

Chapter 22: Gastroenterology

1. d. *Streptococcus* spp.
2. b. Belching
3. a. *HLA-DQ2*
4. c. TNF-α
5. d. Stimulates appetite and increases gastric emptying

Chapter 23: Hepatology

1. d. The patient should be informed that curative tablet-based therapy is available.
2. d. An HIV test should be performed before initiation of therapy
3. d. An endoscopic retrograde cholangiopancreatography examination
4. b. In patients with alcoholic liver disease and steatohepatitis, the ALT is often greater than the AST.
5. c. After liver transplantation for autoimmune hepatitis, the disease never recurs.

Chapter 24: Biliary tract and pancreatic disease

1. e. Biliary colic
2. b. Acute pancreatitis
3. e. Carcinoma localized to the body or tail of the pancreas
4. c. Carcinoma of the head of the pancreas
5. a. Pseudocyst

Chapter 25: Nutrition

1. d. Re-feeding results in hypokalaemia, hypomagnesemia, hypophosphatemia, and thiamine deficiency.
2. b. B_3
3. c. Acute severe ulcerative colitis
4. c. Parenteral nutrition with or without intravenous fluids
5. b. Metastatic colorectal cancer

Chapter 26: Kidney and urinary tract disease

1. b. IgA nephropathy
2. b. Pyelonephritis
3. b. Renal cell carcinoma
4. d. Diabetic nephropathy
5. c. Peritoneal dialysis-associated peritonitis

Chapter 27: Water balance, fluids, and electrolytes

1. d. Amiloride
2. b. Increasing the water permeability of the cortical and medullary collecting ducts
3. b. Hyperglycaemia
4. e. Respiratory acidosis

Chapter 28: Neurology

1. c. Herpes simplex encephalitis
2. a. Cluster headache
3. b. Giant cell arteritis
4. c. Schwannoma (acoustic neuroma)
5. a. Idiopathic intracranial hypertension
6. c. Syringomyelia
7. d. Vitamin B_1 deficiency
8. c. Progressive bulbar palsy
9. a. Neurofibromatosis type II

Chapter 29: Stroke medicine

1. d. Hypertension
2. d. Start aspirin and clopidogrel mg for 3 weeks and then single-antiplatelet treatment.
3. e. Aggressive vascular risk factor control
4. c. Commence immediate heparin

Chapter 30: General hospital psychiatry

1. d. Establish if he has the mental capacity to refuse vital signs and examination; document that he does not have capacity to refuse a basic examination and make a *best interests decision* to examine him. This is a common occurrence in out-of-hours hospital work. You must document that your patient does not have capacity, recording which component(s) is missing AND why you are making a best interests decision. A treatable infection seems likely here, and you cannot wait hours for his team to arrive; they have no greater legal or other powers than you. Security would not be your first call; this would be to ignore the least restraint principle. Medicine never sleeps, but lawyers do. Even if psychiatry (illegally) placed him under mental health legal holding powers, these do not allow for physical health treatments to be administered against a patient's wishes. (There is one exception to this: intravenous fluids, supplements, and enteral feeding are given under mental health law in severe emaciation, including anorexia nervosa and severe depression.)

2. c. Fluctuating consciousness, disorientated in time, and fleeting visual hallucinations. Delirium is a clinical diagnosis, not a laboratory or past medical history one. The latter description points to a psychosis diagnosis, and these voices (delusions, too) are stable over many years, unlike delirium. Although most delirium is related to drugs and illegal substances or alcohol, the cause is not the diagnosis. Lots of people get angry in hospital; this does not establish a delirium or mental disorder diagnosis.

3. b. Changing to haloperidol from newly prescribed olanzapine because the hospital has run out. Patients who are new to antipsychotic agents struggle with side-effects, especially older agents such as haloperidol. Abrupt switching here also risks the rare neuroleptic malignant syndrome. Consider alternative sedation. Asking for allergies, drug reactions, and intolerances to previous medications and checking hospital and community records, including pharmacy if available, for

current and past medications, are best practices; there are no shortcuts in safe prescribing. The other two answers are cautious, safe prescribing. We are aware of both discontinuation syndrome and the potential interactions of antidepressants from different classes, notably monoamine oxidase inhibitors. This is an agreed protocol when prescribing lithium for the first time. You will need thyroid function testing, too.

4. b. Difficulties with concentration, episodes of the mind going blank, and indecision. Chest tightness, palpitations, and 'latchkey' urinary urgency are physical manifestations of anxiety: urgency of urine is rare in anxiety, and latchkey usually points to bladder problems. Fear of dying, fearing the worst, and a conviction that something bad is going to happen are affective symptoms. Hypervigilance, performance anxiety, and avoiding situations that provoke(d) anxiety are behavioural symptoms. Low mood and anger or irritability are affective symptoms, and flashbacks are a cognitive manifestation of anxiety.

Chapter 31: Ophthalmology

1. b. Diabetic macular oedema
2. e. cataract
3. d. keratitis.
4. a. central retinal artery occlusion.
5. d. Central retinal artery occlusion

Chapter 32: Rheumatology

1. e. Rheumatoid arthritis. Although all of these conditions cause joint pain, only psoriatic arthritis, axial spondylarthritis, and rheumatoid arthritis cause this inflammatory pattern of pain and stiffness. Axial spondylarthritis more typically affects the spine, and there is no reported history or family history of psoriasis. Osteoarthritis and gout are not typical in someone of this age and sex and are less likely to present with symmetrical symptoms. Therefore, rheumatoid arthritis is the most likely diagnosis.

2. b. Marginal osteophytes. The correct answer is marginal osteophytes because this person has mechanical knee pain that is most likely to be caused by osteoarthritis in someone of this age with no other reported joint pain. Other radiologic signs of osteoarthritis include loss rather than preservation of joint space. Marginal erosions are more typically found in rheumatoid arthritis, and syndesmophytes are inflammatory spinal lesions found in axial spondylarthritis and periarticular bone resorption and characteristic of psoriatic arthritis.

3. a. Adalimumab. The correct answer is adalimumab because it is a tumor necrosis factor-α–blocking agent and biologic disease-modifying anti-rheumatic drugs (bDMARDs) should be considered to treat patients with persistently active axial spondylarthritis when two different non-steroidal anti-inflammatory drugs (NSAIDs) have failed. Although etoricoxib (a cyclo-oxygenase-2 selective NSAID) and hydrotherapy may also be recommended, the patient must first be offered bDMARD therapy. Methotrexate and prednisolone are not helpful for managing spinal arthritis.

4. b. Fibromyalgia. The correct answer is fibromyalgia because of the history of chronic widespread pain with unrefreshing sleep and normal investigations. Polymyalgia rheumatica is almost never seen before the age of 50 years and usually has a short history and raised inflammatory markers. In osteoarthritis and rheumatoid arthritis, the pain is usually more localized to joints, and tender spots over the body are not usually a feature. Polymyositis typically presents with weakness rather than pain and with abnormal blood test results such as raised creatine kinase.

5. e. Systemic lupus erythematosus. The correct answer is systemic lupus erythematosus (SLE) because although all the other conditions listed can cause multiorgan effects, only SLE fits all the clinical information given. The patient has an acute kidney problem requiring urgent treatment and is likely to need a renal biopsy. The combination of oedema, proteinuria, and hypoalbuminaemia constitutes nephrotic syndrome, which is very unlikely to occur in systemic sclerosis, rheumatoid arthritis, or Sjögren's syndrome but can be a presenting feature of SLE. Granulomatosis with polyangiitis frequently involves the kidneys but does not cause nephrotic syndrome. The butterfly rash over the face and recurrent mouth ulcers are typical features of SLE, which can also cause both pleural and pericardial inflammation. Antibodies to double-stranded DNA are highly specific for SLE.

Chapter 33: Bone disease

1. e. All of the above.
2. b. fracture.
3. e. a, b, or c.
4. c. Teriparatide
5. d. Continue denosumab

Chapter 34: Endocrinology

1. d. Magnetic resonance imaging (MRI) of the pituitary. MRI is preferable to CT for imaging of the pituitary fossa. Lateral skull X-ray is now considered old-fashioned.
2. d. Thyrotoxicosis
3. e. It is commonly seen in patients with multiple endocrine neoplasia type 1 (MEN 1). Hyperparathyroidism is seen in most patients with MEN 1. Histology usually shows an adenoma rather than hyperplasia.
4. e. Phaeochromocytoma
5. b. Cushing's syndrome

Chapter 35: Diabetes mellitus

1. c. Islet autoantibodies
2. d. SGLT-2 inhibitor
3. b. Pancreatic
4. a. Pump therapy reduces HbA_{1c} in those with elevated HbA_{1c}.
5. e. All of the above

Chapter 36: Metabolic disease

1. d. A diagnosis of inherited metabolic disease is very unlikely in patients presenting with symptoms for the first time in adulthood.
2. e. all of the above.
3. e. Porphyria cutanea tarda
4. c. Females are not clinically affected.
5. e. Hyperammonemia

Chapter 37: Obesity

1. a. Hypothyroidism. Her symptoms of weight gain and constipation along with decreased heart rate and low body temperature should prompt you to consider hypothyroidism as a possible cause. Screening thyroid function tests are worth thinking about when seeing such a patient in the obesity clinic. Hypothyroidism does not typically cause marked weight gain but may be a concern for some individuals.

2. a. Monogenic obesity. Monogenic obesity is a rare cause of obesity in which early-onset obesity and hyperphagia are almost always present.

3. e. All of the above.

4. e. Options b, c, and d. All three drugs cause weight gain. The prolactin is raised because of olanzapine; antipsychotics cause hyperprolactinemia. However, raised prolactin in isolation does not cause weight gain.

5. e. Pseudo-Cushing's syndrome. This is associated with hypercortisolism, but the cortisol is suppressed by the overnight dexamethasone suppression test as opposed to in Cushing's syndrome in which endogenously produced cortisol is not suppressed with dexamethasone. This male patient has a high alcohol intake, which is one of the causes of pseudo-Cushing's syndrome. The clinical features overlap with that of Cushing's syndrome.

Chapter 38: Haematology

1. b. A raised reticulocyte count indicates that the bone marrow is working.

2. e. High LDH concentrations suggest increased red cell destruction.

3. e. HbF concentrations can be naturally elevated in patients with sickle cell disease and thalassaemia.

4. d. A delayed transfusion reaction is unlikely to happen after a patient's first ever transfusion.

5. a. Some types of malabsorption can mimic warfarin treatment on clotting tests.

Chapter 39: Haematological oncology

1. b. Blood film. It is a Friday evening. Most specialist investigations are not available out of hours or at the weekend. However, a blood film can be made and examined at any time. Most laboratory staff are trained to look at blood films; a haematologist is on duty to assist if needed. A WCC this high in someone who is otherwise well is almost certainly an incidental finding of haematological malignancy. A blood film (together with the WCC differential) will differentiate between a chronic (chronic lymphocytic leukaemia [CML] or chronic myeloid leukaemia) or acute leukaemia (acute lymphoblastic leukaemia or acute myeloid leukaemia) in almost all cases and may also determine the subtype. If the blood film shows likely CML, the patient can go home and attend an emergency haematology outpatient appointment after the weekend. If the blood film shows acute leukaemia, then hospital admission is required. The patient will very likely need a bone marrow investigation and immunophenotyping, but they are not instantly available, and results can take some time. Imaging (other than for the wrist fracture) has no role in this scenario. Although infection markers will likely be done, the WCC is too high for an isolated infection. This patient was diagnosed with CML on blood film, confirmed after attending a haematology outpatient clinic after the weekend.

2. e. Request imaging of her neck and refer via the suspected cancer pathway. Underlying malignancy needs to be excluded. Nontender lymphadenopathy is worrying (infective and reactive lymph nodes are often tender or painful). She has persistent classical B-symptoms: fever and weight loss. She should be referred via an urgent suspected cancer pathway to the local hospital. This action should be explained so she is aware to attend for further tests. This patient was diagnosed with classical Hodgkin lymphoma on excision biopsy.

3. e. You give her an urgent appointment for face-to-face review and blood tests to take place the next day. You ask her to come in the morning so that her results are available the same day. There are numerous red flags in this scenario:

- Multiple contacts, including out-of-hours contact for persistent symptoms in someone who has rarely sought help in the past.
- Her pain not being in her lower back but higher up is unusual.
- Her mention of her cognition and new constipation are worrying.

This patient had undiagnosed multiple myeloma with anaemia, spinal lesions (causing her back pain), hypercalcaemia (explaining her cognition and constipation), and acute kidney injury.

4. c. You telephone the patient and ask her to come back. You explain that you are worried about a deep vein thrombosis (DVT), and you need to examine her and repeat her blood tests. Clinical assessment is needed. A DVT needs to be excluded with a Doppler scan in a hospital. Her platelet count is significantly raised, and although this could be reactive, it needs to be confirmed and investigated. This patient had an extensive left leg DVT. She was also diagnosed with essential thrombocythaemia (ET), a myeloproliferative neoplasm. ET increases the risk of venous thromboembolism and is often incidentally diagnosed.

5. e. You go to the laboratory to look at the blood film yourself. You contact the patient with the result. This full blood count result is worrying. An analyzer 'flag' regarding a white blood cell count (WCC) differential means that the machine cannot confidently recognize different white blood cells. This happens when the morphology of the cells is abnormal (e.g., leukaemia cells or blasts). A patient with a high WCC without a differential, anaemia, and significant thrombocytopaenia has acute leukaemia until proven otherwise. This is a medical emergency, and responsibility cannot be diverted to someone else. Direct contact with the patient ensures that appropriate urgent assessment occurs. This patient's blood film confirmed she had acute myeloid leukaemia. She was sent to the emergency department, where the haematology team were expecting her.

Chapter 40: Venous thromboembolic disease

1. d. Rivaroxaban and andexanet alfa. Andexanet is licensed for rivaroxaban and apixaban reversal and approved by the National Institute for Health and Care Excellence for use in life-threatening gastrointestinal bleeding.

2. e. Thrombosis in people with antiphospholipid syndrome can be venous, arterial, or microvascular. Heterozygous factor V Leiden is present in around 5% of populations of White European origin. Its discovery changed the concept of thrombophilia from rare disorders with high clinical penetrance to more common disorders with lower clinical penetrance. Antithrombin deficiency is rare. Thrombophilia testing not indicated when venous thromboembolism occurs after major surgery. Most common thrombophilias do not influence duration of anticoagulation.

3. d. Superficial femoral vein thrombosis requires treatment as a DVT. The first step in the diagnostic algorithm for venous thromboembolism is clinical probability score (Wells). Normal

D-dimer excludes DVT and PE only if the Wells score is 'unlikely'. A single negative proximal US excludes DVT if Wells 'unlikely' and D-dimer is positive, but if Wells is 'likely', the scan should be repeated after 1 week.

4. a. The annual incidence of VTE in the population is about 1 per 1000 people. The annual incidence of VTE increases with age. About 40% of VTE are unprovoked. 'Routine' computed tomography scanning of chest, abdomen, and pelvis is not clinically useful. A long-haul flight increases the risk of VTE by two- to threefold. Major surgery increases risk by more than 50-fold.

Chapter 41: Malignant disease

1. c. Metastatic spinal cord compression
2. e. CEA
3. a. Lambert-Eaton syndrome
4. c. *BRCA* mutation
5. c. Start intravenous antibiotics and fluids without waiting for blood results

Chapter 42: Infectious disease

1. a. Send the patient to the nearest emergency department for urgent malaria rapid diagnostic test and blood film. Fever on return from sub-Saharan Africa is malaria until proven otherwise, and patients can deteriorate rapidly without treatment. The turnaround time for malaria diagnostics outside of hospital settings is unlikely to be rapid enough, and assessment in a hospital setting (if the malaria film result is negative) is important to exclude high-consequence pathogens such as Lassa fever and bacterial sepsis (with blood cultures). Symptoms such as cough and diarrhoea can be 'false-localizing symptoms' that should not detract from the need to exclude malaria and other causes of systemic sepsis.

2. d. Ceftriaxone and amoxicillin. This patient appears to have bacterial meningitis, as evidenced by features of meningitis and a neutrophilia with a high CRP. He also may have features of concurrent encephalitis. The most common cause of bacterial meningitis in this age group is *Streptococcal pneumoniae*, although it is important also to cover *Neisseria meningitidis* and *Haemophilus influenzae*; all respond to ceftriaxone. He has diabetes, which also puts him at risk of *Listeria monocytogenes* meningoencephalitis, which does not respond to ceftriaxone, and hence national guidance recommends adding amoxicillin in those in high-risk groups including those with diabetes. Aciclovir is the recommended treatment for patients with herpes simplex encephalitis, which is not associated with features of meningis or neutrophilia and high CRP. Although herpes simplex is a cause of viral meningitis, there is no evidence of benefit of aciclovir in what is usually a self-limiting condition.

Chapter 43: Sexually transmitted infections

1. a. Bacterial vaginosis. Bacterial vaginosis is the most common cause of vaginal discharge in females of reproductive age. It typically presents with an unpleasant-smelling vaginal discharge but no vulval irritation.

2. b. Chlamydia, gonorrhoea, syphilis, and HIV. The minimum investigations are tests for chlamydia, gonorrhoea, syphilis, and HIV. Hepatitis B and C testing should be performed in those at increased risk, but this person had no risk factors for hepatitis.

3. d. About 1 000 000. WHO estimated in 2020 that more than 1 million people acquired an STI every day. Each year, there are an estimated 374 million new infections in people aged 15 to 49 years with one of the four major curable STIs (chlamydia, gonorrhoea, syphilis, and trichomoniasis) worldwide.

4. a. *Chlamydia trachomatis*. The diagnosis is non-gonococcal urethritis (NGU) because no bacteria are seen. There are several causes of NGU, and the most common is *C. trachomatis*. Other causes include *T. vaginalis*, herpes simplex virus, and urinary tract infections, but these are less common.

5. b. Ceftriaxone. Resistance to all these antibiotics has been reported in the UK, but ceftriaxone resistance is very rare. When antimicrobial susceptibility is not known before treatment, single-dose ceftriaxone 1 g intramuscularly is the recommended treatment.

Chapter 44: Human immunodeficiency virus

1. d. Non-Hodgkin's lymphoma
2. d. Acute HIV infection
3. d. Below 350 cells/mm^3 of blood
4. d. Warfarin
5. d. Perimenopause

Chapter 45: Dermatology

1. c. Patients with psoriasis are 2.0 to 2.5 times more likely to consider suicide than patients who do not have psoriasis.

2. d. Moderately potent topical steroid applied daily for 2 to 4 weeks

3. b. Blisters form within the epidermis and present as flaccid erosions with mucosal involvement.

4. d. Mycosis fungoides is the most common form of primary cutaneous lymphoma.

Chapter 46: Young adult medicine

1. c. You establish a rapport with the young person. You judge her to be competent and prescribe the oral contraceptive. You try to persuade her to inform her parents.

2. d. A father only has parental responsibility if he is married to the mother or named as the father on the child's birth certificate.

3. b. A period of preparation, ideally starting in the early teenage years, is an important determinant of improved long-term outcomes.

4. b. It is illegal to sell any vaping product, whether it contains nicotine or not, to those younger than 18 years of age.

5. a. Alcohol use is associated with unplanned pregnancy.

Chapter 47: Sex and gender in medicine

1. d. Intrauterine copper device
2. c. Pelvic ultrasound scan
3. c. painless inability to void.
4. b. It is a time-critical surgical emergency.
5. b. requires ongoing assessment of cardiovascular risk.

Chapter 48: Obstetric medicine

1. d. Chronic hypertension. Chronic hypertension increases the risk of superimposed pre-eclampsia sevenfold. The other medical conditions are not associated with an increased risk of pre-eclampsia.

2. a. Metformin. Metformin is used to treat both pre-existing and gestational diabetes. There are good safety data to support its use in pregnancy. The other medications either lack adequate

safety data or have been proven to be harmful in pregnancy and should therefore be avoided.

3. c. Respiratory rate. Heart rate and cardiac output increase in pregnancy. Blood pressure falls in pregnancy. Renal blood flow increases in pregnancy, leading to an increase in glomerular filtration (such that estimated glomerular filtration rate equations cannot be used in pregnancy). Respiratory rate does not change in pregnancy; an increase in respiratory rate requires clinical assessment.

4. d. Sodium valproate. Sodium valproate carries at least a 10% risk of congenital malformation and is therefore avoided in pregnancy. Carbamazepine is associated with a 4% to 5% risk of malformation and phenytoin with a 6% risk. Lamotrigine and levetiracetam are not thought to increase the risk of malformations above the background rate in the population.

5. d. Lupus anticoagulant antibody. Lupus disease activity at the time of conception or in pregnancy adversely affects pregnancy outcome. High concentrations of double-stranded DNA may indicate disease activity. Anti-Ro/SSA is associated with a 2% risk of congenital heart block and a 5% risk of neonatal cutaneous lupus. Lupus anticoagulant positivity is a marker for thrombosis risk and adverse pregnancy outcome. Protein:creatinine ratio is used to quantify proteinuria in pregnancy. Pre-existing proteinuria may be caused by lupus nephritis, which increases the risk of adverse pregnancy outcome. New proteinuria in pregnancy can be a sign of pre-eclampsia or new lupus nephritis. Alkaline phosphatase is secreted by the placenta, and serum levels are increased in normal pregnancy.

Chapter 49: Medicine of the older adult

1. d. Multifactorial fall. The terms ‘mechanical fall’ and ‘simple fall’ are generally unhelpful, not least because doctors often stop looking for falls risk factors, some of which may be reversible or modifiable. This patient’s risk factors for falling include her medications (the tramadol could make her drowsy or confused; the losartan can cause the blood pressure to drop); her arthritis, which may be painful if she tries bending her knee, so she is less likely to lift her foot over the curb as well, and she may be suitable for a knee replacement; her diabetes, which could cause falls through peripheral or autonomic neuropathy or retinopathy. Patients with diabetes being treated with insulin or gliclazide may also become hypoglycaemic, though metformin would not cause this. Her atrial fibrillation can be associated with syncopal falls if too fast or too slow, but she denies losing consciousness (not always reliable), and the graze on her hand suggests that she reached out to break her fall, which wouldn’t have been seen if she lost consciousness.

2. b. Inappropriate dosing. Always be suspicious of any new medication being started. Buprenorphine is a commonly used opiate. Despite normal-looking kidney function, her creatinine clearance is just 22 mL/min (reference range, >60), reflecting her age and weight. The reduced body fat won’t help but is not the key issue here. Gallstones when causing problems are more likely to increase the liver function test results more than is given here, even if patients don’t have the abdominal classical symptoms. Her thyroid function is only marginally abnormal and would not cause these problems; it may be worth repeating them in 6 months or so.

3. e. Urge incontinence. Her history is typical for urge incontinence. Functional incontinence is when a patient is unable to physically get to the toilet because of reduced mobility. Stress incontinence is associated with briefly raised intra-abdominal pressures, such as with coughing; there is nothing to suggest that here. Mixed incontinence is a combination of stress and urge incontinence. A bladder would be expected to empty to a residual volume of less than 180 mL; her result is normal.

4. a. Comprehensive geriatric assessment. He needs assessment across a range of areas, from his medical issues (breathlessness, mood, cognition) to his mobility (physiotherapy) to his coping within his own environment with appropriate support (social worker and occupational therapist.)

5. b. Financial abuse. Although the main problem is with financial abuse, this case typifies how there is so much overlap with the different types of abuse. There is self-neglect (though this was not a problem until the care package stopped), and there is emotional abuse (withdrawal of interaction). Arguably, from an organizational perspective, social services closed the case too early; an initial visit was appropriate, but further follow-up for this vulnerable patient may also have been appropriate.

Chapter 50: Palliative care and symptom control

1. e. All of the above

2. d. Refer to community palliative care team to consider non-pharmacological interventions (e.g., a handheld fan and breathing exercises).

3. c. should be considered proactively at every healthcare interaction.

4. e. Develop a personalized care plan, in discussion with his family, to support individualized decisions about hydration and nutrition, spiritual care, and symptom control.

5. e. Refer to the hospital palliative care team for assessment of other pain dimensions and consideration of adjuvant analgesics and psychological support.

Index

Note: Page numbers followed by "f" indicate figures; "t" indicate tables; and "b" indicate boxes.
Medical Licensing Assessment (MLA) conditions have been highlighted in green to make them easier to find.

A

AACG. *See* Acute angle-closure glaucoma
Abdomen
  CT of, 131–132
  MRCP of, 138
Abdominal aortic aneurysm, 492
Abdominal IR procedures, 149–151
Abdominal pain, 612–613
  pancreatitis, 755
  STIs and, 1524b
Abdominal paracentesis, 190b, 191f
Abdominal radiography, 122, 123f
Abdominal swelling, 226–227, 226b
Abducens nerve (VI), 896–897
  lesions, 897
Abetalipoproteinaemia, 651
Abnormal chromosome numbers, 63–64
ABO system, 1308, 1308b, 1309f
Absence (petit mal) seizures, 922
Absolute risk reduction (ARR), 46
Absorption, 109–110
  in kidney disease, 828–829
Abuse, in older adults, 1717–1718, 1718b
Acalculous cholecystitis, 743
Acanthosis nigricans, 1605, 1605f
Accelerated automaticity, 412
Accessory nerve (XI), 902
  lesions, 902, 902b
Accidents, 169–170, 170b
Acclimatization
  altitude, 374, 374b
  heat, 379
ACE. *See* Angiotensin-converting enzyme
Acetylcysteine, paracetamol overdose, 359, 359b, 360f
Achalasia, 625, 628f
Achilles bursitis, 1077
Achilles tendonosis, 1077
Achondroplasia, 1082–1083
Aciclovir, 1465, 1466f
Acid-base balance, disturbances of, critical illness, 307–308, 308b
Acid-base disorders, 877–884, 877b
  aetiology of, 879
  assessment, 855, 855f
  diagnosis of, 879, 879b, 880f
Acid-base physiology, normal, 877–879
Acid maltase/acid alpha glucosidase deficiency, 973
Acid-citrate lyase inhibitors, 519
Acidosis, 1244
Acinus, of liver, 682, 683f
Acne, 1574, 1575f, 1576b
Acne conglobata, 1576b
Acne excoriée, 1576b
Acne fulminans, 1576b
Acne inversa, 1576
Acne vulgaris, 1570f
Acoustic neuroma, 946
ACP. *See* Advance care planning
ACPAs. *See* Anti-citrullinated peptide antibodies
Acquired cardiomyopathies, 486–487
Acquired haemolytic anaemia, 1301–1306
Acquired immunodeficiency virus, 1525–1536
Acquired melanocytic naevi (moles), 1614, 1615b
Acquired thrombophilia, 1368
Acromegaly, 1147b, 1185, 1186f–1187f
ACS. *See* Acute coronary syndromes
ACTH, 1145b, 1152b
ACTH stimulation test, 1153
Actinic (solar) keratoses, 1616–1617, 1617f, 1617b
Actinomycosis, 1492
Activated partial thromboplastin time (APTT), 1320, 1324b
Activated protein C, 1318, 1319f
Active immunization, for measles, 1449
Acute abdomen, 674–678
  causes of, 674b–675b, 676
  diagnosis of, 674–677
  history of, 674
  investigations of, 675, 675f–676f
  pain in, 674–675
  peritonitis, 676
  physical examination of, 674–675
  vomiting, 674
Acute angle-closure glaucoma, 1057, 1058f
Acute aortic dissection, management of, 509b
Acute aortic syndromes, 493–494, 493f
Acute bacterial prostatitis, 1663
Acute bronchitis, 545
Acute chest syndrome, 1296
Acute cholangitis, 741
Acute cholecystitis, 739, 739f
Acute colonic pseudo-obstruction, 677
Acute coronary syndromes (ACS), 444–452
  chest pain, 393
  clinical features of, 444
  diagnosis, 444
  investigations, 444–445
  management of, 445–448, 447b, 509b
  NSTEMI and unstable angina, 445
  pathophysiology of, 444
  risk stratification of, 445, 446b
Acute cortical necrosis, 836–837
Acute diarrhoea, 614–615, 615b
Acute disseminated encephalomyelitis, 940
Acute disturbance (AD), 1009b, 1014
Acute diverticulitis, 666–667, 668f
Acute fatty liver of pregnancy (AFLP), 1685
Acute febrile neutrophilic dermatosis, 1603, 1604f
Acute frailty services, 182
Acute generalized exanthematous pustulosis, 277
Acute glaucoma, 1047b
Acute hepatic failure, 704–705, 704b–705b
Acute internal medicine (AIM), 175–277
  clinical presentations in, 192–277
  clinical skills in, 179–182
  conditions of, risk assessment tools for, 180, 181b
  core skills and knowledge in, 175b
  model of, 176f
  patient flow through, 176–178, 177b
  procedures in, 182–190, 182b
  role of, 178
Acute ischaemic stroke, 983
Acute kidney injury (AKI), 228–231, 228b, 327–328, 835–837
  causes of, 228, 228b, 229f
  classification of, 228b
  clinical assessment of, 228–229, 229b
  definition of, 228
  intrarenal parenchymal, 230–231
  investigations of, 229–230
  kidney replacement therapy for, 837
  management of, 230, 231b
  measures for, 230
  postrenal, 230
  pregnancy, 1692
  pre-renal, 230
  staging of, 228
  surgery and, 348
Acute lower gastrointestinal bleeding, 641, 641f–642f
Acute lower limb ischaemia, 491
Acute lymphoblastic leukaemia (ALL), 1342, 1342b
Acute mountain sickness, 375
Acute myeloid leukaemia (AML), 1339f, 1340, 1341b
Acute nephritis, 809
Acute-on-chronic liver failure, 710
Acute otitis media, 586–587, 587f
Acute painful neuropathy, 1230
Acute pancreatitis, 749
  alcohol-related, 749
  causes, 749b
  clinical features, 749–750
  complications, 752–753, 753b
  diagnosis, 750–751
  disease severity, 750–751, 751b
  gallstone-related, 749, 752
  management, 751–752
  pathogenesis, 749
  prognosis, 753
Acute pericarditis, 487–490, 487b, 488f
Acute peritonitis, 676
Acute phosphate nephropathy, 836
Acute promyelocytic leukaemia (APML), 1341–1342
Acute pulmonary oedema, management of, 509b
Acute pyelonephritis, 831, 831f
Acute rejection, in kidney transplantation, 848, 849f
Acute respiratory distress syndrome, 325–327, 326f–327f, 326b
Acute salpingitis, 676
Acute small intestinal ischaemia, 651
Acute stress reaction, 1023
Acute stroke, 978b
Acute tubular necrosis, 828, 835–836, 836f, 847–848
Acute tubulointerstitial nephritis, 828, 833–835, 833b
Adamantanes, 1467
Adaptive immune responses, 82, 82b
Adaptive immune system, 80f, 85–91
  antigen receptors on T and B, 86–87
  respiratory tract, 528–529
Adaptive platform trials (APTs), of evidence-based practice, 39
Addictions, 1030–1032, 1031b
Addison's disease, 1156, 1605
  history and examination of, 1157, 1157f
  investigations, 1157, 1158f
  management, 1157–1158, 1158b
  pathophysiology and aetiology, 1156, 1157b
  patient advice, 1158
Adenocarcinoma
  in gall bladder, 745
  in oesophagus, 630
  in pancreas, 756
  in small intestine, 652
  in stomach, 637, 637f
Adenomyomatosis, of gall bladder, 744
Adenoviral conjunctivitis, 1053, 1055f
Adenovirus, 1451
Adhesive capsulitis, 1070
Adie pupil, 895
Adjustable gastric banding, of bariatric surgery, 1268
Adjustment disorder, 1023–1024
Adjuvant analgesics, 1725–1726, 1726b
Adolescents, diabetes mellitus, 1235–1238
ADPKD. *See* Autosomal dominant polycystic kidney disease
Adrenal gland, 1151–1152, 1152f, 1152b
Adrenaline (epinephrine), 316
Adrenergic stimulation, 390–391
Adult dermatomyositis, 1113, 1113f
Adult Refsum's disease, 1247
Advance care planning (ACP), 1704
Advance decisions, 30, 30b
Advanced cardiac life support, 194–195, 194f
Advanced respiratory support techniques, 183–190
Adverse drug reactions, 114–115, 115b
  classification of, 114–115, 115b
  diagnosis, 115
  hepatotoxicity, 732–733
  management, 115
  older adults, 112b
Aerodigestive tract, upper, malignancy of, 598–600
African histoplasmosis, 1494
African tick typhus, 1490b
African trypanosomiasis, 1502, 1503b
Afterload (outflow resistance), 303, 304f, 430
Ageing, 1695
  applied anatomy and physiology of, 1696, 1696f
  calorie restriction in, 1696
  cell replication theories of, 1695
  cell senescence in, 1696
  economic advantages of, 1719
  evolution theories of, 1695
  male health and, 1669
  societal advantages of, 1719
  well, 1719
Age-related macular degeneration, 1060, 1060f
Agoraphobia, 1022
Agranulocytosis, 1315
Air pollution, as WDH, 158
Air travel, specific considerations for, 373–374
Airflow, 525
Airflow limitation, 545, 546f
Airway clearance, 565
Alberta Stroke Programme Early CT score (ASPECTS), 989f
Albumin, 686, 715b, 1314
Albuminuria, in diabetic nephropathy, 1228
Alcohol
  cancer and, 1381
  myopathy, 971–972
  neuropathy, 967, 967b
  pancreatitis and, 749
  young adults and, 1635
Alcohol dependence, 1030–1031, 1030f, 1031b–1032b
Alcohol-induced hypoglycaemia, 1197
Alcohol liaison service, 182
Alcohol use disorders identification test for consumption (AUDIT-C), 181b
Alcohol withdrawal syndrome, 253, 253b, 1014–1015
Alcoholic hepatitis
  clinical features, 724
  Glasgow score, 725b
  investigations, 725
  Lille score, 725b
  management and prognosis, 725
  pathology, 724
Alcoholic liver disease, 722–726, 723f, 723b–724b
Aldosterone antagonists, 435
Alginate-containing antacids, 623–624
Alkaline phosphatase (ALP), 687
Alkalosis, respiratory, 308
Alkaptonuria (A), 1246.e1b
Alkylating agents, 1389
Allergen provocation tests, 542
Allergic bronchopulmonary aspergillosis, 575–576
Allergic conjunctivitis, 1054
Allergic contact eczema, 1583, 1583f
Allergic disease (immediate hypersensitivity), 101–102, 101b
Allergic reactions
  in anti-retroviral therapy, 1553
  asthma and, 541–542
Allied health professional, early involvement of, patient flow and, 178
Allogeneic stem cell transplant, 1340
Allogeneic stem cell transplantation, chronic lymphocytic leukaemia, 1340
Alloimmune haemolytic anaemia, 1304–1305
Alloimmunization, 1310
Allopurinol, for gout, 1104
Allylamines, 1497
Alopecia areata, 1627, 1627f
ALP. *See* Alkaline phosphatase
Alpha-glucosidase inhibitors, 1221
Alpha$_1$-antitrypsin deficiency, 546
Alpha-thalassaemia, 1294, 1295b, 1296f
Alphaviruses, 1455b
Alport's syndrome, 811

Alternating pulse (pulsus alternans), 395, 395f
Altitude acclimatization, 374, 374b
Alveolar haemorrhage, diffuse, 574
Alveolar oxygen tension, 301–302, 301f
Alveolar stability, 527
Alveoli, 523, 524f
Alzheimer's disease, 954–956, 956f
Amaurosis fugax, 982
Ambulatory care, 177
Ambulatory score (AMB), 180
Ambulatory taped electrocardiography, 399, 402f
Amenorrhoea, 1180–1181, 1181b–1182b
Amikacin
  for infection caused by multidrug-resistant organisms, 296
  for sepsis, 282b, 292
Amino acids, 767–768
Aminoglycosides, for sepsis, 291b, 292, 293f
Aminosalicylates, 660
Aminotransferases, 687
Amiodarone
  hepatotoxicity, 733
  thyrotoxicosis, 1168
Ammonium, 879, 879f
Amnesic shellfish poisoning, 370.e1
Amoebiasis, 1506–1508
Amoebic abscess, 727
Amylin analogues, 1221
Amyloidosis, 651, 806, 967, 1605–1606
  diagnosis of, 807
  management of, 807
  pathophysiology of, 806–807, 807f
  systemic effects, rheumatoid arthritis, 1091
Anaemia, 1157, 1277–1279, 1279f
  of chronic disease, 1282–1283
  chronic kidney disease and, 838–839
  classification of, 1277f
  clinical features of, 1277–1279
  investigations of, 1278–1279
Anaesthesia, 334–352
  airway management and, 345–346
  clinical skills for, 334
  core skills for, 334b
  local, 344–345, 344f
  machine, 346
  preoperative assessment of, 335b
  type of, in surgery, 343
Anal cancer, 1405
Anal fissures, 667–668
Analgesic nephropathy, 834–835
Analgesics, 1081b
Anaphylactic shock, 312
Anaphylaxis, 198
  allergen sources, 198b
  for allergic disease (immediate hypersensitivity), 102
  post-transfusion, 1312
  treatment of, 198, 198b
Anastomotic leak, surgery and, 348
ANCA-positive small-vessel vasculitis, 812–813, 812f
Androgens, 1152
Aneurysm dilation, of myocardium, 450, 451f
Aneurysmal disease, 492–494
Angina, 440–444
  chest pain, 393
  classical, 440
  classification of, 440b
  diagnosis of, 440
  epidemiology of, 440
  examination of, 440–441
  investigations of, 441, 441b
  management of, 441–444, 441f, 442b
Angiogenesis, 1379–1380
Angiography, 146b
  in acute lower gastrointestinal bleeding, 641
  digital subtraction, 490, 491f
  liver disease, 689
Angioplasty, 146b, 148f
Angiotensin II receptor antagonists, 433
Angiotensin-converting enzyme inhibitors, for heart failure, 433
Angiotensin receptor neprilysin inhibitor, for heart failure, 433
Anion gap, 879–880
Ankle pain, 1077
Ankylosing spondylitis, 579, 1074, 1099, 1099b–1100b, 1100f–1101f
Anogenital warts, 1530, 1531f
Anorchia, 1185b
Anorectal abscess, 668
Anorectal disorders, 667–669
Anorexia, palliative care, 1727
Anorexia nervosa, 1036–1039, 1037b
  management, 1037–1039
Anosmia, 591b–592b, 593
Antagonistic pleiotropy theory, ageing and, 1695
Anterior circulation clinical stroke syndromes, 981
Anterior ischaemic optic neuropathy, 894
Anterior knee pain syndrome, 1076
Anterior pituitary, 1146, 1148b
Anterior spinal artery occlusion, 951
Anterior uveitis (iritis), 1047b, 1058, 1058f
Anthracyclines, adverse effects, 1391f
Anthrax, 1485
Anti-citrullinated peptide antibodies (ACPAs), 1089
Anti-GBM glomerulonephritis, 811–812, 811f
Anti-glomerular basement membrane disease, 574
Anti-herpesvirus drugs, 1465–1467
Anti-inflammatory drugs
  for asthma, 544
  for bronchiectasis, 566
  for rhinitis, 538
Anti-influenza drugs, 1467
Anti-ischaemia agents, in acute coronary syndromes, 447–448
Anti-retroviral therapy (ART), 1547–1554, 1549b
  adherence to, 1552
  complications and long-term safety of, 1552–1553
  drug interactions with, 1552
  drug resistance, 1551–1552, 1552b
  failure, 1552
  initial, 1548, 1550b
  monitoring, 1551, 1551b
  post-exposure prophylaxis, 1554
  starting, 1548
  stopping, 1552
  switching, 1552
Anti-seizure medications (ASM), 926–927
Anti-thyroid antibodies, 1165
Anti-thyroid drugs, 1169–1170, 1170b
Anti-TNF agents, for Crohn's disease, 660
Anti-tuberculous chemotherapy, hepatotoxicity of, 732–733
Antiarrhythmic drugs, 426–427, 426b, 427f
Antibiotic chemoprophylaxis, 288b
Antibiotics
  for asthma, 545
  for atopic eczema, 1580
  for bacterial meningitis, 248, 248b
  clinical uses of, 291b
  for COPD, 548
  for Crohn's disease, 660
  therapies, 1493
Antibody deficiencies, 98b, 99–100
Anticholinesterases, 973
Anticoagulants/anticoagulation
  for acute coronary syndromes, 446–447
  loss of function, 1367–1368, 1367f
  oral, 1370f, 1371–1374, 1372b–1373b
  parenteral, 1369–1371, 1370b
  poisoning, 363
  prosthetic heart valves and, 464
  for stroke, 995–997
  venous thrombosis, 1369–1374, 1370f
Anticonvulsants, poisoning, 363, 363b
Anticytokines, 106
Antidepressants, overdose, 361
Antidiabetic drugs, overdose, 363
Antidotes, to poisoning, 358, 358b
Antiemetics, 345, 613b
Antiepileptic drugs, 926b, 927
Antifungal drugs, 1496–1497, 1496b
Antigen presentation, 93–95, 95f
Antigen presenting cells (APC), defects in, 100
Antiglobulin (Coombs') test, 1302f
Anti-hepatitis B drugs, 700, 700b
Anti-hepatitis C drugs, 703
Antihistamines, 538
Antihypertensive agent, 507b
Anti-inflammatory drugs, for asthma, 544
Anti-ischaemia agents, for acute coronary syndromes, 447–448
Antimalarials, overdose, 364
Antimetabolites, 1389
Antimicrobial stewardship, 164, 286–289
  allergy risk, assessment of, 287–288
  antibiotic chemoprophylaxis, 288b
  decision-making at 72 hours, 286–287
  empirical antibiotics, considerations before starting, 286
  OPAT, 287
  UK-based targets, 286
Antimuscarinic bronchodilators, for asthma, 544
Anti-neutrophil cytoplasmic antibody (ANCA), 574
Antinuclear antibodies, 1066, 1066b
Antiphospholipid syndrome, 1110, 1368
Antiplatelet drugs, 995
  for acute coronary syndromes, 445–446
Antipsychotics, 1028b
  overdose, 362
Anti-SARS-CoV-2 drugs, 1467
Antiseptics, for atopic eczema, 1580
Antisynthetase syndrome, 1113
Antithrombin, 1317
α_1–Antitrypsin deficiency, 722
Antitubulin agents, 1389
Anti-venom, 370.e5
Antiviral drugs, 1465–1468, 1466f, 1466b
Antral gastritis, 633
Anxiety disorders, 1021–1022
  aetiology of, 1022
  assessment of, 1024
  epidemiology of, 1022
  treatment of, 1024–1025
  types of, 1022b–1023b
Aortic aneurysm
  abdominal, 492
  rupture, CT angiogram of, 136f
Aortic coarctation, 473, 473f–474f
Aortic dissection
  clinical features of, 393
  CT angiogram of, 136f
  investigations of, 407
Aortic regurgitation, 459f, 460–462, 461f, 461b
Aortic stenosis, 456–458, 458f–459f
Aortic valve
  bicuspid, 458, 458f
  replacement, 460
Aortic valve disease, 456–462
Aortography, 493, 800
APACHE II scoring system, 330, 331b, 752b
Aphasia, 889–890
Apheresis, 519
Aplastic anaemia, 1287–1288, 1287b, 1288f
Apolipoprotein C3 inhibitors, 519
Apoprotein B100, 514
Appendicitis, acute, 675
Apraxia, 891
Aquaporins, 865, 865f
Arabinosides, 1389
Arboviruses, 1455b
Arenaviruses, 1445b
Arginine vasopressin (AVP), effect of, on kidney, 865–866, 865f
Argyll Robertson pupil, 895–896
Arrhythmia. *See* Cardiac arrhythmias
Arrhythmogenic right ventricular cardiomyopathy, 484–485, 485f
Arsenic
  cancer and, 1381
  poisoning, 366
Arterial gas embolism, 381
Arterial thrombosis, 1364–1365
Arterial ulcers, 1609
Arteriovenous malformation
  in brain, 243–244, 244f
  in spinal cord, 951, 951f
Arthropathies, 1119–1120
Arthropod ectoparasites, 1517–1518, 1517f
Articular cartilage, 1064–1065
Artificial intelligence (AI), medicine and, 12–13
Asbestos, 578, 579b
Asbestosis, 578, 579b
Ascariasis, 1511
Ascaris lumbricoides, 1512f
Ascending limb of Henle, 872–873
Ascent, in high altitude, 374b
Ascites, 226–227, 692
  causes, 226b
  clinical features, 715
  investigations, 227, 715–716
  management, 227, 716
  serum-ascites albumin gradient, 715b
Aseptic technique, 183
Aspergilloma, 576
Aspergillosis, 1494
  HIV-associated, 1556
Aspergillus fumigatus, diseases caused by, 575–576, 576f
Aspiration pneumonia, 216
Aspirational ethics, 26–27
Aspirin, 393
  diabetic patients, 1233
  metabolism, 366f
  overdose, 365
Assisted dying, 31
Asteatotic eczema, 1581
Asthma, 540–551
  acute severe, 205–206, 206b
  aetiology, 540–542, 541f
  airway hyper-responsiveness, 540
  atopy and allergy, 541–542
  attacks, 545
  classification, 540
  clinical features, 542
  environmental factors, 541
  genetic factors, 541
  investigations, 542
  management, 543–545
    bronchial thermoplasty, 545
    control of extrinsic factors, 543
    drug treatment, 543–545, 543b, 544f
  non-specific factors, 541–542
  occupational sensitizers, 541, 541b
  pregnancy, 1686
  prevalence, 540
  specialist teams for, 182
Asthmatic bronchopulmonary eosinophilia, 575, 575b
Asthmatic pulmonary eosinophilia, 575–576
Astrovirus, 1445b
Ataxia telangiectasia, 99
Atherosclerotic renovascular disease, 820
Atherosclerotic vascular disease, 1232, 1233b
Atonic seizures, 922
Atopic dermatitis, 1577
Atopic eczema, 1577–1584
  aetiology, 1577–1579
  clinical features, 1578–1579, 1578f
  exacerbating factors, 1578
  investigations, 1579
  management, 1579–1580, 1579b
  secondary infection, 1579, 1579f
Atopy, in asthma, 541–542
Atrial ectopic beats, 423
Atrial fibrillation, 202–203, 420, 421f
  aetiology, 420
  clinical features of, 420
  long-term management of, 421–423
    ablation, 423
    antiarrhythmic medications, 422–423
    rate control, 421–422
    rhythm control, 422
    stroke prevention, 421, 422f, 422b
  management of, 420–423
  pathophysiology, 420
  with pre-excitation, 203
Atrial flutter, 203–204, 421f, 423
Atrial myxomas, 480, 483f
Atrial natriuretic peptide (ANP), 826, 858
Atrial septal defect, 471, 471f–472f
Atrial tachyarrhythmias, 420–423
Atrial tachycardia, 423
Atrioventricular (AV) block, 414–415
  first-degree, 414, 414f
  second-degree, 414, 415f
  third-degree (complete), 414–415, 416f
Atrioventricular (AV) junctional tachycardia, 419–420, 419f
Atrioventricular nodal re-entrant tachycardia (AVNRT), 419
Atrioventricular (AV) node, 388–389
Atrioventricular re-entrant tachycardia (AVRT), 419–420
Atropa belladona, poisoning, 370.e6
Atrophic hypothyroidism, 1165
Atrophic vaginitis, 1649
Attention deficit hyperactivity disorder, 1040
Atypical haemolytic uraemic syndrome, 817–818
Aura, 922
Auscultation
  chest, 530–531
  first heart sound (S1), 396–397
  heart sounds, 396–397
    aortic stenosis, 458–459
    mitral stenosis, 453, 453f
    second, 397
    third and fourth, 397
Austere environments, 376–377
Autism spectrum disorders, 1040b
Autoantibodies
  autoimmune hepatitis, 705–706
  factor VIII, 1327
  serum, 688
Autoimmune cholangitis, 744
Autoimmune chronic pancreatitis, 754
Autoimmune disease, 102–105
  CD4 T lymphocyte activation failure, 102–103, 103b
  common, 104–105
  diabetes mellitus type 1, 1206, 1208f
  endocrine effects, 1143–1144
  mechanisms of tissue damage, 103–104, 104f
  restoring tolerance in, 106
Autoimmune encephalitis (AE), 940–941, 1027
Autoimmune endocrine disease, 1143–1144, 1143b

Autoimmune gastritis, 632
Autoimmune haemolytic anaemia, 1301–1304, 1302f
Autoimmune hepatitis, 705–706, 706f
Autoimmune hypothyroidism, 1165–1166
Autoimmune pancreatitis, 754, 756
Autoimmune polyendocrinopathy, 1198
Autoimmune rheumatic disease (ARD), 651, 1107–1114
Autoimmune thyroid disease, 1173
Autologous transfusion, 1313
Autonomic neuropathy, 968, 1230–1231, 1230b
Autonomy, respect for, 29–32
Autosomal dominant disorders, 66–69
Autosomal dominant inheritance, 74–76
Autosomal dominant polycystic kidney disease (ADPKD), 849–851, 850f, 850b–851b
Autosomal recessive disorders, 66
Autosomal recessive inheritance, 76
Avalanche medicine, 377
Avascular necrosis of femoral head, 1075
Aviation and space medicine, 372–374
Avoidant/restrictive food intake disorder, 1039
AVP. *See* Arginine vasopressin
AVP-resistance, 866, 1192
Axial spondyloarthritis, 1099
Axonal degeneration, 962
Azoles, 1496–1497
Azoospermia, 1179

B

B lymphocytes
  antigen receptors on, 86–87
  deficiency, 98b
B vitamins, 777, 778b
  deficiency, 777
Babesiosis, 1506
Bacillus Calmette-Guérin (BCG) vaccination, 557
Bacillus cereus, 1475
Bacteria
  classification of, 1469b
  colonization and invasion of, 1443
  culture, 1429–1431
  epithelial attachment, 1442–1443
  in urine, 799
Bacterial arthritis, 1106
Bacterial conjunctivitis, 1052
Bacterial gastroenteritis
  clinical syndromes, 1473, 1473b–1474b
  management of, 1476, 1477f
  mucosal adherence, 1472–1473
  mucosal invasion, 1473
  toxin production, 1473
Bacterial infections, 1468–1469
  bone and joint infections, 1481
  cardiovascular system, 1478
  gastrointestinal tract, 1472–1478
  HIV-associated, 1559–1560
  low-/middle-income countries, 1483–1493
  nervous system, 1478–1480
  prostatitis, 832
  respiratory tract, 1470–1472
  skin and soft tissue, 1469–1470, 1469b
  systemic/multisystem, 1481–1483
  transfusion-transmitted, 1312
  urinary tract, 1481
Bacterial meningitis, 246–247, 247f, 1478–1479
Bacterial overgrowth, 647
Bacterial peritonitis, 1478
Bacterial vaginosis, 1529, 1529f
Bacteriuria, 799
  in pregnancy, 832, 1691
Balantidiasis, 1508
Balkan nephropathy, 835
Balloon tamponade, 714
Bannayan-Riley-Ruvalcaba syndrome, 671
Bariatric surgery
  adjustable gastric banding of, 1268
  endoscopic procedures, 1269
  for obesity, 1267–1269
  other surgical techniques, 1268
  Roux-en-Y gastric bypass, 1268, 1269f
  selection of, 1268
  sleeve gastrectomy, 1268
Barium swallow, 125, 125f–126f
Barotrauma, 381, 381b
Barrett's oesophagus, 625, 625f–627f
Barrier methods, contraception, 1653
Bartonella bacilliformis, 1491–1492
Bartonellosis, 1491–1492, 1492b
Bartter's syndrome, 870–871, 873
Basal blood levels, 1144–1145
Basal cell carcinoma, 1618–1619, 1619f
Basal cell papilloma, 1615–1616, 1615f
Basal metabolic rate (BMR), 766, 768b, 1250
Basic calcium phosphate deposition disease, 1105
Basic life support, for cardiac arrest, 192–194, 193b
Basophils, 84, 84b, 1315
Bazett's square root formula, 398, 399b
B-cell lymphoma, 817, 1352–1356, 1352f, 1352b
Beau's lines, 1390f, 1626
Becker's muscular dystrophy, 974
Bed bug infestations, 1600, 1600b
Behçet syndrome, 1603
Behçet's disease, 941, 1117–1118
Bejel (endemic syphilis), 1486
Bell's palsy, 899
Benign bile duct strictures, 744
Benign oesophageal stricture, 629, 629f
Benign paroxysmal positional vertigo, 586f, 590, 900
Benign prostatic enlargement, 1663–1664, 1663f, 1664b
Benign prostatic hypertrophy, 828
Benign recurrent intrahepatic cholestasis, 690
Benign skin tumours, 1614–1615
Benzodiazepines, overdose, 361
Beriberi, 777
Berry aneurysms, 243, 243f
Berylliosis, 578
Best interests
  assisted dying, 31
  lack of capacity, 30–31
  provision or cessation of life-sustaining treatment, 31
$Beta_2$-adrenoceptor agonists, 543
Beta-blockers
  for heart failure, 435
  for Marfan's syndrome, 476
  overdose, 364
  variceal haemorrhage, 715
Beta-lactams, for sepsis, 289–292
Beta-thalassaemia, 1292, 1293b
Bicarbonate, renal reabsorption of, 877–878, 877f
Bicuspid aortic valve, 458, 458f
Bigeminal pulse (pulsus bigeminus), 395, 395f
Bilateral adrenalectomy, 1156
Bilateral weakness/paralysis, 580
Bile, 684–686
Bile acid receptors, 685–686
Bile acids/salts
  malabsorption of, 648
  metabolism, 684–686, 685f
  recirculation, 685f
Bile duct obstruction, malignant, 274
Bile duct strictures, 744, 746f
Bile pigment stones, 737–738
Bi-level positive airway pressure (BiPAP), 186–188, 192b
Bilharzia, 1514–1515
Biliary colic, 738–739
Biliary cysts, 744
Biliary drainage, 151
Biliary system, 683, 736–761
Biliary tract, cancer of, 1405
Biliary tract disorders, 735–761, 735b
Bilirubin, 686
Biochemical markers, of bone formation and resorption, 1127
Bioelectrical impedance analysis, of hip-to-waist ratio, 1253
Biological therapies
  for Crohn's disease, 660–661, 661b
  for psoriasis, 1586
  for rheumatoid arthritis, 1097–1098
Biologics, for myasthenia gravis, 974
Biomarkers, for acute coronary syndromes, 445, 446b
Bipolar affective disorder, 1019–1021, 1021b
Bird fancier's lung, 573b
Birth defects, in epilepsy, 1687
Birth marks, 1628–1629
Bisphosphonates, 1400
  for osteoporosis, 1130b, 1131
  for Paget's disease, 1134
Blackouts, 236–239, 236b–237b, 927–928, 928b
Bladder
  anatomy, 911–912
  functions of, 911–912
  nerve supply in, 913b
  neuropathy, 1230–1231
  overactive, management of, 1657b
Bladder cancer, survival, 1396b
Bladder scanning, 1661f, 1661b
Bladder stones, 823
Blast cells, 1338b
Blastocystis hominis, 1508
Blastomycosis, 1495
Bleeding
  inexplicable, 1321
  surgery and, 347
Bleeding disorders, 1320–1321
Bleeding time, 1321, 1324b
Blepharitis, 1051–1052, 1053f
Blepharospasm, 899, 934
Blinding, of evidence-based practice, 39
Blister cells, 1301f
Blistering skin disease, 1611–1614
Blisters, poisoning, 357
Blood
  haemostasis, 1316–1321
  oxygenation of, 301–303
  results of, abnormal, 227–228, 227b
  in rheumatoid arthritis, 1093
  specimen collection, 1432b
  in urine, 798, 798f
  whole, 1313–1314
Blood-borne infections, prevention of, 1439
Blood disease, rheumatological problems in, 1119
Blood film
  haematological malignancies, 1333, 1334f
  in iron-deficiency anaemia, 1279f, 1282
  leukaemia, 1336f, 1338, 1341f, 1343
Blood gas analysis, for respiratory failure, 318–319
Blood gases
  COPD, 547
  critical illness, 308b
  respiratory failure, 318–319, 535
Blood glucose, 1157
Blood groups, 1308
Blood pressure, 395, 396b
  cardiovascular morbidity and, 497f
  control in special populations, 509
  critical illness, 304, 304f, 304b
  management of, in hospital, 508–509
  mechanisms affecting, 498f
  variable, 502
  on wards, 508
Blood products, 1313–1314
Blood salvage, 1313
Blood tests
  for Crohn's disease, 658
  for gastrointestinal disease, 617–619
  in older adults, 1700b–1701b
  for pneumonia, 213
  for rheumatic disease, 1065
Blood transfusion, 314–315, 1307–1308
  autologous, 1313
  avoidance of unnecessary, 1312–1313
  blood ordering, 1309
  compatibility testing, 1308–1309
  complications of, 1309–1312, 1310b, 1311f
    immunological, 1310–1312
    non-immunological, 1312
  prevention of wrong transfusions, 1310, 1311b
  in upper gastrointestinal bleeding, 225b
Blue naevus, 1615b
Blunt trauma, 1048, 1050f
BMR. *See* Basal metabolic rate
BODE index, 549b
Body dysmorphic disease, 1630
Body mass index, of hip-to-waist ratio, 1253, 1255f, 1257b
Body packers/stuffers, 357
Body weight, energy requirements and, 767
Boils (furuncles), 1590
Bone
  anatomy and physiology of, 1123–1125, 1124f
  biopsy, 1127, 1127f
  growth and remodelling, 1124, 1125f
  matrix components of, 1123–1124
  MRI of, 138, 143f
  scan of, 144, 145f
  structure, 1123–1124
Bone cells, 1124
Bone densitometry, 646
Bone density, 1128f, 1129
Bone disease, 1121–1136
  biochemical markers for, 1127
  biochemical tests for, 1122–1123, 1122b
  chronic kidney disease and, 839–840
  clinical skills for, 1121–1123, 1123f
  core skills and knowledge in, 1121b
  diagnostic imaging, 1127
  examination of, 1121
  history of, 1121–1123
  infections, 1135
  investigation of, 1122b, 1126–1127
  osteonecrosis, 1132–1133
  osteoporosis, 1128–1132
  Paget's disease, 1133–1134
  rickets and osteomalacia, 1134–1135
  sarcoidosis, 570
  Scheuermann's disease, 1136
Bone marrow
  anaemia, 1277b, 1278–1279
  aplasia, 1297
  biopsy, 1274f–1275f, 1277b, 1278–1279
  dysfunction, 839
  myeloma, 1359
  in pernicious anaemia, 1285
  suppression, 1390, 1391f
Bone marrow aspirate, morphology, 1333–1334, 1334f
Bone marrow transplantation, 1294
Bone marrow trephine, 1334
Bone metabolism, in anti-retroviral therapy, 1553
Bone mineral disease, 839–840, 840f
  chronic kidney disease and, 839–840, 840f
Borrelia infection, 1590–1591
Botulism, 1479–1480
Bowel, visualization of, 602–603, 602b
Bowel lengthening procedures, 787
Bowel obstruction, palliative care, 1726–1727
Bowel sounds, of abdomen, 675
Bowen disease (intraepithelial carcinoma), 1617, 1618f
Bradford Hill criteria, 155b
Bradycardia, 413–418
Bradykinesia, 930
Brain
  biopsy, 915, 946
  epilepsy, 926
  imaging, 236
    CT, 131, 132f–133f
    DaTScan, 143, 144f
    MRI, 138, 139f–140f
  vascular anatomy, 981, 982f–983f
Brain abscess, 943f, 944
Brain death, 330–333
Brain infarction/malignant hemispheric cerebral infarction, 991, 994f
Brain metastases, 1412
Brain natriuretic peptide, in heart failure, 430
Brain tumours, 944–946, 944b, 1412–1413, 1412f
  clinical features of, 945–946
  CT of, 133f
  gliomas, 944–946, 945f
  investigations of, 946
  management of, 946
  meningiomas, 945, 945f, 950f
  primary, 1412–1413
  schwannomas, 945
  survival, 1396b
  tuberculoma, 944
Brainstem, 987b
  anatomy, 987f
  compression, 233
  demyelination, 936
  function tests for, 234–235
  infarction, 981, 984b–986b
  lesion, 233
Breast
  aberrations of development and involution, 1641–1642, 1641b
  benign neoplasms and proliferations of, 1642–1643
  congenital abnormalities, 1641
  infection of, 1644–1646, 1645b
  mastalgia, 1643–1644, 1644b
Breast cancer, 1399–1402, 1399f, 1640–1641
  advanced, 1400–1402, 1401f
  early, 1399–1400
  familial, 1411b
  genetics, 75–76
  histology, 1396b
  HRT and, 1656b
  survival, 1396b
Breast cysts, palpable, 1641–1642, 1642b
Breast disease, 1640, 1640b
Breast-feeding, direct oral anticoagulants for, 1373
Breathing, 525
Breathlessness, 204–218, 204b, 1727, 1728f, 1729b–1730b, 1730f
Brevetoxin poisoning, 370.e1
Brincidofovir, 1467
Broad-based gait, 908
Broca's aphasia, 890
Bromocriptine, 1221
Bronchi, 523, 523f
Bronchial adenoma, 564
Bronchial carcinoid tumours, 563
Bronchial carcinoma, 559, 599
  assessing fitness for treatment, 560, 563b
  clinical features, 559–560, 559b, 563b
  histology and cytology, 560, 563b
  investigations, 560
  large-cell, 559b
  non-metastatic extrapulmonary manifestations of, 560
  non-small-cell, 559b
  pathophysiology, 559, 559b
  screening, 563
  secondary tumours, 563

Bronchial carcinoma *(Continued)*
small-cell, 559b
solitary pulmonary nodules, 563
staging and diagnosis, 560, 561b
Bronchial provocation testing, 537
Bronchial thermoplasty, 545
Bronchiectasis, 565–567
aetiology, 565–567, 565b
clinical features, 565b
investigations, 565, 566f
management, 565–566
prognosis, 566
Bronchioles, 523
Bronchiolitis, obliterative, 575
Bronchodilators
antimuscarinic, 544
for COPD, 547
Bronchomotor tone, 525
Bronchoscopy, fibreoptic, 563b, 564f
Brucellosis, 1106, 1481
Brugada's syndrome, 423, 424f
Brugia malayi, 1509b
Brugia timori, 1509b
Bubonic plague, 1489
Budd-Chiari syndrome, 726
Buerger's disease, 494
Bulbar palsy, 902
Bulimia nervosa, 1039, 1039b
Bulla, 1568f–1570f
Bullous pemphigoid, 1612–1613, 1613f
Bullous skin disease, 1611–1614, 1612f
Bullying, young adult mental health and, 1633
Bundle branch block
bifascicular block, 416
complete block, 415, 416b
conduction delay, 415
hemiblock, 416
left, 415, 417f
right, 415, 417f
trifascicular, 416
Bunyavirus, 1445b
Bupropion, for obesity, 1266
Burkitt lymphoma, 1352b, 1354
Bursitis
Achilles, 1077
calcaneal, 1078
calcific, 1069–1070
causes and clinical features of, 256b–257b
pre- and infrapatellar, 1076
trochanteric, 1075
Buruli ulcer, 1485–1486
Byssinosis, 578

C

C1q nephropathy, 814
Café-au-lait macules, 1626
Calcaneal bursitis, 1078
Calcific tendonosis and bursitis, 1069–1070
Calcification
cornea, 1194
coronary artery, 406
in kidney tract, 800f
Calciphylaxis, 841, 841f
Calcitonin, 1125
for osteoporosis, 1132
Calcitriol, 1132
Calcium, 777, 1130
absorption, 1124–1125
distribution, 1124–1125
homeostasis, 1124, 1125f
total plasma, 1126
urinary, 1127
Calcium-channel blockers, overdose, 364
Calcium metabolism disorders, 971, 1193–1196
Calcium phosphate deposition disease, 1105
Calcium pyrophosphate dihydrate deposition arthropathy (CPPD), 1105
Calf pain, 1077
Caliciviruses, 1445b, 1462
Campbell de Morgan spots, 1616
Campylobacter jejuni, 1473
Cancer, 1378f
aetiology of, 1380–1382, 1380b–1381b
biology of, 1378–1380, 1379f–1380f
Cancer *(Continued)*
biopsy and histology of, 1385–1386
in diabetic patients, 1233
epidemiology of, 1382
familial cancer syndromes, 1383b
genetic basis of, 74–78
genetic diagnostics for, 77
genetic factors of, 1380
geographical distribution, 1382, 1382f
haematological oncology, 1328–1362, 1328b, 1329f
imaging of, 1383–1385
immunosuppression, 1383, 1385f
inherited syndromes, 74–76
low-molecular-weight heparin for, 270
metastatic cancer of unknown primary, 1415–1416, 1415f
neuropathy in, 968
palliative care, 1386
paraneoplastic syndromes, 1383, 1385f
patients living with, 272–274
performance status, 1386, 1387b
pre-implantation genetic diagnosis, 73–74, 75f
prevention of, 1381b
screening of, 1382–1386, 1383f, 1383b
secondary malignancies, 1392
staging of, 564f, 1386
surgery and, 350–351
survival, 1396b
symptoms and signs of, 1384b
targeted treatments for, 77
treatment of, 1386–1396
adjuvant therapy, 1386
aims of, 1386
assessment before, 1386, 1387b
assessment of benefits, 1386–1387, 1387f, 1388b
biological therapies, 1392–1393
curative, 1386
endocrine therapy, 1392
gene therapy, 1393
neoadjuvant therapy, 1386
radiotherapy, 1394–1396
for solid tumours, 1386, 1396b
targeted therapies, 1392–1393
venous thrombosis and, 1366
Cancer pain, 1724, 1724f, 1726b
Candida auris, 1494
Candida infections, 1597, 1597f
Candidiasis, 1493–1495, 1529, 1530f
HIV-associated, 1556
Cannabis, poisoning, 363
Cannulae, in dialysis, 843f
Cannulation, radial artery, 304b
Capacity, 29
lack of, 30–31
Capillary blood glucose, self-monitoring of, 1222–1223
Capnography, 319
Carbamate insecticide poisoning, 366
Carbapenem-resistant Enterobacteriaceae, 296
Carbapenem-resistant *Pseudomonas*, 295
Carbapenems, for sepsis, 289–292, 291b
Carbohydrate metabolism disorders, 1246, 1246.e3b
Carbohydrates, 769
dietary fibre, 769
glycaemic index, 1214
healthy diet, 769
metabolism, 683, 684f
Carbon monoxide poisoning, 366
Carbonic anhydrase inhibitors, 862
Carboxyhaemoglobin, 356b, 366
Carbuncles, 1590
renal, 832–833
Carcinoid tumours, 652
Cardiac arrest, 192–198, 193f
Cardiac arrhythmias, 411–437
heart block, 414–418
mechanism of, 398f, 412–413
palpitations, 394
post-STEMI, 451
sinus node function, 400f, 412–413
Cardiac catheterization, 409
for aortic regurgitation, 461
for aortic stenosis, 459
for mitral regurgitation, 456
for mitral stenosis, 455
Cardiac cycle, 390–391, 391f
Cardiac glycosides, 436
Cardiac output, 303, 303f, 307b
Cardiac resynchronization therapy, 410, 410f, 436, 436f
Cardiac syncope, 394
Cardiac transplantation, 437, 437b
Cardiac tumours, 480
Cardiac vectors, 398, 401f
Cardioembolism, 979–981, 980f
Cardiogenic shock, 311
Cardiology, 384–495, 384b–385b
ECG interpretation in, 386–387, 386b
examination of, 386, 387f
history of, 385–387, 385b
Cardiomyopathy, 481–487, 482f
Cardiopulmonary resuscitation (CPR), 180, 193f
Cardiorespiratory arrest, 1704
Cardiorespiratory physiology, 300–307
Cardiotoxicity, chemotherapy, 1391, 1391f
Cardiovascular disease
autonomic neuropathy, 1230
chronic kidney disease, 841–842
obesity and, 1257–1258
in rheumatoid arthritis, 1093
SLE, 1109
Cardiovascular examination, 394–397
arterial pulse, 394–395
auscultation, 396–397, 397f
blood pressure, 395, 396b
blood tests, 397
carotid sinus massage, 400, 403b
chest radiograph, 397
CT, 406
echocardiography, 400–404
electrocardiography, 397–399, 398f
general, 394
jugular venous pressure, 395–396, 396b
MRI, 406–407, 408f
nuclear imaging, 406–407
PET, 407–408
precordium, 396–397
tilt testing, 400, 403f
Cardiovascular syphilis, 494
Cardiovascular therapeutic procedures, 409–411
Carnitine palmitoyltransferase 2 (CPT2) deficiency, 1246.e4b
Caroli's disease, 727
Carotid artery
aneurysms, 1149
dissection, 981
Carotid sinus massage, 400, 403b
Carpal tunnel syndrome, 964, 1071
Carpometacarpal osteoarthritis, 1071
Carrion's disease, 1491–1492
Case-control study, of evidence-based practice, 40
Case series or case reports, of evidence-based practice, 40
Cataplexy, 928
Cataracts, 1055, 1056f, 1056b
Catatonia, 1017, 1017b
Catheter
ablation, 427–428
urinary infections and, 832
Catheterization
male, 1661f, 1661b
in peritoneal dialysis, 846f
Cat-scratch disease, 1470
Cauda equina syndrome, 273
CD classification, 81b
CD4 T lymphocyte functions, 88–90, 89b
Cell culture models, 72
Cell cycle, 65f, 1388f
Cell lysis, 1443
Cell migration, 91, 92f
Cell senescence, 1696
β-cells, of pancreas, 1204f–1205f
Cellular genetics, review of, 61–62
Cellular signalling, 390–391
Cellulitis, 256b–257b, 275–276, 275f, 1589, 1645f
Central nervous system syndrome, 382
Central retinal artery occlusion, 1059, 1059f
Central retinal vein occlusion, 1059, 1059f
Central venous cannulation (CVC), 184f, 184b
Cephalosporins, for sepsis, 289, 291b, 292f
Cerebellar ataxia, 908
Cerebellum, 905–906
lesions, 905–906, 906b
Cerebral abscess, 1479
Cerebral amyloid angiopathy (CAA), 998
Cerebral blood flow, 329f
Cerebral-evoked potentials, 914–915
Cerebral infarction, 991, 994f
Cerebral oedema, high-altitude, 375
Cerebral palsy, 924, 960
Cerebral venous sinus thrombosis, 998
Cerebrospinal fluid (CSF)
low-CSF-volume (low-pressure) headache, 920
meningitis, 941b
sterile pleocytosis, 941, 941b
Cervical and lumbar degeneration, 969
Cervical cancer, 1410, 1410f, 1659b
HIV-associated, 1561
HPV and, 1462
localized, 1410
management of, 1410
metastatic, 1410
MRI of, 141f
prognosis of, 1410
survival, 1396b
Cervical spine, rheumatoid arthritis, 1092–1093
Cervical spondylotic myelopathy, 969, 970f
Cestodes (tapeworms), 1516–1517
CFTR, 566b, 754f
CHA_2DS_2-VASc score, 181b
Chancroid, 1535
Channelopathies, 975
Charcot-Marie-Tooth disease, 968
Charcot neuroarthropathy, 1231
Charcot's joints, 1119
Chédiak-Higashi syndrome, 100
Cheese washer's lung, 573b
Chemokines, 81, 81b–82b
Chemoprophylaxis, HIV-associated, 1554
Chemotherapy, 1387–1392, 1388f–1389f
cell cycle, 1388f
classification of cytotoxic drugs, 1389, 1389b
side-effects of, 1389–1392, 1390b
Cherry angioma, 1616
Chest, examination
auscultation, 530–531
crackles, 531
inspection, 530, 530b
palpation and percussion, 530, 531b
pleural rub, 531
vocal resonance, 531
wheeze, 531
Chest pain, 198–201, 530
central, 393
clinical assessment of, 198–200
differential diagnosis of, 199b–200b, 394b
investigations of, 200–201
management of, 201
musculoskeletal, 1078–1079
Chest physiotherapy, for respiratory conditions, 182
Chest X-ray (CXR), 118–120, 118.e1b
for aortic stenosis, 459
for asthma, 542
for bronchial carcinoma, 560, 561b
for cardiac disease, 397
mitral regurgitation, 456
normal, 118, 119f
for pneumonia, 213
Cheyne-Stokes respiration, 394
Chiari malformation, 947, 947f
Chickenpox, 1447, 1447f
Chikungunya virus, 1456
Child health, 169
Child labour, 169
Childhood dermatomyositis, 1113
Children
arthritis in, 1118
consent and, 32
health, 169
immune thrombocytopenic purpura in, 1322
nutrition, 771f, 771b
rheumatic disease, juvenile idiopathic arthritis, 1118
Chimeric antigen receptor T-cell therapy, 106
Chinese herb nephropathy, 835
Chlamydia trachomatis, 1525
Chlamydial conjunctivitis, 1053
Chlamydophila pneumoniae infection, 841
Chloracne, 1576b
Chloride, 777
Chloroquine, 1500–1501
overdose, 364
Cholangiocarcinoma, 730, 745, 745f, 1405
Cholecystectomy, 740
Choledochal malformation, 744
Cholera, 1487, 1488f
Cholestatic jaundice, 691, 776
Cholesteatoma, 588, 588f
Cholesterol embolization, 821
Cholesterol gallstones, 737, 738b
Cholesterolosis, 743
Chondroma, 564
Chondromalacia patellae, 1076
Chorea, 932
Christmas disease, 1325
Chromium, 777
Chromoblastomycosis, 1496
Chromosomal disorders, 62–64
Chromosomes
abnormal structures, 63, 64b
analysis of, 64
autosomes, 63
centromere, 57–58
constitution, 57f
human, 57–58
mitochondrial, 58, 59f
translocation, 76
Chronic abacterial prostatitis, 1663
Chronic actinic dermatitis, 1583–1584
Chronic bacterial prostatitis, 1663
Chronic cholecystitis, 744
Chronic cold haemagglutinin disease, 1303–1304
Chronic daily headache, 919
Chronic diarrhoea, 615, 615b–616b
Chronic discoid lupus erythematosus, 1607, 1607f
Chronic fatigue syndrome (CFS), 1029–1030, 1458
Chronic granulomatous disease (CGD), 100
Chronic idiopathic intestinal pseudo-obstruction (CIIP), 644
Chronic inflammatory demyelinating polyradiculoneuropathy (CIDP), 966
Chronic kidney disease (CKD), 837–849, 882
causes of, 837b
classification of, 838f
complications of, 838–843
early referral of, 843
heart failure and, 207
management of, 843

Chronic kidney disease (CKD) *(Continued)*
prevalence of, 837–838, 838b
progression of, 838, 839f
staging of, 837–838
Chronic lower limb ischaemia, 490
Chronic lymphocytic leukaemia (CLL), 1357
Chronic meningitis, 247
Chronic mountain sickness, 375
Chronic myeloid leukaemia (CML), 1343, 1343f
Chronic obstructive pulmonary disease (COPD), 545
clinical features, 546–547
diagnosis, 547
epidemiology and aetiology, 545
exacerbation of, 206–207, 207f–208f
investigations, 547
management, 547–549, 548f
drug therapy, 547–548
oxygen therapy, 548–549, 548f, 549b
pulmonary rehabilitation, 549
smoking cessation, 547
pathogenesis, 546
pathophysiology, 545–546, 546f, 550f
prognosis, 549
signs of, 546–547, 547b
surgery of, 342
symptoms of, 546
Chronic pain, 911
Chronic pain syndromes, 1079–1081, 1079f
Chronic pancreatitis, 753
aetiology, 753, 753b
autoimmune, 754, 756
clinical features, 754
complications, 756
differential diagnosis, 755
genetics, 753–754, 754f
investigations, 754–755, 755f
management, 755–756
pathogenesis, 753–754, 754f
Chronic peptic ulcer, 641
Chronic radiation enteritis, 649
Chronic reflux nephropathy, 831–832
Chronic regional pain, 1080–1081
Chronic renal disease, pregnancy and, 1691
Chronic respiratory disease, 1731
Chronic rhinosinusitis, 593, 594b
Chronic sensorimotor neuropathy, no cause found, 968
Chronic small intestinal ischaemia, 651
Chronic spontaneous urticaria, 1588
Chronic tophaceous gout, 1105, 1105f
Chronic tubulointerstitial nephritis, 828, 834–835, 834b
Chronic (work-related) upper-limb pain syndrome, 1080
Chronic ventilatory failure, 551, 551b
Chronic widespread pain, 1079–1080
Chvostek's sign, 1196
Chylomicrons, 511, 511f
disorders of, 513
Chylothorax, 559
Cicuta spp., poisoning, 370.e6
Cidofovir, 1466–1467, 1466f
CIDP. *See* Chronic inflammatory demyelinating polyradiculoneuropathy
Ciguatera fish poisoning, 370.e1
CIIP. *See* Chronic idiopathic intestinal pseudo-obstruction
Circadian rhythms, 1143
Cirrhosis, 708–711
aetiology, 708–711, 708b
complications and effects, 712–718, 712b
hepatic, 859
investigations, 709–710
management, 710
pathogenesis, 708, 708f
pathology, 708, 709f
prognosis, 710–711, 710b
scoring systems, 710b
secondary biliary, 719
Cisgender, 1670b
CJD. *See* Creutzfeldt-Jakob disease
Classical homocystinuria (A), 1246.e1b
Clergyman's knee, 1076
Climate change, poverty and, 171, 171f–172f
Clindamycin
for bacterial vaginosis, 1529
for sepsis, 295
Clinical consultation, of diagnosis, 1–2
Clinical equipoise, 39
Clinical Frailty Scale (CFS), 179, 1697, 1698f
Clinical genetics, 54–56
Clinical Institute Withdrawal Assessment for Alcohol Scale (CIWA), 181b, 253
Clinical negligence, 33–34
Clinical transplantation, organ rejection in, 105, 105b
Closed valvotomy, 455
Clostridium difficile
colitis, 222
diarrhoea, 665
gastroenteritis, 1475–1476
toxin, 662
Clostridium perfringens, 1476
Clozapine, 1029
Cluster headache, 918–919
CMV. *See* Cytomegalovirus
Coagulation, 1316
activation, in shock, 311
inhibitors of, 1327
leukaemia, 1338–1339
pathway, 1316–1319, 1318f
physiological limitation, 1317–1318
Coagulation cascade, 1319f
Coagulation disorders (coagulopathies), 1320
acquired, 1325–1327
inherited, 1324–1325
Coagulation factors
assays, 1321
concentrates, 1314
synthesis, 1318f
Coagulation screens/tests, 1320–1321
Coal-worker's pneumoconiosis, 577–578
Cocaine, poisoning, 362
Coccidioidomycosis, 1494–1495
Cochlear lesions, 900
Cochlear nerve, 900–901
Coeliac disease, 645, 732
aetiology of, 645
clinical features of, 645
complications of, 646–647
diagnosis of, 646
environmental factors in, 645
genetic factors in, 645
investigations in, 646
management of, 646
non-responsive, 646
refractory, 647
small intestinal tumours and, 652
Co-factors, supplementation with, 1244
Cognitive impairment, 1717
Cognitive state, 1005b, 1026
Cohort studies, of evidence-based practice, 39–40
Cold, 376–378
Cold air, asthma and, 541
'Cold' autoimmune haemolytic anaemia, 1303
Cold injury, 376–378
classification of, 378
clinical features of, 378
freezing, 378
management of, 378
non-freezing, 378
peripheral, 378
phases of, 378
surgery of, 378, 378b
Cold water immersion, 380
in hypothermia, 376–377
Colesevelam, 1221
Colitis
Clostridium difficile, 222
ischaemic, 651, 666
pseudomembranous, 665
ulcerative, 223b, 662
Collagen disorders, 1082–1083, 1082b
Collapse, 236b
Collapsing pulse, 395, 395f
Collateral circulation, 979f
Colles' fracture, 1129
Colloids, 315
Colon, 611–612, 665–674
ischaemic disease of, 666
physiology of, 611, 611b
Colon polyps and polyposis syndromes, 669, 669f, 670b
Colonic stenting, 149–150, 149f
Colonic tumours, 669–673
Colonoscopy, 617, 618b, 662–663
Colorectal cancer (CRC), 671–672, 672b, 1403
adjuvant chemotherapy and radiotherapy for, 1404
clinical features of, 673
families, 672, 672f–673f, 672b
follow-up, 674, 1404
genetics, 672
investigations of, 673–674, 673f
lifetime risk of, 672b
management of, 674, 1403
metastatic, 1404–1405
neoadjuvant chemotherapy and radiotherapy for, 1404
pathology of, 673
prognosis of, 1403, 1404b
risk factors in, 672b
screening for, 673
surgery for, 1404
survival, 1396b
Colorectal polyps, 669, 670b
Colour flow Doppler, 402–403, 404f
Coma
causes of, 233, 233b, 949
definition of, 232
diagnosis and investigations of, 236
general examination of, 234–235, 234f
Glasgow Coma Scale for, 233b
immediate assessment of, 233–234
locked-in syndrome, 949f
look-alikes, 236
management of, 233–234, 236
mechanisms of, 233, 233b
neurological examination of, 234, 234f
poisoning and, 357
vegetative state of, 949f
Combined oestrogen and progesterone contraceptives, 1653, 1655b
Commence opioid replacement therapy, 1032
Commensal organisms, UTI and, 829–830
Common bile duct stones, 741
Common cold (acute coryza), 537, 1451
Common peroneal nerve palsy, 964
Common variable immunodeficiencies (CVID), 99
Communication
clinical, 14–24
breaking bad news in, 21–22, 21b
with colleagues, 23–24
core skills in, 14
defined, 14–15
difficult conversations in, 22, 22b–23b
discussing uncertainty and risk in, 20, 20b
ending consultation in, 19, 19b
evidence-based skill, 15–16
fundamental elements of, 16–18, 16b–18b, 17f
gathering information in, 19, 19b
initiating consultation in, 19, 19b
learning, 18–24
patient-centred care in, 14, 15b
remote consulting in, 22–23, 23b
Communication *(Continued)*
shared decision-making in, 20–21, 21b
sharing information in, 19–20, 20b
cortical disorders of, 889–890
diagnosis, 11
healthcare-associated infections and, 1439
mitigation strategies, for acute viral illness, 212
Community-acquired infection, 283
Community-acquired pneumonia, 212–215, 213b
Compartment syndromes, 1077
Compensated euthyroidism, 1167
Complement, 82–83, 82f, 82b, 799, 1066
cascade, activation of, 310
deficiency of, 98b, 100
lytic pathway defects, 98b
regulatory proteins, 100
Complement cascade, 82
Complete eventration, 580
Complete external ophthalmoplegia, 897
Complex discharge planning, in older adults, 1715–1719, 1715b–1716b
Complex post-traumatic stress disorder (CPTSD), 1024
Complex regional pain syndrome, 1080
Complex traits, 69, 69b
Comprehensive Geriatric Assessment (CGA), 1699–1703, 1702b
Compression, 962
Computed tomographic pulmonary angiography (CTPA), in pulmonary embolism, 267, 269f
Computed tomography (CT), 131–137, 131b
of abdomen, 131–132
cardiac, 406
dual/multienergy, 151
in gastrointestinal disease, 618, 619f
of kidney disease, 799
liver and biliary tract, 688, 689f
of pelvis, 131–132, 134f–135f
respiratory disease, 532, 532f
Conduction aphasia, 890
Confidence intervals, 43, 44b
critical appraisal and, 41–43, 42b
relative risk of, 46
Confidentiality, 32–33
breach of, 32–33
psychiatry and, 1010–1011
respect for, in practice, 32
Confounding factors, of evidence-based practice, 38–39
Congenital adrenal hyperplasia, 1158, 1185b
Congenital anomalies, prenatal diagnosis of, 65–66
Congenital heart disease, 469–495
Congenital intrahepatic biliary dilatation, 727
Congenital nephrotic syndrome, 804
Congenital syphilis, 1533–1534
Congenitally corrected transposition of the great arteries, 475
Conium maculatum, poisoning, 370.e6
Conjugate gaze deviation, 235f
Conjugate lateral gaze, 896
Conjunctivitis, 1052–1054, 1054b
allergic, 1054
bacterial, 1052
chlamydial, 1053
gonococcal, 1052
phthiriasis palpebrarum, 1053, 1055f
red eye *vs.*, 1047b
trachoma, 1053, 1054b
viral, 1053, 1055f
Conn's syndrome, 1160, 1691
Conscientious objection, 33
Consciousness, 232
disorders of, 948–950
disturbed, 232–233
Consensus definition (Sepsis-3), 281
Consent, 29, 29b
children and, 32
in educational settings, 29–30
mental health and, 31–32
psychiatry and, 1010–1011
in surgery, 337–338, 338f
in young adult, 1633b
Constipation, 615–617, 992
causes of, 616b
defecatory disorders, 616–617
diagnosis of, 615–617
management of, 617, 617b
normal-transit, 615–616
palliative care, 1726
prevention of, 300
slow-transit, 616f, 617
Constrictive pericarditis, 407f, 489
Contact dermatitis, 1582
Contact eczema, 1583, 1583f
Contact lens-related keratitis, 1054, 1056f
Continuous ambulatory peritoneal dialysis (CAPD), 845
Continuous glucose monitoring, 1223
Continuous positive airway pressure (CPAP), 186, 192b, 323, 323f
Contraception
combined oestrogen and progesterone, 1653, 1655b
emergency, 1653
epilepsy, 1687
female health and, 1653, 1654b
hormonal, 1653, 1654b
long-acting reversible, 1653
non-hormonal, 1654b
sterilization and, 1653
stroke and, 997
Contrast studies, in gastrointestinal disease, 619
Contrast-associated nephropathy, 837
Controlled mechanical ventilation, 320, 322
Convergence, 895
Convulsions, overdose/poisoning, 357
Coombs' test, 1302f
COPD assessment test (CAT), 547b
Copper, 777–778
Copper sulphate poisoning, 367
Corneal abrasions, 1048, 1050f
Corneal dystrophy, 1054
Corneal foreign body, 1048, 1050f
Corneal transplantation, 1054–1055
Coronary angiography, for acute coronary syndromes, 448
Coronary artery bypass grafting (CABG)
for angina, 442, 443f
for STEMI, 450
Coronary artery calcification, 406, 841
Coronary artery disease (CAD), 437–452
diagnosis of, 439–440
MRI for, 406–407, 409f–410f
pathophysiology of, 438–439, 438f–439f
risk factors for, 439, 439b
Coronary circulation, 391, 392f
Coronaviruses, 1445b, 1452, 1452b
Corrosive agents, 367
Cortical disease, 889–891
Cortical necrosis, acute, 836–837
Cortical venous thrombosis, 246
Cortical visual loss, 891
Corticospinal (pyramidal) system, 903–904, 903f
Corticosteroids
adverse reactions of, 1159b
biochemistry, 1152
biosynthetic pathways, 1153f
for COPD, 207, 548
Cushing's syndrome and, 971
inhaled, 543
for meningitis, 248
oral, 543
physiology, 1152–1153
problems with, 1159

Corticosteroids *(Continued)*
for rheumatoid arthritis, 1095, 1095b
for rhinitis, 538
for SLE, 1110
supervision of therapy, 1159
surgery and, 1159, 1160b
Corticotrophin-releasing hormone (CRH), 1152–1153, 1152b
Cortisol, 1142–1143, 1143f, 1152b
Coryza, 1451
Cost-utility calculations, 165b
Cotswolds modification, 1350b
Cotton wool spots, 1226
Cough, 539, 540b, 1727
Countercurrent system, 794f
COVID-19
critical care medicine, 327
death, 1258
HIV-associated, 1559
palliative care, 1731, 1732f
pneumonitis, 552–553
skin manifestations in, 1595, 1595b
Cowden's syndrome, 671
Cowpox, 1449
Cox regression, 47–48
Coxibs, 1094–1095
Cranial nerves, 891–902, 892b
Craniopharyngioma, 1147b, 1149, 1149f
CRAO. *See* Central retinal artery occlusion
CRC. *See* Colorectal cancer
Creatinine clearance, 796, 796f
Creutzfeldt-Jakob disease (CJD), 958
CRH. *See* Corticotrophin-releasing hormone
Cricothyroidotomy, 322
Crigler-Najjar syndrome, 690
Crimean-Congo haemorrhagic fever, 1464, 1518.e1, 1518.e1f, 1518.e2b, 1518.e3b
Criminal justice system, 1041
Critical care medicine, 297–333, 297b
acid-base balance, 307–308
acute kidney injury, 327–328
acute respiratory distress syndrome, 325–327
arterial pressure variation, 306–307
arterial waveform analysis, 307
blood pressure, 304, 304f, 304b
brain death, 330–333
cardiac output, 306
central venous pressure, 305, 305f, 305b
COVID-19, 327
echocardiography, 307, 307f
indications for admission, 298b
left atrial pressure, 305, 305f
Modified Early Warning Score, 298, 298b
monitoring, 304–306, 304b
oesophageal Doppler, 307, 307f
outcomes of, 330–333
pulmonary artery pressures, 305–306, 305b, 306f
recognition and diagnosis of, 298, 299f
respiratory failure, 318–325
tissue perfusion, 304–306
withholding and withdrawing treatment, 330
Critical care outreach teams, 182
Critical illness, neuromuscular weakness complicating, 968
Crohn's disease, 97
anal and perianal complications of, 657b
clinical features of, 658
conventional maintenance therapies for, 660
examination of, 658
extraintestinal manifestations, 656–657, 657b
inflammatory bowel disease and, 653
investigations of, 658–659, 659f
macroscopic changes of, 655
management of, 222–223
medical management of, 659–661, 660b
Crohn's disease *(Continued)*
microscopic changes of, 655, 657b
pathology of, 657f
prognosis of, 662
refractory or fulminant, 660
sites of, 657f
small intestinal tumours and, 652
surgical management of, 661–662
Cross-sectional surveys, of evidence-based practice, 40
Croup, 539, 1451
Cruciate ligament tear, 1076
Crush fractures, 1074
Crusted scabies (Norwegian scabies), 1598f, 1599
Crusting, 1568f–1570f
CRVO. *See* Central retinal vein occlusion
Cryoglobulinaemic kidney disease, 816
Cryoprecipitate, 1314
Cryptococcosis, 1494
HIV-associated, 1556
Cryptorchidism, 1179, 1668
Cryptosporidiosis, 649, 1507
HIV-associated, 1557
Crystal arthritis, 1103–1106, 1103f
Crystal-associated osteoarthritis, 1086, 1086f
Crystalloids, 315
CSF. *See* Cerebrospinal fluid
CT. *See* Computed tomography
CT angiography, 132, 136f
CT colonography, 133
CT coronary angiography (CTCA), 136, 406, 406f
CT urography, 132, 137f
Cullen's sign, 750f
Culture-negative endocarditis, 466
CURB-65 score, 181b, 213, 213b
Cushing's disease, 1119, 1156
Cushing's syndrome, 1147, 1147b, 1153–1154, 1156
clinical features of, myopathy, 971
confirmation of, 1154–1155, 1154b
cutaneous manifestations, 1605
differential diagnosis, 1155–1156
history and examination of, 1154, 1155f
investigations of, 1154–1156
management, 1156
pathophysiology and aetiology, 1154, 1154b
Cutaneous angiosarcomas, 1625
Cutaneous diphtheria, 1471
Cutaneous larva migrans, 1513, 1600
Cutaneous leishmaniasis, 1504–1505, 1505f
Cutaneous leucocytoclastic vasculitis, 1117
Cutaneous malignancy, 1601
Cutaneous squamous cell carcinoma (cSCC), 1620, 1620f
Cutaneous warts, 1593
Cyanide poisoning, 367
Cyanosis, 394, 530
Cyclical vomiting syndrome, 637, 637b
Cyclophosphamide, 1110
Cyclospora cayetanensis, 1508
Cyclothymia, 1018–1019
CYLD cutaneous syndrome, 68, 68f
Cylindroma, 564
Cystatin C, 795–796
Cystic fibrosis (CF), 566, 730
clinical features and complications, 566, 567b
diagnosis, 566–567
gene therapy for, 73, 74f
lung/heart-lung transplantation, 577
management, 567
pancreatic insufficiency and, 756
pathogenesis, 566, 566b, 567f
pregnancy, 1687
Cystic gland polyps, 640
Cystic kidney disease, 849–851
Cysticercosis, 1516
Cystinuria, in kidney calculi, 822
Cytoid bodies, 1226
Cytokines, 81, 106, 310, 311f
Cytomegalovirus (CMV), 1457
clinical features of, 1457, 1457f
diagnosis of, 1457
hepatitis and, 704
HIV-associated, 1558, 1558f
management of, 1457
Cytotoxic antibiotics, 1389
Cytotoxic CD8 T lymphocytes (CTLs), 90
Cytotoxic drugs, 1389
Cytotoxic mushroom, poisoning, 370.e6

D

Dacryocystitis, 1051
Damage accumulation theory, ageing and, 1695
Dantrolene, cathinone poisoning, 362
DaTScan, of brain, 143, 144f
Datura stramonium, poisoning, 370.e6
D-dimer, measurement of, 266–267, 1365
de Quervain's thyroiditis, 1168
Decision-making, patient flow and, 178
Decompression illness (DCI), 380–381, 380b
Decongestants, 538
Deep brain stimulation
for depressive disorders, 1029
for Parkinson's disease, 931
Deep vein thrombosis (DVT), 264–272
clinical assessment of, 265
definition of, 264
diagnosis of, 266–267, 1363
differential diagnosis of, 265b
emergency presentations of, 267
epidemiology of, 264–265
imaging, 267, 1363
investigations of, 265–266, 268f
management of, 267–272
prevention, 299
risk factors for, 264–265, 264b
timescale of investigation, 267
upper limb, 267, 1365
Defecation
disorders, 616–617
physiology of, 611–612
Defibrillation, 195, 195f
Dehydration, 258f
Delirium, 250–253
causes of, 251b
clinical assessment of, 251–252, 252b
definition of, 233
diagnostic criteria for, 251b
investigations of, 252
management of, 252–253
in older adults, 1710–1712, 1713b–1714b
predisposing factors in, 251b
surgery and, 348
Delirium tremens (DT), 1015
Delusional infestation, 1630
Dementia, 952–959, 953b
causes of, 954, 954b, 955f
clinical assessment of, 952–953
epidemiology of, 952
frontotemporal, 957
investigations of, 953–954, 953b, 954f
management of, 958–959
mild cognitive impairment, 954
palliative care, 1733
risk factors for, 955b
syndromes, 954–957
Demyelination, 962
Dendritic cells, 84–85, 85f
Dengue, 1454–1458, 1457f, 1518.e3
classification of, 1518.e3, 1518.e3f, 1518.e4f, 1518.e5f
differential diagnosis of, 1518.e6, 1518.e6b
expanded, 1518.e5, 1518.e6b
Dengue *(Continued)*
investigations, 1518.e6
management, 1518.e6
pathophysiology of plasma leakage, 1518.e5
prevention, 1518.e7
warning signs of, 1518.e7b
Dengue fever, 1454–1458, 1454f–1455f
Denosumab, for osteoporosis, 1130b, 1131–1132
Dent's disease, 876
Depolarization, 388, 389f
Depression, 1017
aetiology of, 1017–1019
clinical features of, 1018
diabetic patients, 1234
diagnosis of, 1018
genetic factors, 1017–1019
in rehabilitation, 1716
risk factors, 1018b
treatment and prognosis of, 1018–1019, 1019b–1020b
Deprivation, 156–157, 156f
Dermatitis artefacta, 1630
Dermatitis herpetiformis, 647, 1613
Dermatofibroma (benign fibrous histiocytoma), 1616
Dermatofibrosarcoma protuberans (DFSP), 1625, 1625f
Dermatology, 1564–1630, 1564b–1565b
Dermatomyositis, 1606
Dermatophyte infections, 1596–1597, 1596f
Dermatophytosis, 1496
Dermatoscope, 1565, 1565f, 1565b
Dermatoses, pregnancy, 1692, 1692b
Dermatosis papulosa nigra, 1615–1616
Dermis, 1572
Descriptive statistics, 43
Desensitization, 101–102, 102b
Dexamethasone suppression test, 1145f, 1153, 1154b
Diabetes
medications, 1266
stroke and, 997
type 1 and 2, 1688–1689. *See also* Gestational diabetes
Diabetes Control and Complications Trial, 1213
Diabetes insipidus, 1191–1193
aetiology, 1191, 1191b
AVP-deficiency, 866
biochemistry, 1191
history and examination of, 1191
management, 1191–1192
pregnancy, 1691
water deprivation test, 1192b
Diabetes mellitus, 966–967, 1200–1239, 1201b, 1202f, 1604–1605, 1604f
acute presentation, 1211
adolescents, 1235–1238
asymptomatic, 1211
classification, 1204–1211, 1207b
clinical skills for, 1200–1239
complications, 1224–1235
core skills and knowledge in, 1200b
diagnosis and investigations, 1211–1212, 1211b–1212b
examination, 1201, 1203b
gestational, 1238, 1238b
inpatient, 1236, 1236b
management, 1212–1214, 1213b
metabolic control, 1222–1223
metabolic emergencies, 1224
monogenic, 1210–1211, 1210b
older adults, 1235–1236, 1235b–1236b
organization, 1238–1239, 1239b
pancreatitis and, 755–756
patients living with, 257–264
physical examination, 1211
pregnancy, 1236–1237
Diabetes mellitus *(Continued)*
prevalence, 1203f
psychosocial implications, 1234–1235, 1234b
rheumatological problems in, 1119
routine checks, 1200–1201
specialist teams for, 182
subacute presentation, 1211
surgery for, 340–341, 1236
type 1, 1206–1211, 1208b
aetiology, 1206–1209
autoimmunity, 1206, 1208f
environmental factors, 1208–1209
epidemiology, 1206
genetic susceptibility and inheritance, 1206
genome-wide association studies, 1206
HLA system, 1206
type 2, 1208b, 1209
aetiology, 1209
epidemiology, 1209
non-insulin treatment, 1219–1221
prevention, 1212, 1222b
Diabetic amyotrophy, 1230
Diabetic eye disease, 1227
cataracts, 1227
clinical features, 1225f
examination, 1227
macular area, 1227
management, 1229
peripheral retina, 1227
screening, 1227, 1228b
Diabetic foot, 1231–1232, 1231f, 1232b
Diabetic ketoacidosis, 257–261, 258f, 258b, 260b
Diabetic mastopathy, 1643
Diabetic nephropathy, 807–808, 808f, 808b, 1227–1229, 1228f–1229f
Diabetic neuropathy, 1229–1231, 1230b
acute painful neuropathy, 1230
Diabetic retinopathy, 1224–1227, 1226f, 1226b
Diagnostic reasoning, 7
bias and error, 8–9, 9b
clinical intuition, 8
diagnostic criteria, 7, 10
interpretation of test results, 7
investigation of, 5–7
levels of diagnostic depth, 9, 10f
models of diagnosis of, 7–8, 8f
overdiagnosis, 10, 11b
role of watchful waiting, 9–10, 10b
Dialysis, withdrawal of, 1731
Diaphragm, 524
disorders of, 580
Diarrhoea, 220–223, 221b, 613–615
abnormal motility and, 614
acute, 614–615, 615b
bile acid malabsorption and, 648
causes of, 613b
chronic, 615, 615b–616b
Clostridium difficile-associated, 665
functional, 644
haemolytic uraemic syndrome and, 817
osmotic, 613
pathophysiology of, 613–614
secretory, 613–614
traveller's, 1476, 1476b
watery, 1473b
Diarrhoeic shellfish poisoning, 370.e1
DIC. *See* Disseminated intravascular coagulation
Dichotomous outcomes, 45–48, 45b
Dicrotic pulse, 395, 395f
Diet
for asthma, 542
cancer association, 1381
for diabetes mellitus, 1213–1214, 1214b
for gout, 1104
for hypertension, 497
lipid-lowering, 518

Dietary Approaches to Stop Hypertension (DASH) diet, 1263b
Dietary requirements, 764–769, 768f
Diethylene glycol, poisoning, 367
Diffuse alveolar haemorrhage, 574
Diffuse brain disorders, 914
Diffuse brain dysfunction, 233
Diffuse cutaneous scleroderma (DcSSc), 1112
Diffuse goitre, 1173
Diffuse idiopathic skeletal hyperostosis (DISH), 1074
Diffuse large B-cell lymphoma (DLBCL), 1352, 1353f, 1353b
DiGeorge's syndrome, 98–99, 1195
Digital rectal examination, 1661f
Digital subtraction angiography (DSA), 490, 491f
Digitalis purpurea, poisoning, 370.e6
Digoxin, overdose, 364
Dilated cardiomyopathy, 483–484, 484f
Dimercaptosuccinic acid (DMSA), renal function and, 144, 145f
Dipeptidyl peptidase-4 inhibitors, 1220
Diphtheria, 1471, 1471b
Diphyllobothrium latum, 1517
Dipstick tests, 798
Direct current cardioversion (DCC), 195–196, 196f
Disability, 1696–1698, 1728–1729
Discharge planning, 182
Discharge policy, in upper gastrointestinal bleeding, 226
Discoid eczema (nummular eczema), 1581
Disease-causing genes, discovery of, 71–72
Disease-modifying anti-rheumatic drugs (DMARDs), in rheumatic disease, 1095–1097, 1096b–1097b
Disease-modifying drugs, for multiple sclerosis, 938, 939b
Disorders of sex development (DSD), 1185–1190, 1185b
Disposable soma theory, ageing and, 1695
Dissecting cellulitis, 1628, 1628f
Disseminated intravascular coagulation (DIC), 1326, 1326f, 1326b–1327b
Dissociative (non-epileptic) attacks, 927–928
Dissociative seizures, 238, 254
Distal convoluted tubule, 873
Distribution, 109–110
Diuretics, 860–863, 861b
  effects on kidney function, 862–863
  for heart failure, 435–436
  resistance, 862
Diversity, 28–29
Diverticular disease, 666–667, 667f–668f
Diverticulitis, acute, 666–667
Diverticulosis, 666
Diving, 380
Dix-Hallpike test, 586f
Dizziness, 236–239, 236b–237b, 900
D-lactic acidosis, 785
DMARDs. *See* Disease-modifying anti-rheumatic drugs
DMSA. *See* Dimercaptosuccinic acid
DNA, 53
  drugs damaging, 1389
  genetic code and, 57, 57f
  repair inhibitors, 1389
  transcription, 58, 60f
  transfection, 72, 73f
DNA polymerase mutants, of hepatitis B, 697
DNA viruses, 1444b
Do not attempt resuscitation (DNAR) orders, in palliative care, 1723
Dobutamine, 316
Doctor
  core skills of, 1b
  diagnosis: art of being, 1–13
    artificial intelligence and medicine, 12–13
    clinical consultation of, 1–2, 2f
    communicating a diagnosis, 11
    conceptualization of, 2f
    diagnostic reasoning, 7
    medical consultation of, 2, 2f
  knowledge of, 1b
Doll's eye reflexes, 235f, 896
Domoic acid poisoning, 370.e1
Donovanosis, 1536
Dopamine, 316, 1189b
Dopamine agonists, 1188
Dopamine antagonist prokinetic agents, 624
Dopamine antagonists, 931, 931f
Dopexamine, 316
Doppler echocardiography, 401–404
Doppler ultrasound, 688
Dorsal columns, 908
  lesions of, 909
Dorsal tenosynovitis, 1070
Dose-response curve, 109f
Down-beat jerk nystagmus, 897
Down's syndrome, 63–64
Dracunculiasis, 1511
Dracunculus medinensis, 1511
Dravet syndrome, 925
DRE. *See* Digital rectal examination
Dropped head syndrome, 902
Drowning, in water environment, 380
Drug eruptions, 1629, 1629b
Drug-exacerbated dermatoses, 1629, 1629b
Drug-induced alloimmune haemolytic anaemia, 1304–1305
Drug-induced conditions
  acneiform eruptions, 1576b
  dermatoses, 1629, 1629b
  diarrhoea, 222
  haemolytic anaemia, 1304
  hepatitis, 706
  hepatotoxicity, 732–733, 733b
  hypoglycaemia, 1197
  kidney disease, 833, 834f
  myopathy, 972
  neuropathy, 967b
  parkinsonism, 932
  psychosis, 1027–1028
  respiratory reactions, 576b
  thrombocytopenia, 1322–1323
Drug metabolism, 110, 732
Drug rashes, 1601
Drug reaction with eosinophilia and systemic symptoms, 277
Drug-induced conditions
  tubulointerstitial nephritis, 833, 834f
Drugs
  adverse reactions of, severe cutaneous, 277, 277b
  choice of, 107–108
  dose, 108–111, 109f
  hypertension co-morbidities and, 497–498
  interactions, 115, 116b
  kidney disease and, 822
  liaison service, 182
  need for, 107
  pharmacodynamics, 110–111
  pharmacokinetics, 109–110, 110b
  renal handling and, 796
  routes of administration, 109b
  in upper gastrointestinal bleeding, 225–226
Dual-energy X-ray absorptiometry (DXA)
  osteoporosis, 1129, 1130f, 1130b
  scanning, 1067
Dubin-Johnson syndrome, 690
Duchenne muscular dystrophy, 974
Duct ectasia, 1642
Duct papilloma, 1642
Duke criteria, 466b
Duodenal ulcer, 633, 633f
Duodenum, 604–605, 632–640
Dupuytren's contracture, 1071
Dural venous sinus thromboses, 246
Duroziez's sign, 461
DVT. *See* Deep vein thrombosis
DXA. *See* Dual-energy X-ray absorptiometry
Dying patients, care of, 1733–1734, 1734b
Dynamic endocrine tests, 1138–1139, 1138b
Dysarthria, 906
Dysentery, bacterial causes of, 1473b
Dyshormonogenesis, 1166
Dyskinesia, tardive, 899
Dyspepsia, 612, 636–637
Dysphagia, 596b, 597–598, 612, 612b, 991
Dysphonia, 598–599
Dyspnoea, 393–394, 529, 529b–530b
Dysraphism, 960
Dystonias, 933–934, 934b
Dystrophic epidermolysis bullosa, 1614
Dysuria, 830

E

Ear, 583–585
  anatomy of, 583–585, 584f
  discharging, 587–588
  examination of, 583b, 584f–585f, 585
  hearing, 584–585
  sensory innervation of, 588f
Ear disorders, 585–591, 586b
Early inflammatory polyarthritis, 1087
Early-onset ataxia, 961
Ears, nose, throat (ENT), 582–600, 582b, 583f
Eastern Cooperative Oncology Group (ECOG) Performance Status, 563b, 1387b
Easy bruising syndrome, 1321
Eating disorders, 1036, 1037b
  in diabetic patients, 1235
  young adult mental health and, 1633
Ebola virus, 1465, 1518.e7, 1518.e8b
EBV. *See* Epstein-Barr virus
Echinocandins, 1497
Echocardiography, 400–404
  diagnostic
    for aortic regurgitation, 461
    for aortic stenosis, 459, 459b, 460f
    for mitral regurgitation, 456, 457f
    for mitral stenosis, 453–454, 454f, 455b
  Doppler, 401–404
  intravascular (coronary) ultrasound, 404
  modalities, 401–404, 403f
  myocardial perfusion, 404
  physics of, 400–401
  stress, 404
  three-dimensional, 401, 404f
  transoesophageal, 404, 457f
Economics, global health and, 173
Ecthyma, 1589
Ectopic beats, 423
Ectopic hormone secretion, 1199
Ectopic pregnancy, ruptured, 676
Ectropion, 1050–1051, 1052f
Eczema, 1577–1584, 1577b
Education
  global health and, 171
  patient consent in, 29–30
Ehlers-Danlos syndrome, 1611
Ehrlichiosis, 1492, 1492b
Ejection fraction, 401
Elapidae, poison of, 370.e3f
Elastase, 748
Elbow pain, 1070, 1070f
Elective surgery, in blood transfusion, 1309
Electric shock, 383
Electrocardiogram (ECG), 397–399, 398f–400f, 403f
  for acute coronary syndromes, 444–445, 445f
  diagnostic
    for aortic regurgitation, 461
    for aortic stenosis, 459
    for mitral regurgitation, 456, 457f
    for mitral stenosis, 453, 454f
  exercise, 398–399, 401f
  Holter (24-h ambulatory), 399, 402f
  for pulmonary embolism, 265, 265f
  sinus node function, 413f
  for STEMI, 448–449, 448b, 449f–450f
  tilt testing, 400, 403f
  waveform, 398, 399f–400f, 399b
Electroconvulsive therapy (ECT), 1029
Electroencephalography (EEG), 913–915, 914f
  of brain death, 914
  in diffuse brain disorders, 914
  of epilepsy, 914, 925–926
  multi-electrode, 329
  of unconscious patients, 236
Electrolyte balance, 764
  in acute kidney injury, 230
Electrolytes, 853–884, 853b–854b, 854f
  clinical skills for, 854–855
  intracellular and extracellular fluid, 856b
  nutritional requirement, 781, 781b
  in urine, 799
Electromyography (EMG), 914, 1067
Embden-Meyerhof pathway, 1299, 1300f
Embolectomy, for venous thrombosis, 272
Embolization, 146b
Embryology, 1175–1176
Emergency contraception, 1653
Emergency department (ED), in-reach to, 178
EMG. *See* Electromyography
Emphysema, 545–546
'Empty sella' syndrome, 1150–1151
Empyema, 215, 559
Encephalitis, 941
  anti-NMDA receptor antibody, 940
  anti-voltage-gated potassium channel (VGKC) complex limbic, 940
  autoimmune, 941
  epilepsy and, 925
  HIV-associated, 942
  Japanese, 1459
  paraneoplastic limbic, 940
  progressive rubella, 943
  tick-borne, 1459–1460
  viral, 1460b
  West Nile, 1459
Encephalocele, 960
Encephalopathy, 1242–1243, 1242f–1243f
End-organ sensitivity, 829
Endemic treponematoses, 1486
Endemic (murine) typhus, 1490b, 1491
Endocrine disease, 1604–1605
Endocrine function tests, 1144–1145
  radiological investigations, 1145
  saliva, 1144
  stimulation and suppression tests, 1144, 1145f, 1145b
  stress-related hormones, 1144
  urine collections, 1144
Endocrine myopathy, 971–973
Endocrine nuclear imaging, 144–145, 145f
Endocrine system, 1140f
Endocrinology, 1137–1199, 1139f
  aetiology of diseases in, 1143–1144
  clinical skills for, 1138–1199
  core skills and knowledge in, 1137b
Endocrinology *(Continued)*
  ectopic hormone secretion, 1199
  emergencies in, 1138b
  examination, 1138–1139
  history, 1138–1139, 1138b
Endometrial biopsy, for heavy menstrual bleeding, 1649, 1651f
Endometrial cancer, 1410–1412, 1659b
Endometriosis, 1528
Endoscopic retrograde cholangiography (ERC), 742, 742f
Endoscopic retrograde cholangiopancreatography (ERCP), 689
  bile duct stricture, 742f
  common bile duct stones, 742f
  pancreas, 749
Endoscopy
  of acute upper gastrointestinal bleeding, 640–641
  capsule, 642
  gastrointestinal disease, 617–618
  of *H. pylori*, 633f
  liver disease, 689
  of oesophageal tumours, 631, 631f
  in upper gastrointestinal bleeding, 225
  variceal banding, 713–714
  variceal haemorrhage, 714f
Endothelin, in heart failure, 392
Endotoxins, 1443
Enemas, 617b
Energy, 764–767, 768f
  balance, 764–766
  expenditure, 765–766, 768f
  intake, 764
  requirements, 766, 781
  stores, 766–767, 769f
Enhanced liver fibrosis (ELF) test, 688
Enoximone, 317
Enteral nutrition, 781–782
  critical illness, 300
  in Crohn's disease, 660
  feeding regimen, 782
  PEG insertion, indications for, 782f, 782b
  post-pyloric feeding, 782
Enteric fever, 1478, 1488
Enterobius vermicularis (threadworm), 1511–1512
Enterohaemorrhagic *Escherichia coli*, 1474–1475, 1475f
Enteroinvasive *Escherichia coli*, 1474
Enteropathic arthritis, 1102
Enterotoxigenic *Escherichia coli*, 1475
Enthesitis, 1077
Envenoming, systemic, 1518
Environmental medicine, 371–383, 371b
  aviation and, 372–374
  clinical skills for, 371–372, 372f
  diving, 380
  electric shock, 383
  examination of, 372, 373f
  fatal conditions in, 371b
  high altitude, 371b, 374–375
  history of, 371
  ionizing radiation, 381, 382b
  life-threatening conditions in, 371b
  lightning strike, 382–383
  patient assessment in, 372b
  physical world, 381–383
  space travel and, 372–374
  water, 379–381
Enzyme replacement therapy (ERT), 1244
Eosinophilia, causes, 1316b
Eosinophilic gastroenteritis, 651
Eosinophilic granulomatosis with polyangiitis (EGPA), 574
Eosinophilic oesophagitis, 629–630, 629f
Eosinophils, 83–84, 1315, 1316b
Ephelides. *See* Freckles
Epicondylitis, 1070, 1070f
Epidermal naevi, 1615
Epidermis, 1570–1572

Epidermodysplasia verruciformis, 1593
Epidermoid cyst ('sebaceous cyst'), 1616
Epidermolysis bullosa, 1614
Epididymal cyst, 1668, 1668f
Epididymitis, 1667–1668
Epididymo-orchitis, 1528
Epidural haemorrhage/haematoma, 950
Epigenetic techniques, 70
Epigenetics, 59
Epiglottitis, 1472
  acute, 539
Epilepsia partialis continua, 250, 923
Epilepsy, 921–927, 1027, 1687
  aetiology of, 923–925, 923b
  birth defects, 1687
  classification of, 921–923, 921f, 921b
  contraception, 1687
  diagnosis of, 925–926, 925b
  driving and, 927
  drug withdrawal and, 927
  epidemiology of, 921
  first fit, 925–926
  investigations of, 925–926, 926b
  lifestyle and safety and, 927
  management of, 926–927, 926b
  pregnancy, 1687
  refractory, 927
  syndromes, 923–925
Epipodophyllotoxins, 1389
Epistaxis, 591b–592b, 592f, 593–595, 594b
Epithelial hyperplasia, of breast, 1642
Epstein-Barr virus (EBV), 1456
Epworth sleepiness scale, 550, 550b
Equality, principles of, 28–29
Erectile dysfunction, 913, 1669, 1669b
Erosions, 633, 1568f–1570f
Erosive osteoarthritis, 1086
Erysipelas, 275–276, 275f
Erythema multiforme, 1602–1603, 1602f, 1603b
Erythema nodosum, 1118–1119, 1602–1604, 1602f, 1602b
Erythema nodosum leprosum, 1486b
Erythrasma, 1590, 1590f
Erythroblasts, 1276
Erythrocyte sedimentation rate (ESR), 1276
Erythrocytes. *See* Red blood cells
Erythroderma, 275–277, 276b, 1574, 1611b
Erythroid hyperplasia, 1288
Erythropoiesis, 1276
Erythropoietin, 797, 1276
Erythrovirus, 1450
Erythrovirus B19, arthritis, 1107
Essential amino acids, 767
Essential hypertension, 819
Essential thrombocythaemia, 1344, 1345b
Essential tremor, 932
Estimated average requirement (EAR), 764
Ethanol, poisoning, 361b, 362
Ethical issues
  in clinical genetics, 55
  in medical practice, 29–34
Ethics
  guidance and codes of practice, 27–28, 27b
  of inequality, 28–29
  medical, 25–34
  of self-care, 27
Ethylene glycol
  metabolism, 367, 367f
  poisoning, 367
Euthyroid goitre, 1174
Evidence, challenges of implementing, 49–51
  barriers and facilitators to change, 50–51
  intervention context, 50–51
  leadership, 51
  patient-centred outcome measures, 50
  pragmatic RCTs, 50
  trial design, 49–50
Evidence-based practice, 35–51.e3
  applying, 49–51
  appraising the evidence, 37–43
  asking answerable questions, 36
  barriers to, 36
  combined approaches of, 41
  core skills and knowledge of, 35b
  critical appraisal, 41–43
  hierarchy of, 37–41, 38f
  need for, 35
  other factors limiting implementation of, 36
  primary sources of, 36–37
  searching for, 36–37, 37b
  secondary sources of, 37
Excessive daytime sleepiness (EDS), 928
Excitation-contraction coupling, 389–390, 389f
Excoriation, 1568f–1570f
Excretion, 110
Excretion urography, 826, 826f
Exotoxins, 1443
Extended-spectrum β-lactamase-producing Enterobacteriaceae, 296
External ocular palsies, diabetic eye disease and, 1227
Extracellular volume
  decreased, 863–865, 863b, 864f
  disorders of, 855–865, 856f
  increased, 859–863
  regulation of, 856–857, 857f
Extracerebral arteries, sites of stenoses in, 979f
Extracorporeal gas exchange, 323, 324f
Extracorporeal shock wave lithotripsy, 740, 823
Extradural haemorrhage, 245, 245f
Extramedullary haemopoiesis, 1307
Extrapyramidal system, 904–905, 905f
Extravasation, 1390
Extubation, 325
Exudate, haemorrhagic intra-alveolar, 326
Exudative (wet) age-related macular degeneration, 1060
Eye
  applied anatomy and physiology, 1044–1046
  as hypertensive target organ, 503, 504f
Eye disorders, 1050–1060
  examination for, 1046–1048, 1047f, 1047b
  history, 1046, 1046b
  in rheumatoid arthritis, 1093
  in SLE, 1109
Eye movements
  control, 896
  examination of, 896
  of unconscious patients, 235, 235f
Eye trauma, 1048
Eyeball, cross-section of, 1045f

F

Fabry's disease, 816, 1246.e2b
Facet joint syndrome, 1073
Facial nerve (VII), 898–899
Facial pain, 915–921
Facioscapulohumeral dystrophy, 974–975, 974b
Factitious disorder, 1041, 1042b
Factitious disorder imposed on another (FDIA), 1041–1042
Factitious hypoglycaemia, 1197
Factor V, 1317
Factor VIII, 1317, 1319f, 1324b
Faecal culture, 1431, 1432b
Faecal incontinence, 665–666, 665b, 666f
Faecal-oral infections, prevention of, 1439
Falls (instability), 1706b, 1707f, 1708b–1709b, 1709f
Familial adenomatous polyposis (FAP), 670, 1383b
Familial amyloidoses, 807
Familial cancer syndromes, 1383b
Familial combined hyperlipidaemia (FCHL), 516
Familial dilated cardiomyopathy, 483–484, 483f
Familial hemiplegic migraine, 917
Familial hypercholesterolaemia, 439, 517
Familial hyperkalaemic hypertension, 872
Familial hypocalciuric hypercalcaemia, 1195
Familial hypomagnesaemia, hypercalciuria and nephrocalcinosis, 873
Familial Mediterranean fever, 1119–1120
Fanconi's syndrome, 795, 1544
FAODs. *See* Fatty acid oxidation disorders
FAP. *See* Familial adenomatous polyposis
Farmer's lung, 573b
Fat necrosis, breast, 1643
Fatigue, 236b
  chemotherapy-related, 1390
  diaphragmatic, 580
  heart disease and, 394
  as neurological disease, 906
  palliative care, 1728
Fats, 768–769
Fat-soluble vitamins, 775–776
Fatty acid oxidation disorders (FAODs), 1246, 1246.e4b
Fatty liver
  acute, of pregnancy, 1685
  clinical features, 724
  investigations, 725
  management and prognosis, 725
  pathology, 724, 724f
FCHL. *See* Familial combined hyperlipidaemia
Febrile convulsion, 924
Febuxostat, for gout, 1104
Feeding, hormones and, 1143
Feet, rheumatoid arthritis in, 1092, 1092f
Felty's syndrome, 1093
Female genital mutilation, 1660, 1660b
Female health, 1646–1660, 1648f
  adolescence, 1649
  contraception and, 1653, 1654b
  domestic abuse and violence, 1659
  examination, 1646b, 1647f
  genital mutilation and, 1660, 1660b
  gynaecological malignancy in, 1658–1659, 1659b
  heavy menstrual bleeding in, 1649–1650, 1650f
  history, 1646b
  leiomyomata and, 1657–1658, 1658f
  life course approach in, 1649
  menopause and, 1655–1656
  miscarriage in, 1651–1652, 1652b
  pregnancy and, 1652
  pre-pregnancy counselling in, 1650–1651
  stillbirth in, 1651–1652
  urinary incontinence and, 1657
  uterine fibroids and, 1657–1658, 1658f
Feminizing surgeries, 1678–1679
Femoral head, avascular necrosis, 1075
Femoral neck fracture, 1075
Fetal circulation, 469–470, 469f
α-fetoprotein (AFP), 687
FEV. *See* Forced expiratory volume
Fever, 1443–1444
FI. *See* Frailty Index
Fibrates, 519
Fibreoptic bronchoscopy, 536, 536f, 536b, 564f
Fibrinogen, in DIC, 1327b
Fibrinolysis, 1318–1319, 1320f
Fibrinolytic system, 1319
Fibroadenoma, 1641, 1641b
Fibrocartilaginous joints, 1064–1065
Fibroids, uterine, 1657–1658, 1658f
Fibromuscular dysplasia, of renal arteries, 820–821
Fibromyalgia, 1029, 1079, 1080f
Fibronectin glomerulopathy, 814–815
Fibropolycystic liver disease, 726–727
Fibrotic bands, 1226
Fibrotic goitre, 1173
Fibrotic lung diseases, 536
Fidaxomicin, 1476
Fifth disease, 1450
Filariasis, 1509, 1509b
Filovirus, 1445b
Finger clubbing, 530, 531f, 531b
First carpometacarpal osteoarthritis, 1071
First generation antipsychotics, 1028
Fiscal interventions, in behaviour change, 159, 160b
FISH. *See* Fluorescence in situ hybridization
Fissures, 1568f–1570f
Fistula in ano, 668, 668f
Fitness, air travel and, 372–373
Fitness-to-fly, 372–373
Fitz-Hugh-Curtis syndrome, 676
5Ms of geriatric medicine, 1694, 1695b
Fixed drug eruptions, 1629
Flap failure, surgery and, 348
Flaviviruses, 1445b, 1455b
Flea-borne spotted fever, 1490b
Flexible sigmoidoscopy (FS), 615
Flexor tenosynovitis, 1070
Flow cytometry, 1335, 1335f–1336f
Flow-volume loops, 526, 527f, 534
Flucytosine, 1497
Fluid balance
  in acute kidney injury, 230
  surgery and, 349–350
Fluid resuscitation, 284
Fluids, 853–884, 853b–854b, 854f
Fluorescence in situ hybridization (FISH), 1385
Fluoride, 778
Fluoroscopy, 125–126
Flushing, 1605
Focal epilepsy, 924–925
Focal motor seizures, 922–923
Focal non-motor seizures, 922
Focal segmental glomerulosclerosis (FSGS), 804, 805f
Folic acid, 1285–1286
  antagonists, 1389
  deficiency, 1285–1286, 1286b
  dietary intake of, 1286
Follicular lymphoma, 1353–1354
  chromosome translocations, 1352b
  clinical features and course, 1353–1354
  management, 1354
  prognosis, 1354, 1354b
Follicular occlusion triad, 1576b
Folliculitis, 1589–1590
Fomepizole, 367–368
Fondaparinux, 1371
Food, fortification of, 780
Food-borne infections, prevention of, 1439
Food poisoning, 1476, 1477b–1478b
Foot pain, 1077–1078, 1078b
Forced expiratory volume (FEV), 318, 534f
Foreign body
  inhaled, 539, 539b
  in nose, 592
  in throat, 597, 598f
Fortification, of foods, 780
Fosfomycin, for sepsis, 295
Fourth heart sound, 397
Foxglove, poisoning, 370.e6
Fractional exhaled nitric oxide (FENO), 535, 542
Fragility fractures, falls and, 1707–1709
Frail patients, dialysis in, 846
Frailty, 1696–1698, 1698f
  cycle, 1699f
  phenotype, 1697, 1697b
  recognition, 1697
  surgery and, 340
Frailty Index (FI), 1697
Frank-Starling curve, 303f, 390, 390f
FRC. *See* Functional residual capacity
Freckles (ephelides), 1626
Free gas, 120, 121f
Free radical theory, ageing and, 1695
Free-living amoebae, 1508
Freezing cold injury, 378
Fresh frozen plasma, 1313
Fridericia's cube root formula, 398, 399b
Fried's phenotypic frailty assessment, 1697b
Frontotemporal dementia, 957
Frostbite, 378
Frostnip, 378
Frozen shoulder, 1070
Fructosamine, 1223
FSGS. *See* Focal segmental glomerulosclerosis
Fulminant colonic disease, 655
Functional capacity assessment, 339–340, 340f
Functional diarrhoea, 644
Functional incontinence, in older adults, 1715
Functional movement disorder, 254
Functional neurological disorder (FND), 253–254, 1029–1030
Functional residual capacity (FRC), 525, 535f
Fungal infections, 1493, 1493b
  bone and joint, 1107
  HIV-associated, 1554–1557
  skin and soft tissue, 1595–1598
  subcutaneous, 1495–1496
  superficial, 1496
  systemic, 1493–1495
Fungal rhinosinusitis, 593
Furuncles (boils), 1590

G

GAD. *See* Generalized anxiety disorder
Gadolinium chelates, MRI and, 137b
Gaisböck's syndrome, 1306
Gait, 907–908, 907b
Gait abnormalities, in Parkinson's disease, 930
Gait apraxia, 908
Galactosaemia, 1246.e3b
Gall bladder
  carcinoma of, 745
  motility, 737
  polyps, 744–745
  ultrasound of, 130f
Gall bladder disease, 743–744
Gallstones, 736–743
  clinical features, 738–739, 739f
  complications, 740, 743
  differential diagnosis, 739–740
  formation, 736f
  investigations, 739
  management, 740–741
  pancreatitis and, 749, 752
  types of, 737
Gambling, young adults and, 1636
Gamma-glutamyl transpeptidase, 687
Ganciclovir, 1465–1466
Ganglion, 1071
Gangrene, gas, 276
Gardner's syndrome, 671
Gas gangrene, 276
Gas transfer, 535
Gastric antral vascular ectasia, 225f
Gastric cancer, 1396b, 1402
Gastric carcinoma, 641
Gastric emptying, 146, 605
Gastric lavage, 358
Gastric outlet obstruction, 635, 635f
Gastric polyps, 640
Gastric protection, 299

Gastric tumours, 637–640
  clinical features of, 638–639
  diagnosis of, 639
  early, 638
  epidemiology and pathogenesis of, 637–638
  H. pylori infection and, 637, 638f
  management of, 640
  pathology of, 638
  screening of, 638
  staging of, 639, 639b
Gastric ulcer, 633, 633f
Gastrinoma, 760
Gastritis, 632
Gastroenteritis, 1472–1478
  bacterial, 1472–1478, 1472b
  viral, 1462b
Gastroenterology, 601–678, 601b–602b
Gastrografin, 619
Gastrointestinal bleeding, 640–642
  acute lower, 641, 641f–642f
  acute upper, 640–642
  chronic, 641
Gastrointestinal disease
  acute abdomen, 674–678
  chronic kidney disease and, 842
  clinical features, 612–617
  complications of obesity, 1258–1259
  functional, 619–620, 620f
  history and examination of, 602, 602b
  investigation of, 617–619
  peritoneum, 677
  pharynx and oesophagus, 621–632
  in rheumatic disease, 1118
  salivary glands, 621, 622b
  in SLE, 1109
  tuberculosis, 650
Gastrointestinal infections
  acute, 222
  bacterial, 1472–1478
  low-/middle-income countries, 1487–1489
  protozoal, 1506–1508, 1506b
  viral, 1461–1462
Gastrointestinal stromal tumours (GISTs), 639, 1403
Gastrointestinal syndrome, 382
Gastrointestinal tract, anatomy and physiology of, 604–612
Gastro-oesophageal reflux disease (GORD), 621–625, 622f–624f, 622b
Gastroparesis, 636
Gastropathy, 632
Gastroscopy, 617, 618b, 639, 639f
Gastrostomy tube placement, 146b
Gate theory of pain, 911
Gaucher disease, 1246.e2b
GAVE. *See* Gastric antral vascular ectasia
G-band karyotyping, 1335
GCA. *See* Giant cell arteritis
Gender, 1670b
  in medicine, 1639–1679, 1639b
  in osteoarthritis, 1084
  science of, 1670
  terminology, 1670, 1670b
Gender dysphoria, 1670b, 1673b
Gender identity, in young adult, 1638
Gender incongruence, 1670b, 1672f, 1673–1679, 1673b, 1675b
Gender-affirming care, 1670b
Gene editing, 53
Gene expression, control of, 58–59
Gene function, identification of, 72
Gene therapy, 72–73, 74f
General hospital psychiatry, 1001–1043
  core skills and knowledge in, 1001b
  clinical skills in, 1002–1009
  ethical and legal issues in, 1010–1012
  legal framework for mental health treatment, 1008f, 1011–1012
  patients wanting to leave hospital, 1011
General paralysis of insane, 942
Generalized anxiety disorder (GAD), 1022, 1022b
Generalized epilepsy, 923–924
Generalized vascular calcification, 841
Genetic anticipation, 67–68
Genetic diagnosis, pre-implantation, 73–74, 75f
Genetic disorders, 62–69, 62b–63b
  chromosomal, 62–64
  different diseases caused by same gene, 68–69
  different genes cause same disease, 68–69
  mutations, 64–65
  nonsense mutations, 65
  point mutation, 65
  prevalence of, 62b
  single-gene disease, patterns of inheritance of, 66–69
  special presentations of, 67–68
  splicing mutations, 65
  testing for, 55–56
Genetic factors
  in Alzheimer's disease, 956–957
  in asthma, 541
  in coeliac disease, 645
  in colorectal cancer, 672
  in depression, 1017–1019
  in diabetes mellitus, 1206, 1209
  in hypopituitarism, 1150
  in inflammatory bowel disease, 653–654
  in multiple sclerosis, 935
  in pancreatitis, 753–754, 754f
  in Parkinson's disease, 929, 929b
  in rheumatoid arthritis, 1088
Genetic family history, taking, 52–53, 53b
Genetic medicine, 52–53
Genetic neuropathies, 968
Genetic polymorphisms, 71, 111b
Genetic testing, 58b
Genetics
  cancer, 1380
  cell and, 61, 61f
  cellular basis of, 57–62
  clinical, 54–56
  techniques for studying genetic code, 69–72
Geneva Score, 181b
Genital herpes, 1531, 1532f
Genital itching, 1525b
Genital lumps, 1524b
Genital mutilation, female, 1660, 1660b
Genital ulceration, 1525b
Genitalia, nerve supply in, 913b
Genome databases, 70–71
Genome detection assays, for HIV, 1545
Genome-wide association studies (GWAS), 71, 72f
Genomic medicine, 72–74
Genotype, 55f
Geographic tongue, 621, 621f
Geriatric medicine, 1694–1719
  ageing, 1695, 1696f
  Comprehensive Geriatric Assessment (CGA) in, 1699–1703, 1702b
  core skills and knowledge, 1694b
  frailty in, 1697, 1698f
  investigations in, 1699, 1700b–1701b
  multimorbidity in, 1697
  sarcopenia in, 1697, 1699f, 1699b
Gestational diabetes, 1236–1237, 1689, 1689b–1690b
Gestational hypertension, 1683
GFR. *See* Glomerular filtration rate
GH. *See* Growth hormone
Giant cell arteritis (GCA), 920, 1115, 1116f
Giant cell myocarditis, 481
Giardiasis, 1507, 1507f
Gigantism, 1147b, 1185
Gilbert's syndrome, 690
Gingivostomatitis, 1446f
GISTs. *See* Gastrointestinal stromal tumours
Gitelman's syndrome, 871
Glands, primary and secondary failure, 1142
Glasgow Admission Prediction score (GAP), 180
Glasgow alcoholic hepatitis score, 725b
Glasgow Coma Scale, 233b, 328–329
Glasgow scoring system for pancreatitis, 752b
Glasgow-Blatchford score, 181b, 225b
Glaucoma, 1055–1057, 1057f
Glioblastoma, MRI of, 140f
Gliomas, 944–946, 945f, 1412–1413
Global aphasia, 890
Global burden of disease, 168
Global health, 167–174
  accidents, 169–170, 170b
  conflict and catastrophe, 172
  core skills and knowledge in, 167b
  economics and politics, 173
  education, 171
  environmental degradation and pollution, 171
  health spending and resources, 169f
  human migration, 172
  human rights, 173
  infectious diseases, 168
  mental health, 169
  mother and child, 168–169
  oral disease, 170
  organizations involved, 172–173, 173b
  pandemics, 168
  poverty, 170–171, 170b
  scale of problem, 167–168
  social determinants of, 170–172
  surgery, 170
  Sustainable Development Goals, 168f, 173
  trauma, 170, 170b
  value of engagement in, 173
  water and sanitation, 171
Global Registry of Acute Coronary Events (GRACE) Score, 181b
Globulin, 687
Globus pharyngeus, 597–598
Glomerular basement membrane, 793f
Glomerular disease, 800–817, 802b
Glomerular filtration rate (GFR), 794–796
  creatinine clearance in, 796, 796f
  estimated, 796
  measurement of, 795–796, 796b
  pressures controlling, 795f
Glomerulonephritis, 808–813
Glomerulopathy, associated with pre-eclampsia, 817
Glomerulus, 800–817, 801f
Glossitis, 621
Glossopharyngeal nerve (IX), 902, 902b
GLP-1 receptor agonists, 1220
GLP2 analogues, 787
Glucagon test, 1148b
Glucagon-like peptide 1 receptor agonists, 1265, 1266f–1267f
Glucagonoma, 760
Glucocorticoids, 1152, 1152b
  actions of, 1152b
  basal levels, 1153
  osteoporosis and, 1132
  physiology, 1152–1153
  production, 1153f
  suppressible hyperaldosteronism, 1161
Glucocorticosteroids, 106
  for Crohn's disease, 660
Glucose
  fasting, diabetes mellitus, 1212
  hormonal regulation, 1204
  metabolism, 1204, 1300f
  production, 1204
Glucose *(Continued)*
  profile, 1205f
  transport, 1204, 1205f
  in urine, 799
  utilization, 1204
Glucose-6-phosphate dehydrogenase deficiency, 1299, 1300f–1301f, 1301b
Glucose tolerance test (GTT), 1212b
Glucose transporter 1 (GLUT-1) deficiency (AD), 1246.e3b
1a (glucose-6-phosphatase deficiency), 1246.e3b
1b (glucose-6-phosphatase translocase deficiency), 1246.e3b
Glue ear, 587f, 588–589
Gluten intolerance, non-coeliac, 647
Gluten-sensitive enteropathy, 645
Gluteus medius tendinopathy, 1075
Glycogen storage disorders, 1246, 1246.e3b
Glycopeptides, for sepsis, 291b, 292–293
Goitre, 1173–1174, 1173b
Golfer's elbow, 1070
Gonadal dysgenesis, 1185b
Gonococcal arthritis, 1106
Gonococcal conjunctivitis, 1052
Gonorrhoea, 1526, 1527f
Goodpasture's syndrome, 574, 811
GORD. *See* Gastro-oesophageal reflux disease
Gordon's syndrome, 872
Gottron's papules, 1113, 1113f
Gout, 256b–257b, 842, 1103, 1103f
Graft failure, surgery and, 348
Graft-*versus*-host disease, transfusion-related, 1311–1312
Grand mal (tonic-clonic) seizures, 922
Granular cell tumours, of breast, 1643
Granulocyte concentrates, 1313
Granuloma annulare, 1587
Granulomatous lobular mastitis, 1645
Granulomatous tubulointerstitial nephritis (TIN), 835f
Graves' dermopathy, 1168
Graves' disease, 1167–1168
Graves' orbitopathy, 1169f, 1171–1173, 1172f
Gravitational eczema, 1581
Grief, 1024
Growth hormone (GH), 1186b
  antagonists, 1188
  deficiency, hypopituitarism and, 1149
  physiology, 1185, 1186f
Growth hormone-releasing hormone (GHRH), 1186b
GTT. *See* Glucose tolerance test
Guillain-Barré syndrome, 965–966
Gums, 621, 622b
Gut-acting cholesterol-lowering drugs, 519
Gut decontamination, 358
Gut-liver axis, 710–711
Gut motility, 606
Gut-brain axis, 1251
GWAS. *See* Genome-wide association studies
Gynaecological cancers, 1410–1412
Gynaecological malignancy, in female health, 1658–1659, 1659b
Gynaecomastia, 1179–1180
Gyromitrin poisoning, 370.e6

H

H_2-receptor antagonists, 624
Haemangioma, liver, 730
Haemarthrosis, 256b–257b, 1076
Haematemesis, 223–226
Haematinic deficiency, 839
Haematological disease, 1273f, 1605
Haematological oncology, 1328–1362, 1328b, 1329f, 1330b
  aetiology, 1330–1337, 1331f
  biochemistry, 1334–1335
  clinical skills for, 1328–1329
  cytogenetics, 1335
  history of, 1328
  immunophenotyping, 1335
  investigations, 1332–1337
  minimal or measurable residual disease testing, 1335–1336
  molecular genetics, 1335
  morphology, 1333–1334, 1334f
  multidisciplinary team, 1337
  myeloma, 1358–1362, 1359f
  pathogenesis, 1331–1332, 1333f–1334f
  radiological imaging, 1337
  specialist integrated haematological malignancy diagnostic service, 1337
  tissue typing, 1336–1337
  treatment of, 1328–1329, 1330b
Haematological tests, 687
Haematology, 1270–1327, 1270b, 1272f
Haematuria, 798, 798f
Haemobilia, 744
Haemochromatosis, hereditary, 720, 720f
Haemodialysis, 844–845
  access for, 844, 845f
  for acute kidney injury, 231
  aims of, 844–845
  complications of, 845
  dialysate concentrations in, 844b
  initiation of, 843–846
  in poisoning, 359
  principles of, 844, 844f
Haemodynamic changes, 311, 311f
Haemofiltration, 231, 845, 845f
Haemoglobin
  alterations related to plasma, 1279f
  electrophoresis, 1294f
  function, 1276–1277, 1279f
  model of molecule, 1278f
  synthesis, 1276, 1278f
  types of, 1292b
Haemoglobin A_{1C}, 1211, 1223, 1223b
Haemoglobinopathies, 1291–1292, 1292f–1293f
Haemolysis
  consequences of, 1288–1289
  diagnostic and haematological features of, 1289f
  evidence for, 1289
  extravascular, 1288, 1303f
  intravascular, 1288–1289
  sites of, 1288–1289
Haemolytic anaemia, 1288–1289, 1290b
Haemolytic disease of newborn (HDN), 1304
Haemolytic jaundice, 690–692
Haemolytic uraemic syndrome, 817–818, 817f
Haemopexin, 1288–1289
Haemophilia, 72–73, 74f, 1119
Haemophilia A, 1324–1325, 1324b
Haemophilia B (Christmas disease), 1325
Haemopoiesis, 1271
Haemopoietic growth factors, 1271–1273
Haemopoietic stem cell transplantation, aplastic anaemia, 1287–1288, 1287b
Haemopoietic syndrome, 382
Haemoptysis, 218–219, 529, 529b
  causes of, 218, 218b
  clinical assessment of, 219
  investigations of, 219
  management of, 219, 219b–220b
  massive, 218
Haemorrhagic intra-alveolar exudate, 326
Haemorrhoids, 667
Haemosiderin, 1281, 1288–1289

Haemostasis, 1316–1321, 1369
  platelets. *See* Platelets
  vessel wall, 1316, 1317f
Hair, chemotherapy effects, 1390, 1391f
Hair disorders, 1627–1628
Hair loss (alopecia), 1627–1628, 1627b
Hairy cell leukaemia, 1358
Hald's rings, 1663f
Hallpike manoeuvre, 239, 240b
Hallucinogen mushroom poisoning, 370.e6
Hallux rigidus, 1078
Hallux valgus, 1077–1078
Hamartoma, 1643
Hamartomatous polyps, 671
Hand, foot and mouth disease, 1448
Hand and wrist pain, 1070–1071, 1071b
Hand eczema, 1581–1583, 1581f–1582f
Hand hygiene, 1438
Handover processes, 181
Hansen's disease. *See* Leprosy
Hantaviruses, 1464
HAS. *See* Human albumin solution
Hasenclever score, 1350b
Hashimoto's thyroiditis, 1165, 1173
HAV. *See* Hepatitis A virus
Hazard ratio, 47–48
HBV. *See* Hepatitis B virus
HCV. *See* Hepatitis C virus
HDV. *See* Hepatitis D virus
Head and neck cancer, 1413–1414, 1414f
Head impulse test, 901f
Head lice (pediculosis capitis), 1599
Headache, 239–249, 915–921
  causes of, 241b
  chronic daily, 919
  clinical assessment of, 239–242, 241b
  history for, 916, 916b
  investigations for, 242, 916
  secondary, 916b, 919–920
Health
  ageing and, 1719
  improvement in, 154f, 158–159, 158b
  inequalities in, 155–157, 156f, 156b
  protection of, 159–163
Health conditions, in young adult, 1634b, 1636–1637, 1636f–1637f, 1637b
Health literacy, of young adult, 1638
Health needs assessments (HNAs), 154f, 155, 156f
Healthcare-associated infections, 1438–1440
  care bundle approach to, 1438, 1438b
  eradication of reservoir in, 1439
  prevention and control measures for, 1438–1439
  protection for travellers to developing/tropical countries, 1440, 1441b
Healthcare public health, 163–166
  antimicrobial resistance and, 164–165, 164b
  health economics and, 165
  screening for, 163–164, 164b
  service delivery for, 165–166
Healthcare workforce, 173, 174f
Healthy life expectancy, 155, 156f
Hearing loss, 586b, 588–589, 589b, 900b
Heart, 388–393
  adrenergic stimulation of, 390
  blood vessel control, 391–393
  cardiac cycle of, 390–391, 391f
  conduction system of, 388–390, 388f
  coronary circulation of, 391, 392f
  as hypertensive target organ, 503
  nerve supply of, 390
  in rheumatoid arthritis, 1093
  Starling's law, 390, 390f
  vascular endothelium, functions of, 391–393
Heart block, 414–418, 418b, 451
Heart disease
  acute coronary syndromes, 444–452
  clinical features of, 393–394, 394b, 432b
  ischaemic, surgery of, 341–342
  Marfan's syndrome, 475–477
  pericardial, 487–490
  pregnancy, 1685–1686, 1685b
  pulmonary, 477–480
  sarcoidosis, 570
  SLE, 1109
  valvular, 207, 452–469
Heart failure, 429–437, 429b, 450
  acute, 207–211
    clinical assessment of, 207–209
    clinical syndromes, 209b, 442b
    discharge of, 211
    investigations of, 209
    management of, 209–211, 209f, 210b, 443f
    pathophysiology of, 207
  arginine vasopressin in, 431
  causes of, 432b
  classification of, 437b
  clinical features of, 431, 432b
  clinical syndromes of, 429b, 431–432
  compensatory physiological response, 438f
  diabetes mellitus and, 1233
  diagnosis of, 431, 432f, 432b
  endothelial function in, 430–431
  management of, 432–437
    drug, 433, 434b, 435f
    lifestyle advice, 432–433
    monitoring, 433
    multidisciplinary team approach, 433
    non-pharmacological, 436–437
  palliative care, 1729–1731
  pathophysiology of, 429–431, 429b
  stages of, 439f
Heart murmurs, 397
Heart sounds
  first, 396–397
  fourth, 397
  second, 397
  third, 397
Heartburn, 622
Heart-lung transplantation, 577
Heat
  acclimatization, 379
  cramps, 379
  rash, 379
  stroke, 379
  syncope, 379
Heat illness, 379
Heat stroke, 379
Heavy chain diseases, 1362
Heavy drinker, 1032
Heavy menstrual bleeding, 1649–1650, 1650f
Hedonic hunger, 1251, 1252f
Heel pain, 1078, 1078b
Helicobacter pylori infection, 632–637, 632f, 1478
  antral gastritis, 633
  diagnosis of, 633–634, 634f
  epidemiology of, 632
  eradication therapy for, 634–635
  management of, 634–635
  measures for, 635
  pathogenesis of, 633
  regimens for, 635
  results of, 633
HELLP syndrome, 1684
Helminthic infections, 1508–1509, 1508b–1509b
Hemiballismus, 933
Hemifacial spasm, 899
Hemiparesis, 891, 904, 907
Hendra virus, 1460
Henoch-Schönlein syndrome (purpura), 816, 1321
Heparin, 1370–1371
Heparin-induced thrombocytopenia, 1371
Hepatic mitochondrial injury syndromes, 730–732
Hepatic sinusoidal obstruction syndrome, 726
Hepatic stiffness, 688
Hepatitic damage, from drugs, 732–733, 733b
Hepatitis, 692–704, 1462
  acute, 692–694, 694f, 703–704
  associated conditions
    cytomegalovirus, 704
    hepatocellular carcinoma, 1462–1463
    herpes simplex, 704
    toxoplasmosis, 704
    yellow fever, 704
  Brunt scoring system, 694
  chronic, 692, 694b
  clinical features, 695b
  Ishak score, 694
  Knodell Scoring System, 694
  METAVIR Scoring System, 694
  non-A-E, 703
  viral, 694–703
Hepatitis A virus (HAV), 694, 695b, 696f
Hepatitis B virus (HBV), 695, 1525
  acute, 698, 699f
  arthritis, 1107
  cancer and, 1381
  chronic, 699, 700b
  clinical features, 695b, 698
  epidemiology, 695–697
  genome, 697f
  HIV-associated, 1558
  horizontal transmission, 696
  investigations, 698
  management of, 698–699
  mutants, 696–697
  pathogenesis, 697–698, 698f
  prevention/prophylaxis, 699
  prognosis, 698
  screening for, 1521
  vaccination for, 1521
  vertical transmission, 696
  viral markers, 699b
  viral proteins, 697b
Hepatitis C virus (HCV), 701, 702f
  chronic, 702
  clinical features, 695b, 701
  course, 702f
  epidemiology, 701
  HIV-associated, 1558
  investigations, 701
  management, 701
  prognosis, 701
  screening for, 1522
Hepatitis D virus (HDV), 695b, 701
Hepatitis E virus (HEV), 695b, 703, 703f
Hepatobiliary disease, ultrasonography of, 130, 130f
Hepatocellular carcinoma, 707, 729, 731f–732f, 1405
  aetiology, 729
  clinical features, 730
  hepatitis virus and, 1462–1463
  investigations, 730
  management of, 1405
  pathology, 730
  prevention, 730
  prognosis of, 1405
Hepatology, 679–734, 681f
  clinical skills for, 680–734
  core skills and knowledge for, 679b
  drug prescribing, 733–734
  examination for, 680
  history for, 680, 680b
  investigations, 680, 680f, 686–689, 687b, 707
Hepatomegaly, 691, 692b
Hepatopulmonary syndrome, 717
Hepatorenal syndrome, 717, 836
Hepatosplenomegaly, 570
Hepatotoxicity, drugs, 732–733
Hepcidin, 1281
Hepevirus, 1445b
HER2-positive disease, breast cancer, 1400
Hereditary angio-oedema, 1588
Hereditary ataxias, 961–962
Hereditary diffuse gastric cancer syndrome, 1383b
Hereditary elliptocytosis, 1291
Hereditary fructose intolerance, 1246.e3b
Hereditary haemorrhagic telangiectasia, 1321
Hereditary neuropathy with pressure palsies, 964
Hereditary spastic paraparesis, 951
Hereditary spherocytosis, 1289–1290, 1290f–1291f
Heritable thrombophilias, 1367–1368
Hermaphroditism, 1185b
Herpangina, 1448
Herpes simplex conjunctivitis, 1053, 1055f
Herpes simplex keratitis, 1054
Herpes simplex virus (HSV), 1445–1446, 1446f, 1592, 1592f
  diagnosis and management of, 1446–1447
  diseases caused by, 1446b
  hepatitis, 704
  HIV-associated, 1558–1559
  HSV-1, 1445, 1446f
  HSV-2, 1446
  management, 1592
  urethral discharge and, 1527
Herpes zoster (shingles), 942, 1385f, 1447, 1447f, 1592–1593, 1593f
Heterozygotes, hereditary haemochromatosis, 721
Heterozygous familial hypercholesterolaemia, 513–514, 514b
HEV. *See* Hepatitis E virus
HGD. *See* High-grade dysplasia
HHV-6, 1450
HHV-7, 1450
HHV-8, 1463
Hiatus hernia, 622, 622b
Hibernating myocardium, 437
Hiccups, 580
Hidradenitis suppurativa, 1576, 1645
High altitudes, 374–375
High-density lipoprotein (HDL), 511f, 512
  disorders of, 516
High-dependency unit (HDU), 297
  discharge from, 300
High-frequency oscillation, 323
High-grade dysplasia (HGD), 625
High-stepping gait, 908
High-flow nasal oxygen (HFNO), 184–186, 192b
High-intensity focused ultrasound (HIFU), 152
High-velocity trauma, 1048, 1050f
HINTS examination, 239, 240b, 901
Hip fracture, surgery for, 350
Hip pain, 1074–1075, 1074b
Hippocampal sclerosis, 924
Hirsutism, 1181–1182, 1628
His bundle, 388–389
Histiocytoma. *See* Dermatofibroma
Histology
  of breast cancer, 1396b
  of cancer, 1385–1386
  of coeliac disease, 646, 646b
  of Crohn's disease, 655, 657b
  of IgA nephropathy, 810, 810b
  of SLE, 1109
  of tuberculosis, 651
  of ulcerative colitis, 655, 657b
Histoplasmosis, 1494
History-taking, 3–4, 4b
  drug history, 4
  family history, 4
  golden minute, 3, 4b
  meeting of two experts, 3
  past medical and surgical history, 4, 5b
  presenting symptom, 3–4, 4b–5b
  social history, 4
HIV. *See* Human immunodeficiency virus
Hoarseness, 596b, 598–599
Hodgkin lymphoma (HL), 1349–1351
  aetiology, 1349
  classification, 1349b
  clinical features, 1349, 1349f
  diagnosis, 1349, 1349f
Hodgkin lymphoma (HL) *(Continued)*
  investigations, 1349
  long-term follow-up, 1351
  management, 1349–1351
  prognostic groups, 1350b
  staging, 1350b
  survival, 1396b
Homozygotes, hereditary haemochromatosis, 721
Homozygous familial hypercholesterolaemia, 514
Hookworm, 1512
Hormonal contraception, 1653, 1654b
Hormone-binding proteins, 1142–1143
Hormone receptors, 1140–1141
Hormone replacement therapy
  hypogonadism, 1179, 1179b
  hypopituitarism, 1149b
  hypothyroidism, 1167
  for menopause, 1656, 1656b
  for osteoporosis, 1132
  stoke and, 997
  in transgender women, 1675–1676
  in transgender men, 1676–1678, 1677b
Hormones, 1140–1143
  control and feedback, 1141–1142, 1141f
  deficiency, 1148
  ectopic secretion, 1199
  excess, 1142, 1147–1148
  measurement, 1142–1143, 1142f
  primary and secondary gland failure, 1142
  resistance, 1142
  secretion, 1143
  stress-related, 1144
  synthesis, storage, and release of, 1140–1143
  transport, 1140–1141
Horner's syndrome, 895, 896b
Hospital-acquired infection, 283
Hospital-acquired pneumonia, 215, 216b
Host response to infection, 1443
Hot injury, 378–379
Household products, poisoning with, 368
Housemaid's knee, 1076
Howell-Jolly bodies, 1291f
HSV. *See* Herpes simplex virus
HTLV. *See* Human T-cell leukaemia virus
Human albumin solution (HAS), 315
Human chromosomes, 57–58
Human genetics, 52–78
  cellular basis of, 57–62
  clinical, 54–56
  clinical skills for genetic medicine, 52–53
  core skills and knowledge of, 52b
Human immunodeficiency virus (HIV), 1458, 1525–1563
  acquisition, 1539–1540
    biomedical prevention of, 1524
    contaminated blood, blood products, and organ donation in, 1540
    contaminated needles in, 1540
    routes of, 1539
    sexual transmission in, 1539–1540
    vertical transmission in, 1540
  AIDS-dementia complex, 942
  antibodies, 1545
  associated cancers, 1561
  associated disease, 1554–1561, 1555b
  changing face of epidemic, 1539
  clinical approach to persons with, 1545–1547
    genotyping, 1546
    initial assessment, 1545–1546, 1546b
    monitoring, 1546–1547
  clinical latency for, 1542, 1543f

Human immunodeficiency virus (HIV) *(Continued)*
clinical skills for care of people living with, 1538
conditions associated with immunosuppression, 1544–1545, 1545b
core skills and knowledge, 1537b
end-organ disease, 1543–1544
cardiac complications, 1544
endocrine complications, 1544
eye disease, 1543
gastrointestinal effects, 1544
haematological complications, 1544
kidney complications, 1544
mucocutaneous manifestations, 1543–1544, 1543b
neurological disease, 1543
respiratory complications, 1544
enteropathy, 1544
epidemiology of, 1538–1539
global strategies in, 1563
in high-income countries, 1539
history of, 1538, 1538b
management of, 1547, 1548b
musculoskeletal involvement, 1107
pathogenesis of, 1540–1543, 1541f, 1542b
pre-exposure prophylaxis in, 1563
pregnancy, 1553–1554
prevention and control of, 1561–1563
seroconversion for, 942
skin and, 1600–1602, 1601f
strategies for reducing transmission of, 1562–1563
in sub-Saharan Africa, 1539
symptomatic, 1543–1545
symptoms relevant to, 1538b
testing, 1545, 1561–1562, 1562b
tuberculosis, 557
viral structure of, 1540, 1540f
viral testing for, 1542
Human leucocyte antigens (HLA) molecules, 92–95, 93f–94f
class II, 92
class III, 92
classical class I, 92
disease associations, 91b
disease associations, diabetes mellitus, 1206
genotypes and range of protein products, 92–93
non-classical class I, 92
typing, 646
Human major histocompatibility complex, 92–93, 93f
Human papillomavirus, 1449, 1530, 1531f, 1593, 1594f
cervical cancer and, 1381, 1462
clinical features of, 1530, 1531f
diagnosis of, 1530–1531
HIV-associated, 1559
management of, 1531
prevention/vaccination for, 1531
Human rights, 173
Human T-cell leukaemia virus (HTLV), cancer and, 1381, 1463
Human T-cell leukaemia/lymphotropic virus infection, 1458
Humidifier fever, 573b
Humoral defence, 609, 610f
Huntington's disease, 905b, 932–933
Hydatid cysts, 728, 729f, 1517
Hydatid disease, 728, 729f, 1517
Hydrocephalus, 946–947, 946f
Hydrocoele, 131f, 1668, 1668f
Hydrofluoric acid poisoning, 368
Hydrogen
breath test, 648
renal excretion of, 878, 878f
Hydrogen fluoride poisoning, 368
Hydronephrosis, ultrasound of, 127f
Hydroxychloroquine, for rheumatoid arthritis, 1095–1097
25-hydroxyvitamin D, 1127, 1132
Hyper IgE syndrome, 100
Hyper IgM syndrome, 99
Hyperaldosteronism, 1160, 1160b, 1691
aetiology, 1160
glucocorticoid-suppressible, 1161
history and examination of, 1160
investigations, 1160–1161
management, 1161
pathophysiology, 1160
pregnancy, 1691
syndrome of apparent mineralocorticoid excess, 1161
Hyperbilirubinaemia, congenital, 690
Hypercalcaemia, 145f, 1193
acute, 274, 274b
causes of, 1193b
clinical features, 1194
familial hypocalciuric, 1195
investigations and differential diagnosis, 1194
kidney calculi, 822
of malignant disease, 1199
management, 1194, 1195b
mild, 1194
pathophysiology and aetiology, 1193–1194
sarcoidosis, 570
severe, 1194
Hypercalcaemic nephropathy, 818
Hypercalciuria, 822, 873
Hypercholesterolaemia
familial, 439
with hypertriglyceridaemia, 516–517
polygenic, 516
primary prevention of, 518
risk prediction tables of, 518
secondary prevention of, 518
Hypercoagulable states, 802
Hypercortisolism, pituitary-dependent, 1156
Hyperemesis gravidarum, 1684
Hypereosinophilic syndrome, 575, 575b
Hyperglycaemia, 262–263, 263b
Hyperhomocysteinaemia, 841
Hyperkalaemia, 231–232, 871–872
causes of, 231b, 871, 872b
clinical features of, 872, 872f
follow-up for, 232
Gordon's syndrome, 872
hyperkalaemic periodic paralysis, 872
hyporeninaemic hypoaldosteronism, 871
management of, 231–232, 232b, 843, 872
pseudohypoaldosteronisms, 872
risk states of, 231b
suxamethonium, 872
Hyperkalaemic periodic paralysis, 872, 975
Hyperkinetic movement disorders, 932–934
Hyperlipidaemia, 513–517, 1183
combined, 516–517
cutaneous manifestations, 1605
familial combined, 516
HDL disorders, 516
hypercholesterolaemia, 513–516
management of, 517–520
primary, 513
remnant, 517, 517f
screening of, 512–513
secondary, 513, 513b
treatment of, 520b
Hypermagnesaemia, 873–874, 873b
Hypermobile Ehlers-Danlos syndrome, 1082–1083, 1082b
Hypermobility spectrum disorders, 1082–1083, 1082b
Hypernatremia, 869, 869b
Hyperosmolar hyperglycaemic state, 258b, 261–262, 261b
Hyperoxaluria, in kidney calculi, 822
Hyperparathyroidism, 1119
management, 1194–1195
pregnancy, 1690
primary, 1193
secondary, 840, 1193–1194
tertiary, 840, 1194
Hyperphosphataemia, 876–877, 876b
Hyperpigmentation, 1626
Hyperprolactinaemia, 1187f, 1188
Hypersensitivity pneumonitis, 572–573, 573b
Hypersplenism, 1307
Hypertension, 496–509, 819–821
aetiology and pathophysiology of, 497
blood pressure measurement for, 500–502, 502f
causes of, secondary, 499b–500b
choice of therapy for, 506–507, 507b, 508f
clinical approach to, 498–500, 501f
core skills and knowledge in, 496b
drugs for, 505–506
emergencies in, 508, 509b
endocrine causes of, 1160b
essential, 819
examination and investigations for, 500
familial hyperkalaemic, 872
history of, 498
intracerebral haemorrhage and, 998
isolated systolic, 502
lifestyle changes for, 505, 506b
management of, 819
orthostatic hypotension and, 502
phenotypes of, 502, 502b
poisoning, 357
porto-pulmonary, 718
pregnancy, 1683–1684
primary, 497–498
pulmonary, VTE and, 1366
referral to specialist care, 506–507
renal, 819
renal sympathetic denervation and, 507
secondary, 498
surgery of, 341
target organ damage, 503
targets in, 503
thresholds in, 503, 505b
treatment of, 505–506
Hypertensive encephalopathy, management of, 509b
Hyperthermia
malignant, 973
poisoning, 357
Hyperthyroidism (thyrotoxicosis), 1167–1171, 1168b
causes, 1167b, 1168
differential diagnosis, 1169
history and examination of, 1168–1169, 1169f
investigations, 1169
long-term consequences, 1171
management, 1169–1171
pregnancy and, 1690
special situations, 1171
Hypertrichosis, 1628
Hypertriglyceridaemia, 517–518
Hypertrophic cardiomyopathy, 481–483, 483f
Hypertrophic pulmonary osteoarthropathy, 1118
Hypertrophic scars, 1616
Hyperuricaemia, 1103
causes of, 1104b
control of, 1339–1340
in kidney calculi, 822
pathogenesis of, 1103
Hyperuricaemic nephropathy, 835
Hyperuricosuria, 822
Hyperventilation, 394, 529
Hypervolaemia, 859
Hypoadrenalism, 1150
Hypoalbuminaemia, 802
Hypobetalipoproteinaemia, 517
Hypocalcaemia, 1195, 1196b
Hypoglossal nerve (XII), 902
lesions, 902, 902b
Hypoglycaemia, 263–264, 263b, 1196–1197, 1244
in diabetic patients, 1234
endocrine causes of, 1197
hepatic and renal causes of, 1197
insulin-related, 1218
management of, 1197
with tumours, 1197
Hypogonadism, 1150, 1178, 1178b
men, 1178, 1178b
women, 1180
Hypokalaemia, 870–871
cardiopulmonary resuscitation and, 195
causes of, 870, 870b
clinical features of, 871
management of, 871, 871b
Hypokalaemic periodic paralysis, 975
Hypolipidaemia, 517
Hypomagnesaemia, 873, 873b
Hypomania, 1019–1021, 1021b
Hyponatraemia, 866–869
Addison's disease, 1157
with euvolaemia, 867, 867b–868b
with hypervolaemia, 868–869, 869b
with hypovolaemia, 866–867, 866b–867b
Hyponatraemic encephalopathy, 867–868, 868b
Hyponatremia, 379
Hypoparathyroidism, 1196
Hypoperfusion, 981
Hypophosphataemia, 874–876, 875b
Hypophysitis, 1149, 1191f
Hypopigmentation, 1625–1626
Hypopituitarism, 1149
aetiology, 1150, 1150b
clinical features, 1150–1151, 1151f
genetics, 1150
investigations, 1151, 1151b
management, 1149b, 1151
pathophysiology, 1149–1150
pregnancy, 1691
Hyporeninaemic hypoaldosteronism, 871
Hypothalamic amenorrhoea, 1181
Hypothalamic-pituitary function tests, 1148b
Hypothalamic-pituitary-gonadal axis, 1175–1190
Hypothalamic-pituitary-thyroid axis, 1141f, 1163–1164, 1164b
Hypothalamo-pituitary-adrenal axis, 1151–1162, 1152b, 1154f
anatomy and physiology, 1151–1152, 1152f, 1152b
androgens. *See* Androgens
cortex, 1152–1161
glucocorticoids. *See* Glucocorticoids
medulla, 1161–1162, 1161f
mineralocorticoids, 1152
Hypothalamus, 1145–1146, 1145f–1146f
Hypothermia, 372f, 376–377
accidental, 376
avalanche medicine for, 377
clinical features, 376
cold water immersion in, 376–377
diagnosis, 376, 377b
field management for, 376–377
hospital treatment of, 377
management, 377
poisoning, 357
prevention, 378
Swiss staging of, 376, 377b
urban, 376
wrap, 377f
Hypothesis tests, 43–49
Hypothyroidism, 1165–1167
aetiology, 1165–1166, 1165b
atrophic, 1165
borderline, 1167
history and examination of, 1166, 1166f
investigations, 1166
management, 1166–1167
pathophysiology, 1165, 1165b
pregnancy, 1690
radioactive iodine, 1170
screening, 1167
Hypotonia, 906
Hypovolaemia, 859, 863b
Hypovolaemic shock, 311
Hypoxia, 302–303, 381
Hypoxic-ischaemic brain injury (HIBI), 948

I

IBD. *See* Inflammatory bowel disease
IBS. *See* Irritable bowel syndrome
ICD. *See* Implantable cardioverter-defibrillator
Idiopathic adult ductopenia, 730
Idiopathic fibrillary glomerulopathy, 814
Idiopathic interstitial pneumonias, 570–572, 571b–572b
Idiopathic intracranial hypertension, 919–920, 1259
Idiopathic normal pressure hydrocephalus, 947
Idiopathic Parkinson's disease, 929
Idiopathic pulmonary fibrosis (IPF), 570, 571f
IDL. *See* Intermediate-density lipoprotein
IGF-1, 1186f, 1186b
IgG4-related disease, 103–104, 103b–104b
IgM nephropathy, 814
Ileal resection, 648f
Ileo-anal pouch, 664, 664f
Ileus, paralytic, 348
Imidazoles, 1496–1497
Immersion, in water environment, 380
Immune-based therapies, 105–106
Immune checkpoint inhibitors, 1392
Immune checkpoints, 89f
Immune-mediated neuropathies, 965–966
Immune mediators, 309–311, 310f
Immune reconstitution inflammatory syndrome (IRIS), 1553, 1602
Immune system, 81
acute inflammation, 95–96, 96f
antigen presentation, 92–95, 95f
autoimmune disease, 102–105
cell migration, 91, 92f
cells involved in, 79–81, 80f
cells of, 81
chronic inflammation, 96–97
HLA molecules, 92–95, 93f
innate, 81–85
laboratory investigations of, 94b, 97
molecules of, 79–81
organ rejection in clinical transplantation, 105, 105b
tissues of, 79–81
Immune thrombocytopenic purpura, 1322
Immunity, 79–106, 79b
Immunization, for HIV, 1554, 1556b
Immunoassay, 1142, 1142f
Immunobullous diseases, 1612–1614
Immunocompromised patients
infectious diseases, 1427, 1428b
pneumonia in, 216
Immunodeficiency
classification of, 98b
clinical, 98
clinical features of, 98, 98b
secondary (acquired), 97–98
severe combined, 99

Immunodiagnostic tests, 1432
Immunoglobulin A (IgA)
  nephropathy, 810, 810f, 810b
  selective deficiency, 99–100
Immunoglobulins, 86–87, 86f, 87b
  intravenous, 106
  liver disease, 687–688
  normal, 1314
  specific, 1314
Immunological monitoring, 1546
Immunology, in renal function, 799
Immunomodulatory drugs, 1392
Immunophenotype, 1335
Immunoproliferative small intestinal disease (IPSID), and small intestinal tumours, 652
Immunoprophylaxis, 1439–1440
Immunosuppressants, for myasthenia gravis, 973
Immunosuppression
  cancer-associated, 1383, 1385f
  transfusion-related, 1312
Immunosuppressive drugs, 106
  for kidney transplantation, 847, 848b
  for SLE, 1110
Immunotactoid glomerulopathy, 814
Immunotherapy, 1439–1440
  for rhinitis, 538
Immunotherapy-induced thyrotoxicosis, 1168
Impetigo, 1589, 1589f
Implantable cardioverter-defibrillator (ICD), 428–429, 428f
Implantable loop recording, 400
Imprinting, 68
IMV. *See* Intermittent mandatory ventilation
Inattention, 891
Incarcerated umbilical hernia, CT of, 135f
Incidentalomas, 1159–1160
Inclusion body myositis, 1114
Inclusivity, principles of, 28–29
Incontinence, faecal, 665–666, 665b, 666f
Increased extracellular volume, 859–863
Indian childhood cirrhosis, 730
Individual interventions, in behaviour change, 158–159, 159f
Infantile hydrocephalus, 947
Infarction, 962
Infections, 1421
  chest X-ray, 118–119, 120f
  clinical approach to, 1421–1433, 1423f
  COPD and, 546
  in diabetic patients, 1233
  epidemiology and prevention of, 1433–1440
    acquisition of, 1434, 1436b
    changing patterns of, 1433–1434, 1434b
    goals in international development, 1434
    increased population mobility, 1433
  host response to, 1443
  kidney transplantation and, 849
  metabolic/immunological consequences of, 1443–1444
  purpura due to, 1321
  starvation-related under-nutrition and, 772, 772b
  symptoms of, and host response, 1442
  vectors of, 1518
Infectious agents, 1421, 1442f
Infectious colitis, 662
Infectious diarrhoea, 642
Infectious disease, 1417–1518. e9, 1420f
  age and, 1425
  allergies/intolerances and, 1427
  animal contact and, 1426
  antimicrobial chemotherapy for, 1439
  classification of outbreaks of, 1436–1437, 1437b
  clinical examination of, 1427–1428, 1428f, 1428b
Infectious disease *(Continued)*
  clinical skills for, 1418–1518
  co-morbidities of, 1427
  core skills and knowledge in, 1418b
  in developing world, 1434
  development of infection by commensal organisms in, 1441–1442
  devices/prostheses/implants and, 1427
  diagnostic assessment of, 1421–1427, 1424b–1425b
  food and drink history and, 1426
  global distribution of, 1422f
  healthcare-associated infections, 1438–1440
  history of, 1418, 1419b, 1421–1427
  host-organism interactions in, 1441
  human infection contact and, 1425–1426
  immunization for, 1439–1440, 1440b, 1441f
  immunization history and, 1427
  immunocompromised patients, 1427, 1428b
  intravenous drug use and needle exposure and, 1426
  investigations of, 1428–1433, 1429b, 1430f, 1431b
  leisure activities, 1426
  medications and, 1427
  mortality of, 1434b
  notifiable, 1437b
  occupational history and, 1425
  prevention and control of, 1437–1440
  returning travellers, 1426, 1426b
  routes of transmission of, 1434–1436
    air-borne spread, 1434
    consumption of infected material, 1436
    direct inoculation, 1436
    direct person-to-person spread, 1435–1436
    endogenous infection, 1434–1436
    faecal-oral spread, 1435
    indirect person-to-person spread, 1436
    vector-borne disease, 1435, 1436b
  sex and, 1425
  sexual activity and, 1426
  specimen collection in, 1432b
  symptoms of, 1425
  travel history and, 1426
Infectious mononucleosis, 703–704
Infective endocarditis, 465–469, 998, 1106
  aetiology of, 465–466, 465f
  clinical features of, 466–467, 467f, 467b
  diagnosis of, 466b, 467
  high clinical suspicion, 467
  investigations of, 467, 468b
  management of, 468, 468b
  prevention of, 469
  rare causes of, 465, 466b
Infective stones, in kidney calculi, 822
Inferential statistics, 43
Inferior vena cava filters, for venous thrombosis, 272
Infiltration, 962
Inflammation
  acute, 95–96, 96f
  chronic, 96–97, 103
  obesity and, 1259
Inflammatory arthritis, 1087
  early inflammatory polyarthritis, 1087
  hand and wrist, 1071
  of hip, 1075
  joint involvement, 1087b
  of knee, 1076
  subgroups of, 1087b
Inflammatory bowel disease (IBD), 222–223, 653–665
  aetiology and pathogenesis of, 653–655, 654f, 654b
    environmental factors, 654
    genetic factors, 653–654
    intestinal immune system, 655, 656f
    intestinal microbiota, 654–655
  cancer in, 664
  differential diagnosis of, 657–658
  enteropathic arthritis associated with, 1102
  epidemiology of, 653–665
  mortality in, 664
  pathology of, 655–657, 657f, 657b
    extraintestinal manifestations, 656–657
    macroscopic changes, 655, 657f
    microscopic changes, 655–656, 657b
Inflammatory dermatoses, 1601
Inflammatory diarrhoea, 614, 614f
Inflammatory myopathies, 971
Inflammatory optic neuropathy, 894, 895b
Inflammatory response, 1444
Influenza, 539, 1453, 1453b
  clinical features, 539
  complications, 539
  diagnosis and management, 539
  HIV-associated, 1559
  prophylaxis, 539
Influenza virus, 1453b
Information-gathering, 3
  history-taking, 3–4
  physical examination of, 4–5, 6f
Inhalers, 543b
Inherited haemolytic anaemia, 1289–1301
Inherited metabolic disease (IMDs), 1241–1244
Inherited polyposis syndromes, 670–671
Innate host defence, in UTI, 829–830
Innate immune responses, 82b
Innate immune system, 81–85
Innate lymphoid cells, 85
Inotropic agents, 316b
  for heart failure, 436
Insect bites, 370.e2
Institutionalization, falls and, 1709
Insulin, 1214–1218
  actions, 1203–1204, 1209
  administration, 1216
  β-blocker overdose, 364
  challenges, 1217–1218, 1218b
  in chronic kidney disease, 842
  deficiency, 1224
  dose adjustment, 1218b
  effects, 1206f
  inhaled, 1217
  injection, 1216
  intermediate- and longer-acting, 1215
  management, whole-pancreas transplantation, 1222
  overdose, 363
  poisoning, 364
  profile, 1205f
  pumps, 1216–1217
  regimens, 1215–1216, 1216f–1217f
  secretion, 1203–1204, 1204f–1205f
  short-acting, 1215–1218
  signaling, 1205f
  strengths of, 1215
  structure, 1205f
Insulin/glucagon ratio, 770
Insulin receptor, 1203–1204
Insulin tolerance test, 1148b
Insulin treatment, critical illness, 300
Insulinomas, 1196–1197, 1197b
Integrase strand transfer inhibitors, 1549–1550
Intention tremor, 906
Intention-to-treat analysis, of evidence-based practice, 39
Intercostal chest drain insertion, 184f, 186b, 187f
Interferons (IFNs), 1392, 1468
Interleukins (ILs), 1392
Intermediate uveitis, 1058
Intermediate-density lipoprotein (IDL), 511, 511f
Intermittent fasting, 1263b
Intermittent mandatory ventilation (IMV), 323
Intermittent positive pressure ventilation (IPPV), 320
Internuclear ophthalmoplegia, 896, 896f
Intersectionality, 28, 28b
Intersex, 1670b
Interstitial lung diseases, 568–576
Intervention context, evidence-based practice and, 50–51, 51f
Interventional oncology, 146b
Interventional radiology (IR), 146–151, 146b
Intestinal failure, 783–787, 783b
  functional classification of, 784, 785b
  type 1, 784
  type 2, 784
  type 3, 784
Intestinal ischaemia, 651
Intestinal lymphangiectasia, 651
Intestinal obstruction, 677, 677b
Intestinal transplant, 787, 788b
Intra-aortic balloon pumping, 411, 412f
Intracellular fluid, 856, 856b
Intracellular signal inhibitors, 1393, 1393b
Intracerebral haemorrhage, 998–1000
  aetiology of, 998–999
  clinical features of, 999, 999f
  investigations for, 999–1000, 1000f
  medical management of, 1000
  secondary prevention for, 1000
  surgical management of, 1000
Intracranial bleeds, 243–245
Intracranial compliance curve, 328f
Intracranial epidural abscess, 944
Intracranial pressure
  measurement of, 329f
  raised, 273–274, 329b
Intrahepatic cholestasis, of pregnancy, 1684–1685
Intraluminal monitoring, 623, 624f
Intraosseous (IO) access, 185f, 185b
Intravascular coronary ultrasound, 404
Intravenous catheter insertion, 146b
Intravenous immunoglobulin, for myasthenia gravis, 974
Intravenous iron, 436
Intravenous thrombolysis (IVT), 988, 991f, 992b
Intravitreal injection, for diabetic nephropathy, 1227
Intussusception, 668–669
Invasive bladder cancer, 1407
Iodine, 778
  deficiency, 779f, 1165–1166
  dietary requirement, 1164
  indications, 1171b
  long-term surveillance, 1170
  radioactive, 1170
Ionizing radiation, 381, 382b
Iris, diabetic retinopathy, 1224–1227
Iritis, 1047b, 1058, 1058f
Iron, 778–779, 1279–1283
  absorption, 1279–1281, 1280f, 1281b
  in blood transportation, 1281
  dietary intake of, 1279
  overload, 1294
  poisoning, 365
  requirements, 1281
  storage, 1281
Iron-binding capacity, 1282–1283
Iron-deficiency anaemia, 1279f, 1281, 1282b
Irritable bowel syndrome (IBS), 642, 642b–643b
Ischaemic colitis, 651, 666
Ischaemic heart disease, surgery of, 341–342
Ischaemic stroke, 979–980, 980f
  investigative work-up for, 996b
  MRI of, 139f
Islet transplantation, 1222
Isolation, healthcare-associated infections and, 1439
Isoxazole poisoning, 370.e7
IV thrombolysis windows, 988, 992b, 993f
Ivabradine, 436
IVC filter insertion, 146b

J

Japanese encephalitis, 1459
Jarisch-Herxheimer reaction, 1534
Jaundice, 690–692, 691f, 693f
Jelly fish stings, poisoning, 370.e2
Jerk nystagmus, 897
Jiggers, 1517
Job's syndrome, 100
Joint infections, 1106–1107
  bacterial arthritis, 1106
  fungal, 1107
  HIV-related, 1107
  septic arthritis, 1091
  viral arthritis, 1106–1107
Joints, 1064
  anatomy and physiology of, 1064–1065
  aspiration, 1064b, 1066, 1067f, 1067b, 1086, 1094
  blood vessels and nerves, 1065
  Charcot's, 1119
  examination of, 1064b, 1067f
  fibrous and fibrocartilaginous, 1064–1065
  juxta-articular bone, 1064
  MRI of, 138, 142f
  neuropathic, 1119
  swollen, 254–257
  synovial, 1064, 1065f
Jugular veins, in mitral stenosis, 453
Jugular venous pressure, 395–396, 396b
  measurement of, 396, 396b
  wave, 396, 396f
Junctional epidermolysis bullosa, 1614
Juvenile hypertrophy, 1641
Juvenile idiopathic arthritis, 1118
Juvenile myoclonic epilepsy, 924, 924b
Juvenile nephronophthisis, 851
Juxta-articular bone, 1064
Juxtaglomerular apparatus, 792, 794f, 796

K

Kala azar, 1504
Kallmann's syndrome, 1150, 1179
Kaplan-Meier plot, 47, 48f
Kaposi's sarcoma, 632, 1624, 1624f
  HIV-associated, 1561, 1561f
  human herpesvirus type 8 and, 1463
Kawasaki's disease, 1117
Kearns-Sayre syndrome, 64
Keloids, 1616, 1616f
Keratinocyte cancers, 1618
Keratoacanthoma, 1617–1618, 1618f
Keratoconus, 1054
Keratoderma blennorrhagica, 1102
Ketoacidosis, 882
Ketogenesis, 258f
Ketonuria, 258b
Ketosis, 258b
Kidney
  anatomy and physiology, 792–797, 793f
  calculi, 821–824
    aetiology of, 821–822, 822b
    clinical features of, 822–823, 823b
    CT of, 134f
    hypercalcaemia and, 1194
    investigations of, 823, 824f
    management of, 823
    type and frequency of, 821b
  carbuncle, 832–833
  glucose reabsorption, 1210

Kidney *(Continued)*
as hypertensive target organ, 503
ultrasonography of, 127–128, 127f–128f
Kidney disease, 789–852, 789b
asymptomatic, 808–809
atheroembolic, 821
clinical skills for, 790–852
diagnostic approach to, 790, 792b
examination in, 790
history in, 790, 790b
proteinuria quantification in, 790, 790b
conditions presenting with
antiphospholipid syndrome, 818
hypertension, 819–821
multiple myeloma, 818–819, 818f
thrombotic thrombocytopenic purpura, 818
cystic, 849–851
drug-induced, 828b
drugs used in, 828–829, 829b
hyperuricaemic nephropathy, 835
investigations of, 797–800
involvement in other diseases, 817–819
mixed nephritic/nephrotic, 813–816
pregnancy, 1691–1692
prescribing in, 112–113, 829b
tubulointerstitial nephritis, 833–835
Kidney replacement therapy, 837, 843–846
Kidney transplantation, 846–849
anatomy, 847f
complications of, 847–849
considerations in successful, 847–849
donor and recipient HLA mismatches in, 847
immunosuppression for, 847, 848b
living donation, 847
pregnancy, 1691–1692
Klinefelter's syndrome, 1179, 1185b
Knee jerk reflex, 907f
Knee joint effusions, 1075
Knee pain, 1075–1077, 1075b
around knee, 1076–1077
within knee, 1075–1076
rheumatoid arthritis, 1092
Knodell Scoring System, 694
Koilonychia, 1626
Krebs cycle, obesity and, 1259–1260
Kwashiorkor, 771
Kyphoscoliosis, 579

L

Labyrinthitis, 589b, 590
β-lactam/β-lactamase inhibitor combinations, for sepsis, 292
Lactating adenoma, 1641
Lactation, low-molecular-weight heparin for, 268–270
Lactic acidosis, 258b, 262b, 882
Lactose intolerance, 648, 787–788
LAD. *See* Leucocyte adhesion deficiency
LADA. *See* Latent autoimmune diabetes in adults
Lambert-Eaton myasthenic-myopathic syndrome (LEMS), 974
Lambert-Eaton syndrome, 1383
Langerhans cell histiocytosis, 573
Language impairment, in Alzheimer's disease, 955
Language therapy, in transgender women, 1678
Lanreotide
acromegaly, 1187–1188
pancreatic neuroendocrine tumours, 760
Laparoscopy
of acute abdomen, 675
of gastric cancer, 639
of oesophageal tumours, 631
Large artery atherosclerosis, 978–979
Large bowel obstruction, CT of, 135f
Large intestine
disorders of
colorectal cancer, 671–672, 672b
diverticular disease, 666–667
ischaemic colitis, 666
structure of, 611f
Laryngeal cancer, survival, 1396b
Laryngeal carcinoma, 599, 599f
Laryngeal diphtheria, 1471
Laryngopharyngeal reflux disease (LPRD), 623
Laryngotracheobronchitis, acute, 539
Laser photocoagulation therapy, for diabetic nephropathy, 1227
Lassa fever, 1464
Lasting power of attorney, 30
Latent autoimmune diabetes in adults (LADA), 1206
Late-onset ataxia, 961–962
Lateral cerebellum, 906
Lateral cervical disc protrusion, 969, 969f
Lateral cutaneous nerve of thigh, compression, 964
Lateral hip pain syndrome, 1075
Lateral lumbar disc protrusion, 969
Lateral medullary syndrome, 981–982
Law, for young adult, 1633b
Laxatives, 617, 617b
LBP. *See* Lipopolysaccharide-binding protein
LDL. *See* Low-density lipoprotein
Lead poisoning, 368
Leadership, evidence-based practice and, 51
Lean-to-fat tissue, proportion of, 770–771
Learning disability, 1039
Leber's hereditary optic neuropathy (LHON), 64
Leflunomide, for rheumatoid arthritis, 1097
Left atrial appendage occlusion, 998
Left parasternal heave, 396
Left ventricular assist device, 438f
Left ventricular non-compaction, 486
Leg ulcers, 1608–1609, 1608f, 1608b
Legal interventions, in behaviour change, 159
Legal issues
advance decisions, 30, 30b
advance statements, 30
assisted dying, 31
capacity, 29
confidentiality, 32–33
consent, 29, 29b
lack of capacity, 30–31
lasting power of attorney, 30
in medical practice, 29–34
professional competence and mistakes, 33–34
provision or cessation of life-sustaining treatment, 31
self-determination, 29b
statutes and common law, 29
Leiomyomas, 631
Leiomyomata, female health and, 1657–1658, 1658f
Leishmania spp., 1503–1504
Leishmaniasis, 1503–1505, 1504b
cutaneous, 1504–1505, 1505f, 1600
HIV-associated, 1504, 1557
mucocutaneous, 1505
visceral, 1504
LEMS. *See* Lambert-Eaton myasthenic-myopathic syndrome
Lenacapavir, 1551
Lenalidomide, cancer therapy, 1392
Lenegre-Lev's disease, 415
Lennox-Gastaut syndrome, 925
Lens, 1045
Lentigines, 1626
Lentigo maligna, 1618, 1619f, 1621
Lepromatous leprosy, 96–97, 1484, 1591
Leprosy (Hansen's disease), 1106, 1483–1484, 1484f, 1591, 1626
clinical features of, 1484
diagnosis of, 1484–1485
indeterminate, 1591
lepromatous, 1484, 1484f, 1591
management of, 1485, 1485b–1486b
multibacillary, 1484
paucibacillary, 1484
prevention of, 1485
reactions of, 1486b
tuberculoid, 1484, 1484f, 1591
Leptin, 737
Leptospira interrogans, 1481
Leptospirosis, 1481–1483
Letermovir, 1467
Letrozole, 1184
Leucocyte adhesion deficiency (LAD), 100
Leucocytoclastic vasculitis, 1608
Leucocytosis, neutrophils, 1314–1315, 1315b
Leuconychia, 1626
Leucoplakia, 621
Leukaemia, 817, 1337–1346
acute, 1337–1342
clinical features, 1337–1338, 1338b
investigations, 1338–1339
management, 1339–1340
chronic, 1343
classification, 1332b
genetics, 1335
hairy cell, 1358
supportive care, 1339–1340
survival, 1396b
Leukotriene antagonists, rhinitis, 538
Leukotriene receptor antagonists, for asthma, 544
Levodopa
intestinal gel infusion, 932
for Parkinson's disease, 930–931
Levofloxacin, for sepsis, 292
Levosimendan, 317
Lewy body dementia, 957
Lhermitte's phenomenon, 909
LHON. *See* Leber's hereditary optic neuropathy
Liaison psychiatry, 1009–1010
team, 1010, 1010f
Libido, in male, 1669
Lice, 1599
Lichen planus, 1586–1587, 1587f
Lichen sclerosus (LS), 1587
Lichen simplex, 1583, 1583f
Lichenification, 1570f
Liddle's syndrome, 871
Lids, disorders of, 1050–1052, 1052f
Li-Fraumeni syndrome, 1383b
Ligaments, 1065
Lightning strike, 382–383
Lille alcoholic hepatitis score, 725b
Limb ischaemia, 490–491
Limb-girdle muscular dystrophy, 974–975, 974b
Limbs, rapid examinations of, 1064b
Limited cutaneous scleroderma (LcSSc), 1111–1112
Linear IgA disease, 1612f–1613f, 1613
Lipid disorders, 510–520, 510b
Lipid-lowering diet, 518
Lipid-lowering drugs, 518–519
Lipid metabolism disorders, 510
Lipids
metabolism, 683–684
physiology of, 510–513, 511f
unabsorbed, 608
Lipoatrophy, 1217
Lipodystrophy, 1553
Lipohypertrophy, 1217
Lipoma, 564, 1616, 1643
Lipopeptides, for sepsis, 293
Lipopolysaccharide-binding protein (LBP), 309, 310f
Lipoprotein(a), 512
inhibitors, 519
Lipoprotein lipase pathway deficiencies, 513
Listeriosis, 1482
Lithium, overdose, 365
Livedo reticularis, 1608
Liver
anatomy, 682–683, 682f
drug hepatotoxicity, 732–733, 733b
drug metabolism, 732
functions, 683–686
MRI of, 138, 140f
radiofrequency ablation, 148b, 151
ultrasound of, 130f, 688
Liver abscess, 727–728
Liver assist devices, 710
Liver biopsy, 689, 690b
Liver disease
bile acid receptors in, 685–686
chronic, 1606
coagulation disorders and, 1325–1326
complications of obesity, 1258
palliative care, 1731–1733
pregnancy, 1684–1685
rheumatological problems in, 1118
Liver failure, acute-on-chronic, 710
Liver flukes, 1516
Liver function tests, 686–687
Liver transplantation, 711–712
Liver tumours, 728–730
Loa loa, 1509b
Locked-in syndrome, 949, 949f
LogMAR chart, 1047–1048, 1048f
Loiasis, 1510
Loneliness, in older adults, 1718–1719
Long QT syndrome, 424, 424b, 425f
Loop diuretics, 860, 861f
Loop of Henle, 858b
Losartan, for gout, 1105
Louse-borne relapsing fever, 1490
Low back pain, 969, 1071–1074, 1072f, 1072b
Low-CSF-volume (low-pressure) headache, 920
Low-density lipoprotein (LDL), 511–512, 511f–512f
disorders of, 513–516
Low-molecular-weight heparin, 1371
Low-calorie and very-low-calorie diets, 1263b
Low-carbohydrate diet, 1263b
Lower abdominal pain, 1524b
Lower gastrointestinal cancer, 1403–1405
Lower limb ischaemia, 490–491
Lower limb weakness, 908
Lower motor neuron (LMN) lesions, 906–908, 907b
Lower oesophageal rings, 629
Lower respiratory tract infections
bacterial, 1472
viral, 1451–1454
Lower urinary tract, symptoms, male health, 1660b
Low-GI diet, 1263b
LPRD. *See* Laryngopharyngeal reflux disease
Lumbar disc prolapse, 1073, 1073b
Lumbar puncture, 188b, 189f, 913
Lumbar spondylosis, 1072–1073
Lung abscess, 215–216
Lung cancer, 1396–1399
CT of, 134f, 560, 563b
management, 1397–1398
prognosis of, 1396, 1396b
staging, 1397, 1397f
survival, 1396b
Lung collapse, 120, 122f
Lung cysts, 578–580
Lung disease
occupational, 577–578, 577b
restrictive, in pregnancy, 1687
rheumatoid arthritis, 1093
SLE, 1109
Lung function tests, 533b
'Lung-protective' ventilation, 323
Lung transplantation, 577
Lung volumes, 534–535, 535f
Lungs
airways of, 525–526
anatomy, 523–524, 524f
collapse, 531b, 561b
nerve supply, 524–525
vasculature and lymphatics, 524
Lupus erythematosus (LE), 1607
Lupus nephritis, 815, 815b, 821f
Lyme arthritis, 1106
Lyme disease, 1482, 1483f, 1590–1591
Lymph nodes
cancer staging, 562b
in rheumatoid arthritis, 1093
tuberculosis, 554, 555b–556b
Lymphangitis, in breast, 1645f
Lymphatic disease, 1609–1610
Lymphatic filariasis, 1509
Lymphocytes, 1315–1316
Lymphocytic choriomeningitis virus, 1460
Lymphoedema, 1609–1610
Lymphogranuloma venereum, 1534
Lymphoma, 1348–1358
classification, 1332b, 1351.e1b
HIV-associated, 942, 1561
Hodgkin, 1349–1351
investigations, 1348–1349, 1348b
malabsorption and, 650
management, 1348–1349
marginal zone, 1356
mycosis fungoides, 1357–1358
non-Hodgkin, 1351–1352
primary central nervous system, 1356
of small intestine, 652
T-cell, enteropathy-associated, 647
Lymphoplasmacytic lymphoma, 1355
Lynch syndrome, 670–671, 671b, 1383b
Lysosomal storage disorders, 1246, 1246.e2b
Lysozyme nephropathy, 815

M

Macrocytic anaemia, 1283–1287
Macrocytosis, without megaloblastic changes, 1286–1287
Macrolides, for sepsis, 291b, 293–294
Macrophages, 84
Macroprolactinoma, 1190f
Macular degeneration, age-related, 1060, 1060f
Macules, 1568f–1570f
Maculopapular (morbilliform) exanthems, 1629, 1629f
Madura foot, 1492–1493, 1496
Magnesium, 779
Magnetic resonance cholangiography (MRC), 741, 742f
Magnetic resonance cholangiopancreatography (MRCP), 138, 140f–141f, 689
Magnetic resonance imaging (MRI), 137–138, 137b
cardiac, 406–407, 407f–408f, 407b
of cervical spine, 1069f
gadolinium chelates and, 137b
of gastrointestinal disease, 618
of hereditary haemochromatosis, 721
of joints, 138, 142f
of kidney disease, 799
of knee, 1076f, 1094f
of lumbar spine, 1071, 1073f
safety consideration, 138b
safety warning signs, 138f
sequences, 138b
of spastic paraparesis, 142f
of testicular and ovarian germ cell tumours, 1413
Malabsorption, 644–650, 644b
Malaise, 1727

Malaria, 1497–1502
  clinical features of, 1498–1499, 1499b, 1500f
  diagnosis of, 1499
  epidemiology of, 1497, 1498f
  management of, 1499–1501, 1500b–1501b
  parasitology of, 1498, 1499f
  pathogenesis of, 1498
  prevention and control of, 1501–1502, 1502b
Malassezia infection, 1598
Malassezia spp., 1496
Male health, 1660–1669, 1662f
  ageing and, 1669
  bladder scanning, 1661f, 1661b
  catheterization, 1661f, 1661b
  digital rectal examination, 1661f
  erectile dysfunction in, 1669, 1669b
  history, 1660b
  loss of libido and, 1669
Malignant bile duct obstruction, 745–746, 746f
Malignant disease, 1375–1416, 1375b
  clinical skills for, 1376–1377
  examination of, 1377, 1377f
  history in, 1376, 1376b
  implications of, 1377b
  investigations in, 1376, 1376b, 1387b
Malignant hypertension, management of, 509b
Malignant hyperthermia, 973
Malignant melanoma, 1414, 1415f
  survival, 1396b
Malignant meningitis, 941
Malignant pericarditis, 488
Mallampati score, 341f
Mallory-Weiss tear, 640–641
Malnutrition, 763–764, 767b
  assessment and diagnosis of, 762–763, 763f–766f
  childhood, 771f, 771b
  management of, 763, 767f
Malnutrition Universal Screening Tool (MUST), 763f
Maltwork's lung, 573b
Mammary duct fistula, 1645
Mammography, 124–125, 125f
Manganese, 779
Mania, 1019–1021, 1021b
Manometry, 627, 628f
Mansonella perstans, 1509b
Mansonellosis, 1511
Mantle cell lymphoma, 1352b, 1355
Maple syrup urine disease (MSUD), 1244, 1246.e1b
MAPLE-V, 1710–1711, 1713f
Marburg virus, 1465
Marfan's syndrome, 475–477, 476f–477f, 1082
Marginal keratitis, 1053f
Marginal zone lymphoma, 1356
Maribavir, 1467
Marmot review, 156b
Marsh-Oberhuber classification, of coeliac disease, 646b
Masculinizing surgeries, 1678
Massive blood transfusion, 315, 1327
Massive cortical damage, 233, 233f
Mast cells, 84, 84b
Mastalgia, breast, 1643–1644, 1644b
Mastitis, granulomatous lobular, 1645
Maternal health, 168–169
McArdle's syndrome, 972–973
McCune-Albright syndrome, 1199
MCGN. *See* Mesangiocapillary (membranoproliferative) glomerulonephritis
MDS/MPN overlap syndrome, 1348
Measles (rubeola), 1449, 1449f
Mechanical cardiac support, temporary, 411, 411f
Mechanical haemolytic anaemia, 1306
Mechanical low back pain, 1071
Mechanical thrombectomy, 988–991, 993f
Mechanical ventilation
  complications of, 322
  controlled, 320
  extracorporeal gas exchange, 323
  high-frequency oscillation, 323
  indications for, 320–321
  intermittent mandatory ventilation, 323
  intermittent positive pressure ventilation (IPPV), 320
  lung-protective, 323
  positive end-expiratory pressure, 322–323
  pressure support, 323
  sedation, analgesia and muscle relaxation, 321
  weaning, 324–325
Mechanobullous diseases, 1614
Meckel's diverticulum, 650
Medial knee pain, 1076
Medial temporal lobe herniation, 234f
Mediastinal lesions, 580–581, 580f–581f
Mediastinoscopy, 536–537
Mediastinum, 580
Medical clerking, 179–180, 179b
Medical consultation, 2
  information-gathering, 3
  initial interaction of, 2–3, 3b
  investigation, 5–7
Medical electives, 173–174
Medical emergencies in eating disorders (MEED), 1037, 1038b
Medical ethics, 25–34
Medically unexplained symptoms, 1029–1030
Medication review, 114
Medicine
  artificial intelligence and, 12–13
  cancer, 1381
  inequalities in, 11–12
  young adult, 1631–1638
Mediterranean diet, 1263b
Mediterranean spotted fever, 1490b
Medium-chain acyl-CoA dehydrogenase deficiency (MCADD), 1246.e4b
Medullary cystic disease (juvenile nephronophthisis), 851
Medullary sponge kidney, 822, 823f, 851
Megacolon, 665
Megaloblastic anaemia, 1283–1286, 1283f–1284f
Meglitinides, 1221
Melaena, 223–226
Melanocytic naevi, 1614–1615
Melanoma, 1383b, 1620–1621, 1621f, 1621b–1622b
Melasma (chloasma), 1626
Melioidosis, 1492
Membranous glomerulopathy, 805, 806f
Memory, 891, 892f, 892b
Memory loss, 891, 892b
Men
  puberty, 1176
  reproductive disorders, 1178–1180
  reproductive physiology, 1176, 1176f
Ménière's disease, 590, 900
Meninges, biopsy, 915
Meningiomas, 945, 945f, 950f, 1149, 1149f, 1413
Meningitis, 246–248, 941
  bacterial, 246–247, 247f, 941, 1478–1479
  causes of, 246b, 941b
  chronic, 247, 941
  clinical assessment of, 246, 247b
  cryptococcal, 942
  CSF changes in, 941b
  investigations of, 247
  malignant, 941
  management of, 247–248, 248b
  meningococcal, emergency treatment for, 248b
  Mollaret's, 943
  tuberculous, 941
  viral, 247, 1460b
Meningococcaemia, emergency treatment for, 248b
Meningococcal arthritis, 1106
Meningococcal sepsis, 1479
Meningovascular syphilis, 942
Meniscal tear, 1076, 1076f
Menopause, 1649
  female health and, 1655–1656
  hormone replacement therapy for, 1656, 1656b
  premature, 1181, 1391–1392
Menstruation
  hormonal and follicular changes, 1177f
  in PCOS, 1183
  in young adult, 1638
Mental disorders
  aetiology of, 1012–1013
  biological mechanisms of, 1012, 1012f
  epidemiology of, 1012–1013
  prescribing in, 1007–1009, 1009b
  treatment approaches in, 1006–1007, 1008b–1009b
  wider context of, 1012, 1013b
Mental health, 169
  issues, in young adult, 1632–1634
Mental state examination (MSE), 1004b–1005b, 1026b
  cognitive state, 1005b
  diagnostic information, 1002–1003, 1005b–1007b
  formulation, 1003
  risk assessment, 1003–1006, 1005b
Meralgia paraesthetica, 964, 1075
Mercaptopurine, for Crohn's disease, 660
Mercury poisoning, 368
Merkel cell carcinoma, 1624, 1625f
MERS. *See* Middle East respiratory syndrome
MERS-CoV. *See* Middle East respiratory syndrome coronavirus
Mesangial cells, 792
Mesangiocapillary (membranoproliferative) glomerulonephritis (MCGN), 813–814, 813f
Mesothelioma, 560–563, 579b
Messenger RNA (mRNA) codes, 58
Meta-analysis, 47–48, 49f
Metabolic acidosis, 879–880
  with anion gap, 879–880, 880b, 882b
  critical illness, 308
  mixed, 882
  poison-induced, 356b
Metabolic alkalosis, 883
  causes of, 883b
  chloride-resistant, 884
  chloride-responsive, 883–884
  classification and definitions of, 883
  clinical features of, 883
  critical illness, 308
  management of, 883–884
'Metabolic block,' dietary manipulation to bypass, 1244
Metabolic disease, 510–520, 1240–1247
  amyloidosis, 806
  approach to the patient, 1240–1241
  carbohydrate metabolism disorders, 1246
  in chronic kidney disease, 842
  core skills and knowledge, 1240b
  enzyme replacement therapy, 1244
  epilepsy and, 924–925
  erythrocytes, 1299–1301
  examination of, 1241
  general principles of management of, 1244
  history of, 1240–1241
  investigations of, 1241, 1241b
  lysosomal storage disorders, 1246, 1246.e2b
Metabolic disease *(Continued)*
  maintenance of regular and adequate calorie intake, 1244
  newer therapies, 1244
  peroxisomal, 1247
  porphyrias, 1246, 1247f
  protein, 1245–1246, 1246.e1b
  special considerations of, 1241
  supplementation with vitamins, minerals, and co-factors, 1244
  supportive management and surveillance, 1244
Metabolic myopathy, 971–973
Metabolic neuropathies, 966–967, 967b
Metabolic syndrome, in anti-retroviral therapy, 1553
Metabolism
  erythrocytes, 1299–1301, 1301f
  in kidney disease, 828
  regulation of, 770
Metaphase, 62f, 65f
Metapneumovirus, 1454
Metastases, of unknown primary, 1415–1416, 1415f, 1415b–1416b
Metastatic disease
  confined to regional lymph nodes, 1401
  indolent, 1401
  limited, 1401
Metastatic urothelial cancer, 1407
Metatarsalgia, 1078
METAVIR Scoring System, 694
Metered-dose inhalers, 543b
Metformin
  associated lactic acidosis, 262, 262b
  PCOS, 1184
  type 2 diabetes, 1219
Methanol
  metabolism, 369f
  poisoning, 369
Methylmalonic aciduria (MMA), 1246.e1b
Meticillin-resistant *Staphylococcus aureus* (MRSA), 295, 1470
Microbiological investigations, 1429–1432
Microcytic anaemia, 1279–1283, 1282b
Micronutrients, deficiencies and, 774–780, 780b
Microscopic colitis, 664–665
Microscopic polyangiitis, 574
Microsporidiosis, 1508
  HIV-associated, 1557
Micturating cystourethrography, 800
Micturition disorders, 912–913
Middle East respiratory syndrome coronavirus (MERS-CoV), 160b, 1452
Midgut dysmotility, 643–644
Midline cerebellum, 906
Migraine, 242, 242b, 915–921
  clinical features of, 917
  diagnostic criteria for, 917b
  management of, 917–918
  mechanisms of, 916–917
  pregnancy, 1687
Migraine-related dizziness, 917
Mild cognitive impairment, 954
Miliaria rubra, 379
Miliary tuberculosis, 554, 555b–556b
Milrinone, 317
Mineral and bone disorder, in chronic kidney disease, 839–840, 840f
Mineralocorticoids, 1152
Minerals, 777–780, 1244
Mini-Mental State Examination, 953
Minimal-change nephropathy, 803, 803f
Miscarriage, 1651–1652, 1652b
Mitochondrial chromosome, 58, 59f
Mitochondrial chromosome disorders, 63–64
Mitochondrial disease, 975–976, 1244–1245, 1245b
Mitochondrial replacement therapy, 73, 75f
Mitochondrial toxicity, in anti-retroviral therapy, 1553
Mitosis, 63, 65f
Mitotic spindle, 65f
Mitral prolapse, 456
Mitral regurgitation, 451, 453f, 455, 455f
Mitral stenosis, 452–456
  clinical features of, 452–453
  investigations of, 453–455
  management of, 455
  pathophysiology of, 452
  symptoms of, 452–453
Mitral valve disease, 452–456, 452f
Mitral valve replacement, 455
Mixed metabolic acidosis, 882
Mixed urinary incontinence, 1657b
Mixed venous oxygen saturation, 302
Mixed venous oxygen tension, 302
M-mode echocardiography, 401, 403f
Modified Early Warning Score, 298, 298b
Modified Marshall scoring system, 752b
Mollaret's meningitis, 943
Molluscum contagiosum, 1449, 1531, 1593–1594, 1595f
Molnupiravir, 1467
Molybdenum, 779
Mondor's disease, 1643
Monoarthritis, acute, causes and clinical features of, 256b–257b
Monobactams, for sepsis, 289–292, 291b
Monoclonal antibodies
  for asthma, 544–545
  cancer therapy, 1393, 1393b
Monoclonal antibody therapy (targeted therapy), 105–106
Monoclonal gammopathy of clinical significance, 1362
Monoclonal gammopathy of renal significance, 814, 1362
Monoclonal gammopathy of undetermined significance (MGUS), 1362
Monocytes, 84, 1315
Monogenic diabetes mellitus, 1210–1211, 1210b
Mononeuritis, 1230
Mononeuritis multiplex, 962, 964, 1230
Mononeuropathies, 962, 964
Monro-Kellie doctrine, 328
Mood (affective) disorders, 1017, 1017f
Moral injury, ethics and, 27
Morbidity, young adult, 1632
Morisky-8 Medication Adherence Scale, 506, 506b
Morphine, in symptom control, 1724, 1725b
Mortality
  surgery and, 349
  of venous thrombosis, 1365
  of young adult, 1632, 1634f
Morton's neuroma, 1078
Mosaicism, 67
Motility disorders, of oesophagus, 625–628
Motivational interviewing (MI), in obesity, 1252, 1253b
Motor control systems, 902–906
Motor neuron disease (MND), 959–960, 960b
  palliative care, 1733
Mouth, 604, 620–621
  care, 299
  gums, 621, 622b
  neoplasia of, 621
  non-neoplastic lesions of, 621
  pigmented neoplastic lesions of, 621
  recurrent aphthous ulceration, 620, 620b
  squamous cell carcinoma of, 621
  teeth, 621
  tongue, 621
  white patches of, 621

Movement disorders, 929–934
Mpox, 1448–1449, 1535, 1594–1595
  HIV-associated, 1559
MRCP. *See* Magnetic resonance cholangiopancreatography
MRFIT. *See* Multiple Risk Factor Intervention Trial
MRI. *See* Magnetic resonance imaging
MRSA. *See* Meticillin-resistant *Staphylococcus aureus*
MSE. *See* Mental state examination
Mucinous cystadenomas, 759
Mucocutaneous leishmaniasis, 1505
Mucolytic agents, 548
Mucormycosis, 1495
Mucositis, 1391
Multidisciplinary care model, 178
Multidisciplinary team (MDT)
  acute care, liaison with, 182
  in CGA, 1703, 1703b
  meetings, regular, patient flow and, 178
  in obesity care, 1261, 1262f
Multifocal motor neuropathy with conduction block, 966
Multimorbidity, 1696–1698
Multinodular goitre, 128f, 1173, 1174f
Multiorgan failure, 317b
Multiple endocrine neoplasia (MEN), 1198, 1198b
  management, 1198–1199
  screening, 1198–1199
  type 1, 1198, 1198f, 1383b
  type 2, 1383b
  type 2a and 2b, 1198
Multiple mononeuropathy, 1230
Multiple myeloma, 1358–1362, 1359f
  asymptomatic, 1362
  bone disease, 1360, 1360f
  classification, 1361b
  clinical features, 1359–1360, 1359f–1360f
  complications, 1359b
  diagnosis, 1360
  investigations, 1360–1361
  management, 1361–1362
  renal involvement of, 818–819, 818f
  staging and prognosis, 1361
  survival, 1396b
Multiple organ dysfunction syndrome, 312–313, 312f
Multiple Risk Factor Intervention Trial, 513f
Multiple sclerosis (MS), 934–941
  aetiology and pathogenesis of, 935
  clinical features of, 935–937, 935f–936f
  clinically isolated syndrome, 937–938
  diagnosis of, 937
  epidemiology of, 934–935
  investigations of, 937–938, 937f
  late-stage, 937
  management of, 938, 939b
  MRI of, 139f
  palliative care, 1733
  pathology of, 935, 935f
  pregnancy, 1688
  prognosis for, 938–939
  radiologically isolated syndrome, 938
  symptoms of, 936–937, 936b
  types of, 935–936
  unusual presentations of, 937
Mumps, 1457, 1458f
Mupirocin, for sepsis, 295
Muscle
  atrophy (wasting), marasmus, 771
  biopsy, 915, 971
  chronic kidney disease and, 842
  infection, 1458
Muscle diseases, 970–976
  classification of, 971b
  clinical approach to, 971, 972b
  definitions of, 970–971
  diagnosis of, 970–971
Muscle diseases *(Continued)*
  inflammatory myopathies, 971
  metabolic and endocrine myopathies, 971–973
  pathophysiology of, 970
  SLE, 1108
Muscle relaxants, for surgery, 344
Muscular dystrophies, 974–975
Musculoskeletal disorders
  complications of obesity, 1259
  regional, 1067–1079, 1068f
  SLE, 1108
Mushroom, poisoning, 370.e6
Mushroom worker's lung, 573b
MUST. *See* Malnutrition Universal Screening Tool
Mutation accumulation theory, ageing and, 1695
Mutations, 76
Myasthenia gravis, 973, 1688
Mycetoma, 1492–1493, 1496
Mycobacterial infections, 1489b, 1591–1592
Mycobacterial ulcer (Buruli ulcer), 1485–1486
Mycobacterium avium-intracellulare, HIV-associated, 1560
Mycobacterium leprae, 96
Mycobacterium tuberculosis, HIV-associated, 650, 1560
Mycoplasma genitalium, 1527
Mycosis fungoides, 1357–1358
Myelin oligodendrocyte glycoprotein antibody-associated disease, 940
Myelitis, 942
Myelodysplasia, 1346–1348, 1347f, 1347b
Myelodysplastic syndrome, 817, 1347f
Myelofibrosis, 1345, 1345f
Myeloproliferative disorders
  essential thrombocythaemia, 1323
  polycythaemia, 1306
Myeloproliferative neoplasms, 1343–1346
MYH-associated polyposis, 670
Myiasis, 1518, 1600
Myocardial action potentials, 388, 388f
Myocardial contractility, 303, 315–317, 430
Myocardial/endocardial disease, 480–495
Myocardial infarction, 348
Myocardial injury, surgery and, 348
Myocardial perfusion and viability, 405, 405f
Myocardial remodelling, 430, 431f
Myocardial stunning, 437
Myocarditis, 480–481, 1458
  causes of, 480b
  Chagas' disease, 481
  clinical features of, 480–481
  giant cell, 481
  investigations of, 481, 481f
  management of, 481
  pathology of, 480, 481f
Myocardium
  hibernating, 437
  mechanical support of, 317–318
  nerve supply of, 390
Myoclonic seizures, 922
Myoclonus, 933
Myofibrils, 389f
Myopathies, 63
Myophosphorylase deficiency (McArdle's syndrome), 972–973
Myositis, inclusion body, 1114
Myotonia congenita, 975
Myotonic dystrophy, 975
Myotonic syndromes, 975
Myxoedema coma, 1167
Myxoedema madness, 1167

N

Nail disorders, 1626–1627
Nails, 1572
  chemotherapy effects, 1390, 1390f–1391f
Nairoviruses, 1455b
Naltrexone, for obesity, 1266
Narcolepsy, 928
Nasal polyps, 538
Nasogastric tube, 119–120, 121f
National Institute of Health Stroke Scale (NIHSS), 983, 988b
Natriuretic peptides, 430, 826
Natural killer cells, 90–91, 90b
Nausea and vomiting, 220, 220b
Neck circumference, 1253, 1257f
Neck of femur fractures, falls and, 1707–1709, 1707f, 1710b, 1711f
Neck pain, 1067–1069, 1068b
Necrobiosis lipoidica, 1604f
Necrotizing fasciitis, 276, 1574
Needle biopsies, 146b, 147
Neglect, 891
Negligence, clinical, 33–34
Nelson's syndrome, 1147, 1147b, 1156
Nematodes, 1509–1513
  intestinal, 1512b
  tissue-dwelling, 1509–1511
  zoonotic, 1513
Neonatal rashes, 1628–1629
Neonates
  alloimmune thrombocytopenia, 1322–1323
  haemolytic disease of newborn, 1304
  hyperthyroidism, 1690
  tetanus in, 1480
  transient diabetes mellitus, 1211
Neoplastic disease, of bone, 1135–1136, 1136b
Neovascularization, 1226
Nephritic syndrome, 809
Nephroblastoma (Wilms' tumour), 851, 1383b
Nephrocalcinosis, 824, 825f, 825b, 873
Nephrogenic systemic fibrosis, 842, 1606
Nephrons, 792, 793f, 858f, 858b
Nephrostomy, 150, 151f
Nephrotic syndrome
  causes of, 803–808
  glomerular lesions in, 802–803
  glomerulopathies associated with, 803b
  management of, 802–803
  oedema, 859
Nephrotoxicity, of chemotherapy, 1391
NERD. *See* Non-erosive reflux disease
Nerium oleander, poisoning, 370.e6
Nerve agent poisoning, 369
Nerve conduction studies, 914, 915f, 1067
Nerve regeneration, 962
Nerve root entrapment
  lumbar, 1073b
  neck, 1068, 1068b, 1069f
Nervous system infections, 941–944
  encephalitis, 941
  HIV-related, 942
  meningitis, 941, 941b
  neurocysticercosis, 942–943
  neurosyphilis, 942
  progressive rubella encephalitis, 943
  subacute sclerosing panencephalitis, 943
  toxin-mediated, 1479–1480
  viral, 1458–1461
  Whipple's disease, 943–944
Neuralgia, 909
Neuralgic amyotrophy, 970
Neuraminidase inhibitors, 1467
Neuroanatomy, 888
  cranial nerves, 891–902, 892b
  essential elements of, 889
  neuron and synapse, 888
  neurotransmitters, 888
Neurocritical care, 328–329
Neurocutaneous syndromes, 961
Neurocysticercosis, 942–943
Neurodiversity, 1039
  of young adult, 1638
Neuroendocrine peptides, 606, 607b
Neuroendocrine response, to shock, 308–309
Neuroendocrine tumours, of pancreas, 760
Neurofibroma (schwannoma), 1413
Neurofibromatosis
  genetics, 76
  type 1, 961, 1383b, 1610–1611
  type 2, 961
Neurogenetic disorders, 961–962
Neuroimaging, 913, 980f
Neuroinflammation, in systemic disease, 940–941
Neuroinflammatory disorders, 934–941
Neuroleptic malignant syndrome, 1016–1017
Neuroleptics, 362, 934
Neurological autoantibody tests, 914b
Neurological disease
  bladder control and sexual dysfunction, 911–913
  brain and spinal abscess, 944
  brain lesions, focal, 889
  brain tumours, 944–946
  complications of obesity, 1259
  congenital disorders, 960
  cranial nerves, 891–902, 892b
  epilepsy, 921–927
  functional, 253–254
  HIV-associated, 942
  hydrocephalus, 946–947
  migraine, 916
  movement disorders, 929–934
  muscle diseases, 970–976
  neurogenetic disorders, 961–962
  palliative care, 1733
  paraneoplastic syndromes, 962, 1383, 1385f
  paraplegia, 951–952
  patterns of, 889–915, 890f–891f
  peripheral nerve disease, 962–970
  plexus and nerve root lesions, 969–970, 969b
  in pregnancy, 1687–1688
  rheumatological problems in, 1119
  sleep disorders, 928–929
  spinal cord disease, 950–952
  spinal cord lesions, 909
  stroke, 977–1000
  traumatic brain injury, 947–948
Neurological examination, 888b
Neurological investigations, 913–915
  biopsy, 915
  cerebral-evoked potentials, 914–915
  electroencephalography, 913–915, 914f
  electromyography, 914
  genetic tests, 913
  immunological tests, 913, 914b
  lumbar puncture, 913
  nerve conduction studies, 914, 915f
  neuroimaging, 913
  psychometric assessment, 915
Neurology, 885–976
  background, 888
  clinical skills, 886–888
  core skills and knowledge, 886b
  diagnosis in, 889b
  examination, 886
  history in, 886–887, 886b–887b, 887f
Neuromuscular junction, Lambert-Eaton syndrome, 1383
Neuromuscular junction disorders, 973–974
Neuromyelitis optica, 940
Neuron, 888, 890f
Neuronal ceroid lipofuscinosis, 933
Neuropathic pain, 911, 913b
Neuropathic ulcers, 1609
Neuropathies, 962
  autonomic, 1230–1231, 1230b
  diabetic, 1229–1231
  drug-induced, 967b
Neurosarcoidosis, 940
Neurosyphilis, 942, 1534
Neurotoxic mushroom poisoning, 370.e7
Neurotoxic shellfish poisoning, 370.e1
Neurotoxicity, of chemotherapy, 1391
Neurotransmitters, 888
  in pain perception, 911
Neutropaenic sepsis, 272, 272b
Neutropenia, 98b, 1315, 1315b
Neutropenic patients, sepsis in, 284
Neutrophil granules, 83b
Neutrophils, 83, 829, 1314–1315
  defects, 98b
  function of, 1314
  leucocytosis, 1314–1315, 1315b
  migration, 93f
  opsonization, 83f
Next-generation sequencing, 70, 71f
NHL. *See* Non-Hodgkin lymphoma
Niacin, 519
Niemann-Pick disease, 1246.e2b
Nipah virus, 1460
Nipple
  adenoma, 1643b
  benign breast condition of, 1643, 1643b
  discharge of, 1644, 1644b
  eczema of, 1643b
  Raynaud's disease, 1643b
Nirmatrelvir, 1467
Nitrates, for heart failure, 436
Nitric oxide (NO), 392, 392f
  biochemical action of, 311f
  fractional exhaled, 535, 542
  inhaled, ARDS, 327
  renal effects of, 797
  synthesis of, 311f
Nitrofurantoin, for sepsis, 295
Nitroimidazoles, for sepsis, 294–295, 294f
Nitrous oxide, overdose, 365
NNT. *See* Number needed to treat
Nocardia spp., infections, 1492–1493
Nociceptive pain, 911
Nodal osteoarthritis, 1071, 1084, 1085b, 1086f
Nodular goitre, 1173
Nodular prurigo, 1583
Nodularity, of breast, 1641
Nodules/polyps, 598–599, 599f, 1568f–1570f
Noise trauma, 589
Nominal aphasia, 890
Non-alcoholic fatty liver disease, 707
Non-binary people, treatment of, 1679
Non-coding RNAs (ncRNAs), 58
Non-coeliac gluten intolerance, 647
Non-coeliac gluten sensitivity, 788
Non-epileptic attack disorder, 927–928
Non-erosive reflux disease (NERD), 624, 624f
Non-freezing cold injury, 378
Non-functioning tumours, 1147b
Non-gonococcal urethritis, 1527
Non-haem iron absorption, 1280
Non-haemolytic (febrile) transfusion reactions, 1311
Non-Hodgkin lymphoma (NHL), 1351–1352
  aetiology, 1351
  B-cell, 1351.e1b, 1352–1356
  Burkitt lymphoma, 1352b, 1354
  chromosome translocations, 1352b
  clinical features, 1352
  cytogenetic features, 1352b
  genetic features, 1351–1352
  indications for treatment, 1354b
  pathogenesis, 1351–1352
  small intestine, 652
  survival, 1396b
  T/NK cell, 1351.e1b
Non-hormonal contraception, 1654b
Non-immune haemolytic anaemia, 1305–1306

Non-invasive ventilation, 320, 323–324, 324b
Non-malignant disease, palliative care, 1729–1733
Non-nucleoside analogues, 1550
Non-opioid analgesia, for surgery, 345
Non-pulsatile tinnitus, 586b
Nonsense mutations, 65
Non-small-cell lung cancer (NSCLC), 1396–1397, 1396b
Non-steroidal anti-inflammatory drugs (NSAIDs)
  effects and side-effects of, 1081
  for *H. pylori* infection, 636
  for musculoskeletal problems, 1081, 1081b
  overdose, 361
  in renal damage, 828b
  for rheumatoid arthritis, 1094–1095
  surgery and, 345
  uses of, 1081
Non-streptococcal infective glomerulonephritis, 809
Non-sustained ventricular tachycardia, 425, 426f
Non-transfusion-dependent thalassaemia, 1293
Non-tuberculous mycobacterial infection, 557, 558b, 1489
Noradrenaline (norepinephrine), 316
Normal-transit constipation, 615–616
Normocytic anaemia, 1283
Normosmic idiopathic hypogonadotropic hypogonadism, 1179
Nose, 591
  anatomy and physiology of, 591, 591f–592f
  disorders of, 591–595
  examination of, 591
Notifiable diseases, 1437b
NSAIDs. *See* Non-steroidal anti-inflammatory drugs
NSCLC. *See* Non-small-cell lung cancer
Nuclear imaging, 143–146, 144b
  for amyloid disease, 406
  endocrine, 144–145, 145f
  of heart, 405–406
  image type, 405–406
Nucleic acid amplification, 554
Nucleic acid detection, 1432
Nucleoside analogues, 1465–1467
Nucleoside/nucleotide reverse transcriptase inhibitors, 1548–1549
Nucleotide analogues, 1466–1467
Number needed to treat (NNT), 46–47
Numeric outcomes, 45
Nutrition, 762–788, 762b
  for acute kidney injury, 230
  ageing and, 774
  chronic intestinal failure and, 786–787
  clinical skills for, 762–763
  dietary requirements, 764–769, 768f
  FODMAPs and, 788
  food intolerances and, 787–788
  fortification of foods, 780
  intestinal failure and, 783–787, 783b
  minerals, 777–780
  short bowel syndrome and, 784–786, 786f
  starvation-related under-nutrition, 769–774
  surgery and, 342–343
  water and electrolyte balance, 764
  of young adults, 1634–1635
Nutritional support, 780–783
  critical illness, 300
  enteral nutrition, 300, 781–782
  intravenous (parenteral) nutrition, 300
  parenteral nutrition, 782–783
  principles, 780–781
  re-feeding syndrome, 300
Nystagmus, 897, 906

O

Obesity, 1248–1269
  aetiology of, 1248–1249, 1249b
  anthropometric measurements of, 1252–1253
  bariatric surgery for, 1267–1269
  biology of, 1249, 1250f
  care
    e-health in, 1264
    implementation of physical activity in, 1264
    multidisciplinary team in, 1261, 1262f
    physical activity in, 1263–1264, 1264f–1265f
    specific dietary strategies, 1262–1263, 1263b
  clinical approach to, 1251–1253, 1252b–1253b
  complications of, 1257–1260
  control of appetite of, 1251
  core skills and knowledge, 1248b
  diagnosis, of hip-to-waist ratio, 1253
  dietary strategies for, 1261–1262
  epigenetics in, 1249–1250
  examination of, 1252, 1256f
  future therapies for, 1267
  genetic investigations, 1257
  gut-brain axis, 1251
  hedonic hunger, 1251, 1252f
  history of, 1252, 1255b
  investigations, 1255–1257, 1258b–1259b
  lifestyle intervention in, 1261–1264
  management of, 1260–1269, 1260b, 1261f
  medical therapy in, 1265–1269
  obesogenic environment, 1249
  pathophysiology of, 1259–1260
  physiological mediators of, 1250–1251
  prevalence of, 1248
  staging of, 1254–1255, 1258f
  through the life course, 1261
  type 2 diabetes and, 1209
  young adults and, 1634–1635
Obesity-focused medical history, components of, 1254b
Obesogenic environment, obesity, 1249
Observational studies, of evidence-based practice, 39–40
Obsessive-compulsive disorder (OCD), 1024–1025, 1024b
Obstetric and perinatal obesity, 1259
Obstetric medicine, 1680–1693, 1682f
  clinical skills for, 1681
  core skills and knowledge in, 1680b
  examination in, 1681
  history for, 1681
  investigations in, 1681, 1681b
  screening for, 1681, 1681b
Obstructive megaureter, 826–827
Obstructive shock, 311
Obstructive sleep apnoea, 549–550, 550f, 550b
Occipital cortex, 895
Occupational asthma, 541, 541b
Occupational cancers, 1381
Occupational lung disease, 577–578, 577b
Occupational sensitizers, 541, 541b
Occupational therapy, in Parkinson's disease, 932
OCD. *See* Obsessive-compulsive disorder
Oculocutaneous albinism, 1626
Oculomotor nerve (III), 896–897
  lesions, 897, 897b
Odds ratios (OR), interpreting, 45–46
Odonophagia, 596–597, 596b
Oesophageal cancer, 630, 1402
  adenocarcinoma, 630
  advanced or metastatic, 1402
  aetiology of, 630
  clinical features of, 630
Oesophageal cancer *(Continued)*
  diagnosis of, 631, 631f
  early, 1402
  epidemiology of, 630
  investigations of, 631
  management of, 1402
  palliative approaches of, 1402
  prognosis of, 1402
  risk factors for, 630b
  squamous cell carcinoma, 630
  staging of, 631
  survival, 1396b
Oesophageal disorders, 628–630
Oesophageal diverticulum, 628
Oesophageal infections, 629
Oesophageal perforation, 630
Oesophageal rings/webs, 628–629
Oesophageal rupture, 630
Oesophageal stenting, 149
Oesophageal tumours, 630–632, 631f
Oesophagitis, 623f
Oesophagus, 604, 621–632
Oestrogen receptor-positive disease, breast cancer, 1400
Oestrogens
  deficiency, 1128, 1180, 1180b
  effects of, 1180b
  hirsutism, 1184
  liver damage produce, hepatotoxicity, 732
  in PCOS, 1184
  for prostate cancer, 1408
  sodium retention and, 860
Oil acne, 1576b
Okadaic poisoning, 370.e1
Older adults
  adverse drug reactions, 112, 112b
  clinical approach to, 1696
  diabetes mellitus, 1235–1236, 1235b–1236b
  dialysis for, 846
  giant cell arteritis in, 920, 1115
  gout in, 1103
  investigations in, 1699
  kidney disease in, 852
  mental problems in, 1040
  osteoporosis on, 1128, 1131f
  prescribing in, 112, 112b
  UTI in, 852
Olfactory nerve (I), 891–892
Oligospermia, 1179
Omega-3 fatty acids, 515b–516b, 519
Onchocerca, 1509b
Onchocerciasis, 1510
Oncogenes, 76, 76b
Oncology, interventional, 146b
One and a half syndrome, 896
Onychomycosis, 1596f, 1597
Open valvotomy, 455
Ophthalmology, 1044–1060
  applied anatomy and physiology, 1044–1046
  core skills and knowledge, 1044b
  emergencies, 1048, 1049f
  eye trauma, 1048
  sudden onset sight loss, 1049–1050, 1051b
Opioid receptors, 911
Opioids/opiate, 1032–1033, 1032b–1034b, 1033f, 1724–1725
  dose titration and route of, 1724–1725, 1725b
  overdose, 361
  side effects of, 1725
  for surgery, 344
  titration in non-malignant respiratory disease, 1731
  toxicity, 1725
Opsonization, 83f
Optic atrophy, 894
Optic chiasm, 894
Optic disc, 1057f
Optic nerve (II), 892–896, 893f
  head oedema, 893
  lesions, 893–894, 894f
  pupils, 895–896
  visual acuity, 893–896
  visual field defects, 893
Optic neuritis, 894, 895b
Optic radiation, 894–895
Optic tract lesions, 894–895
Optical coherence tomography (OCT), 894
Optical imaging, 151
Oral anticoagulants, 1370f, 1371–1374, 1372b–1373b
Oral contraceptives, 1165
Oral corticosteroids, 543
Oral disease, 170
Oral rehydration solutions (ORS), 1463b, 1487
Orchitis, 1667–1668
Orellanin poisoning, 370.e7
Orf, 1449, 1594
Organ dysfunction, 281
Organ rejection, in clinical transplantation, 105, 105b
Organ support, in critical illness, 300
Organophosphorus insecticide poisoning, 369
Oriental hepatolithiasis syndrome, 738
Orlistat, 649, 1265
Oropharyngeal squamous cell carcinoma, 597
Oropharynx, 604
ORS. *See* Oral rehydration solutions
Orthobunyaviruses, 1455b
Orthomyxovirus, 1445b
Orthopnoea, 394, 529–530
Osgood-Schlatter disease, 1077
Osmolality, urine, 798
Osmoregulation, clinical disorders of, 866
Osmotic demyelination syndrome, 868
Osmotic diarrhoea, 613
Osmotic pressure, 856–859, 856f, 856b
Osteoarthritis, 1083–1087, 1083f, 1084b
  aetiology and pathogenesis of, 1083–1084, 1085f
  clinical features of, 256b–257b, 1084
  clinical subsets of, 1084–1086
  crystal-associated, 1086, 1086f
  epidemiology of, 1083–1087
  erosive, 1086
  factors predisposing to, 1084b
  first carpometacarpal, 1071
  investigations in, 1086
  localized, 1084–1086
    hip, 1074–1075, 1084–1085
    knee, 1075–1076, 1085–1086, 1085f
    nodal, 1071, 1084, 1085b, 1086f
  management of, 1086–1087
  medication for, 1087
  physical measure for, 1086–1087
  primary generalized, 1086
Osteoblasts, 1124
Osteochondritis dissecans, 1076
Osteochondromatosis, 1119
Osteoclasts, 1088, 1124
Osteocytes, 1124
Osteogenesis imperfecta, 1082
Osteomalacia, 840, 1134–1135, 1134b
Osteomyelitis
  acute and chronic, 1135
  diagnosis of, 1135
  digital, 124f
  management of, 1135
  tuberculous, 1135
Osteonecrosis, 1132–1133
Osteoporosis, 840, 1128–1132
  associated disease, 1129
  clinical features of, 1129
  crush fractures, 1074
  definition of, 1128
  glucocorticoid-induced, 1128, 1132
  investigations of, 1129–1130
  in men, 1132
  pathogenesis of, 1128
  prevention and management of, 1130–1131
  risk factors of, 1128–1132, 1128b
  treatment for, 1129–1130, 1130b–1131b
Ostium secundum atrial septal defect, 472f
Ostium secundum defects, 472f
Otalgia, 586b
Otitis media, 587f, 588–589
Otorrhoea, 586b, 587–588
Otosclerosis, 589
Ottawa Subarachnoid Haemorrhage (SAH) Rule, 181b
Outcomes Registry for Better Informed Treatment (ORBIT) score, 181b
Outflow resistance (afterload), 430, 431f
Outpatient parenteral antimicrobial treatment (OPAT), 287
Ovarian cancer, 1411, 1659b
  advanced, 1411
  early, 1411
  genetics, 75–76
  management of, 1411–1412
  MRI of, 142f
  palliative approaches, 1411–1412
  prognosis, 1411, 1411b
  relapsed, 1411
  risk of malignancy index, 1411b
  survival, 1396b
Ovary, ultrasound of, 128–129
Overactive bladder, in older adults, 1714–1715
Overflow incontinence, in older adults, 1715
Overlap syndromes, 1114
'Owl's eye' inclusion, 1457f
Oxazolidinones, for sepsis, 291b, 293
Oxygen, toxicity, 319
Oxygen cascade, 302, 302f
Oxygen consumption, 300–307, 301f
Oxygen delivery, 300–307, 301f
Oxygen therapy
  administration of, 319, 319f
  COPD, 548–549, 548f, 549b
  for respiratory failure, 319, 319f–320f
Oxygenation of blood, 301–303
Oxyhaemoglobin dissociation curve, 301, 301f
Oxytocin, 1190b

P

p arm, 57–58
p24 antigen, viral, 1545
p53 gene, 76–77
Pacemaker
  permanent, 409–411, 409f
  types of, 409f, 410
Pacing
  temporary cardiac, 196, 196f
  transcutaneous, 196
  transvenous, 196
Paget's disease, 1133–1134, 1133f
PAH. *See* Pulmonary artery hypertension
Pain, 910–911, 1062
  breast, 1641
  gate theory of, 911
  management of, 299
  neurotransmitters involved in, 911
  perception, 912f
  peripheral and central sensitization, 911
  physiology of, 910–911
  plasticity and receptor changes in, 911
  sports and performing arts-associated, 1079
  surgery and, 347–348
Pain/gas/bloat syndrome, 643–644
Palliative care, 1721b
  access to, 1721
  advance care planning, 1721b, 1722–1723
  assessment of, 1721–1722, 1721b–1722b
  cancer, 1386
  care provision, 1721
  constipation, 1726, 1727b
  core skills in, 1720b

Palliative care *(Continued)*
difficult conversations in, 1722–1723
disability, 1728–1729
dying patients, 1733–1734, 1734b
fatigue, 1728
gastrointestinal symptoms, 1726–1727
general aspects of, 1720–1722
goal of, 1720
knowledge in, 1720b
loss of function, 1728–1729
malaise, 1727
mouth symptoms, 1728
nausea, 1726
in non-malignant disease, 1729–1733
pain, 1723–1726, 1724f
patient's needs, 1722, 1722b
physical symptoms, 1728–1729
psychosocial issues, 1729
rehabilitation, 1728–1729
respiratory symptoms, 1727, 1728f, 1729b–1730b, 1730f
symptom control, 1720–1734, 1723b
use of, 1721–1722, 1721f, 1721b–1722b
vomiting, 1726
weight loss, 1727
in young people, 1733
Palpable breast cysts, 1641–1642, 1642b
Palpitations, 201–204, 201b–202b, 394
Pancreas, 746–749
anatomy and structure of, 747f
β-cells, 1204f–1205f
endocrine function, 747–748
exocrine function, 746–747, 748f
investigations, 748–749
structure, 746
transplantation, 1222
Pancreatic adenocarcinoma, 1405, 1406f
Pancreatic cancer, 756–759, 1405–1406
adenocarcinoma, 756
aetiology, 756–757
clinical features, 757–758, 757f
cystic, 759, 759f, 761f
differential diagnosis, 759
investigations, 758–759, 758f–759f
management, 759
neuroendocrine, 760
pathogenesis, 757, 757f
risk of, 757b
signs, 758
survival, 1396b
Pancreatic disease, 735–761, 735b
Pancreatitis, 749–757
acute, 749
chronic, 753
classification, 749–756
Pancytopenia, 1288b
Panic disorders, 1022
P_aO_2/F_IO_2 ratio, 319
Papillary necrosis, 834f
Papilloedema, 893–894, 894b
Papular urticaria, 1599–1600
Papules, 1568f–1570f
PAR. *See* Physical activity ratio
Paracentesis, 716
abdominal, 190b, 191f
Paracetamol
liver damage produce, hepatotoxicity, 732
overdose, 359, 360f, 360b
surgery and, 345
Paracoccidioidomycosis, 1495
Paraganglioma, 1161, 1691
Paragonimiasis, 1515–1516
Parainfluenza, 1451
Paralysed patients, sepsis in, 284
Paralytic ileus, surgery and, 348
Paralytic shellfish poisoning, 370.e1
Paramyxovirus, 1445b
Nipah virus, 1460
Paraneoplastic glomerulonephritis, 817
Paraneoplastic syndromes, 962, 1118, 1383, 1384b, 1385f
Paraphimosis, 1666–1667, 1666f
Parapneumonic effusion, 215
Paraproteinaemic neuropathies, 966
Paraproteins
immunological assessment, 1335
secretion, 1359, 1360f
Parasite infestation, 649, 650b
Parasomnias, 928
Parathyroid hormone (PTH), 1125, 1193–1196
Parenteral anticoagulant, 268, 1369–1372, 1370b
Parenteral nutrition, 782–783
complications of, 783
in critical illness, 300
vascular access for, 783, 783b
Parietal cortex lesions, 910
Parkinsonism, 934
Parkinson's disease, 904, 929–932, 1027
aetiology of, 929
clinical evolution of, 930
clinical features of, 929–930
dementia, 957
diagnosis of, 930
gait in, 907–908
management of, 930–932, 931f
neurotransmitter changes in, 905b
pathology of, 929
Paroxysmal cold haemoglobinuria, 1304
Paroxysmal hemicrania, 919
Paroxysmal nocturnal dyspnoea (PND), 394
Paroxysmal nocturnal haemoglobinuria, 1305
Pasteurellosis, 1470
Patent ductus arteriosus, 472, 473f
Patent foramen ovale, 998
Pathological fractures, 1361
Patient-centred outcome measures, evidence-based practice and, 50
Paul-Bunnell test, 1457
PBC. *See* Primary biliary cholangitis
PCI. *See* Percutaneous coronary intervention
PCOS. *See* Polycystic ovary syndrome
PCR. *See* Polymerase chain reaction
Peak expiratory flow rate (PEFR), 533–534, 534f–535f
Pectus carinatum, 579–580
Pectus excavatum, 579–580
Pediculosis pubis, 1536
Pedigree, drawing, 53, 54f
PEFR. *See* Peak expiratory flow rate
Pegloticase
for gout, 1104–1105
for hyperuricaemic nephropathy, 835
Pellagra, 777, 779f
Pelvic inflammatory disease, 1528
Pelvic malignancies, MRI of, 138, 141f–142f
Pelvic organs, ultrasound of, 128–129, 129f
Pelvis
CT of, 131–132, 134f–135f
MRCP of, 138
Pelviureteric junction obstruction, 826, 827f
Pemphigus, 1612
Pendred's syndrome, 1166
Penicillins
for leptospirosis, 1481
for sepsis, 289, 291b
for syphilis, 1534
for Whipple's disease, 649
Penile cancer, 1409, 1667
Penis, diseases of, 1666–1667
Peptic stricture, 625
Peptic ulcer disease, 633
Peptide YY (PYY), 747, 1251
Percussion, chest, 530
Percutaneous coronary intervention (PCI)
for angina, 441–442, 443f
bleeding after, 641
for STEMI, 449–450
Percutaneous transluminal coronary angioplasty (PTCA), 443f, 491
Perennial rhinitis, 537
allergic, 537, 538f
non-allergic, with eosinophilia, 537
Peri-aortitis, 678
Peri-arrest rhythms, therapeutic procedures for, 195–196
Pericardial disease, 487–490
Pericardial effusion, 411f, 488, 489f
Pericardiocentesis, 410–411, 411f
Pericarditis, 842
acute, 487–490
clinical features of, 393
constrictive, 489
dialysis, 842
incessant or chronic, 487–488
malignant, 488
post-cardiac injury syndromes, 488
tuberculous, 488
uraemic, 842
Perinatal brain injury, 924
Periodic paralysis, hyperkalaemic, 872
Perioperative medicine, 334–352
Perioral dermatitis, 1577
Periorbital cellulitis, 593
Peripartum cardiomyopathy, 487, 1685
Peripheral arterial disease, 490–494
Peripheral blood, 1273–1276, 1274b, 1277f
in anaemia, 1278, 1278f–1279f
Peripheral nerve compression and entrapment, 964, 964b
Peripheral nerve conduction, 914
Peripheral nerve disease, 962–970, 962f
chronic sensorimotor neuropathy, no cause found, 968
clinical approach to, 963, 963b
genetic, 968
HIV-associated, 942
vitamin deficiency neuropathies, 967–968
Peripheral nerve lesions, 908
Peripheral nerves, 908
biopsy, 915
conduction, 914, 915f
Peripheral neuropathies, autonomic, 1230–1231, 1230b
Peripheral non-lactating breast abscess, 1644–1645
Peripheral oedema, 394
Peripheral oxygen saturation, 535
Peripheral vascular disease, 490–494
Peripheral venous disease, 494–495
Peritoneal adhesions, 678
Peritoneal dialysis, 845–846, 846f
adequacy of, 846
catheter siting in, 846f
complications of, 845–846, 846b
continuous ambulatory, 845
dialysate concentrations in, 846b
in frail patients, 846
nightly intermittent, 845
Peritoneum, 677, 678b
Peritonitis
CAPD, 846b
generalized, 676
localized, 676
spontaneous bacterial, 716
tuberculous, 678
Peritonsillar abscess, 597
Permanent pacing, 409–411, 409f
Pernicious anaemia (PA), 638, 1285
Peroxisomal diseases, 1247
Persistent generalized lymphadenopathy (PGL), 1543
Personal protective equipment, for healthcare-associated infections, 1438
Personality disorder, 1033–1036, 1035b–1036b
Pertussis (whooping cough), 161f, 161b
PESI. *See* Pulmonary embolism severity index
PET. *See* Positron emission tomography
Petechiae, 1568f–1570f
Peutz-Jeghers syndrome, 653, 653f
pH, urine, 798, 829
Phaeochromocytoma, 1161, 1199, 1691
history and examination of, 1161–1162, 1162b
investigations, 1162, 1162f
management, 1162, 1162b
pathology, 1161
pregnancy, 1691
Phagocytosis, 1306–1307
Pharmacodynamics, 110–111
Pharmacokinetics, 109–110
Pharyngeal diphtheria, 1471
Pharyngeal pouch, 597–598, 597f–598f
Pharyngitis, 538–539, 596–597, 597b
Pharynx, 604, 621–632
Phenothiazines, hepatotoxicity, 732
Phenotype, 55f
frailty, 1697, 1697b
Phentermine, for obesity, 1266
Phenylketonuria (PKU), 1245–1246
Philadelphia chromosome (Ph), 76, 77f, 1335
Phimosis, 1666, 1666f
Phleboviruses, 1455b
Phobic disorders, 1023
Phosphate, 779
in bone disease, 1126
concentration of, disorders of, 874–877, 875f
Phosphate nephropathy, acute, 836
Phosphide poisoning, 370
Phosphodiesterase inhibitors, 317
Phosphodiesterase type 4 inhibitors, 547–548
Photoprotection, 1622–1623, 1623b
Photosensitivity disorders, 1611, 1611b
Phototherapy, 1573–1574, 1585–1586
Phthiriasis palpebrarum, 1053, 1055f
Phyllodes tumours, 1642, 1642f
Physical activity
diabetes mellitus, 1214
in obesity, 1251, 1263–1264, 1264f–1265f
young adults and, 1635
Physical activity ratio (PAR), 766, 768b
Physical deterioration, 375
Physical health, mental illness and, 1013
Physiotherapy
for Parkinson's disease, 932
for pneumonia, 214
specialized, 299
PICO approach, 36, 37b
Picornaviruses, 1445b, 1448, 1448b
Pigmentation, disorders of, 1625–1626
Pigmented villonodular synovitis, 1120
Pinta, 1486
Pituitary-adrenal axis, 1152b
Pituitary apoplexy, 1150, 1151f
Pituitary-dependent hypercortisolism, 1156
Pituitary gland, 1145–1151, 1145f
Pituitary-thyroid axis, 1164b
Pituitary tumours, 1146, 1147b
differential diagnosis, 1149, 1149f
investigations, 1147–1148, 1147f
management, 1148–1149, 1148b
Pityriasis rosea, 1586
PKU. *See* Phenylketonuria
Plague, 1489
Plain radiographs, 118–125, 118b
Plantar fasciitis, 1078
Plantar spurs, 1078
Plants, poisoning, 370.e6
Plaques, 1568f–1570f
Plasma
osmolality, 866
protein, complications of, 1312
transport of, 1140
viscosity, 1276
Plasma cell neoplasms, 1359b
Plasma potassium, hypomagnesaemia and, 873
Plasmacytoma, 1362
Plasmapheresis, in myasthenia gravis, 974
Plasmin, 1318
Platelet concentrates, 1313
Platelet disorders, 1321–1324, 1321b
Platelets, 1316
adhesion, 1316
aggregation, 1316
function disorders, 1323, 1323b
release, 1316, 1318f
transfusion, 1326
Platinum compounds, 1389
Pleocytosis, sterile CSF, 941, 941b
Pleura, 524
Pleural aspiration, 536
Pleural biopsy, 536
Pleural effusion, 119, 120f, 557–559
chest X-ray, 558f
diagnosis, 557–558
management of, 559
management, 559
pancreatitis and, 756
Pleural plaques, 579b
Pleural pneumonia, 536
Pleural rub, 531
Pleurisy, 580
Plexiform neurofibroma (schwannoma), 961f
Plexus and nerve root lesions, 969–970, 969b
Plummer-Vinson syndrome, 628–629
PML. *See* Progressive multifocal leucoencephalopathy
Pneumocystis jirovecii infection, 1496
HIV-associated, 1554–1555, 1555b
pneumonia, 216, 552–553
Pneumocytes, 523, 524f
Pneumonia, 212–215, 212f, 552–553, 1472
clinical features, 212–213, 213b
complications, 215–218, 215b
in immunocompromised patients, 216
initial assessment, 213
investigations, 213, 213b
management, 214–215, 214f
pregnancy, 1687
prevention, 215
Pneumothorax, 119, 121f, 216–218
aspiration, 218, 218b
clinical features of, 216
investigations of, 217
management of, 217–218, 217f, 218b
tension, 195
POAG. *See* Primary open-angle glaucoma
Podophyllotoxin, 1531
POEMS syndrome, 966, 1362
Point mutations, 65
Point-of-care ultrasound (POCUS), 183, 192b
Poisoning, 355, 355b
approach, 355–356
examination of, 353–354, 353b, 355b
general management, 355–356
history, 353–354, 353b
investigations, 355b–356b
management strategy in, 356b
metabolic acidosis, 356b
in natural world, 370
plants, 370.e6
principles of management of, 356–359, 356b
specific, 359–370
specific management of, 358–359

Poliomyelitis, 1458–1461
Politics, global health and, 173
Polyarteritis nodosa, 817–819, 1116
Polychromasia, 1288
Polycyclic hydrocarbons, 1381
Polycystic liver disease, 726–727
Polycystic ovary syndrome (PCOS), 1180–1181, 1183, 1183f
  clinical features, amenorrhoea, 1180–1181
  diagnosis, 1184
  differential diagnosis, 1184
  history and examination of, 1183
  investigations, 1184
  management, 1184
Polycythaemia, 1306
  causes of, 1306b
  'relative'/'apparent' (Gaisböck's syndrome), 1306
  secondary, 1306, 1306b
Polycythaemia vera, 1306, 1343–1344, 1344b
Polydipsia, primary, 1192
Polyenes, 1496
Polygenic hypercholesterolaemia, 516
Polygenic inheritance, 69
Polyglandular autoimmune syndromes, 1198
Polymerase chain reaction (PCR), 69, 69f
Polymixins, for sepsis, 291b
Polymorphic light eruption, 1586
Polymyalgia rheumatica, 1075, 1115
Polymyositis and dermatomyositis, 1113
Polymyxins, for sepsis, 294
Polyneuropathies, 963–968, 965b
Polyol pathway, 1224
Polyomavirus infection, HIV-associated, 1559, 1559f
Polypharmacy, in older people, 1704–1705
  benefits of, 1705, 1706f
  clinical decision-making in, 1705, 1706f
  drug interactions, 1704
  evidence base in, 1705
  patient choice in, 1705, 1705b
  prescribing, 1704, 1704b
Polyposis syndromes
  classification of, 670b
  familial adenomatous polyposis, 670
  inherited, 670–671
  MYH-associated, 670
Polysymptomatology, 254
Polyuria, 1192
Pomade acne, 1576b
Pomalidomide, 1392
Pompe disease, 973, 1246.e2b
Pompholyx eczema, 1582f
Pons, 898, 904
Pontine lesions, 909
Pontine tumours, 898
Popliteal (Baker's) cyst, 1077
Popliteal vein thrombosis, 130f
Porphyrias, 1246, 1247f
Portable high-altitude chamber, 375, 375f
Portal hypertension, 712–718, 712b
Porto-pulmonary hypertension, 718
Portosystemic encephalopathy (PSE), 716, 717b
Port-wine stain (naevus flammeus), 1628–1629
Positive end-expiratory pressure, 322–323
Positron emission tomography (PET), 143–146
  cardiovascular, 407–408
  of gastrointestinal disease, 618
  of Hodgkin lymphoma, 1350f
  of kidney disease, 799
  of respiratory disease, 532–533
  of rheumatic disease, 1067
Positron emission tomography-computed tomography (PET-CT), 144f, 145, 146b
Post-cholecystectomy syndrome, 740
Posterior circulation clinical stroke syndromes, 981–982
Posterior circulation stroke, 590, 590b
Posterior knee pain, 1077
Posterior pituitary, 1146, 1148b
Posterior uveitis, 1058
Post-exposure prophylaxis, 1554
Post-inflammatory hyperpigmentation, 1626
Post-inflammatory hypopigmentation, 1625–1626
Postpartum haemorrhage, 1692
Postpartum thyroiditis, 1165
Postprandial hypoglycaemia, 1197
Post-streptococcal glomerulonephritis, 809, 810f
Post-thrombotic syndrome, 1366, 1366f
Post-traumatic headache, 920
Post-traumatic stress disorder (PTSD), 1024
Postural back pain, 1072
Postural hypotension, 863, 863b
  in older adults, 1710, 1712b
  syncope, 238
Postural orthostatic tachycardia syndrome (POTS), 418–419
Postural tremor, 906
Posture, in Parkinson's disease, 930
Postviral/chronic fatigue syndrome, 1458
Potassium, 779
  absorption of, 795
  concentration, disorders of, 869–872
  hypomagnesaemia and, 873
  urinary loss, 1161
Potassium-sparing diuretics, 862, 862f
POTS. *See* Postural orthostatic tachycardia syndrome
Pott's disease, 1135
Pouchitis, 664, 664f
Poverty, 170–171, 170b
Poxvirus infections, 1448–1449
PPIs. *See* Proton pump inhibitors
PR3-ANCA positivity, 812
Pragmatic RCTs, evidence-based practice and, 50
Prague classification, 625, 626f
Pralidoxime, 366, 369
Pramlintide, 1221
Pre- and infrapatellar bursitis, 1076
Precordium, 396–397
Prediabetes, 1211
Pre-eclampsia, 509b, 1683–1684, 1683b–1684b
Pregnancy
  adrenal disease, 1691
  bacteriuria in, 832, 1691
  cardiac disease, 1685–1686, 1685b
  complications of, 1652
  dermatoses, 1692b
  diabetes mellitus, 1236–1237
  direct oral anticoagulants for, 1373
  ectopic, ruptured, 676
  endocrine disease, 1688–1691
  female health and, 1652
  genital herpes in, 1532–1533
  HELLP syndrome, 1684
  HIV and, 1553–1554
  hypertension, 1683–1684, 1684b
  infectious disease and, 1426
  kidney disease, 1691–1692
  liver disease, 1684–1685
  low-molecular-weight heparin for, 268–270
  mental illness and, 1652
  neurological disease, 1687–1688
  postnatal visit, 1652–1653
  prescribing in, 111–112, 112b, 1693
  respiratory disease and, 1686–1687
  rheumatoid arthritis, 1692, 1693b
  skin disorders in, 1692, 1692b
  SLE and, 1110
Pregnancy *(Continued)*
  syphilis and, 1533–1534
  teenage, 1652
  termination of, 1653–1655
  test, in acute abdomen, 675
  thromboembolic disease, 1686, 1686b
  thyroid function tests and, 1165
Preload, 303, 303f, 429–430
Premature menopause, 1181
Prenatal diagnosis, congenital adrenal hyperplasia, 1159
Presbycusis, 585f, 589
Prescribing, 108b
  act of, 113
  breast-feeding women, 112
  calculation skills, 113
  electronic, 113
  ethical issues relating to therapeutics and, 113–114
  in kidney disease, 829b
  liver disease, 113, 733–734
  older adults, 112, 112b
  pregnancy, 111–112, 1693
  special population, 111–113
Pressure support ventilation, 323
Pressure ulcers, 1609, 1609b–1610b
Priapism, 1667
Prickly heat rash, 379, 1586
Primaquine, overdose, 364
Primary biliary cholangitis (PBC), 718, 718f
Primary central nervous system lymphoma, 1356
Primary cutaneous T-cell lymphoma (mycosis fungoides), 1623, 1623f
Primary gastric lymphoma, 639
Primary hyperaldosteronism, 859
Primary hyperlipidaemia, 513
Primary hypertension, 497–498
Primary immunodeficiency, 98–100
Primary kidney disease, 822, 823f
Primary open-angle glaucoma, 1055–1056, 1057f
Primary sclerosing cholangitis (PSC), 719, 719f, 744
Prion disease, 1468
Procedural ethics, 26
Procoagulants, gain of function, 1367
Proctitis, 663
Proctoscopy, 617–618, 618b
Prodrome, 238, 922
Professional competence and mistakes, 33–34
Progressive familial intrahepatic cholestasis, 690
Progressive multifocal leucoencephalopathy (PML), 942, 1461
Progressive myoclonic epilepsy-ataxia syndromes, 933
Progressive rubella encephalitis, 943
Proinsulin, 1204f
Prolactin, 1188f, 1189b
Prolactinoma, 1147, 1147b
  pregnancy, 1690–1691
Prophylaxis, for meningitis, 248
Propionic aciduria (PA), 1246.e1b
Propofol, for surgery, 343–344
Proprotein convertase subtilisin kexin-9 (PCSK9), 802
Prostaglandin (PG), 797, 1318f
Prostate
  anatomy, 852
  benign conditions of, 1662–1664
  benign enlargement of, 1663–1664, 1663f, 1664b
Prostate cancer, 1407, 1664–1665, 1664b–1665b, 1665f
  clinical features of, 1407
  diagnosis of, 1407
  disease localized to, 1408
  endocrine therapy for, 1408
  locally advanced, 1408
  management of, 1408–1409
  non-hormonal therapy, 1408–1409
  prognosis of, 1408, 1408f
  screening, 163–164
  survival, 1396b
Prostatitis, 1663
Prostheses, infected, management of, 257
Prosthetic heart valves, 464–465, 464f
  complications of, 464
  interruption of anticoagulant therapy, 464
  mechanical *vs.* tissue, 464
  para-prosthetic leak, 464
  pregnancy, 464–465, 1685–1686
Protease inhibitors (PIs), 1550–1551
Proteasome inhibitors, 1393
Protein
  structure, 60–61
  synthesis and secretion, 60–61
  translation, 60–61, 60f
  in urine, 798–799
Protein binding, 828
Protein C, 1367–1368
Protein-losing enteropathy, 650, 650f
Protein metabolism, disorders of, 1245–1246, 1246.e1b
Protein S, 1318, 1367–1368
Proteins, 767–768
Proteinuria, 802
Proteomics, 1385–1386
Prothrombin complex concentrates, 1313
Prothrombin time (PT), 686–687, 1320, 1324b, 1327b, 1372–1373
Proton pump inhibitors (PPIs), 624
Protozoal infections, 1497
  blood and tissue, 1497–1506
  gastrointestinal, 1506–1508, 1506b
  HIV-associated, 1557
Proximal femur, fractures of, 1129
Proximal tubule, 858b
Pruritus, 1604, 1604b
Pruritus ani, 667
PSC. *See* Primary sclerosing cholangitis
PSE. *See* Portosystemic encephalopathy
Pseudobulbar palsy, 902
Pseudogout, causes and clinical features of, 256b–257b
Pseudomembranous colitis, 665
Pseudoseizures, 927
Pseudoxanthoma elasticum, 1611
Psoriasis, 1584–1588
  aetiology, 1584
  associated features, 1585, 1585f
  clinical features, 1584–1585, 1584f
  erythrodermic and pustular, 1585
  exacerbating factors, 1584–1586
  flexural, 1585
  guttate, 1585
  management, 1585–1586, 1585f
Psoriatic arthritis, 1101
  children, 1118
  clinical features of, 1101, 1101f
  management and prognosis of, 1101–1102
Psychiatric disorders
  acute disturbance, 1009b, 1014
  alcohol withdrawal, 1014–1015
  anxiety disorders, 1021–1022
  chronic fatigue syndrome/myalgic encephalopathy, 1029–1030
  eating disorders, 1036
  mood (affective) disorders, 1017, 1017f
  obsessive-compulsive disorder, 1024–1025, 1024b
  personality disorder, 1033–1036, 1035b–1036b
  in selected patient groups, 1039–1041, 1040b
  self-harm and suicide, 1013–1014
Psychiatric history, 1002b–1003b
Psychodermatology, 1630
Psychological obesity, 1259
Psychometric assessment, 915
Psychosis, 1025–1027
Psychosocial issues, in palliative care, 1729
Psychotic disorders, 1026–1027
  aetiology of, 1027
  management of, 1028, 1028b
  other medical conditions, 1027–1029
PTCA. *See* Percutaneous transluminal coronary angioplasty
PTH. *See* Parathyroid hormone
PTSD. *See* Post-traumatic stress disorder
Pubertal gynaecomastia, 1180
Pubic lice (phthiriasis pubis), 1536, 1599
Public health, 153–166
  causality, 155
  core skills and knowledge in, 153b
  epidemiology, 155
  global health and, 153–154
  health needs assessments, 154f, 155, 156f
  healthcare, 163–166
  inequalities in, 155–157, 156f, 156b
  pillars of, 157–166
    health improvement, 154f, 158–159, 158b
    health protection, 159–163
  politics and advocacy for, 157
  teams, 155b
Public mental health, 1042–1043
Publication bias, 48, 49f
Pulmonary alveolar proteinosis, 573–574
Pulmonary artery hypertension (PAH), 478
  clinical features of, 478
  epidemiology of, 478
  investigations of, 479, 479f
  management of, 479–480
  pregnancy and, 479, 1685
Pulmonary embolism, 216, 264–272
  clinical assessment of, 265
  CT angiogram of, 136f
  CT of, 136f
  definition of, 264
  diagnosis of, 266–267
  epidemiology of, 264–265
  imaging for, 267
  investigations of, 265–266, 269f–270f
  risk assessment, 1369b
  risk factors for, 264–265
Pulmonary Embolism Rule-out Criteria (PERC) Score, 181b, 267b
Pulmonary Embolism Severity Index (PESI) Score, 181b, 270, 271b
Pulmonary embolus, 273, 273f
Pulmonary eosinophilia, 575–576
  asthmatic, 575, 575b
  chronic, 575
  hypereosinophilic syndrome, 575, 575b
  prolonged, 575b
  simple, 575, 575b
  tropical, 575, 575b
Pulmonary fibrosis, 120, 122f
Pulmonary gas exchange, 302, 302f
Pulmonary hamartoma, 563–564
Pulmonary heart disease, 477–480
Pulmonary hypertension, 477–480, 478b, 1366
  ARDS and, 326
  clinical classification of, 478b
  pathophysiology of, 478
  sickle cell anaemia and, 1296
Pulmonary infiltration, sarcoidosis, 569, 569f
Pulmonary lymphangioleiomyomatosis, 573
Pulmonary oedema, 118–119, 119f
  acute heart failure and, 207
  high-altitude, 375
  management of, 326
  non-cardiogenic, 326
Pulmonary regurgitation, 463–464, 463f
Pulmonary rehabilitation, 549

Pulmonary stenosis, 463, 463f
Pulsatile tinnitus, 586b
Pulse, 394–395
  character of, 395, 395f
  mitral stenosis, 453
  rate, 395
  rhythm, 395
Pulse oximetry, 318
Pulsed-wave (PW) Doppler, 402
Pulsus alternans, 395, 395f
Pulsus bigeminus, 395, 395f
Pulsus bisferiens, 395, 395f
Pupil, 895–896
  light reflex, 895f
  relative afferent pupillary defect, 895
  of unconscious patients, 234, 234f
Pure red cell aplasia, 1288
Purgative abuse, 644
Purine antagonists, 1389
Purkinje fibres, 388–389
Purpura, 1568f–1570f
  due to infections, 1321
  due to steroids, 1321
  immune thrombocytopenic, 1322
  post-transfusion, 1323
  senile, 1321
  thrombotic thrombocytopenic, 1323
Pus cells, in urine, 799
Pustules, 1568f–1570f
p-values, 43–49
  critical appraisal and, 41–43, 42b
  relative risk of, 46
Pyelonephritis
  acute, 831, 831f
  xanthogranulomatous, 833
Pyoderma gangrenosum, 1603, 1603f
Pyogenic abscess, 727–728, 727f
Pyogenic granuloma, 1616
Pyrimidine 5′ nucleotidase deficiency, 1301
Pyrimidine antagonists, 1389
Pyrophosphate analogues, 1467
Pyruvate kinase deficiency, 1300–1301

Q

q arm, 57–58
Q fever, 1482
Quadrantanopia, bitemporal, 894
Quality of life, 155
Quantitative polymerase chain reaction, 1335–1336
Quinagolide, 1189
Quincke's sign, 461
Quinine, overdose, 364
Quinolones, for sepsis, 291b, 292
Quinsy, 597

R

Rabies, 1460–1461
  clinical features of, 1461
  control of, 1461
  diagnosis of, 1461
  management of, 1461
  prevention and control of, 1461
Radial artery cannulation, 304b
Radial nerve compression, 964
Radiation
  cancer and, 1381–1382, 1382b
  dermatitis, 382
  enteritis, 649
  ionizing, 381, 382b
  plexopathy, 970
  proctitis, 649
Radiation sickness
  acute, 382
  haemopoietic syndrome, 382
  management, 383
Radiation therapy, 1394–1396
  for acromegaly, 1187
  adjuvant, 1394
  for breast cancer, 1400, 1400b
  cancer and, 1381–1382
  clinical application of, 1394–1395
  in combination with systemic therapy, 1395
  curative, 1395b
  for hyperprolactinaemia, 1190
  palliative, 1394, 1395b
  radical, 1394
Radiation therapy *(Continued)*
  side-effects of, 1395–1396, 1395b
  theoretical background of, 1394
  types of, 1395
  x-rays with matter, 1394
Radiculopathy, 963
Radioactive iodine, 1170
Radiofrequency ablation (RFA), 625
  liver metastasis, 151
Radiography
  of gastrointestinal disease, acute abdomen, 675, 675f
  of kidney disease, 800
  osteoporosis, 1129, 1129f
  Paget's disease, 1133f, 1134
  pancreatitis, 750, 751f
  of spine, 1071, 1072f
Radio-iodine, 1170, 1171b
Radioisotopes, in gastrointestinal disease, 619
Radiologically inserted gastrostomy (RIG), 150, 150f
Radiology, 117–152, 117b
  background of, 117–118, 118b
  interventional, 146–151, 146b
Ramsay Hunt syndrome, 899
Randomization, of evidence-based practice, 38
Randomized controlled trials (RCTs), 108, 108b
  of evidence-based practice, 38
Ranson scoring system for pancreatitis, 752b
Rapid tranquilization, adverse events from, 1009b, 1015f
Rapidly progressive glomerulonephritis, 809f, 809b
Rashes, 274–277
Raynaud's phenomenon/disease, 494, 1111
Reactive arthritis, 256b–257b, 1102, 1102f
Reactive systemic (secondary AA) amyloidosis, 807
Reassurance and support, constant, 299
Receptor-mediated endocytosis, 512f
RECIST criteria, 1388b
Recombinant human erythropoietin, 797
Recombinant human parathyroid hormone, osteoporosis, 1132
Recompression therapy, 381
Recreational drugs, young adults and, 1635–1636
Rectal cancer, MRI of, 141f
Rectal examination, of acute abdomen, 675
Rectal prolapse, 668–669
Rectal ulcers, 668–669
Rectum, 611–612, 665–674
Recurrent ischaemic stroke, 998
Recurrent miscarriage, 1652
Red blood cells (erythrocytes), 1276–1277
  haemoglobin. *See* Haemoglobin
  haemolytic transfusion reactions, 1310
  immune destruction of, 1302, 1303f
  metabolism, 1299–1301, 1300f
  phagocytosis, 1306–1307
  production and breakdown, 1279f
  sequestration, 1306–1307
  in urine, 798, 798f
Red cell concentrates, 1313
Red cell scan, 145–146
Red eye
  differential diagnosis of, 1047b
  red flags for, 1046b
5α–reductase deficiency, 1185b
Re-feeding, 772
Re-feeding syndrome, 300, 772–773, 876
  management and prevention of, 773
  pathophysiology of, 773, 774b
  risk factors of, 773
Referred otalgia, 587, 588f
Reflux nephropathy, 831–832, 831f, 834f
Refractive errors, 1050, 1052f
Refractory defects, diabetic eye disease and, 1227
Regulatory ethics, 26
Regulatory T lymphocytes, 90
Rehabilitation
  malnourished children, 772
  in older adults, 1715–1716
  in palliative care, 1728–1729
  in paraplegia, 952
  pulmonary, 549
  in traumatic brain injury, 948
Rehydration solutions, oral (ORS), 1463b, 1487
Reinke's oedema, 598, 599f
Relapsing fevers, 1490
Relapsing polychondritis, 1120
Relative afferent pupillary defect, 895
Relative risk, 46
Remdesivir, 1467
Remnant hyperlipidaemia, 517, 517f
Renal arteries
  coil embolization of, 148f
  fibromuscular dysplasia, 820–821
  occlusion of, 821
Renal arteriography, 800
Renal calcification, 800f
Renal cancer, 1406
  survival, 1396b
Renal cell carcinoma, 851–852
Renal disease
  complications of obesity, 1259
  palliative care, 1731
  vascular, cholesterol embolization, 736f
Renal function
  autocrine, 797
  countercurrent system, 794f
  DMSA and, 144
  endocrine, 796–797
  GFR, 794–796
  hypercalcaemia, 1194
Renal hypertension, 819
Renal mass lesion, ultrasound of, 128f
Renal replacement therapy, 231
Renal scintigraphy, 800, 800f
Renal stones, 785–786
Renal tubular acidosis, 880–881
  causes of, 881b
  diagnosis of, 881b
  in kidney calculi, 822
  type 3, 881
  type 4, 880–881, 881b
Renal tubules
  associated disease, 858b
  collecting ducts, principal cells, 878
  distal, 858b, 873
  loop of Henle, 858b
  proximal, 858b
  sodium transport, 858b
Renal tumour, CT urogram of, 137f
Renal vein thrombosis, 821
Renin-angiotensin system, 796–797
Renoprotection, 843
Renovascular disease, 820
Reoviruses, 1445b
Replacement fluids, 864, 864b
Reproduction
  definitions, 1175–1190, 1175b
  embryology, 1175–1176
  physiology, 1176–1177, 1176f
Reproductive disorders, 1177, 1177b–1178b
  complications of obesity, 1259
  disorders of sex development, 1185–1190, 1185b
  gonadal function, investigation of, 1177, 1178b
  men, 1178–1180
  women, 1180–1184
Resident ectoparasite infections, 1517–1518
Respiration, 525
  Cheyne-Stokes, 394
  control of, 525, 526f, 526b
  mechanical process, 525
  normal values, 526b
Respiratory acidosis, 308, 879
Respiratory alkalosis, 308, 879
Respiratory disease, 529–537
  ARDS and, 327f
  clinical features, 529–530
    breathlessness, 529–530, 530b
    chest pain, 530
    cough, 529
    haemoptysis, 529, 529b
    nasal symptoms, 529
    sputum, 529
    wheezing, 530
  complications of obesity, 1258
  diffuse parenchymal lung disorders, 568, 568f
  hypersensitivity pneumonitis, 572–573
  investigation of, 532–537
  lower respiratory tract
    bacterial infections, 1472
    coronavirus, 1452, 1452b
    influenza, 1453, 1453b
    viral infections, 1451–1454
  pregnancy, 1686–1687
    asthma, 1686
    cystic fibrosis, 1687
    pneumonia, 1687
    sarcoidosis, 1687
    severe restrictive lung disease, 1687
    tuberculosis, 1687
  upper respiratory tract
    bacterial infections, 1470–1472
    epiglottitis, 1472
    influenza, 539
    laryngotracheobronchitis, 539
    rhinitis, 537–540
    viral infections, 1451
Respiratory failure, 318–325
  classification and aetiology of, 318–325
  clinical features, 318
  COPD, 207, 208f
  investigations, 318
  management of, 319–325
  monitoring in, 318–319
Respiratory medicine, 521–581, 521b, 522f, 523b
Respiratory microbiome, 528
Respiratory support
  invasive, 321, 321b
  for respiratory failure, 320–325, 321b
  techniques for, 322–323
Respiratory syncytial virus (RSV), 1453
Respiratory system
  airflow, 525
  airways, 525–526
  alveolar stability, 527
  anatomy, 523–525
  defence mechanisms, 527–529, 528f
  examination, 530–532
  function of, 523–529
  physiology, 525–529
  ventilation and perfusion relationships, 526–527, 528f
Respiratory tract, viral infections, 1451–1454, 1451b
Respiratory tumours, 559–565
  benign, 563–565
  malignant, 559–563
Rest tremor, 906
Restless leg syndrome, 928–929
Restrictive cardiomyopathy, 486, 486f
Resuscitation, 772
Retina
  anatomy, 1045, 1045f
  disorders of, 1059–1060
Retinal detachment, 1059, 1059f
Retinitis pigmentosa, 1059
Retinoblastoma, 74–75, 1383b
Retroperitoneal fibrosis, 678, 827, 827f
Retrosternal/intrathoracic thyroid, 580
Retroviruses, 1445b, 1458b
Revascularization, for heart failure, 436–437
RFA. *See* Radiofrequency ablation
RFs. *See* Rheumatoid factors
Rh blood group system, 1308
Rhabdomyolysis, 836, 1243
  in poisoning, 357
Rhabdoviruses, 1445b
RhD immunization, prevention of, 1305
Rheumatic fever, 1478, 1479b
Rheumatoid arthritis (RA), 1088–1098
  ACR/EULAR 2010 criteria for, 1091b, 1097b
  aetiology and pathogenesis of, 1088, 1089f
  antinuclear antibodies, 1089f
  clinical features of, 1089–1093, 1090b–1091b
  complications of, 1091
  diagnosis and investigations in, 1093–1094, 1094f
  differential diagnosis of, 1090b
  disease activity score, 1095f
  epidemiology of, 1088–1098
  immunology for, 1088, 1089f
  joint involvement in, 1091–1093, 1092f
  management of, 1094–1098, 1094b
    drug therapy, 1094–1098, 1095f
    physical measures, 1098
    surgery, 1098
  non-articular manifestations in, 1093, 1093f
  pathology of, 1088–1089, 1090f
  pregnancy, 1692, 1693b
  prognosis of, 1090b, 1098
  seronegative, 1089, 1091
Rheumatoid factors (RFs), 1065, 1066b, 1088
Rheumatology, 1061–1120, 1063f
  clinical conditions in, 1067
  clinical skills for, 1062–1120
  investigations, 1065–1067
  joint examination, 1064b, 1067f
  musculoskeletal history, 1062, 1062b
Rhinitis, 537–540
  common cold (acute coryza), 537, 1451
  investigations and diagnosis, 537–538
  management, 538
  nasal polyps, 537
  perennial, 537
  seasonal, 537
  vasomotor, 537
Rhinorrhoea, 591b–592b, 593
Rhinosinusitis, 593, 593b
Rickets/osteomalacia, 1134–1135, 1134b
Rickettsia spp., 1490–1491, 1490b, 1491f
Rickettsial pox, 1490b
Rickettsial spotted fever, 1491f
Riedel's thyroiditis, 1173
Rifaximin, for sepsis, 295
Rift Valley fever, 1464
Rigidity, in Parkinson's disease, 930
Rilonacept, for gout, 1105
Rinne test, 585f
Risk assessment
  of ADPKD, 850b
  of osteoporosis, 1129–1130, 1130b–1131b
  psychiatric, 1003–1006, 1005b
  pulmonary embolism, 1369b
RNA interference (RNAi), 72
Rockall score, for upper gastrointestinal bleeding, 224, 224b
Rocky Mountain spotted fever, 1490b
Romosozumab, 1132
Rosacea, 1576, 1576f
Rotator cuff
  tendonosis, 1069
  torn, 1069
Rotavirus, 1461–1462, 1462f
Rotor's syndrome, 690
Round opacities, 120, 122b
Roux-en-Y gastric bypass (RYGB), 1268, 1269f
RSV. *See* Respiratory syncytial virus

Rubella, 1450
  arthritis, 1107
  clinical features of, 1450, 1450f
  complications of, 1450
  diagnosis and management of, 1450
  prevention of, 1450
  progressive encephalitis, 943
Rubeola. *See* Measles
Ruptured ectopic pregnancy, 676

S

Saccular (berry) aneurysms, 243, 243f
Safeguarding, in older adults, 1717–1718, 1718b
SAH. *See* Subarachnoid haemorrhage
Salicylates, overdose, 365, 366f
Saliva, hormone assays, 1144, 1155
Salivary glands, 621, 622b
Salmonella spp., 1473
  HIV-associated, 1560
Same day emergency care (SDEC), 177
Sanger sequencing, 69–70, 70f
Sanitation, 171
SAPHO syndrome, 1119
Sapovirus, 1462
Sarcoid lung disease, CT of, 134f
Sarcoidosis, 568–576, 1118, 1603–1604
  clinical features, 568–570, 569f–570f, 569b
  epidemiology and aetiology, 568
  immunopathology, 568
  investigations, 570
  pregnancy, 1687
  prognosis and management, 570
Sarcopenia, 1698, 1699f, 1699b
SARS-CoV-2 infection
  HIV-associated, 1559
  skin manifestations in, 1595, 1595b
Saxitoxin poisoning, 370.e1
SBAR (Situation, Background, Assessment, Recommendation), 180–181
SBS. *See* Short bowel syndrome
Scabies, 1517, 1560, 1598–1599, 1598f, 1599b
Scaling, 1568f–1570f
Scaphoid fractures, 1071
Scarlet fever, 1470, 1471f
Schatzki's ring, 629, 629f
Scheuermann's disease, 1136
Schistosoma haematobium, 1514–1515
Schistosoma japonicum, 1515, 1515b
Schistosoma mansoni, 1515
Schistosomiasis, 728, 1513–1514
  clinical features of, 728, 1514–1515
  diagnosis of, 1515
  geographical distribution of, 1514f
  investigations, 728
  management of, 728, 1515
  parasitology and pathogenesis of, 1514, 1514f
  prevention of, 1515
Schwannomas, 945
SCID. *See* Severe combined immunodeficiency
Scintigraphy, 533
SCLC. *See* Small-cell lung cancer
Scleritis, 1093
Scleroderma, 1606–1607, 1607f
Sclerosing lesion, of breast, 1642
Sclerotic bone metastases, 124f
Scombroid fish poisoning, 370.e1
Scorpions, poisoning, 370.e2
Screening
  of ADPKD, 850b
  of cancer, 1382–1386, 1383f
  of diabetic eye disease, 1227, 1228b
  of haemochromatosis, 721
  of hyperlipidaemia, 512–513
  of hypothyroidism, 1167
  of multiple endocrine neoplasias, 1198–1199
Screening *(Continued)*
  neonatal, for haemoglobinopathies, 1298–1299
  for public health, 163–164, 164b
  for STIs, 1520–1522
Scrub typhus, 1491
Scurvy, 776–777, 777f
Seasonal rhinitis, 537
Seasonal/perennial conjunctivitis, 1054
Seborrhoeic eczema, 1580, 1581f
Second generation antipsychotics, 1028
Secondary biliary cirrhosis, 719
Secondary hyperlipidaemia, 513
Secondary peristalsis, 604
Secondary thrombocytosis, 1345
Secretions, excessive respiratory, 1727
Secretory diarrhoea, 613–614, 614f
Sedating antihistamines, for atopic eczema, 1580
Seizures, 249–250, 921f
  brain tumours, 924, 946
  causes of, 249b
  clinical assessment of, 250
  focal, 922–923
  generalized, 922
  management of, 250
  types of, 249–250
Seldinger technique, 184f, 186b, 187f
Selection bias, critical appraisal and, 42
Selective catheterization, for Cushing's syndrome, 1156
Selective IgA deficiency, 99–100
Selective oestrogen-receptor modulators (SERMs), for osteoporosis, 1132
Selective serotonin reuptake inhibitors (SSRIs)
  bulimia nervosa, 1039
  overdose, 361
Selenium, 779
Self-determination, 29b
Self-harm, 1013–1014
Self-management education, for diabetes mellitus, 1213–1214, 1214f
Senile purpura, 1321
Senior review, early, patient flow and, 178
Sensation, loss of, 910f
Sensory ataxia, 908
Sensory cortex, 908, 909f
Sensory loss, 910f
Sensory pathway lesions, 908–910
Sensory pathways, 908–911, 909f
Sensory polyneuropathy, 1543
Sepsis
  acute kidney injury and, 230
  antibiotic therapies, 289–296, 290f
  antimicrobial stewardship, 286–289
  clinical features of, 283–284
  clinical skills for, 278–281
  core skills and knowledge in, 278b
  critical care medicine, 308–318
  examination of, 280–281
  history of, 278, 279b
  management of, 278, 279f, 284–286, 313–318, 314f
  meningococcal, 1479
  multiple organ failure and, 312–313, 312f
  for nephrotic syndrome, 802
  neutropaenic, 272, 272b
  origin of, 283
  pathophysiology of, 281
  recognition of, 282–284, 283b
  in special situations, 284
  symptoms and signs, 312
  treatment of bacterial infection, 278–296, 280f–281f
Sepsis Six, 284, 284b
Septal haematoma, 594f
Septic arthritis, 255–257, 256b–257b, 1091
Septic discitis, 1074
Septic shock, 281, 312
Septo-optic dysplasia, 1150
Sequential Organ Failure Assessment (SOFA) score, 281, 282b
Sequestration, erythrocytes, 1306–1307
SERMs. *See* Selective oestrogen-receptor modulators
Seroconversion rash, 1601–1602
Serotonin syndrome, 1016–1017
Serum amylase, 675
  common bile duct stones, 750
  elevation, 750b
Serum-ascites albumin gradient, 715b
Serum biochemistry, in renal function, 799
Serum tumour markers, 1383–1386, 1385b
Sessile serrated adenomas, 669–670
Setmelanotide, 1266
Severe combined immunodeficiency (SCID), 99
Severe cutaneous adverse drug reactions, 1629
Sever's disease, 1077
Sex, 1670b
  in medicine, 1639–1679, 1639b, 1646b
Sex-linked disorders, 66–67, 67f
Sexual dysfunction
  diabetic patients, 1231
  erectile dysfunction, 1180
Sexual health, of young adult, 1638
Sexually transmitted infections (STIs), 1519–1536, 1519b
  asymptomatic, screening for, 1520–1522
  examination of, 1519, 1521f–1522f
  history of, 1519, 1520b
  investigation of, 1519, 1523b–1526b
  management, prevention and control of, 1523–1524
  symptomatic, screening for, 1523
  trichomoniasis, 1530
Sézary's syndrome, 1357–1358
Sheehan's syndrome, 1150
Shigella spp., gastroenteritis, 1474
Shin pain, 1077
Shingles, 942, 1447, 1447f
Shock, 308–318
  aetiology of, 308–318
  anaphylactic, 312
  cardiogenic, 311
  causes of, 308b
  clinical features of, 311–313, 312b
  hypovolaemic, 311
  management of, 313–318, 314f
  metabolic response to, 313, 313b
  pathophysiology of, 308–311
  septic, 312
Shock wave lithotripsy, 740
Short ACTH stimulation test, 1145f, 1145b, 1148b, 1157, 1192b
Short bowel syndrome (SBS), 648–649, 784–786, 786f, 786b–787b
Short QT syndrome, 425
Short-lasting unilateral neuralgiform headache with conjunctival injection and tearing (SUNCT), 919
Shoulder girdle pain, 1068
Shoulder pain, 1067–1070, 1068b–1069b, 1070f, 1091–1092
Shuffling gait, 907–908
Shwachman-Diamond syndrome, 100
SIADH. *See* Syndrome of inappropriate antidiuretic hormone secretion
Sicca syndrome, 1093
Sick sinus syndrome, 413
Sickle cell anaemia, 1296
  clinical features, 1296
  complications of, 1298b
  investigations of, 1294f, 1297
Sickle cell anaemia *(Continued)*
  long-term problems in, 1297
  management of, 1297–1298
  prognosis for, 1298
Sickle cell crisis, 1119, 1298b
Sickle cell disease, 65
Sickle cell trait, 1298
Sickle nephropathy, 816–817
Sickle syndromes, 1295–1299, 1296f
Sideroblastic anaemia, 1282b, 1283
Sievert, 382
Sigmoidoscopy, 618
Sildenafil, for PAH, 479
Silicosis, 578
Simplified acute physiology score, 330, 332b
Single-gene disease, patterns of inheritance of, 66–69
Sinoatrial disease, 413
Sinonasal tumours, 592, 593f
Sinus arrhythmia, 412
Sinus bradycardia, 412–413, 413f
Sinus node, 388–389
Sinus tachycardia, 412
Sinusoids, of liver, 682
Sirolimus, drug-eluting stents, 441–442
Sjögren syndrome, 1114
Skeletal muscle, 1065
Skeletal radiography, 122, 124f
Skin
  chemotherapy-related, 1390, 1391f
  complications of obesity, 1259
  dermis, 1572
  epidermis, 1570–1572
  functions of, 1570–1572, 1571b
  structure, 1570–1572, 1571f
  subcutis, 1572
Skin and soft tissue infections
  bacterial, 1469–1470, 1469b
  low-/middle-income countries, 1483–1487
  toxin-mediated, 276
  viral, 1444
Skin disease, 1566–1570, 1570f, 1571b
  age distribution of, 1570f
  diagnostic approach to, 1567f–1570f
  distribution and pattern, 1566, 1566f–1567f, 1566b
  investigation of, 1572, 1572b
  lesion morphology, 1568f–1570f
  sarcoidosis, 568–576, 570f
  systemic disease-related, 1602–1611
  tropical dermatoses, 1600
Skin infections, 1589–1600
  bacterial, 1589–1591
  of breast, 1645, 1645f
  fungal, 1595–1598
  mycobacterial, 1591–1592
  viral, 1592–1595
Skin-prick tests, 537
Skin tumors, malignant, 1618–1625
Skinfold measurements, 1252–1253, 1257f
Skull
  base of, 899
  fractures, 947
  MRI of, 143f
SLE. *See* Systemic lupus erythematosus
Sleep, young adults and, 1636
Sleep disorders, 928–929
Sleeve gastrectomy, of bariatric surgery, 1268
Slit lamp, 1047f
Slow-transit constipation, 617
Small bowel obstruction, CT of, 135f
Small-cell lung cancer (SCLC), 1398
Small intestinal cancer, 1403, 1403f–1404f
Small intestine, 642–653
  absorption of
    active transport, 607, 608f
    calcium, 608
    carbohydrate, 607–608
    facilitated diffusion, 606
Small intestine *(Continued)*
    fat, 608, 609f
    iron, 608
    metals and trace elements, 608
    nutrients, 607–608
    protein, 608
    water and electrolytes, 608
    water-soluble vitamins, 608
  anatomy of, 605–611, 606f
  to antigens and pathogens, response of, 608–611
  biopsy of, 646
  gut microbiome, 611
  mechanisms of, 609–611
  oral tolerance in, 611
  structure of, 606f
  tumours of, 651–653
Small vessel disease, 980–981
Small-cell lung cancer, 1398
Smallpox (variola), 1448
Small-vessel vasculitides, 574
Smoking, 551–552
  cessation, 547, 552
  dangers, 551–552, 551b
  diabetic patients, 1214
  prevalence, 551–552
  toxic effects, 551
  young adults and, 1635
Smoking bans, 157
Snellen chart, 1047–1048, 1048f
Snoring, 600
Snow, John, 153, 154b
Social anxiety disorder, 1023
Social determinants of health, 170–172
Social media, young adult mental health and, 1633–1634
Sodium, 779
  absorption of, 795
  concentration, disorders of, 865–869
  retention, 860
Sodium calcium edetate, 368
Sodium fusidate, for sepsis, 294
Sodium-glucose co-transporter-2 (SGLT2) inhibitors, 862, 1219–1220
  diabetic ketoacidosis and, 261, 261b
  for heart failure, 435–437
Sodium valproate, liver damage produce, 733
SOFA score, 330
Sofosbuvir, 703
Soft tissues, MRI of, 138
Solar keratitis, 375, 376f
Solitaire stentriever device, 993f
Solitary cysts, 727
Solitary nodular goitre, 1173
Solitary plasmacytoma, 1362
Solitary rectal ulcer syndrome, 668–669
Solitary toxic adenoma, 1168
Somatostatin, 747, 1186b
Somatostatin receptor agonists, 1187–1188
Somatostatinoma, 760
South American trypanosomiasis (Chagas' disease), 1503
Spacers, 543b
Spastic paraparesis, 142f, 904
Spasticity, 907
Specific gravity, of urine, 798
Specimen collection, 1432b
Speculum examination, 1647f
Speech therapy, in transgender women, 1678
Spermatic cord disease, 1667–1668
Spherocytosis, hereditary, 1289–1290, 1290f–1291f
Sphincter of Oddi dysfunction, 740–741
Spiders, poisoning, 370.e2
Spinal canal stenosis, 1074
Spinal cord, 904, 908
  compression, 273, 909, 911f, 950–951, 950f, 950b, 1328
  dorsal columns, 908
  lesions of, 909
  spinothalamic tracts, 908–909
  subacute combined degeneration, 968
Spinal cord disease, 936, 950–952
Spinal cord injury, 188b

Spinal cord tumours, 950, 950b
Spinal epidural abscess, 944
Spinal reflex arc, 907–908, 907f, 907b
Spinal roots, 908, 908f
  lesions of, 909
Spinal stenosis, 969
Spinal vacuolar myelopathy, 942
Spine
  cervical, rheumatoid arthritis, 1092–1093
  imaging, MRI, 1067
    cervical spine, 1069f
    knee, 1076f
    lumbar spine, 1073f
    spinal metastasis, 1072f
  rapid examinations of, 1064b
Spinoreticular tracts, pain perception, 912f
Spinothalamic tracts, 908
  lesions of, 909
  pain perception, 912f
Spirometry, 533, 534f, 534b
Spleen, 1306–1307
  atrophy of, 1307
  function of, 1306–1307
  in rheumatoid arthritis, 1093
Splenectomy
  haematology after, 1291f, 1307
  immune thrombocytopenic purpura, 1323
  infection prophylaxis, 1307, 1307b
  problems after, 1307
Splenic sequestration, 1297
Splenomegaly, 691, 1307
Splicing mutations, 65
Spondyloarthritis, 1099–1106, 1099b
Spondylolisthesis, 1074
Spondylosis, lumbar, 1072–1073
Spontaneous bacterial peritonitis, 716
Spontaneous osteonecrosis of knee (SONK), 1076, 1077f
Sporadic adenomas, 669, 670b, 671f
Sporotrichosis, 1495–1496
Spotted fevers, 1490b, 1491
Sputum, 529, 535–536
Squamous cell carcinoma (SCC)
  of mouth, 621
  of oesophagus, 630
  oropharyngeal, 597
SSRIs. *See* Selective serotonin reuptake inhibitors
Stabilization, 772
Stamping gait, 908
Standardized protocols, patient flow and, 178
Staphylococcal scalded skin syndrome, 276, 1470
Staphylococcus aureus
  associated conditions, 1443b
  gastroenteritis, 1475
  invasive infection, 1469–1470
  virulence factors, 1470
Starling's law, 390, 390f
Starvation-related under-nutrition, 769–774
  clinical features, 771–772, 771f, 771b
  insulin/glucagon ratios, 770
  investigations, 772
  lean-to-fat tissue, proportion of, 770–771
  management of, 772
  pathophysiology of, 769–770, 770f
  prevention, 772–773, 773b
  prognosis, 772
  regulation of metabolism, 770
  role of cytokines, 770–771
  WHO classification, 771b
Stasis, 1581
STAT3 syndrome, 69
Statin intolerance, 518–519
Statins, 515b–516b, 518, 737
  for coronary artery disease, 440
Statistical significance, 44
Statistics, 43
Status epilepticus, management of, 250, 251b
Status migrainosus, 918
Steatorrhoea, 755
ST-elevation myocardial infarction (STEMI), 448
  clinical features of, 448, 449f–450f
  complications of, 450–451
  management of, 449–452
  pathophysiology of, 448, 448b
Stem cell therapy, 73
Stem cells, 61, 62f
Stent placement, 146b
Sterile CSF pleocytosis, 941, 941b
Sterility, chemotherapy-induced, 1391–1392
Sterilization, contraception and, 1653
Steroid compounds, hepatotoxicity of, 732
Steroid-sparing agents, 543
Steroids, 327
  for anti-GBM glomerulonephritis, 812
  purpura due to, 1321
Steroid-sparing agents, for asthma, 543
Stertor, 600
Stevens-Johnson syndrome (SJS), 277, 1574, 1603b, 1629b
Stiff person syndrome, 975
Stiffness, 1062
Stillbirth, 1652
Stimulants, overdose, 362
Stings, poisoning, 370.e2
STIs. *See* Sexually transmitted infections
Stokes-Adams attacks, 415
Stomach, 632–640
  acid secretion of, 605, 605f
  anatomy of, 604–605, 604f
  emptying, 605
  physiology of, 605
Stomach cancer, survival, 1396b
Stomatocytosis, hereditary, 1291
Stone dissolution, 740
STOP BANG tool, 550, 550b
Streptococcus pneumoniae, meningitis, 1478–1479
Stress cardiomyopathy, 486–487, 486f
Stress (march) fractures, 1078, 1078f
Stress incontinence, 1657b
  in older adults, 1714, 1715b
Stress-related hormones, 1144
Stress ulceration, 357
Stridor, 596b, 599–600, 599b
Stroke, 977–1000, 1027, 1243–1244
  brainstem infarction, 981, 984b–986b
  cardioembolism and, 979–981, 980f
  cerebral infarction, 991, 994f
  clinical syndromes, 984b–986b
  core skills and knowledge, 977b
  decompressive craniectomy for, 994f
  definitions of, 977
  imaging, 983–991, 989f, 990b
  ischaemic, 979–980, 980f
  management, 994b
  pathogenesis and aetiology of, 977–979, 979f
  pregnancy, 1688, 1688b
  primary prevention of, 993, 996b
  risk factors of, 993, 995f
  secondary prevention of, 993
  TOAST classification, 978, 979b
  vascular anatomy of, 981, 982f–983f
  venous, 981
Stroke unit care, 991–992
  bowel and bladder dysfunction, 992
  brain infarction/malignant hemispheric cerebral infarction, 991, 994f
  dysphagia, 991
  goals of therapy, 992
  impairment of mobility, 991
  ongoing disability, 992
  post-stroke complications, 991–992
  spasticity and pain, 991
Stroke volume, 303
Strongyloidiasis, 1513
Study design, key messages about, 41b
Subacute cutaneous lupus erythematosus, 1607
Subacute sclerosing panencephalitis, 943
Subarachnoid haemorrhage (SAH), 243–245, 243b, 244f
Subcutaneous mycoses, 1598
Subcutaneous nodules, 1093
Subcutis, 1572
Subdural empyema, 944
Subdural haematoma, 245, 245f
  CT of, 133f
Substance use disorder, 1031b
Succussion splash, 635f
Sudden onset sight loss, 1049–1050, 1051b
Suicide, 1013–1014
Sulfasalazine, for rheumatoid arthritis, 1095
Sulphonylureas, 1219b, 1221
  overdose, 363
  poisoning, 364
Sunscreens, 1622–1623
Superficial cancers, 1407
Superficial reflexes, 904
Superficial thrombophlebitis, 494–495
Superior vena caval obstruction, 273
Supraventricular tachycardia (SVT), 418–419, 418b
Surgery, 170
  airway assessment of, 340, 341f
  anaesthesia and, 343
  anaesthetic assessment of, 340
  anaesthetic drugs in, 343–345
  blood ordering, 1309
  brain, 924
  consent in, 337–338, 338f
  core skills for, 334b
  corticosteroids and, 1159, 1160b
  of Crohn's disease, 661–662
  diabetic patients, 1236
  direct complications of, 348
  elective, blood transfusion, 1309
  emergency laparotomy, 350
  emergency *versus* elective, 337, 338b
  for epilepsy, 927
  fasting and, 342–343
  fluid balance and, 349–350
  frailty in, 340
  functional capacity assessment of, 339–340, 340f
  hip fracture, 350
  human factors and, 346
  hyperparathyroidism, 1195
  hyperthyroidism, 1170–1171
  intraoperative care of, 343–346
  medicine, 336f
  nutrition and, 342–343
  obstetrics, 351
  optimization of, 342
  orthogeriatrics, 349
  for osteoarthritis, 1087
  osteoporosis, 1132
  Paget's disease, 1134
  pathophysiology of, 337
  patient journey and, 336f
  perioperative care in, 337–343
  perioperative clinics and, 340
  perioperative fluid balance, 349–350
  postoperative care for, 346–350
  prediction models, 339, 339b
  prehabilitation, 342
  for rheumatoid arthritis, 1098
  risk assessment of, 338–340
  scoring systems, 339, 339b
  serum biomarkers in, 340
  shared decision-making in, 338
  specific common disease and, 340–342
  surgical safety checklist, 336f
  theatre safety and, 346
  for trauma, 351–352
  of ulcerative colitis, 663–664, 663b, 664f
Surgical drainage, 147, 147f
Surgical obesity, 1259
Surgical site infections, 350
Surviving Sepsis Campaign (SSC), 284–286, 285b–286b
Sustainable Development Goals, 168f, 173
Sustained ventricular tachycardia, 197, 197f, 197b
Suxamethonium, 872
SVT. *See* Supraventricular tachycardia
Swallowing, 604
  in Parkinson's disease, 930
Sweet syndrome, 1603, 1604f
Symmetrical distal polyneuropathy, 1230
Sympatho-adrenal response, to shock, 308, 309f
Synapses, 888, 890f
Synaptic transmission, 888
Syncope, 238, 927
  cardiac, 394
Syndrome of apparent mineralocorticoid excess, 1161
Syndrome of inappropriate antidiuretic hormone secretion (SIADH or SIAD), 866–868, 1192, 1193b, 1199
Synovial fluid, 1065
Synovial joints, 1064, 1065f
Synovitis, pigmented villonodular, 1120
Synovium, 1065
Syphilis, 1533
  cardiovascular, 494
  clinical features of, 1533–1534
  congenital, 1533–1534
  diagnosis of, 1534
  endemic (bejel), 1486
  HIV and, 1534
  latent, 1533
  management of, 1534
  neurosyphilis, 942, 1534
  pregnancy and, 1533–1534
  primary, 1533
  prognosis of, 1534
  secondary, 1533, 1533f
  tertiary, 1533
  transfusion-transmitted, 1312
Syringomyelia/syringobulbia, 952, 952f
Systemic disorders, causes and clinical features of, 256b–257b
Systemic inflammatory vasculitis, 1114–1118, 1115b, 1116f
  large-vessel, 1115–1116, 1116f
  medium-sized vessel, 1116–1117, 1116f
  small-vessel, 1116f, 1117
Systemic lupus erythematosus (SLE), 575, 815, 941, 1107, 1607, 1608f
  aetiology of, 1107–1108
  classification of, 815b, 821f
  clinical features of, 1108–1109, 1108f
  epidemiology of, 1107
  investigations in, 1109
  management of, 815–816, 1109–1110
  pathogenesis of, 1108
  pathology of, 1108
  pathophysiology of, 815
  pregnancy and, 1110
  prognosis of, 816, 1110
Systemic sclerosis (SSc), 575, 1111, 1111f

T

T cell activation, 85, 86f
T cell deficiency, 98–100
T-cell lymphoma, 1356, 1356f
T-cell receptor (TCR) genes, 87
T-cell-antigen presenting cell responses, 98b
T cells
  activation of, 85, 86f
  CD4, failure of activation, 102–103, 103b
  in rheumatoid arthritis, 1088
T helper cells
  T follicular, 90
  Th1, 88, 89b
  Th2, 88, 89b
  Th17, 88–90, 89b
T lymphocytes
  antigen receptors on, 86–87
  CD4, 88–90
  CD8, 90, 90f
  deficiency, 98b, 99
  development and activation, 87–90, 88f
  identification of, 88b
  other deficiencies, 99
Tabes dorsalis, 942
Tachycardia, 418–423, 863
  cardiomyopathy, 487
Tachypnoea, 393–394, 529
Taenia saginata, 1516–1517
Taenia solium, 942, 1516
Takayasu's arteritis, 1116
Takayasu's disease, 494
Takotsubo syndrome, 486–487, 486f
Tangier disease, 516
Tapeworms. *See* Cestodes
Tarsal tunnel syndrome, 1078
Taxanes, 1389
Tay-Sachs disease, 1246.e2b
TBG. *See* Thyroxine-binding globulin
TCAs. *See* Tricyclic antidepressants
Teeth, 621
Teicoplanin, for sepsis, 293
Telangiectasia, 1568f–1570f
Telehealth, for acute viral illness, 212
Telogen effluvium, 1628
Telomere shortening, ageing and, 1695
Temporal lobe herniation, 234f
Temporary cardiac pacing, 196, 196f
Temporomandibular pain dysfunction syndrome, 1080
Tendon reflexes, 903
Tendonitis, causes and clinical features of, 256b–257b
Tendons, 1065
Tennis elbow, 1070
Tenosynovitis, 1070
Tension pneumothorax, 195, 322
Tension-type headache, 242–243, 243b, 918
Teratogens, 112b
Teriparatide, 1130b, 1132
Testes, undescended, 1668
Testicular cancer, 1668
Testicular disease, 1667–1668
Testicular feminization, 1185b
Testicular germ cell tumours, 1668
Testicular synthetic defects, 1185b
Testicular torsion, 1667
Testis, ultrasound of, 130, 131f
Tetanus, 1480
Tetany, 1196b
Tetracyclines, for sepsis, 291b, 293, 293f
Tetralogy of Fallot, 474–475, 474f
Thalamic lesions, 910
Thalassaemia, 1292–1295, 1293f–1295f
Thalassaemia trait, 1282b, 1293
Thalidomide, 1392
Therapeutic drug monitoring, 114, 114b
Therapeutics, 107–116
  choice of drug, 107–108
  core skills and knowledge, 107b
  dose, 108–111, 109f
  drug interactions, 115, 116b
  information sources, 116
  need for drugs, 107
  pharmacodynamics, 110–111
  pharmacokinetics, 109–110, 110b
Thevetia peruviana, poisoning, 370.e6
Thiazide diuretics, 860–862, 861f
Thiazolidinediones, 1221
Third heart sound, 397
Thirst, 1190–1193
Thoracic inlet X-rays, for nodules, 1174
Thoracic outlet syndrome, 970
Thoraco-abdominal aneurysm, 492
Thorax
  CT of, 131, 134f
  ventilation/perfusion ($\dot{V}/\dot{Q}$) scan of, 143–144, 144f

Threadworm *(Enterobius vermicularis)*, 1511–1512
Throat, 595–596
  anatomy and physiology of, 595–596, 595f
  disorders of, 596–598
  examination of, 596
Thrombectomy, 146b
Thrombin time (TT), 1320–1321
Thromboangiitis obliterans, 494
Thrombocytopenia, 1321–1323, 1322b
Thrombocytosis, 1323–1324
Thromboembolism, pregnancy, 1686, 1686b
Thrombolysis, 146b, 270–272, 450, 450b
Thrombolysis in Myocardial Infarction (TIMI) Score, 181b
Thrombolytic therapy, indications, contraindications, and dosing for, 992b
Thrombophilia testing, 1367–1368
Thrombosis, 981
  cardiopulmonary resuscitation and, 195
  cortical venous, 246
  dural venous sinus, 246
Thrombotic thrombocytopenic purpura, 1323
Thymectomy, 973–974
Thymic education, failure of, 102–103
Thymic tumours, 581
Thymoma, 581, 817
Thyroid cancer, 1174–1175, 1175b
Thyroid colloid cyst, ultrasound of, 128f
Thyroid crisis/storm, 1171
Thyroid disorders, 1604
Thyroid eye disease, 1168
Thyroid function tests, 1164–1165, 1165b
Thyroid gland, 1163–1175
  nuclear imaging of, 145
  ultrasound of, 128, 128f
Thyroid hormones
  physiological effects, 1164, 1164b
  resistance, 1171
  synthesis and metabolism, 1163, 1163f
Thyroid stimulating hormone (TSH), 1141, 1164b
Thyroidectomy, 1170–1171
Thyroiditis, 1173
Thyrotoxic cardiomyopathies, 1171f
Thyrotrophin-releasing hormone (TRH), 1141, 1163–1164, 1164b
Thyroxine (T_4), 1163–1164, 1164b
Thyroxine-binding globulin (TBG), 1163
TIA. *See* Transient ischaemic attack
Tick-borne encephalitis, 1459–1460
Tick-borne relapsing fever, 1490
Tics, 899, 933
Tilt testing, 400, 403f
Time-restricted eating, 1263b
Time-to-event data, 47
TIN. *See* Tubulointerstitial nephritis
Tinea capitis, 1596f, 1597
Tinea corporis, 1596
Tinea cruris, 1596, 1596f
Tinea manuum, 1597
Tinea pedis, 1596, 1596f
Tinea unguium, 1597
Tinel's sign, 1071
Tinnitus, 586b, 591
Tissue perfusion, 304–306
Tissue plasminogen activator, 1319
Titratable acids, 878, 878f
TLRs. *See* Toll-like receptor
TOAST classification, 978, 979b
Tobramycin, for sepsis, 292
Todd's paralysis, 922–923
Togaviruses, 1445b
Tolerance, restoration of, 106
Toll-like receptor (TLRs), 85b
Tongue, 621
Tonic-clonic (grand mal) seizures, 922
Tonic pupil, 895
Tonic seizures, 922
Tonsillitis, 596–597, 597b
Tophaceous gout, 1105, 1105f
Topical therapy, 1572–1573, 1573b
  adverse effects of, 1573
  for atopic eczema, 1579–1580, 1580b
  corticosteroids, 1573
  for psoriasis, 1585, 1586b
Topiramate, for obesity, 1266
Tourette's syndrome, 933, 1024b
Toxic epidermal necrolysis (TEN), 277, 277b, 1574, 1603b
Toxic megacolon, 662, 662f
Toxic multinodular goitre, 1168
Toxic neuropathies, 967, 967b
Toxic nodule, 1174
Toxic pustuloderma, 277
Toxic shock syndrome, 276, 1470
Toxicology, 353–370.e7, 353b, 354f
  clinical skills for, 353–354
    examination of, 353–354, 353b, 355b
    history, 353, 353b
Toxocara canis, 1513
Toxocariasis, 1513
Toxoplasmosis, 1505
Trabecular (cancellous) bone, 1123
Trabeculectomy, for glaucoma, 1057
Trace elements, nutritional requirement, 781
Trachea
  anatomy, 523
  rupture, 579
  tumours, 560, 564–565
Tracheal intubation, complications of, 321b
Tracheostomy, 321–322, 321b–322b
Tracheostomy tubes, 599f
Trachoma, 1053, 1054b, 1486
Traction alopecia, 1628
TRALI. *See* Transfusion-related acute lung injury
Transcatheter aortic valve replacement (TAVR or TAVI), 460, 460f
Transcriptomics, 70
Transcutaneous kidney biopsy, 800, 801b
Transfusion-dependent thalassaemia, 1293
Transfusion-related acute lung injury (TRALI), 1311
Transgender, 1670b
Transgender women, hormone therapy in, 1674–1676, 1675b
Transgender health, 1670–1679
Transgender men, hormone therapy in, 1676–1678, 1677b
Transient ischaemic attack (TIA), 978b, 982–983, 987b
Transjugular intrahepatic portosystemic shunt, 150
Transoesophageal echocardiography, 404
Transposition of the great arteries, 475, 475f
Trans-septal balloon valvotomy, 455
Transudates, 558
Transverse myelitis, 939–940
Trauma, 170, 170b, 256b–257b, 579, 900
Traumatic brain injury (TBI), 924, 947–948, 1027
Travellers' diarrhoea, 1476, 1476b
Trematodes (flukes), 1513–1516
Tremor, 906
Trench fever, 1492
TRH. *See* Thyrotrophin-releasing hormone
Triazoles, 1497
Trichinosis, 1513
Trichomonas vaginalis, 1530
Trichomoniasis, 1530
Trichuris trichiura, 1512
Tricuspid regurgitation, 404f, 462–463, 462f
Tricuspid stenosis, 462, 462f
Tricyclic antidepressants (TCAs), overdose, 361
Trigeminal autonomic cephalgias, 918–919
Trigeminal nerve (V), 898, 898f
Trigeminal neuralgia, 898, 920
Trigeminal sensory neuropathy, 898
Triggered activity, 412
Triiodothyronine (T_3), 1163–1164, 1164b
Trimethoprim
  for sepsis, 291b, 294
  for UTI, 832
Trinucleotide repeat genetic disorders, 68b
Triplet repeat mutations, 67–68
Trochanteric bursitis, 1075
Trochlear nerve (IV), 896–897
  lesions, 897
Tropheryma whipplei, 649
Tropical dermatoses, 1600
Tropical pulmonary eosinophilia, 575, 575b
Tropical sprue, 647
Tropical ulcer, 1486
Trypanosomiasis, 1502–1503
Trypsin, 754f
Trypsinogen, 754f
TSH. *See* Thyroid stimulating hormone
Tuberculin skin test, 557
Tuberculoid leprosy, 1484, 1484f, 1591
Tuberculoma, 944
Tuberculosis, 553–557, 553b, 650, 1489
  BCG vaccination, 557
  clinical features, 553, 650
  diagnosis of, 650–651
  HIV co-infection, 557
  latent, 553, 553b, 557
  management, 554–557, 556b
    DOT therapy, 554, 556b
    drug resistance, 556–557, 557b
  microbiological diagnosis, 554
  Mycobacterium bovis infection, 557
  pathogenesis, 553
  pregnancy, 1687
  prevalence and risk, 553b
  primary, 553
  pulmonary, 553–554, 554f, 555b–556b
  reactivation, 553, 553b
Tuberculosis verrucosa cutis, 1591
Tuberculous arthritis, 1106
Tuberculous meningitis, 941
Tuberculous osteomyelitis, 1135
Tuberculous pericarditis, 488
Tuberculous peritonitis, 678
Tuberous sclerosis, 961, 1611
Tubular adenoma, 1641
Tubulointerstitial nephritis (TIN), 833–835
Tularaemia, 1483
Tumor necrosis factor (TNF), 1251
Tumor necrosis factor-α (F–α) blockers, for rheumatoid arthritis, 1097–1098
Tumor suppressor genes, 76–77
Tumour ablation, 148b
Tumour-induced osteomalacia, 1134
Tumour lysis syndrome, 274, 836
Turcot's syndrome, 671
Turner's syndrome, 1181, 1185b
Typhus, 1490–1491
Tyrosinaemia type 1A, 1246.e1b
Tyrosine kinase inhibitors, 106, 1398, 1398b

U

Ulcerative colitis, 662
  clinical features of, 662
  course and prognosis of, 664
  definition of, 662b
  examination of, 662
  extragastrointestinal manifestations of, 656–657, 657b
  investigations of, 662–663, 662f
  management of, 222, 223b
  medical management of, 663
  microscopic changes of, 655, 657b
Ulcerative colitis *(Continued)*
  pathology of, 655, 657f
  sites of, 657f
  surgical management of, 663–664, 663b, 664f
Ulcerative jejunitis, 647
Ulcers, 1568f–1570f
Ulnar nerve compression, 964
Ultrafiltration rate, 794
Ultrasound, 127–131, 127b
  of gastrointestinal disease, 618
  high-intensity focused, 130–131, 152
  of hydrocoele, 131f
  intravascular coronary, 404
  of kidney disease, 799–800
  of multinodular goitre, 128f
  of pelvic organs, 128–129, 129f
  point-of-care, 183, 192b
  reproductive disorders, PCOS, 1177f, 1184
  respiratory disease, 533
  special procedures of, 130–131
Ultraviolet exposure, 1622, 1622b
Ultraviolet radiation, 1622b
Unconscious patient, 232–236
Under-nutrition, mechanisms of, 764–774
Undescended testes, 1668
Unfractionated (standard) heparin, 1371
Unilateral paralysis, 580
Unknown onset bilateral tonic-clonic seizures, 922
Unstable angina, 440
Unverricht-Lundborg disease, 933
UOS. *See* Upper oesophageal sphincter
Upper gastrointestinal bleeding
  causes of, 223f
  clinical assessment of, 223–224
  Glasgow-Blatchford score for, 224, 225b
  investigations of, 224
  management of, 224–226, 225b
  risk assessment of, 224
  Rockall score for, 224, 224b
Upper gastrointestinal cancer, 1402–1403
Upper motor neuron (UMN) lesions, 903b, 904
Upper oesophageal sphincter (UOS), 604
Upper respiratory tract infections
  bacterial, 1470–1472
  epiglottitis, 1472
  influenza, 539
  laryngotracheobronchitis, 539
  rhinitis, 537–540
  viral, 1451
Upper tract tumours, 1407
Uraemia
  acute tubular necrosis, 828, 835–836, 836f
  in neuropathy, 967
Urban hypothermia, 376
Urea, plasma/serum, 796b
Urea cycle disorders, 1242–1243, 1242f–1243f, 1246.e1b
Urethral discharge, 1526, 1527f
Urethritis, 1523b
Urge incontinence, 1657b, 1714–1715
Uricosuric agents, for gout, 1105
Urinary anion gap, 881–882
Urinary catheters, for healthcare-associated infections, 1438–1439
Urinary incontinence
  female health and, 1657, 1657b
  mixed, 1657b
  in older adults, 1714–1715
  stress, 1657b
  urge, 1657b
Urinary tract, 791f
  disease, 789–852, 789b
Urinary tract calculi, 822–823
Urinary tract infection (UTI), 829–833, 829b
  aetiology and pathogenesis of, 829–833
  clinical features of, 830
  diagnosis of, 830–831, 830b
  indwelling catheter in, 832
  investigations of, 832
Urinary tract infection (UTI) *(Continued)*
  kidney carbuncle, 832–833
  management of, 832
  natural history of, 829f, 831–832, 831f
  in older adults, 852, 1712–1713
    investigations in, 1713, 1714b
    management of, 1713
    recurrent suspected, 1713, 1715b
  pregnancy, 1691
  re-infection, 830f
  relapse, 830f
  tuberculosis, 833, 833f
Urinary tract obstruction, 825–828
  aetiology of, 825–826, 825b
  clinical features of, 825
  investigations of, 826
  management of, 826
  pathophysiology of, 825
  post-obstructive diuresis, 826
  prognosis for, 828
Urine
  alkalinization, 358–359
  calcium, 1127, 1194
  catecholamines, 1162
  examination, 798–799
    in acute abdomen, 675
    for acute kidney injury, 229, 230b
    in carcinoid syndrome, 652
    in coma, 236
    dipsticks, 798
    osmolality, 798
    pH, 798, 829
    for rheumatic disease, 1065
    specific gravity, 798
    specimen collection, 1432b
    volume, 798–799
  hormone assays, 1144
  microalbuminuria, 798–799
Urography, excretion, 826, 826f
Urological cancers, 1406–1409
Urothelial cancer, 1407, 1407f
Urothelial tumours, 851–852
Urticaria, 1587–1588, 1588f
Urticaria pigmentosa (cutaneous mastocytosis), 1626
Urticarial vasculitis, 1588
Uterine fibroids, female health and, 1657–1658, 1658f
Uterus, ultrasound of, 128–129
UTI. *See* Urinary tract infection
Uveitis, 1057–1058

V

Vaccines/vaccination, 161–163, 161b–162b, 162f–163f, 169
  for acute viral illness, 212
  for hepatitis A, 695, 1520–1521
  for hepatitis B, 1521
  for HPV, 1531
  for rabies, 1461
  for rotavirus, 1462
  for tuberculosis (BCG), 557
  young adults and, 1636
Vaccinia virus, 1449
Vaginal bleeding, in early pregnancy, 1652b
Vaginal discharge, 1523b
Vaginal examination, 1646b, 1647f
  in acute abdomen, 675
Vagus nerve (X), 902
  lesions, 902, 902b
Valaciclovir, 1465
Values, in medical ethics, 26
Valvular heart disease, 207, 452–469
  aortic valve, 456–462
  mitral valve, 452–456, 452f
  MRI for, 407
  prosthetic valves, 464–465, 464f
Vancomycin-resistant enterococci, 295
Vaping, young adults and, 1635
Variant (Prinzmetal's) angina, 440
Variant Creutzfeldt-Jakob disease (vCJD), 958
Variceal banding, 713–714
Variceal haemorrhage, 713, 713f
Varicella zoster virus, 1592
Varicocele, 1668, 1668f
Varicose eczema, 1581
Varicose veins, 494–495

Vascular access devices, for healthcare-associated infections, 1439
Vascular anatomy, 981, 982f–983f
Vascular dementia, 957
Vascular disorders, 1320–1321, 1321b
  epilepsy and, 924, 924f
Vascular endothelium, 391–393
  products of, 393b
Vascular IR, 147–148, 148f
Vasculitis, 1093, 1607–1608
Vasodilators
  for heart failure, 436
  for shock, 317
Vasomotor control, of blood vessels, 392–393
Vasomotor rhinitis, 537
Vaso-occlusive crises, 1296
Vasopressin, 317, 859, 1190b
  release, factors affecting, 1191b
  response to increasing osmolality, 1191f
Vasopressin antagonists, 868
Vasopressors
  for heart failure, 436
  surgery and, 345
Vasovagal syncope, 238
Vaughan Williams classification of antiarrhythmic drugs, 426b, 427f
Vector-borne disease, 1435, 1436b
  prevention of, 1439
Vegetative state, 949f
Venesection, 721
Venomous snakes, 370, 370.e2, 370.e2f, 370.e4f, 370.e5f
Venous beading, 1226
Venous eczema, 1581
Venous loops, 1226
Venous stroke, 981
Venous thromboembolic disease, 1363–1374
Venous thromboembolism (VTE)
  anticoagulant management of, 271f
  surgery and, 347
Venous thrombosis, 1364–1365
  clinical skills, 1363–1364, 1363b
  complications, 1366
  diagnosis, 1363, 1364f
  factors influencing, 1365
  investigation, 1366–1368
  management of, 1365
  mechanical prophylaxis, 1369
  pathogenesis, 1364–1365
  pharmacological prophylaxis, 1369
  pregnancy, 1686, 1686b
  prevention, 1368–1369
  recurrence, 1365
  risk factors, 1365, 1369b
Venous ulcers, 1608–1609, 1608f
Venous ultrasound, 129, 129f–130f
Ventilation/perfusion mismatch, 302, 545
Ventilation/perfusion ($\dot{V}/\dot{Q}$) scan
  in pulmonary embolism, 267, 269f
  of thorax, 143–144, 144f
Ventilator-associated pneumonia, 322
Ventricular assist devices (VADs), 437, 438f
Ventricular fibrillation, 197–198, 197f
Ventricular premature beats (ectopics), 425–426, 426f
Ventricular septal defect (VSD), 451, 451f, 470, 470f
Ventricular tachyarrhythmias, 196–198, 423–429
Venturi mask, 206–207, 207f
Vericiguat, for heart failure, 436
Vertebral artery dissection, 981
Vertebral fracture, 1129
Vertebral fracture assessment (VFA), 1129, 1129f, 1129b
Vertical gaze, 897
Vertigo, 236b, 238–239, 586b, 589–590, 900
  causes of, 239b, 900, 900b
    central, 240b, 901
    peripheral, 240b, 900
    red flags for, 239b
  central causes of, 590
  clinical assessment of, 238–239, 240b
  head impulse test for, 901f
  investigations of, 901
  management of, 239
  Ménière's disease, 900
  migraine, 917
  peripheral causes of, 590–591
Very-low-density lipoprotein (VLDL) particles, 511, 511f
  disorders of, 513
Vesical calculi, 821–824, 821b
Vesicles, 1568f–1570f
Vessels, dissection of, 997, 997f
Vestibular disorders, 900
Vestibular migraine, 591
Vestibular neuritis, 589b, 590
Vestibular neuronitis, 900
Vestibular schwannoma, 589
Vestibular system, 900
Vestibulocochlear nerve (VIII), 900–901
Vestibulo-ocular reflexes, 896
Vestibulo-ocular (doll's eye) reflexes, 235f, 896
VFA. *See* Vertebral fracture assessment
Video-assisted thoracoscopic lung biopsy, 537
Vinca alkaloids, 1389
Viperidae, poison of, 370.e2, 370.e3f
VIPoma, 760
Viral activation, 76
Viral arthritis, 1106–1107
Viral conjunctivitis, 1053, 1055f
Viral encephalitis, 249, 1460b
Viral exanthems, 1592
Viral gastroenteritis, 1472
Viral haemorrhagic fevers, 1463–1465, 1464b, 1518.e3b
Viral illness, acute, 211–212, 211b–212b
Viral infections, 1444, 1445b
  cellular outcomes of, 1444
  gastrointestinal tract, 1461–1462
  HIV-associated, 1458, 1557–1559
  and malignant disease, 1462–1463
  nervous system, 1458–1461
  postviral/chronic fatigue syndrome, 1458
  respiratory tract, 1451–1454, 1451b
  skin and mucous membranes, 1444–1450
  systemic, 1454–1458
  transfusion-transmitted, 1312
Viral markers, 687
Viral meningitis, 247, 1460b
Virchow's triad, 1365
Virilizing tumour, 1185b
Virological monitoring, 1546
Virtual wards, 177
  for acute viral illness, 212
Viruses, 1421
Visceral larva migrans, 1513
Visceral leishmaniasis (kala azar), 1504
Visual acuity, 893–896, 1047–1048, 1048f–1049f
Visual field defects, 893
Visual loss, 1060
Visual pathway, 893f
Vital capacity, 318
Vitamin A, 772, 775
  adverse effects, 775
  deficiency, 773b, 775, 775f
Vitamin B_1 (thiamine)
  daily requirement, 778b
  deficiency, in neuropathy, 967
Vitamin B_2 (riboflavin), 778b
Vitamin B_3 (niacin), 778b
Vitamin B_5 (pantothenic acid), 778b
Vitamin B_6 (pyridoxine), deficiency, 778b, 967
Vitamin B_7 (biotin), 778b
Vitamin B_9 (folate), 778b
Vitamin B_{12} (cobalamin), 778b, 1284
  absorption and transport of, 1284
  daily requirement, 778b
  deficiency, 778b, 1284, 1285b
    differential diagnosis of, 1285
    management of, 1286
    in neuropathy, 968
    in spinal cord lesions, 951
Vitamin C (ascorbic acid), 776–777
Vitamin D, 775–776
  binding protein, 1127
  deficiency, 1127, 1195
    management of, 1131b
    obesity care and, 1265
  insufficiency, 1127
  metabolism of, 797, 971, 1125, 1126f
Vitamin E, 776
Vitamin K, 776
  antagonists, 776
  deficiency, 776, 1324b, 1325–1327
  function, 776
Vitamin K antagonist, oral, 1372–1374
Vitiligo, 1625, 1625f
  management of, 1625
Vitreoretinal surgery, for diabetic nephropathy, 1227
Vitreous haemorrhage, 1226
Vitreous humour, 1045
VLDL. *See* Very-low-density lipoprotein
Vocal resonance, 531
Voiding micturating cystourethrogram (VMCUG), 125–126, 126f–127f
Volatile anaesthetic gases, for surgery, 343
Volume-controlled ventilation, 322
Volume replacement, 313–315
Volutrauma, 322
Volvulus, 677
Vomiting, acute abdomen and, 674
von Hippel-Lindau (VHL) disease, 851
von Hippel-Lindau (VHL) syndrome, 961, 1161, 1383b
von Recklinghausen's disease, 961
von Willebrand disease, 1319f, 1324b, 1325
von Willebrand factor, 1317, 1319f, 1324b
Vote counting, 47
VSD. *See* Ventricular septal defect
VTE. *See* Venous thromboembolism

W

Waddling gait, 908
Waist-to-height ratio, 1252, 1257f
Waldenström's macroglobulinaemia, 1355
Warfarin, 1373, 1373b
'Warm' autoimmune haemolytic anaemia, 1302, 1303b
Warts, anogenital, 1530, 1531f
Water
  absorption of, 795
  excretion, regulation of, 865–866
  loss of, 866
  retention, 430, 430f
Water balance, 764, 853–884, 853b–854b, 854f
Water deprivation test, 1148b, 1192b
Water excretion, regulation of, 865–866
Water hammer pulse, 395, 395f
Water-soluble vitamins, 776–777
WDHs. *See* Wider determinants of health
Weal, 1568f–1570f
Weight-related amenorrhoea, 1181
Wellbeing, ageing and, 1719
Wells Score, 181b
  for deep vein thrombosis, 266b
  for pulmonary embolism, 266b
Wernicke-Korsakoff syndrome, 777, 1015, 1016f
Wernicke's aphasia, 890
Wernicke's encephalopathy, 897
West Nile encephalitis, 1459
Wheezing, 530
Whiplash injury, 1068–1069
Whipple procedure, 745
Whipple's disease, 649, 943–944, 1478
Whipworm, 1512
White blood cells (leucocytes), 1314–1316
White cell count, 1275
White cell scan, 145–146
Whole-bowel irrigation, 358
Whole-exome and whole-genome sequencing (WES/WGS), 71, 161b
Whole-pancreas transplantation, 1222
Whooping cough, 161f, 161b
Wider determinants of health (WDHs), 154f, 158
Wilms' tumour, 1383b
Wilson and Jungner criteria, 164b
Wilson's disease, 721, 934
'Windscreen wiper' eyes, 235f
Winemaker's lung, 573b
Wiskott-Aldrich syndrome, 99
Withholding and withdrawing treatment, 330
Wolff-Parkinson-White syndrome, 419
Wolfram's syndrome, 1191
Women's health
  reproductive disorders, 1180–1184
  reproductive physiology, 1176–1177, 1176f–1177f
Wool-sorter's disease, 1485
World Health Organization (WHO) analgesic ladder, 1724, 1724f
Wound infection, surgery and, 348
Writer's cramp, 934
Wuchereria bancrofti, 1509b

X

X chromosome, 59
X-adrenoleukodystrophy (ALD), 1247
Xanthogranulomatous pyelonephritis, 833
Xeroderma pigmentosa, 76, 1383b
Xerophthalmia, 775, 775f, 775b
XLA. *See* X-linked agammaglobulinaemia
X-linked agammaglobulinaemia (XLA), 99
X-linked dominant disorders, 67
X-linked recessive disorders, 66
X-rays, 118–125, 118b, 618
Xylometazoline, 538

Y

Y chromosome, 66
Yaws, 1486
Yellow fever, 704, 1463–1465
Yellow oleander, poisoning, 370.e6
Yellow-nail syndrome, 1626
Yersiniosis, 1475
Young adult, 1632f
  consent for, 1633b
  core skills and knowledge, 1631b
  developmental changes during, 1632f
  gender identity in, 1638
  health literacy of, 1638
  law for, 1633b
  lifestyle choices in, 1634–1636
  long-term health conditions, 1634b, 1636–1637, 1636f–1637f, 1637b
  medicine, 1631–1638
  menstruation in, 1638
  mental health issues in, 1632–1634
  morbidity, 1632
  mortality, 1632, 1634f
  needs of, consideration in, 1638
  neurodiversity of, 1638
  sexual and reproductive health of, 1638
  shared decision-making of, 1638

Z

Zafirlukast, 544
Zanamivir, 1467
Zika virus, 1456
Zinc, 779–780
  deficiency, 780f
Zoledronate
  for osteoporosis, 1130b
  for Paget's disease, 1134
Zoonoses, 1434, 1435b
Zygomycosis, 1496

This page intentionally left blank

REFERENCE INTERVALS *(These vary: please check with your local laboratory)*

Haematology

Haemoglobin
- Male 135–175 g/L
- Female 115–155 g/L

Mean corpuscular haemoglobin (MCH) 27–32 pg

Mean corpuscular haemoglobin concentration (MCHC) 320–360 g/L

Mean corpuscular volume (MCV) 80–96 fL

Packed cell volume (PCV) or haematocrit
- Male 0.40–0.54 L/L
- Female 0.37–0.47 L/L

White blood count (WBC) 4–11 × 10^9/L
- Basophil granulocytes 0.01–0.1 × 10^9/L
- Eosinophil granulocytes 0.04–0.4 ×10^9/L
- Lymphocytes 1.5–4.0 ×10^9/L
- Monocytes 0.2–0.8 ×10^9/L
- Neutrophil granulocytes 2.0–7.5 × 10^9/L

Platelet count 150–400 × 10^9/L

Serum B_{12} 160–925 ng/L (150–675 pmol/L)

Serum folate 4–18 μg/L (5–63 nmol/L)

Red cell folate 160–640 μg/L

Reticulocyte count 50–100 × 10^9/L

Erythrocyte sedimentation rate (ESR) <20 mm/h

Plasma viscosity 1.5–1.72 mPa.s

Coagulation

Activated partial thromboplastin time (APTT) 26–37 s

Prothrombin time 11.5–13.5 s

International Normalized Ratio (INR) 1.0–1.3

D-dimer <500 μg/L (0.50 μg/mL) pulmonary embolism unlikely

Biochemistry

(Serum/plasma in alphabetical order)

Alanine aminotransferase (ALT) <40 U/L

Albumin 35–50 g/L

Alkaline phosphatase 35–129 U/L

Amylase 25–125 U/L

Angiotensin-converting enzymes 10–70 U/L

α_1-antitrypsin 2–4 g/L

Aspartate aminotransferase (AST) 12–40 U/L

Bilirubin <17 μmol/L (0.3–1.5 mg/dL)

Brain natriuretic peptide (BNP) threshold 100 pg/mL

Caeruloplasmin 1.5–2.9 μmol/L

Calcium 2.15–2.55 mmol/L

Chloride 98–107 mmol/L

Complement
- C3 0.75–1.65 g/L
- C4 0.20–0.60 g/L

Copper 11–20 μmol/L (100–200 mg/dL)

C-reactive protein 0–4 mg/L

Creatine kinase (CPK)
- Female 20–170 U/L
- Male 30–200 U/L

Creatinine 45–90 μmol/L (0.5–1.0 mg/dL)

Ferritin
- Female 30–200 μg/L
- Male 30–300 μg/L

α-fetoprotein <10 k U/L

Fructosamine 205–285 μmol/L

Gamma-glutamyl transpeptidase (γ-GT)
- Male 4–72 IU/L
- Female 4–60 IU/L

Glucose (fasting) 3.9–5.6 mmol/L (70–100 mg/dL)

Glycosylated (glycated) haemoglobin (HbA_{1c}) 20–41 mmol/mol (4.0–5.9%)

Immunoglobulins (11 years and over)
- IgA 0.8–4.0 g/L
- IgG 6.0–16.0 g/L
- IgM 0.5–2.0 g/L

Iron 13–32 μmol/L (50–150 μg/dL)

Iron binding capacity (total) (TIBC) 42–80 μmol/L (250–410 μg/dL

Lactate 0.6–2.0 mmol/L

Lactate dehydrogenase 240–480 U/L

Magnesium 0.7–1.1 mmol/L

β_2-microglobulin 1.0–3.0 mg/L

Osmolality 285–295 mOsm/kg

Phosphate 0.8–1.5 mmol/L

Potassium 3.5–5.0 mmol/L

Prostate-specific antigen (PSA) up to 4.0 μg/L

Procalcitonin 0.00–0.05 μg/L (bacterial infection likely if ≥0.25 [or ≥0.50 in ITU])

Protein (total) 62–77 g/L

Sodium 135–145 mmol/L

Tryptase <13 ng/mL

Urate 0.18–0.42 mmol/L (3.0–7.0 mg/dL)

Urea 2.5–6.7 mmol/L (8–25 mg/dL)

Vitamin A 0.5–2.01 μmol/L

Vitamin D (seasonal variation) 25-hydroxy 37–200 nmol/L (0.15–0.80 ng/L)

Zinc 11–24 μmol/L

Lipids and lipoproteins

Cholesterol <5 mmol/L

HDL cholesterol
- Male >1 mmol/L
- Female >1.2 mmol/L

LDL cholesterol <3.0 mmol/L

Triglycerides
- Fasting <1.7 mmol/L
- Non-fasting <2.3 mmol/L

Thyroid function tests
- Thyroxine (free T4) 9–21 pmol/L
- Tri-iodothyronine (free T3) 0.9–2.4 nmol/L
- Thyroid stimulating hormone 0.2–4.2 mU/L

Blood gases (arterial)

P_aCO_2 4.8–6.1 kPa (36–46 mmHg)

P_aO_2 10.6–13.3 kPa (80–100 mmHg)

$[H^+]$ 35–45 nmol/L

pH 7.35–7.45

Bicarbonate 22–26 mmol/L

Urine values

Calcium 7.5 mmol/24 h or less (<300 mg daily)

Copper 0.2–1.0 μmol/24 h (15–40 mg/24 h)

Creatinine 0.13–0.22 mmol/kg body weight per day

5-hydroxyindole acetic acid (5HIAA) <47 μmol daily; amounts lower in females than males

Protein (quantitative) <0.15 g/24 h

Sodium 60–180 mmol/24 h

Serum/urine values

Estimated glomerular filtration rate (eGFR)
- Male 90–140 mL/min
- Female 80–125 mL/min

Albumin mg/Creatinine mmol (ACR)
- Female <3.5
- Male <2.5